z/OS Communications Server

IP Configuration Guide

Version 1 Release 7

SC31-8775-08

z/OS Communications Server

IP Configuration Guide

Version 1 Release 7

SC31-8775-08

Note:
Before using this information and the product it supports, be sure to read the general information under "Notices" on page 1245.

Ninth Edition (April 2006)

This edition applies to Version 1 Release 7 of z/OS (5694-A01) and Version 1 Release 7 of z/OS.e (5655-G52) and to all subsequent releases and modifications until otherwise indicated in new editions.

IBM welcomes your comments. You may send your comments to the following address.

International Business Machines Corporation
Attn: z/OS Communications Server Information Development
Department AKCA, Building 501
P.O. Box 12195, 3039 Cornwallis Road
Research Triangle Park, North Carolina 27709-2195

You can send us comments electronically by using one of the following methods:

Fax (USA and Canada):

1+919-254-4028

Send the fax to "Attn: z/OS Communications Server Information Development"

Internet e-mail:

comsvrcf@us.ibm.com

World Wide Web:

http://www.ibm.com/servers/eserver/zseries/zos/webqs.html

If you would like a reply, be sure to include your name, address, telephone number, or FAX number. Make sure to include the following in your comment or note:

- Title and order number of this document
- Page number or topic related to your comment

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Figures

Tables

About this document

This document contains guidance material to enable you to configure IP address spaces, servers, and applications for z/OS® Communications Server. This volume is part of a two-volume set:

- *z/OS Communications Server: IP Configuration Guide*, which contains concepts and guidance, explaining an overall approach to IP configuration.
- *z/OS Communications Server: IP Configuration Reference*, which describes parameters, options, and syntax of statements.

The information in this document supports both IPv6 and IPv4. Unless explicitly noted, information describes IPv4 networking protocol. IPv6 support is qualified within the text.

For detailed information about configuration-related data sets and statements, refer to *z/OS Communications Server: IP Configuration Reference*.

For detailed information about commands used during configuration, refer to *z/OS Communications Server: IP System Administrator's Commands*.

This document supports z/OS.e.

Who should read this document

This document is intended for programmers and system administrators who are familiar with TCP/IP, MVS™, z/OS UNIX®, and the Time Sharing Option Extensions (TSO/E).

How this document is organized

This book is divided into the following parts:

Part 1, "Base TCP/IP system," on page 1 contains information on configuring the base TCP/IP stack.

Part 2, "Server applications," on page 403 contains chapters that explain the server applications for z/OS CS.

Part 3, "Appendixes" provides additional information for this document.

"Notices" contains notices and trademarks used in this document.

"Bibliography" contains descriptions of the documents in the z/OS Communications Server library.

How to use this document

Use this document to perform the following tasks:

- Configure z/OS CS
- Customize and administer z/OS CS

Determining whether a publication is current

As needed, IBM® updates its publications with new and changed information. For a given publication, updates to the hardcopy and associated BookManager® softcopy are usually available at the same time. Sometimes, however, the updates to hardcopy and softcopy are available at different times. The following information describes how to determine if you are looking at the most current copy of a publication:

- At the end of a publication's order number there is a dash followed by two digits, often referred to as the dash level. A publication with a higher dash level is more current than one with a lower dash level. For example, in the publication order number GC28-1747-07, the dash level 07 means that the publication is more current than previous levels, such as 05 or 04.
- If a hardcopy publication and a softcopy publication have the same dash level, it is possible that the softcopy publication is more current than the hardcopy publication. Check the dates shown in the Summary of Changes. The softcopy publication might have a more recently dated Summary of Changes than the hardcopy publication.
- To compare softcopy publications, you can check the last two characters of the publication's file name (also called the book name). The higher the number, the more recent the publication. Also, next to the publication titles in the CD-ROM booklet and the readme files, there is an asterisk (*) that indicates whether a publication is new or changed.

How to contact IBM service

For immediate assistance, visit this Web site:
http://www.software.ibm.com/network/commserver/support/

Most problems can be resolved at this Web site, where you can submit questions and problem reports electronically, as well as access a variety of diagnosis information.

For telephone assistance in problem diagnosis and resolution (in the United States or Puerto Rico), call the IBM Software Support Center anytime (1-800-IBM-SERV). You will receive a return call within 8 business hours (Monday – Friday, 8:00 a.m. – 5:00 p.m., local customer time).

Outside of the United States or Puerto Rico, contact your local IBM representative or your authorized IBM supplier.

If you would like to provide feedback on this publication, see "Communicating Your Comments to IBM" on page 1277.

Conventions and terminology used in this document

For definitions of the terms and abbreviations used in this document, you can view the latest IBM terminology at the IBM Terminology Web site.

Clarification of notes

Information traditionally qualified as **Notes** is further qualified as follows:

Note Supplemental detail

Tip Offers shortcuts or alternative ways of performing an action; a hint

Guideline
Customary way to perform a procedure; stronger request than recommendation

Rule Something you must do; limitations on your actions

Restriction
Indicates certain conditions are not supported; limitations on a product or facility

Requirement
Dependencies, prerequisites

Result Indicates the outcome

Prerequisite and related information

z/OS Communications Server function is described in the z/OS Communications Server library. Descriptions of those documents are listed in "z/OS Communications Server information" on page 1255, in the back of this document.

Required information

Before using this product, you should be familiar with TCP/IP, VTAM®, MVS, and UNIX System Services.

Related information

This section contains subsections on:

Softcopy information

Softcopy publications are available in the following collections:

Titles	Order Number	Description
z/OS V1R7 Collection	SK3T-4269	This is the CD collection shipped with the z/OS product. It includes the libraries for z/OS V1R7, in both BookManager and PDF formats.
z/OS Software Products Collection	SK3T-4270	This CD includes, in both BookManager and PDF formats, the libraries of z/OS software products that run on z/OS but are not elements and features, as well as the *Getting Started with Parallel Sysplex®* bookshelf.
z/OS V1R7 and Software Products DVD Collection	SK3T-4271	This collection includes the libraries of z/OS (the element and feature libraries) and the libraries for z/OS software products in both BookManager and PDF format. This collection combines SK3T-4269 and SK3T-4270.
z/OS Licensed Product Library	SK3T-4307	This CD includes the licensed documents in both BookManager and PDF format.
System Center Publication IBM S/390® Redbooks™ Collection	SK2T-2177	This collection contains over 300 ITSO redbooks that apply to the S/390 platform and to host networking arranged into subject bookshelves.

Other documents

For information about z/OS products, refer to *z/OS Information Roadmap* (SA22-7500). The Roadmap describes what level of documents are supplied with each release of z/OS Communications Server, as well as describing each z/OS publication.

Relevant RFCs are listed in an appendix of the IP documents. Architectural specifications for the SNA protocol are listed in an appendix of the SNA documents.

The following table lists documents that might be helpful to readers.

Title	Number
DNS and BIND, Fourth Edition, O'Reilly and Associates, 2001	ISBN 0-596-00158-4
Routing in the Internet , Christian Huitema (Prentice Hall PTR, 1995)	ISBN 0-13-132192-7
sendmail, Bryan Costales and Eric Allman, O'Reilly and Associates, 2002	ISBN 1-56592-839-3
SNA Formats	GA27-3136
TCP/IP Illustrated, Volume I: The Protocols, W. Richard Stevens, Addison-Wesley Publishing, 1994	ISBN 0-201-63346-9
TCP/IP Illustrated, Volume II: The Implementation, Gary R. Wright and W. Richard Stevens, Addison-Wesley Publishing, 1995	ISBN 0-201-63354-X
TCP/IP Illustrated, Volume III, W. Richard Stevens, Addison-Wesley Publishing, 1995	ISBN 0-201-63495-3
TCP/IP Tutorial and Technical Overview	GG24-3376
Understanding LDAP	SG24-4986
z/OS Crytographic Service System Secure Sockets Layer Programming	SC24-5901
z/OS Integrated Security Services Firewall Technologies	SC24-5922
z/OS Integrated Security Services LDAP Client Programming	SC24-5924
z/OS Integrated Security Services LDAP Server Administration and Use	SC24-5923
z/OS JES2 Initialization and Tuning Guide	SA22-7532
z/OS MVS Diagnosis: Procedures	GA22-7587
z/OS MVS Diagnosis: Reference	GA22-7588
z/OS MVS Diagnosis: Tools and Service Aids	GA22-7589
z/OS MVS Using the Subsystem Interface	SA22-7642
z/OS Program Directory	GI10-0670
z/OS UNIX System Services Command Reference	SA22-7802
z/OS UNIX System Services Planning	GA22-7800
z/OS UNIX System Services Programming: Assembler Callable Services Reference	SA22-7803
z/OS UNIX System Services User's Guide	SA22-7801
z/OS XL C/C++ Run-Time Library Reference	SA22-7821
zSeries OSA-Express Customer's Guide and Reference	SA22-7935

Redbooks

The following Redbooks might help you as you implement z/OS Communications Server.

Title	Number
Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 1: Base and TN3270 Configuration	SG24-5227
Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 2: UNIX Applications	SG24-5228
Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 4: Connectivity and Routing	SG24-6516
Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 7: Security	SG24-6840
IBM Communication Controller Migration Guide	SG24-6298
IP Network Design Guide	SG24-2580
Managing OS/390® TCP/IP with SNMP	SG24-5866
Migrating Subarea Networks to an IP Infrastructure	SG24-5957
OS/390 eNetwork Communications Server V2R7 TCP/IP Implementation Guide: Volume 3: MVS Applications	SG24-5229
Secureway Communications Server for OS/390 V2R8 TCP/IP: Guide to Enhancements	SG24–5631
SNA and TCP/IP Integration	SG24-5291
TCP/IP in a Sysplex	SG24-5235
TCP/IP Tutorial and Technical Overview	GG24-3376
Threadsafe Considerations for CICS	SG24-6351

Where to find related information on the Internet

z/OS

This site provides information about z/OS Communications Server release availability, migration information, downloads, and links to information about z/OS technology

http://www.ibm.com/servers/eserver/zseries/zos/

z/OS Internet Library

Use this site to view and download z/OS Communications Server documentation

http://www.ibm.com/servers/eserver/zseries/zos/bkserv/

IBM Communications Server product

The primary home page for information about z/OS Communications Server

http://www.software.ibm.com/network/commserver/

IBM Communications Server product support

Use this site to submit and track problems and search the z/OS Communications Server knowledge base for Technotes, FAQs, white papers, and other z/OS Communications Server information

http://www.software.ibm.com/network/commserver/support/

IBM Systems Center publications

Use this site to view and order Redbooks, Redpapers, and Technotes

http://www.redbooks.ibm.com/

IBM Systems Center flashes

Search the Technical Sales Library for Techdocs (including Flashes, presentations, Technotes, FAQs, white papers, Customer Support Plans, and Skills Transfer information)

http://www.ibm.com/support/techdocs/atsmastr.nsf

RFCs

Search for and view Request for Comments documents in this section of the Internet Engineering Task Force Web site, with links to the RFC repository and the IETF Working Groups Web page

http://www.ietf.org/rfc.html

Internet drafts

View Internet-Drafts, which are working documents of the Internet Engineering Task Force (IETF) and other groups, in this section of the Internet Engineering Task Force Web site

http://www.ietf.org/ID.html

Information about Web addresses can also be found in information APAR II11334.

DNS Web sites: For more information about DNS, see the following USENET news groups and mailing addresses:

USENET news groups
comp.protocols.dns.bind

BIND mailing lists
http://www.isc.org/ml-archives/

BIND Users

- Subscribe by sending mail to bind-users-request@isc.org.
- Submit questions or answers to this forum by sending mail to bind-users@isc.org.

BIND 9 Users (This list might not be maintained indefinitely.)

- Subscribe by sending mail to bind9-users-request@isc.org.
- Submit questions or answers to this forum by sending mail to bind9-users@isc.org.

Note: Any pointers in this publication to Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Using LookAt to look up message explanations

LookAt is an online facility that lets you look up explanations for most of the IBM messages you encounter, as well as for some system abends and codes. Using LookAt to find information is faster than a conventional search because in most cases LookAt goes directly to the message explanation.

You can use LookAt from the following locations to find IBM message explanations for z/OS elements and features, z/VM®, VSE/ESA™, and Clusters for AIX® and Linux™:

- The Internet. You can access IBM message explanations directly from the LookAt Web site at http://www.ibm.com/eserver/zseries/zos/bkserv/lookat/.

- Your z/OS TSO/E host system. You can install code on your z/OS or z/OS.e systems to access IBM message explanations, using LookAt from a TSO/E command line (for example, TSO/E prompt, ISPF, or z/OS UNIX System Services).
- Your Microsoft® Windows® workstation. You can install code to access IBM message explanations on the *z/OS Collection* (SK3T-4269), using LookAt from a Microsoft Windows command prompt (also known as the DOS command line).
- Your wireless handheld device. You can use the LookAt Mobile Edition with a handheld device that has wireless access and an Internet browser (for example, Internet Explorer for Pocket PCs, Blazer or Eudora for Palm OS, or Opera for Linux handheld devices). Link to the LookAt Mobile Edition from the LookAt Web site.

You can obtain code to install LookAt on your host system or Microsoft Windows workstation from a disk on your *z/OS Collection* (SK3T-4269), or from the LookAt Web site (click **Download**, and select the platform, release, collection, and location that suit your needs). More information is available in the LOOKAT.ME files available during the download process.

Using IBM Health Checker for z/OS

IBM Health Checker for z/OS is a z/OS component that installations can use to gather information about their system environment and system parameters to help identify potential configuration problems before they impact availability or cause outages. Individual products, z/OS components, or ISV software can provide checks that take advantage of the IBM Health Checker for z/OS framework. This book may refer to checks or messages associated with this component.

For additional information about checks and about IBM Health Checker for z/OS, see *IBM Health Checker for z/OS: User's Guide*. z/OS V1R4, V1R5, and V1R6 users can obtain the IBM Health Checker for z/OS from the z/OS Downloads page at http://www.ibm.com/servers/eserver/zseries/zos/downloads/.

SDSF also provides functions to simplify the management of checks. See *z/OS SDSF Operation and Customization* for additional information.

How to send your comments

Your feedback is important in helping to provide the most accurate and high-quality information. If you have any comments about this document or any other z/OS Communications Server documentation:

- Go to the z/OS contact page at:

  http://www.ibm.com/servers/eserver/zseries/zos/webqs.html

  There you will find the feedback page where you can enter and submit your comments.
- Send your comments by e-mail to comsvrcf@us.ibm.com. Be sure to include the name of the document, the part number of the document, the version of z/OS Communications Server, and, if applicable, the specific location of the text you are commenting on (for example, a section number, a page number or a table number).

Summary of changes

Summary of changes
for SC31-8775-08
z/OS Version 1 Release 7

This document contains information previously presented in SC31-8775-07, which supports z/OS Version 1 Release 7.

The information in this document includes descriptions of support for both IPv4 and IPv6 networking protocols. Unless explicitly noted, descriptions of IP protocol support concern IPv4. IPv6 support is qualified within the text.

This document refers to Communications Server data sets by their default SMP/E distribution library name. Your installation might, however, have different names for these data sets where allowed by SMP/E, your installation personnel, or administration staff. For instance, this document refers to samples in SEZAINST library as simply in SEZAINST. Your installation might choose a data set name of SYS1.SEZAINST, CS390.SEZAINST, or other high level qualifiers for the data set name.

New information

- LIBPATH environment variable, see:
  - "Starting and stopping the Policy Agent" on page 746
  - "Starting and stopping the Network SLAPM2 subagent" on page 774
- OSA-Express Direct subagent, see:
  - "TCP/IP subagent" on page 1051
  - "OSA-Express Direct subagent" on page 1053
  - "Step 4: Configure the Open Systems Adapter (OSA) support" on page 1074

Changed information

- Configuring the z/OS UNIX snmp command, see "Configure the z/OS UNIX snmp command" on page 1067
- Telnet use of Client Identifiers, see "Client Identifiers" on page 469

Deleted information

- OROUTED

  If OROUTED was used in a prior release and the RIP protocol is still the preferred dynamic routing method in your host configuration, use OMPROUTE to provide RIP support. Additionally, support for the VARSUBNETTING and NOVARSUBNETTING keywords on the IPCONFIG and ASSORTEDPARMS statements has been removed. For more information on migrating from OROUTED to OMPROUTE, refer to *z/OS Migration*.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change. This document also contains the revision bars from SC31-8775-07 for reference purposes.

You might notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Summary of changes
for SC31-8775-07
z/OS Version 1 Release 7

This document contains information previously presented in SC31-8775-05 and SC31-8775-06, which support z/OS Version 1 Release 6.

The information in this document includes descriptions of support for both IPv4 and IPv6 networking protocols. Unless explicitly noted, descriptions of IP protocol support concern IPv4. IPv6 support is qualified within the text.

This document refers to Communications Server data sets by their default SMP/E distribution library name. Your installation might, however, have different names for these data sets where allowed by SMP/E, your installation personnel, or administration staff. For instance, this document refers to samples in SEZAINST library as simply in SEZAINST. Your installation might choose a data set name of SYS1.SEZAINST, CS390.SEZAINST, or other high level qualifiers for the data set name.

New information

- QDIO OSA-Express segmentation offload, see "Checksum offload" on page 91 and "TCP segmentation offload" on page 91.
- Common Information Model (CIM), see "Considerations for Common Information Model (CIM) providers" on page 102 and "CIM provider access control" on page 130.
- Operator commands for DVIPA management, see "Manually initiating takeover for an individual dynamic VIPA" on page 308, "Manually quiescing DVIPA Sysplex Distributor server applications" on page 318, and "Sysplex problem detection and recovery" on page 376.
- Server-specific Workload Manager (WLM) for Sysplex Distributor, see "Configuring distributed DVIPAs — Sysplex Distributor" on page 316 and "Sysplex Distributor" on page 388.
- Sysplex autonomics health monitor for target stacks, see "Configuring distributed DVIPAs — Sysplex Distributor" on page 316, "Verifying Sysplex Distributor workload" on page 356, and "Sysplex Distributor" on page 388.
- Optimized routing for Sysplex Distributor, see "Route selection for distributing packets" on page 319, "Generic routing encapsulation (GRE)" on page 320, "Summary of problems monitored and actions taken" on page 380, and "Steps for setting up Sysplex Distributor to be the service manager for the Cisco MNLB (IPv4 only)" on page 394.
- Subarea VTAM XCF, see "Connectivity in a sysplex" on page 364.
- z/OS Load Balancing Advisor, see "Workload balancing" on page 386 and Chapter 19, "z/OS Load Balancing Advisor," on page 1009.
- TN3270E AES encryption exploitation, see "Transport layer security" on page 461.

- FTP AES encryption exploitation, see “Steps for customizing the FTP server for TLS” on page 558 and “Steps for customizing the FTP client for TLS” on page 566.
- FTP confidence of success level reporting, see “The FTPOSTPR user exit” on page 574.
- IPv4 integrated IPSec/VPN support and NAT traversal support for integrated IPSec/VPN, see Chapter 14, “Policy-based networking,” on page 717 and Chapter 17, “IP security,” on page 819.
- Application Transparent Transport Layer Security (AT-TLS), see Chapter 14, “Policy-based networking,” on page 717 and Chapter 18, “Application Transparent Transport Layer Security (AT-TLS) data protection,” on page 987.
- FTP security server enhancements, see “Enable CSFSERV resources” on page 1172.
- The following areas contain new information pertaining to IPv6 support:
  - HiperSockets™, see “HiperSockets concepts and connectivity” on page 92, “Setting up physical characteristics in PROFILE.TCPIP” on page 190, and “HiperSockets” on page 375.
  - Advanced socket API options, see “IPv6 advanced socket API options” on page 127, “Required configuration in a multilevel secure environment” on page 149, and “Defining security labels on other profiles in the SERVAUTH class” on page 153.

Changed information

- **Promotion of the use of IPv6 global unicast addresses**

  Site-local addresses were designed to use private address prefixes that could be used within a site without the need for a global prefix. Until recently, the full negative impacts of site-local addresses in the Internet were not fully understood. The IETF has deprecated the special treatment given to this site-local prefix. Because of this, it is preferable to use global unicast addresses. This means we are replacing addresses and prefixes that use the site-local prefix (fec0::/10) with ones that use the global prefix for documentation (2001:0DB8::/32).
- The phrase ″shared HFS″ has been changed to ″shared file system″.
- References to OpenEdition have been replaced with ″z/OS UNIX System Services″ or ″z/OS UNIX″.

Deleted information

- OROUTED

  If OROUTED was used in a prior release and the RIP protocol is still the preferred dynamic routing method in your host configuration, use OMPROUTE to provide RIP support. For more information on migrating from OROUTED to OMPROUTE, refer to *z/OS Migration*.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change.

You might notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Summary of changes
for SC31-8775-06
z/OS Version 1 Release 6

This document contains information previously presented in SC31-8775-05, which supports z/OS Version 1 Release 6. The information in this document supports both IPv6 and IPv4. Unless explicitly noted, information describes IPv4 networking protocol. IPv6 support is qualified within the text.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change. This document also contains the revision bars from SC31–8775–05 for reference purposes.

You may notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Summary of changes
for SC31-8775-05
z/OS Version 1 Release 6

This document contains information previously presented in SC31-8775-04, which supports z/OS Version 1 Release 5. The information in this document supports both IPv6 and IPv4. Unless explicitly noted, information describes IPv4 networking protocol. IPv6 support is qualified within the text.

New information

- Introductory chapter, see Chapter 1, “Overview of z/OS Communications Server,” on page 3.
- Environment variables explicitly set by z/OS Communications Server and its applications, see “Environment variables” on page 38.
- General networking hardware attachment information and considerations associated with the IBM zSeries® platform, see “Considerations for networking hardware attachment” on page 85.
- System resource protection, see “Socket option access control” on page 125 and “Required configuration in a multilevel secure environment” on page 149.
- Multilevel security for applications, see Chapter 4, “Preparing for TCP/IP networking in a multilevel secure environment,” on page 145.
- Consistency checking to maintain the security of information, see “Changing your multilevel secure networking environment” on page 172.
- SRCIP and ENDSRCIP profile statement block to allow a job to use a unique IP address as its source address for outbound TCP connections, see “Setting up TCP/IP operating characteristics in PROFILE.TCPIP” on page 187, “Static VIPAs, dynamic VIPAs (DVIPAs), distributed DVIPAs” on page 300, and “Steps for configuring static VIPAs for a z/OS TCP/IP stack” on page 301.
- Sysplex autonomics, see “Sysplex problem detection and recovery” on page 376.
- Telnet as a stand-alone application running in its own address space, see “Telnet in its own address space” on page 443.
- SNA character stream (SCS) format for unformatted system services (USS) tables, see “Using the Telnet USS and INTERPRET support” on page 514.

- FTP multibyte character support, see “FTP code page conversion” on page 555.
- Policy enforcement point (PEP) and the PEPInstance statement, see “Configuring the Policy Agent” on page 741.
- DELETEBADSPOOLFILE statement for SMTP, see “Summary of SMTP configuration statements” on page 1114.
- Select examples, tasks, and conceptual material are enabled for z/OS library center advanced searches.
- The following areas contain new information pertaining to IPv6 support:
  - OSPF for OMPROUTE, see Chapter 6, “Routing,” on page 217.
  - Sysplex, see Chapter 7, “Virtual IP Addressing,” on page 297 and Chapter 8, “TCP/IP in a sysplex,” on page 363.

Changed information

- **Promotion of the use of IPv6 global unicast addresses** - Site-local addresses were designed to use private address prefixes that could be used within a site without the need for a global prefix. Until recently, the full negative impacts of site-local addresses in the Internet were not fully understood. The IETF has deprecated the special treatment given to the site-local prefix. Because of this, it is preferable to use global unicast addresses. This means we are replacing addresses and prefixes that use the site-local prefix (fec0::/10) with ones that use the global prefix for documentation (2001:0DB8::/32). Some samples and examples in this document may display site-local prefixes instead of the now preferred global unicast addresses.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change.

Starting with z/OS V1R4, you may notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Summary of changes
for SC31-8775-04
z/OS Version 1 Release 5

This document contains technical and editorial updates to the information previously presented in SC31-8775-03, which supports z/OS Version 1 Release 5. The information in this document supports both IPv6 and IPv4. Unless explicitly noted, information describes IPv4 networking protocol. IPv6 support is qualified within the text.

New information

- Verifying TCPIP.DATA statement values in the native MVS environment, see “Verifying TCPIP.DATA statement values in the native MVS environment” on page 213.
- Verifying TCPIP.DATA statement values in the z/OS UNIX environment, see “Verifying TCPIP.DATA statement values in the z/OS UNIX environment” on page 213.
- Specifying the subnet mask, see “IPv4 static routing” on page 221.

- Providing community name information, see “Provide community name information” on page 1058.
- Configuring security, see “Provide community-based and user-based security and notification destination information” on page 1061.
- RELAY_DOMAIN() added to .mc file information, see “The minimal mc file” on page 1128.

Changed information

- SMTPJOB variable definition, see “Step 3: Customize the system CLIST and modify parmlib data sets” on page 1103.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change. This document also contains the revision bars from SC31–8775–03 for reference purposes.

Starting with z/OS V1R4, you may notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Summary of changes
for SC31-8775-03
z/OS Version 1 Release 5

This document contains information previously presented in SC31-8775-02, which supports z/OS Version 1 Release 4. The information in this document supports both IPv6 and IPv4. Unless explicitly noted, information describes IPv4 networking protocol. IPv6 support is qualified within the text.

New information

- Using msys for FTP customization, see “z/OS msys for Setup and Wizard” on page 11.
- MVS system symbol resolution enhancements in TCPIP.DATA, see “MVS system symbols” on page 54 and “Configuring TCPIP.DATA” on page 183.
- Access control information for network management programming interfaces, see “TCP/IP resource protection” on page 119.
- Multilevel security for applications, see Chapter 4, “Preparing for TCP/IP networking in a multilevel secure environment,” on page 145.
- DVIPA limits, see “Static VIPAs, dynamic VIPAs (DVIPAs), distributed DVIPAs” on page 300 and “Configuring the unique application-instance scenario” on page 310.
- Multiple static VIPA addresses and firewall support for Enterprise Extender, see “Configuring static VIPAs for Enterprise Extender” on page 303.
- Dynamically assigning Sysplex Distributor ports, see “Dynamic port assignment” on page 322.
- SYSPLEXPORTS performance enhancements, see “SYSPLEXPORTS” on page 323.
- Workload distribution enhancements, see “Timed affinities” on page 324.
- VIPABACKUP, see “Configuring VIPAs for activation with VIPABACKUP” on page 344.

- TN3270 keyboard control enhancements, see “Connection mode choices” on page 454.
- Multilevel security for the Telnet server, see “Network Access Control” on page 467, “LU name assignment user exit” on page 495, and “LU mapping with multilevel security active” on page 503.
- TN3270 network management, see “Connection monitoring mapping statement” on page 487, “Tracing” on page 528, and “Step 3: Configure the SNMP subagents” on page 1073.
- TN3270 IP address range configuration, see “Mapping Objects to Client Identifiers” on page 467.
- TN3270 takeover enhancements, see “Connection and session takeover” on page 506.
- Multilevel security for FTP, see “Security considerations for the FTP server” on page 544.
- Autoconfigure target library for FTP load module transfer, see “Specifying attributes for new MVS data sets” on page 552.
- FTP code page conversion, see “FTP code page conversion” on page 555.
- FTP TLS support enhancements, see “Customizing Transport Layer Security (TLS) and Kerberos security” on page 558.
- Stopping the DHCP server, see “Setting up a DHCP network” on page 682.
- Policy API (PAPI), and policy performance collection, see “Policy components overview” on page 717, “QoS-specific Policy Agent functions” on page 754, and “Policy performance collection configuration” on page 757.
- Network SLA MIB enhancements, see:
  - “Network SLAPM MIBs” on page 718.
  - “Network service level agreement performance monitor MIB subagents” on page 773.
  - “Verification” on page 776.
  - “Provide trap destination information” on page 1060.
  - “Step 3: Configure the SNMP subagents” on page 1073.
- Intrusion Detection Services enhancements (ibm-idsFloodAttackActionsAuxClass auxiliary class and interface flood detection), see “Overview of the object classes” on page 731 and “Attack policies” on page 793.
- IPMAILERNAME statement, see “Summary of SMTP configuration statements” on page 1114.
- The following areas contain new information pertaining to IPv6 support:
  - syslogd, see “Configuring the syslog daemon (syslogd)” on page 175.
  - OMPROUTE (RIP), see Chapter 6, “Routing,” on page 217.
  - Enterprise Extender, see “Configuring static VIPAs for Enterprise Extender” on page 303.
  - Dynamic XCF, see Chapter 8, “TCP/IP in a sysplex,” on page 363.
  - Policy, see Chapter 15, “Quality of service (QoS),” on page 751.
  - TN3270, see “TN3270E Telnet server” on page 441.
  - SNMP applications, see “Provide community-based security and notification destination information” on page 1058 and “Step 3: Configure the SNMP subagents” on page 1073.
  - sendmail, see “Configuring z/OS UNIX sendmail and popper” on page 1122.
  - SNTP, see Chapter 25, “SNTPD daemon,” on page 1147.

– TSO rexec, rsh, and associated MVS daemons, see “Configuring the TSO Remote Execution server” on page 1151.

This document contains terminology, maintenance, and editorial changes. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change.

Starting with z/OS V1R4, you may notice changes in the style and structure of some content in this document–for example, headings that use uppercase for the first letter of initial words only, and procedures that have a different look and format. The changes are ongoing improvements to the consistency and retrievability of information in our documents.

Part 1. Base TCP/IP system

Chapter 1. Overview of z/OS Communications Server

z/OS Communications Server provides a set of communications protocols that support peer-to-peer connectivity functions for both local and wide-area networks, including the most popular wide-area network, the Internet. z/OS Communications Server also provides performance enhancements that can benefit a variety of TCP/IP applications.

What is z/OS Communications Server?

z/OS Communications Server provides both SNA and TCP/IP networking protocols for z/OS. The SNA protocols are provided by VTAM and include Subarea, Advanced Peer-to-Peer Networking®, and High Performance Routing protocols. For more information on z/OS Communications Server SNA protocols, refer to *z/OS Communications Server: SNA Network Implementation Guide*.

Figure 1 on page 4 shows the z/OS V1R7 Communications Server TCP/IP protocol suite (also called stack), whose functions include associated applications, transport- and network-protocol layers, and connectivity and gateway functions. z/OS V1R7 Communications Server contains IPv6 support. Refer to *z/OS Communications Server: IPv6 Network and Application Design Guide* for more detailed information.

The z/OS V1R7 Communications Server protocol suite supports two TCP/IP environments:

- A native MVS environment in which users can exploit the popular TCP/IP protocols in MVS application environments such as batch jobs, started tasks, TSO, CICS® applications, and IMS™ applications.
- A z/OS UNIX System Services environment that lets you create and use applications that conform to the POSIX or XPG4 standard (a UNIX specification).

Note: z/OS Communications Server exploits z/OS UNIX services even for traditional MVS environments and applications. Prior to utilizing TCP/IP services, therefore, a full-function mode z/OS UNIX environment—including a Data Facility Storage Management Subsystem (DFSMSdfp™), a Hierarchical File System (HFS), and a security product (such as Resource Access Control Facility, or RACF®)—needs to be defined and active before z/OS Communications Server can be started successfully.

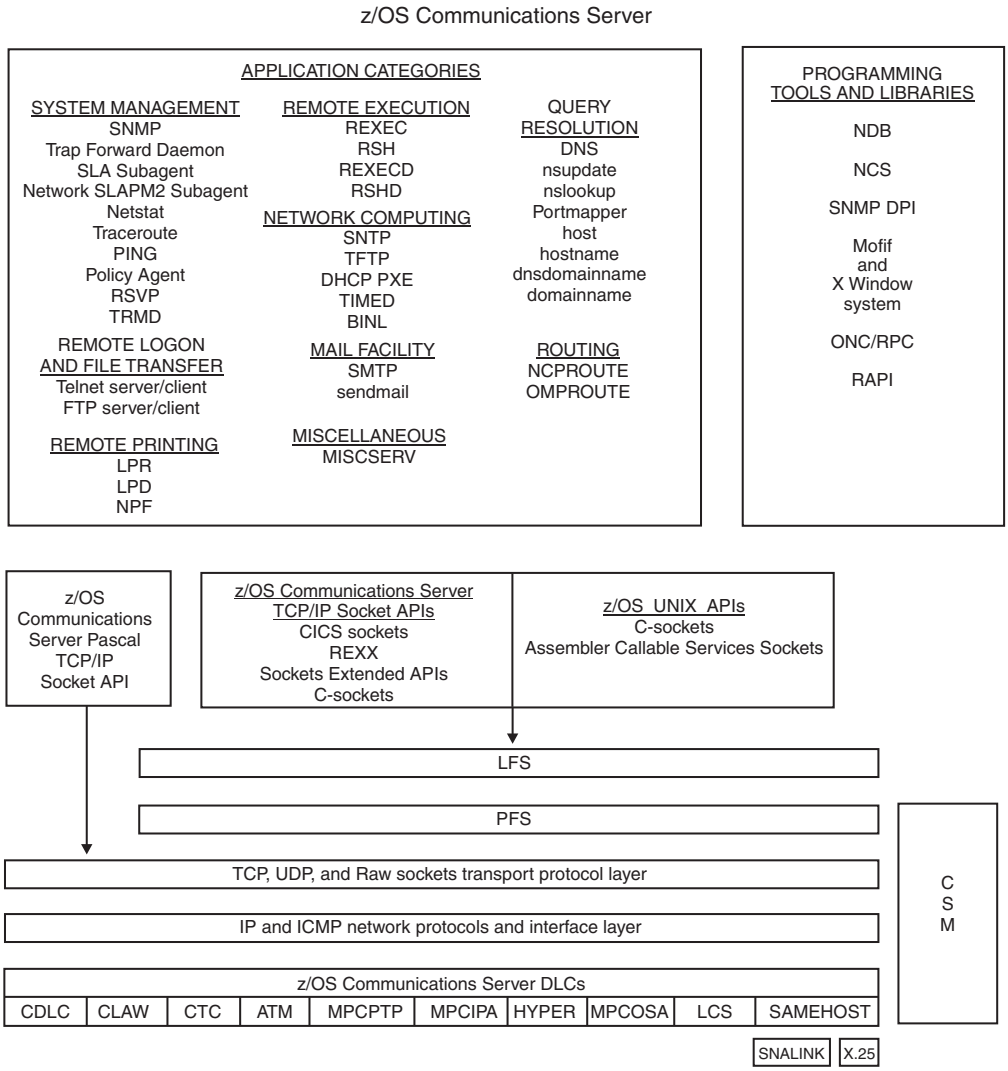


Figure 1. z/OS V1R7 Communications Server TCP/IP protocol suite

The application categories in Figure 1 do not list every application of the TCP/IP protocol suite.

z/OS V1R7 Communications Server TCP/IP protocol-suite functions include the following categories:

- "TCP/IP protocol stack"
- "Connectivity and gateway functions" on page 5
- "Network protocol layer" on page 6
- "Transport layer" on page 7
- "File systems" on page 7
- "Application Programming Interfaces (APIs)" on page 8

The following sections describe each of these functional categories.

TCP/IP protocol stack

The Transmission Control Protocol (TCP) and the Internet Protocol (IP) refer to a non-proprietary protocol suite that enables different packet-switched networks to function as a single entity regardless of underlying network topology.

z/OS V1R7 Communications Server provides robustness and high performance with the following features:

- A fully multiprocessor capable stack
- Exploitation of MVS Reliability, Availability, and Serviceability (RAS) services
- Exploitation of the z/OS architecture to optimize performance, CPU utilization, and throughput
- Exploitation of the z/OS Sysplex functions to maximize availability and scalability of TCP/IP workloads

In addition, z/OS V1R7 Communications Server design includes a tightly integrated storage and I/O model. I/O is provided by multipath channel (MPC) for network communication; storage management is provided by Communications Storage Manager (CSM). MPC and CSM are discussed in the following sections.

Multipath channel (MPC) I/O process

The term multipath channel (MPC) describes all z/OS Communications Server I/O driver support. There are specific I/O drivers under this support that are also referred to as MPC (such as MPCPTP).

MPC is a component of the I/O process model. MPC handles protocol headers and data separately and executes multiple I/O dispatchable units of work. MPC provides support for all devices supported by z/OS V1R7 Communications Server. The MPC I/O process and the CSM facilities that TCP/IP exploits are part of the VTAM component of z/OS V1R7 Communications Server. As a result, VTAM must be configured and active when starting devices on the TCP/IP stack.

Communications Storage Management (CSM)

The CSM facility is used by authorized programs to manage subsystem storage pools. CSM reduces data moves by providing a flat storage model that is accessible at multiple layers of the process model and across MVS address space boundaries. In addition, CSM provides the following technical advantages:

- MVS cellpool-like services
- Automatic handling of the contraction and expansion of storage resources
- Handling of different types and sizes of storage requests (for example, pageable and fixed)

Connectivity and gateway functions

TCP/IP connectivity and gateway functions handle the physical interfaces and the routing of IP data packets called datagrams. The following communication interfaces are supported by z/OS V1R7 Communications Server:

ATM Enables TCP/IP to send data to an asynchronous transfer mode (ATM) network using an OSA-2 or OSA-Express ATM adapter over an ATM virtual circuit.

CDLC Provides connectivity to a Network Control Program (NCP) through 3745/3746 network Front End Processor (FEP) controllers.

CLAW Provides access from IBM RS/6000® workstations directly to TCP/IP hosts over a channel. The CLAW (common link access to workstation) interface can also be used to provide connectivity to the original equipment manufacturer (OEM), such as the Cisco Channel Interface Processor (CIP).

CTC Provides access to TCP/IP hosts by way of a channel-to-channel (CTC) connection established over a zSeries ESCON® channel.

HYPERchannel
Provides access to TCP/IP hosts by way of HYPERchannel series A devices and series DX devices that function as series A devices.

LCS Provides access to TCP/IP hosts using the following devices:

- An IBM 3172 Interconnect Controller, which connects to a token ring, an Ethernet, or FDDI local area network
- An IBM 8232
- An IBM 2216 Multiaccess Connector Model 400
- An FDDI, Ethernet, Fast Ethernet, or Token Ring OSA or OSA-2 adapter
- An ATM OSA-2 in LAN emulation mode

MPCIPA
Provides access to TCP/IP hosts using the following:

- OSA-Express adapter with the Queued Direct I/O (QDIO) interface. OSA-Express supports the QDIO architecture for Gigabit Ethernet, Fast Ethernet, ATM LAN-Emulation attachment, and Token Ring. OSA-Express also supports IPv6 for Gigabit Ethernet and Fast Ethernet.
- HiperSockets using the Internal Queued Direct I/O (iQDIO). HiperSockets provides high-speed, low-latency IP message passing between Logical Partitions (LPARs) within a single IBM @server z/Series z800, z900, or z990 server.

MPCOSA
Provides access to TCP/IP hosts by way of OSA-2 configured in HPDT MPC mode for Fast Ethernet or FDDI.

MPCPTP
Provides access to TCP/IP hosts through multipath channel point-to-point (MPCPTP) links. MPCPTP supports IPv4 and IPv6 protocols. MPCPTP can be used in two ways to provide direct connectivity to other mainframe hosts running z/OS Communications Server:

- By using a set of two or more zSeries channels
- By configuring to utilize XCF services, if the zSeries hosts are part of the same sysplex

MPCPTP can also be used to provide the following connectivity options:

- Direct communication between two z/OS Communications Server TCP/IP Services protocol stacks running on the same MVS host without requiring any network attachments
- Connectivity to network attachments such as the IBM 2216 Multiaccess Controller Model 400 or the IBM RS/6000

SAMEHOST
Provides connectivity between the TCP/IP address space and other program address spaces within the same MVS host. The SAMEHOST Data Link Control (DLC) provides support for the SNA backbone network (SNALINK LU0 and SNALINK LU6.2) and the X.25 network.

Network protocol layer

The network protocol layer provides the support for the IP protocol. All TCP and User Datagram Protocol (UDP) data goes through the IP layer when entering and leaving the host. TCP and UDP will use the IPv4 routing layer or the IPv6 routing layer.

The network layer also provides support for the Internet Control Message Protocol (ICMP) and ICMPv6. This is used by the IP layer to exchange information and error messages with IP layers on other hosts and routers. ICMP is used for the IPv4 protocol and ICMPv6 is used for the IPv6 protocol.

Transport layer

The transport layer provides the support for the TCP, UDP, and RAW protocols. All three protocols use IPv4 or IPv6 as the network layer. The TCP protocol provides a connection-oriented, reliable transport layer, whereas UDP provides a simpler, connectionless and unreliable transport layer. The RAW transport layer provides for a more direct interface to the IP layer, which is primarily used by system-management type applications.

File systems

The file system layer provides the main interface between the application programming interfaces (APIs) and the transport layers. The first component of the file system layer is the z/OS UNIX logical file system (LFS). The LFS provides the API layer with a common interface to access files and sockets. In a POSIX-compliant environment, applications can access both files and sockets in a similar fashion. For example, both files and sockets are represented by a 16-bit integer referred to as a descriptor. Common functions can be used to access both file and socket resources.

The layer beneath the LFS is the physical file system (PFS). The PFS layer is where the distinction between files, sockets, and other resources is made. Based on the resource type, the LFS passes the incoming function requests to one of the following physical file systems:

Hierarchical file system PFS
This PFS handles requests related to resources in the z/OS UNIX hierarchical file system.

zFS The zSeries File System (zFS) is a UNIX file system that can be used in addition to HFS, TFS, and other z/OS UNIX file systems. It contains files and directories that can be accessed by APIs.

AF_UNIX PFS
If a descriptor represents an AF_UNIX socket, the request is handled by this PFS. AF_UNIX sockets, often referred to as local sockets, enable two z/OS UNIX applications within the same system to communicate with each other.

AF_INET PFS
If a descriptor represents an AF_INET socket, it is handled by this PFS. AF_INET sockets are used on an IPv4 TCP/IP-based network and are commonly referred to as network sockets. If an IPv4 application wants to communicate with another program on a different TCP/IP host on an attached IP network, the application opens an AF_INET socket for that purpose.

AF_INET6 PFS
If a descriptor represents an AF_INET6 socket, it is handled by this PFS. AF_INET6 sockets are used on an IPv6 TCP/IP-based network and are commonly referred to as network sockets. If an IPv6 application wants to communicate with another program on a different TCP/IP host on an attached IP network, the application opens an AF_INET6 socket for that purpose.

From a TCP/IP perspective, the AF_INET and the AF_INET6 PFS are of main interest. TCP/IP is enabled for IPv6 by defining an AF_INET6 PFS. Defining the file systems is the responsibility of the installation's z/OS UNIX programmer. The definitions are found in the BPXPRMxx member of SYS1.PARMLIB.

For information about defining AF_INET and AF_INET6 physical file systems, and about customizing BPXPRMxx for INET and CINET systems, refer to *z/OS UNIX System Services Planning*.

The AF_INET and the AF_INET6 PFS can be configured in two ways:

Integrated sockets PFS
The integrated sockets PFS can support the AF_INET PFS alone or AF_INET and AF_INET6 PFS together, but not AF_INET6 PFS alone.

Common INET PFS
This configuration is commonly referred to as the C_INET PFS configuration. It enables multiple AF_INET and AF_INET6 transport providers to be configured and active concurrently. Applications using the z/OS UNIX APIs do not know that multiple transport providers exist. For example, both the AnyNet® and TCP/IP Services components of z/OS V1R7 Communications Server can be configured at the same time. The AnyNet AF_INET PFS could provide access to an SNA network, while the TCP/IP AF_INET PFS could provide access to the TCP/IP network. The C_INET PFS is responsible for selecting the PFS over which to flow the request, based on the IP routing information from each of the AF_INET providers.

Under this configuration, it is also possible for TCP/IP application servers using the z/OS UNIX socket APIs to field incoming client requests from all AF_INET transport providers without knowing the particular transport provider.

Application Programming Interfaces (APIs)

This section provides a short description of each of the programming interfaces that can be used to interface with the TCP/IP Services protocol stack provided by z/OS Communications Server. All of the APIs, with the exception of the PASCAL API, interface with the LFS layer previously described.

The APIs are divided into the following two categories:

- TCP/IP socket APIs provided by z/OS V1R7 Communications Server
- z/OS UNIX APIs

TCP/IP socket APIs provided by z/OS V1R7 Communications Server

z/OS V1R7 Communications Server provides several APIs to access TCP/IP sockets. These APIs can be used in either or both integrated and common INET PFS configurations. In a common INET PFS configuration, however, they function differently from z/OS UNIX APIs. In this type of configuration, the z/OS Communications Server APIs always bind to a single PFS transport provider, and the transport provider must be the TCP/IP stack provided by z/OS V1R7 Communications Server.

The following TCP/IP socket APIs are included in z/OS V1R7 Communications Server:

Pascal API

The Pascal application programming interface enables you to develop TCP/IP applications in Pascal language. Supported environments are normal MVS address spaces. The Pascal programming interface is based on Pascal procedures and functions that implement conceptually the same functions as the C socket interface. The Pascal routines, however, have different names than the C socket calls. Unlike the other APIs, the Pascal API does not interface directly with the LFS. It uses an internal interface to communicate with the TCP/IP protocol stack.

Pascal API only supports AF_INET.

CICS sockets

The CICS socket interface enables you to write CICS applications that act as clients or servers in a TCP/IP-based network. Applications can be written in C language, using the C sockets programming, or they can be written in COBOL, PL/I or assembler, using the Sockets Extended programming interface.

CICS sockets only support AF_INET.

C sockets

The C sockets interface supports socket function calls that can be invoked from C programs. However, note that for C application development, IBM recommends the use of the UNIX C sockets interface. These programs can be ported between MVS and most UNIX environments relatively easily if the program does not use any other MVS specific services.

C sockets only support AF_INET.

IMS sockets

The Information Management System (IMS) IPv4 socket interface supports client/server applications in which one part of the application executes on a TCP/IP-connected host and the other part executes as an IMS application program.

The IMS sockets API supports AF_INET.

Sockets Extended macro API

The Sockets Extended macro API is a generalized assembler macro-based interface to sockets programming. It includes extensions to the socket programming interface, such as support for asynchronous processing on most sockets function calls.

The Sockets Extended macro API supports AF_INET and AF_INET6.

Sockets Extended Call Instruction API

The Sockets Extended Call Instruction API is a generalized call-based, high-level language interface to sockets programming. The functions implemented in this call interface resemble the C-sockets implementation, with some extensions similar to the sockets extended macro interface.

The Sockets Extended Call Instruction API supports AF_INET and AF_INET6.

REXX sockets

The REXX sockets programming interface implements facilities for socket communication directly from REXX programs by using an address rxsocket function. REXX socket programs can execute in TSO, online, or batch.

The REXX sockets programming interface supports AF_INET and AF_INET6.

Refer to *z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference* for complete documentation of the TCP/IP Services APIs.

z/OS UNIX APIs

The following APIs are provided by the z/OS UNIX element of z/OS and are supported by the TCP/IP stack in z/OS V1R7 Communications Server:

z/OS UNIX C sockets

z/OS UNIX C sockets is used in the z/OS UNIX environment. It is the z/OS UNIX version of the native MVS C sockets programming interface. Programmers use this API to create applications that conform to the POSIX or XPG4 standard (a UNIX specification). Applications built with z/OS UNIX C sockets can be ported to and from platforms that support these standards.

The z/OS UNIX C sockets support AF_INET and AF_INET6.

z/OS UNIX assembler callable services

z/OS UNIX assembler callable services is a generalized call-based, high-level language interface to z/OS UNIX sockets programming. The functions implemented in this call interface resemble the z/OS UNIX C sockets implementation, with some extensions similar to the sockets extended macro interface.

The z/OS UNIX assembler callable services support AF_INET and AF_INET6.

Refer to *z/OS XL C/C++ Run-Time Library Reference* for complete documentation of the z/OS UNIX C sockets APIs and refer to *z/OS UNIX System Services Programming: Assembler Callable Services Reference* for information about z/OS UNIX assembler callable services.

Chapter 2. Configuration overview

The objective of this chapter is to help you be better prepared for installation related activities. It is important to understand the terms, relationships, and dependencies presented in this chapter as a prerequisite to installation and customization. For additional background information on TCP/IP, refer to *TCP/IP Tutorial and Technical Overview*, GG24-3376-06, ISBN 0738421650.

IPv6 support

In this document, there might be no statement of support for a particular function in an IPv6 network. To determine whether a given function is applicable to IPv6, refer to the IPv6 support tables in *z/OS Communications Server: IPv6 Network and Application Design Guide*.

z/OS msys for Setup and Wizard

Wizard

IBM provides a Web-based wizard called the z/OS IP Configuration Wizard. Use it at any time to configure a single stack, with simple instances of OMPROUTE, FTP, and TN3270 servers, and with all device types, including static VIPA. If you finish using the wizard and complete the tasks defined in the output checklist, you will be ready to use z/OS TCP/IP to communicate with other hosts in your network. The wizard can be found at: http://www.ibm.com/eserver/zseries/zos/wizards/.

z/OS msys for Setup

A z/OS system is controlled using a multitude of settings, such as parmlib members, /etc files for UNIX System Services, or the RACF database, which are governed by different access methods. There is not a consistent representation of the configuration data; rather system administrators must keep track of various configuration data sets with varying semantics and syntax. While this renders z/OS systems highly flexible, at the same time it makes them complex and laborious to maintain.

Managed System Infrastructure for Setup (Msys for Setup) addresses these difficulties by establishing a central directory for product configuration data and a single interface to this directory. System administrators using msys will configure z/OS by way of GUI panels presented by the msys Windows NT® or Windows 2000 application. Multiple products can be configured through the same msys application. The configuration data will be collected and stored together as a common entity within an LDAP server. When directed by the system administrator, msys code running on z/OS will extract the configuration data stored in LDAP and produce the various configuration data sets on the z/OS system. This removes the system administrator from the details of the actual configuration statements, parameters and data set locations.

Msys for Setup consists of a collection of GUI panels that run on a Windows NT or Windows 2000 workstation and is connected through IP to an LDAP directory and to an MVS driving system. Msys for Setup provides the infrastructure for an msys exploiter to provide the user the following functionality:

- Refresh/Priming processing - parses a customer's existing configuration files and stores this configuration data, transformed into a tree structure, into an LDAP directory.
- Customization - reads the Refresh data stored in LDAP and populates the data fields on the GUI panels from this Refresh data. The data can then be changed by the system administrator as the panels are navigated. It is also likely that no Refresh step was done, and the panels are presented with either default data or no data. When customization changes are completed, the data is transformed into a tree structure and stored into the LDAP directory.
- Update Processing - after the configuration customization is complete, the administrator can make the 'Perform update' selection. Msys for Setup sends an FTP batch job to the mainframe driving system. Msys for Setup invokes a series of the msys exploiter's Java™ methods that ultimately result in updating msys-created configuration files or creating new configuration files from the new customized data that is extracted from the LDAP directory.
- Commit processing - after update processing is complete, the administrator can make the 'Commit update' selection. If update processing resulted in the creation or modification of temporary msys-created configuration files, those temporary files are copied into more permanent configuration files.

When using TCP/IP's msys for Setup, you will be prompted by GUI panels for customization information, which is then stored in an LDAP directory. Only a subset of TCP/IP's total configuration is supported using msys for Setup.

TCP/IP offers two different levels of service in msys:

- Customization and update processing only
- Refresh, customization, update, and commit processing

Included in the *customization and update processing only* level of service is support for all network devices, routing selections of OSPF or RIP, or use of default routers and static routes. The TN3270 server can be configured in a basic setup, or in advanced mode that supports nearly all TN3270 options. The TN3270 server can also be configured to start in its own address space or as part of the TCP/IP stack. Multiple FTP servers and FTP clients are completely configurable. Update processing creates TCP/IP configuration files in a PDS of your choice. The configuration files created are those typically referred to as TCPIP.DATA, PROFILE.TCPIP, FTP.DATA, and OMPROUTE.CONF, as well as TN3270 and PORTS. Supported configuration statements are listed below:

- For TCPIP.DATA

  DATASETPREFIX
  DOMAINORIGIN
  HOSTNAME
  NSINTERADDR
  TCPIPJOBNAME

  Defaults are used for all other configuration statements.
- For PROFILE.TCPIP

  ATMARPSV
  ATMLIS
  ATMPVC
  AUTOLOG / ENDAUTOLOG
  BEGINROUTES / ENDROUTES and ROUTE

DEVICE
HOME
LINK
START
TCPCONFIG RESTRICTLOWPORTS
TRANSLATE
UDPCONFIG RESTRICTLOWPORTS

Defaults are used for all other configuration statements.

- For FTP servers in FTP.DATA

ACCESSERRORMSGS
ADMINEMAILADDRESS
ANONYMOUS
ANONYMOUSFILEACCESS
ANONYMOUSFILETYPEJES
ANONYMOUSFILETYPESEQ
ANONYMOUSFILETYPESQL
ANONYMOUSFTPLOGGING
ANONYMOUSHFSDIRMODE
ANONYMOUSHFSFILEMODE
ANONYMOUSHFSINFO
ANONYMOUSLEVEL
ANONYMOUSLOGINMSG
ANONYMOUSMVSINFO
ASATRAMS
AUTOMOUNT
AUTORECALL
AUTOTAPEMOUNT
BANNER
BLKSIZE
CCXLATE
CHKPTINT
CIPHERSUITE
CONDDISP
CTRLCONN
DATACLASS
DATATIMEOUT
DB2
DB2PLAN
DCBDSN
DCONNTIME
DEBUGONSITE
DIRECTORY
DIRECTORYMODE
DUMPONSITE
EMAILADDRCHECK
ENCODING
EXTENSIONS
FILETYPE (only SEQ)
FTPKEEPALIVE
FTPLOGGING
HFSINFO
INACTIVE
ISPFSTATS
JESENTRYLIMIT

JESINTERFACELEVEL
JESLRECL
JESPUTGETTO
JESRECFM
KEYRING
LISTSUBDIR
LOGINMSG
LRECL
MBDATACONN
MGMTCLASS
MIGRATEVOL
MVSINFO
MVSURLKEY
PASSIVEDATAPORTS
PDSTYPE
PORTCOMMAND
PORTCOMMANDIPADDR
PORTCOMMANDPORT
PRIMARY
RDW
RECFM
REPLYSECURITYLEVEL
RETPD
SBDATACONN
SBSUB
SBSUBCHAR
SECONDARY
SECURE_CTRLCONN
SECURE_DATACONN
SECURE_FTP
SECURE_LOGIN
SECURE_PASSWORD
SMF
SMFAPPE
SMFDEL
SMFEXIT
SMFJES
SMFLOGN
SMFREN
SMFRETR
SMFSQL
SMFSTOR
SPACETYPE
SPREAD
SQLCOL
STARTDIRECTORY
STORCLASS
TLSTIMEOUT
TRAILINGBLANKS
TRUNCATE
UCOUNT
UMASK
UNITNAME
VCOUNT

VOLUME
WRAPRECORD
WRTAPEFASTIO

Note: The FTP server support is designed to create configuration files for the FTP server. Many of the statements also apply to the configuration of FTP clients. If you share the same FTP configuration file for both your server and client, be aware that not all parameters are supported for clients.

- For FTP clients in FTP.DATA

ASATRANS
AUTOMOUNT
AUTORECALL
AUTOTAPEMOUNT
BLKSIZE
CCONNTIME
CCTRANS
CHKPTINT
CHKPTPREFIX
CIPHERSUITE
CLIENTERRCODES
CONDDISP
CTRLCONN
DATACLASS
DATACTTIME
DB2
DB2PLAN
DCBDSN
DCONNTIME
DIRECTORY
DIRECTORYMODE
ENCODING
EPSV4
EXTENSIONS UTF8
FILETYPE
FTPKEEPALIVE
FWFRIENDLY
INACTTIME
ISPFSTATS
KEYRING
LISTSUBDIR
LOGCLIENTERR
LRECL
MBDATACONN
MGMTCLASS
MIGRATEVOL
MYOPENTIME
NETRCLEVEL
PDSTYPE
PRIMARY
RDW
RECFM
RESTGET
RETPD
SBDATACONN
SBSUB

SBSUBCHAR
SECONDARY
SECURE_CRTLCONN
SECURE_DATACONN
SECURE_FTP
SECURE_MECHANISM
SOCKSCONFIGFILE
SPACETYPE
SPREAD
SQLCOL
STORCLASS
TLSTIMEOUT
TRAILINGBLANKS
TRUNCATE
UCOUNT
UMASK
UNITNAME
VCOUNT
VOLUME
WRAPRECORD
WRTAPEFASTIO

- For OMPROUTE

  INTERFACE
  OSPF_INTERFACE
  RIP_INTERFACE

  Defaults are used for all other configuration statements.
- For TN3270

  ALLOWAPPL
  BEGINVTAM / ENDVTAM
  CLIENTAUTH
  CONNTYPE
  DEFAULTAPPL
  DEFAULTLUS
  DEFAULTLUSSPEC
  DEFAULTPRT
  DEFAULTPRTSPEC
  DESTIPGROUP
  DROPASSOCPRINTER
  ENCRYPTION
  EXPRESSLOGON
  HNGROUP
  INACTIVE
  IPGROUP
  KEYRING
  LINEMODEAPPL
  LINKGROUP
  LUGROUP
  LUMAP
  LUSESSIONPEND
  MSG07
  PORT / SECUREPORT
  PRTDEFAULTAPPL

PRTGROUP
PRTMAP
SCANINTERVAL / TIMEMARK
SMFINIT / SMFTERM
SNAEXT
TELNETDEVICE
TELNETGLOBALS / ENDTELNETGLOBALS
TELNETPARMS / ENDTELNETPARMS
TKOSPECLU
USERGROUP
USSTCP

Defaults are used for all other configuration statements.

- For PORTS

  PORT
  PORTRANGE

The *refresh, customization, update, and commit processing* level of service supports only port reservations. Refresh processing is optional, but useful if the administrator's port reservations are in their own configuration file. If no refresh is performed, customization begins with default port reservations. This TCP/IP service can accept port reservations requested by other msys for Setup services such as LDAP. For example, during customization of the LDAP msys for Setup service, the administrator might be asked which port LDAP should use. When the administrator is done customizing the LDAP service, the TCP/IP service would receive a request for that port. Update processing creates or modifies a temporary PORTS configuration file and reserves ports requested by other msys for Setup services. Commit processing copies this temporary file into a more permanent msys PORTS configuration file and can also modify the user's active PROFILE.TCPIP configuration file to use the msys PORTS file. Supported configuration statements are listed below:

- For PORTS

  PORT
  PORTRANGE

For more detailed information on z/OS msys for Setup, refer to *z/OS Managed System Infrastructure for Setup User's Guide*.

z/OS UNIX System Services (z/OS UNIX) concepts

Beginning with MVS/ESA™ Version 4.3 a new type of application program interface was added to the MVS platform with the intent of integrating a UNIX operating system into MVS. Both a C programming API and an interactive environment called the shell were defined to interoperate with UNIX-style files, called Hierarchical File Systems (HFS). Over time, other organizations developed approaches to working with UNIX on various platforms until an organization named X/Open documented standards of what to implement for UNIX interfaces in a series of guides published as the X/Open Portability Guides (XPG). X/Open now owns the term UNIX and certifies different implementations of UNIX according to the UNIX definitions contained in XPG 4.2.

z/OS UNIX System Services, or z/OS UNIX, is the certified z/OS or MVS implementation of UNIX as defined by X/Open in XPG 4.2. z/OS UNIX coexists

with traditional MVS functions and traditional MVS data set types (partitioned data sets, sequential data sets, and so on). It concurrently enables access to UNIX files and to UNIX utilities and commands by means of application programming interfaces (APIs) and the interactive SHELL environment. MVS offers two variants of the UNIX SHELL environment:

- The OMVS shell, much like a native UNIX environment
- The ISHELL, an ISPF interface with access to menu-driven command interfaces

With the APIs, programs can run in any environment including batch jobs, in jobs submitted by TSO/E interactive users, and in most other started tasks, or in any other MVS application task environment. The programs can request:

- Only MVS services
- Only z/OS UNIX services
- Both MVS and z/OS UNIX services

The shell interface is an execution environment analogous to TSO/E, with a programming language of shell commands analogous to Restructured eXtended eXecutor (REXX) language. The shell support consists of:

- Programs that are run interactively by shell users
- Shell commands and scripts that are run interactively by shell users
- Shell commands and scripts that are run as batch jobs

Prior to OS/390 V2R5, OS/390 UNIX required APPC/MVS for programs issuing the fork() or spawn() function. APPC/MVS is no longer required for this purpose. Forked and spawned address spaces are now implemented in z/OS for UNIX processing by the Work Load Manager (WLM) component of MVS.

For a fork(), the system copies one process, called the parent process, into a new process, called the child process, and places the child process in a new address space, the forked address space.

Spawn() also starts a new process in a new address space. Unlike a fork(), in a spawn() call the parent process specifies a name of a program to start the child process.

The types of processes can be:

- User processes, which are associated with a user
- Daemon processes, which perform continuous or periodic functions, such as a Web server

  Daemons are programs that are typically started when the operating system is initialized and remain active to perform standard services. Some programs that initialize processes for users are considered daemons, even though these daemons are not long-running processes. Examples of daemons provided by z/OS UNIX are *cron*, which starts applications at specific times, and *inetd*, which starts applications on demand.

A user or daemon process can have one or more threads. A thread is a single flow of control within a process. Application programmers create multiple threads to structure an application in independent sections that can run in parallel for more efficient use of system resources.

Overview of data sets and HFS files

Data set and *file* are comparable terms. If you are familiar with MVS, you probably use the term data set to describe a unit of data storage. If you are familiar with AIX or UNIX, you probably use the term file to describe a named set of records stored or processed as a unit. In the TCP/IP environment, in addition to the traditional MVS data set organizations (such as sequential, partitioned) the z/OS UNIX files are arranged in a Hierarchical File System (HFS) and are called HFS files.

Some data sets and HFS files have special importance because of their function. For example, certain data sets and HFS files are used when configuring the TCP/IP environment. Other data sets are used by the Telnet server (Telnet daemon) when performing specific communication functions. See Table 1 on page 30 for descriptions of the data sets and HFS files necessary for configuring the TCP/IP environment and the search orders used to find them. A search order can include both HFS files and data sets, and these data sets and HFS files will be collectively referred to as the *configuration files* in this section.

Note: Not all applications support HFS files.

Hierarchical File System concepts

File systems let you set up a file hierarchy that consists of the following:

- **Files**, which contain data or programs. A file containing a program object, shell script, or REXX program is called an *executable file*. Files are kept in directories.
- **Directories** that contain files, other directories, or both. Directories are arranged hierarchically, in a structure that resembles an upside down tree, with root directory at the top and the branches at the bottom. The **root** is the first directory for the file system at the peak of the tree and is designated by a slash (/).
- Named pipes, links, and other UNIX items, such as character special files like /dev/console that are used by applications like syslogd. Refer to *z/OS UNIX System Services Planning* for more information about UNIX items like character special files.

The term *file system* has all of the following meanings:

- A logical collection of files, directories, named pipes, links, and other UNIX items and metadata that are arranged in a hierarchy.
- A particular instance of a logical collection of these items that are arranged in a hierarchy. They might reside on local or remote disks or in computer memory.
- A program that is designed to provide the functions and data of one type of file system.

The context indicates which meaning is intended. Often more than one meaning is intended; this is an industry convention.

To the z/OS system, the file hierarchy is a collection of file systems. Additional instances of local or remote file systems might be mounted (logically connected) on directories of the root file system or of additional file systems.

Several types of file systems are supported by z/OS, including the following:

- zFS (zSeries File System)

  Each instance of zFS resides in a linear data set.

- HFS (hierarchical file system)

  Each instance of HFS resides in an HFS data set.
- TFS (temporary file system)

  Each instance of TFS resides in computer memory.
- NFS (Network File System)

  NFS server provides access to file systems that reside on other computers.

For most application programs, these types of file systems are interchangeable. The root file system is the first file system that is mounted. Subsequent file systems can be logically mounted on a directory within the root file system or on a directory within any mounted file system.

Except for the direction of the slashes, the UNIX file system concept is similar to a PC Disk Operating System (DOS) file system.

References to installation data sets

This document refers to Communications Server installation data sets by their default SMP/E distribution library name. Your installation might, however, have different names for these data sets where allowed by SMP/E, your installation personnel, or administration staff. For instance, this document refers to sample members of the *hlq*.SEZAINST library as SEZAINST(*member*), where the *hlq* value is the high-level qualifier specified during TCP/IP installation. Your installation can choose a data set name of SYS1.SEZAINST, CS390.SEZAINST, or other high level qualifiers for the data set name.

Understanding resolvers

The resolver acts on behalf of programs as a client that accesses name servers for name-to-address or address-to-name resolution. The resolver can also be used to provide protocol and services information. To resolve the query for the requesting program, the resolver can access available name servers, use local definitions (for example, /etc/resolv.conf, /etc/hosts, /etc/ipnodes, HOSTS.SITEINFO, HOSTS.ADDRINFO, or ETC.IPNODES), or use a combination of both. How and if the resolver uses name servers is controlled by TCPIP.DATA statements (resolver directives).

The resolver address space must be started before any application or TCP/IP stack resolver calls can occur. When the resolver address space starts, it reads an optional resolver setup data set pointed to by the SETUP DD card in the resolver JCL procedure. This resolver setup data set enables the following capabilities:

- Specification of a global TCPIP.DATA file that contains settings for the MVS image. The GLOBALTCPIPDATA setup statement identifies the file.

  The global TCPIP.DATA file is being provided to allow the administrator to retain control of which resolver statements are used for name resolution, and to eliminate the complexity of attempting to merge resolver statements from multiple files in a predictable and useful manner.

  This global TCPIP.DATA file, when specified, will become the first TCPIP.DATA file read regardless of the Socket API library being used. Any parameters found in this file will be global settings for this MVS image. If a global TCPIP.DATA file has been specified then all resolver statements will *only* be obtained from this file. The search continues beyond the file specified by GLOBALTCPIPDATA, but any of the resolver statements specified in files lower in the search order will be ignored.

Resolver statements are those required by the resolver to process queries. Resolver TCPIP.DATA statements are:

- DomainOrigin/Domain
- NSInterAddr/NameServer
- NSPortAddr
- ResolveVia
- ResolverTimeOut
- ResolverUDPRetries
- Search
- SortList

Other statements not specified in the global TCPIP.DATA file can still be located in one of the TCPIP.DATA files in the search order for each socket API type. For example, if TCPIPJOBNAME is not specified in the global TCPIP.DATA file, the resolver library will locate the next available file in the search order (the search will depend on the socket API being used) and attempt to find the TCPIPJOBNAME there. Note that once a file is found beyond the global TCPIP.DATA file the searching stops. For example, if the TCPIP.DATA file was found by way of the SYSTCPD DD card and no TCPIPJOBNAME was specified in this file, then the normal defaults for TCPIPJOBNAME are applied (for example, TCPIP if the native MVS API search order is used, or a null character if the z/OS UNIX API search order is used). In effect, you can concatenate up to two TCPIP.DATA files with this approach. Note that the search order for the local hosts table (HOSTS.xxxxINFO, ETC.IPNODES, /etc/hosts, or /etc/ipnodes) remains the same. Depending on the application environment, either the native MVS or z/OS UNIX search order will be in effect.

The ability to specify a global TCPIP.DATA file has several advantages. The administrator can decide on which options are global for the installation and which can be specified on an application basis. For example, it is anticipated that most administrators will prefer to control the resolver statements in TCPIP.DATA at a global level. However, it is quite unlikely that they will want a global setting of the TRACE RESOLVER option. This option would typically not be specified on a global TCPIP.DATA file, rather it would get picked up from the first file found in the search order after the global TCPIP.DATA file. This would allow application programmers to continue to turn on the option. Another advantage of this approach is that the administrator may not be aware of all the private TCPIP.DATA files that may be in use on their systems. This approach allows them to implement global options gradually versus an all or nothing approach.

Also, note that this approach lends itself to a multistack (CINET) environment. The administrator can still set up a global TCPIP.DATA file with the global options for this MVS image and omit specifying the TCPIPJOBNAME keyword. The TCPIPJOBNAME keyword would then be located using the appropriate search order. However, using the global TCPIP.DATA file with CINET requires that the resolver TCPIP.DATA statements are able to be used by all stacks. For example, the IP addresses specified by the NameServer statement must be accessible from all stacks. If they are not, then the GLOBALTCPIPDATA file should not be used and you should continue with multiple TCPIP.DATA data sets.

- Support for user specified default TCPIP.DATA file. The DEFAULTTCPIPDATA setup statement identifies the file.

The user can specify the file to be used as the final location in the search order instead of TCPIP.TCPIP.DATA. This can be used as the replacement for the TCP/IP V3R2 EZAPPRFX sample installation job.

- Specification of the local host file search order for IPv4 and IPv6 name queries. The COMMONSEARCH setup statement identifies that a common local host file search order is to be used for both IPv4 and IPv6 name queries in the native MVS and z/OS UNIX environments. The NOCOMMONSEARCH setup statement identifies that a different local host file search order is to be used for IPv4 and IPv6 name queries in the MVS and UNIX environments.
- Specification of a local host file that contains hard-coded IP addresses and host names that can be used globally. The GLOBALIPNODES setup statement identifies this file.
- Support for a user-specified default local host file. The DEFAULTIPNODES setup statement identifies this file.

Follow the steps in "Setting up a resolver address space" on page 23 to take advantage of the resolver capabilities described above.

If the setup information is not provided, the resolver uses the applicable native MVS or z/OS UNIX search order without any GLOBALTCPIPDATA, DEFAULTTCPIPDATA, GLOBALIPNODES, DEFAULTIPNODES, or COMMONSEARCH information.

To process its definitions, the resolver might need to allocate data sets or HFS files. For those definitions, such as GLOBALTCPIPDATA, DEFAULTTCPIPDATA, /etc/hosts, HOSTS.SITEINFO, and HOSTS.ADDRINFO, allocation messages appear in the JES joblog. For long-running applications that heavily use resolver services, such as Tivoli® NetView®, consider using a started job with MSGLEVEL=(1,0) to eliminate all allocation messages. Note that this could also eliminate allocation messages that might be useful in analyzing a problem. For information on started jobs and the MSGLEVEL parameter, refer to *z/OS MVS JCL Reference*.

Application programs using the gethostbyaddr and gethostbyname resolver calls from the following IBM APIs result in using the z/OS Communications Server resolver.

- z/OS Language Environment® C/C++ API
- z/OS UNIX assembler callable services
- z/OS Communications Server C/C++ API
- z/OS Communications Server Callable and Macro API
- z/OS Communications Server REXX API
- z/OS Communications Server PASCAL API

Application programs using the getaddrinfo, getnameinfo, and freeaddrinfo resolver calls from the following IBM APIs result in using the z/OS Communications Server resolver.

- z/OS Language Environment C/C++ API
- z/OS UNIX assembler callable services
- z/OS Communications Server Callable and Macro API
- z/OS Communications Server REXX API

Application programs using the sethostent, gethostent, and endhostent resolver calls from the following IBM APIs result in using the z/OS Communications Server resolver.

- z/OS Language Environment C/C++ API
- z/OS Communications Server C/C++ API

The z/OS Communications Server SMTP server, BIND 9 DNS and DNS V9 utilities (dig, nslookup and nsupdate) provide their own unique resolver services. When their resolver initializes it will use GLOBALTCPIPDATA and DEFAULTTCPIPDATA information.

Note that the SMTP resolver only uses the first value of the SEARCH TCPIP.DATA statement when resolving host names.

Setting up a resolver address space

There are two ways to start the resolver address space:

- z/OS UNIX initialization will attempt to start the resolver unless explicitly instructed not to. Using z/OS UNIX is the recommended method because it ensures that the resolver is available before any applications can make a resolution request.

  Use the BPXPRMxx statement, RESOLVER_PROC, to specify the procedure name, if any, to be used to start the resolver address space.

  If the RESOLVER_PROC statement is not in the BPXPRMxx parmlib member or is specified with a procedure name of DEFAULT, z/OS UNIX will start a resolver address space with the assigned name of RESOLVER. The resolver will use the applicable search order for finding TCPIP.DATA statements, but without a GLOBALTCPIPDATA or a DEFAULTTCPIPDATA specification.

  If the RESOLVER_PROC statement has been used to specify a start procedure name, then the following are true:

  – For the procedure to be found, it must reside in a data set that is specified by the MSTJCLxx parmlib member's IEFPDSI DD card specification. For MSTJCL considerations, refer to *z/OS MVS Initialization and Tuning Reference*.

  – The procedure must not contain any DD cards that specify SYSOUT=*.

  When z/OS UNIX starts the resolver, it is started so that the resolver does not require JES (that is, SUB=MSTR is used).

  – For SUB=MSTR considerations, refer to *z/OS MVS JCL Reference*.

  – Because z/OS UNIX does not receive any error indication when it tries to start the address space, it will issue an informational message containing the name of the procedure it has started. The message is:

    `BPXF224I THE RESOLVER_PROC, procname, IS BEING STARTED.`

    Rule: If the RESOLVER_PROC statement is not present or is specified with a procedure name of DEFAULT, the *procname* value will be RESOLVER even though no start procedure was used. If you want to use a start procedure named RESOLVER, you must add a RESOLVER_PROC(RESOLVER) statement to your BPXPRMxx parmlib member.

  – If the start procedure is not found or has a JCL error in it, the usual z/OS START command error messages will be issued.

  – If the address space cannot be started, z/OS UNIX initialization continues.

  For more detailed information, refer to *z/OS UNIX System Services Planning*.

- An installation can use its automation tools to start the resolver by use of the MVS START operator command. If this approach to starting the resolver is used, care should be taken to ensure that no applications that need resolver services

(for example, INETD) are started before the resolver address space is initialized. This may mean removing the starting of INETD from the z/OS UNIX /etc/rc file and starting INETD with automation after the resolver has initialized.

Resolver customization

If an installation wants to make use of any resolver setup statement facilities, the following steps will be required. If the facilities are not required, no customization is required and the search order for TCPIP.DATA will be determined by the API being used.

- Create a resolver start procedure

  The procedure requires a //SETUP DD JCL statement that points to a resolver setup file. The z/OS Communications Server provided sample procedure below can be found as member EZBREPRC (alias RESOPROC) in SEZAINST:

```
//RESOLVER PROC PARMS='CTRACE(CTIRES00)'
//*
//*   IBM Communications Server for z/OS
//*   SMP/E distribution name: EZBREPRC
//*
//*   5694-A01 (C) Copyright IBM Corp. 2001, 2005.
//*   Licensed Materials - Property of IBM
//*
//*   Function: Start Resolver
//*
//EZBREINI EXEC PGM=EZBREINI,REGION=0M,TIME=1440,PARM=&PARMS
//*
//*   When the Resolver is started by UNIX System Services it is
//*     started with SUB=MSTR.
//*   This means that JES services are not available to the Resolver
//*     address space. Therefore, no DD cards with SYSOUT can be used.
//*     See the MVS JCL Reference manual for SUB=MSTR considerations in
//*     section "Running a Started Task Under the Master Subsystem".
//*   This Resolver start procedure will need to reside in a data
//*     set that is specified by the MSTJCLxx PARMLIB member's
//*     IEFPDSI DD card specification. If not, the procedure will
//*     not be found and the Resolver will not start.
//*     See the MVS Initialization and Tuning Reference manual for
//*     MSTJCL considerations in section "Understanding the Master
//*     Scheduler Job Control Language"
//*
//*   SETUP contains Resolver setup parameters.
//*     See the section on "Understanding Resolvers" in the
//*     IP Configuration Guide for more information. A sample of
//*     Resolver setup parameters is included in member RESSETUP
//*     of the SEZAINST data set.
//*
//*SETUP   DD   DSN=TCPIP.TCPPARMS(SETUPRES),DISP=SHR,FREE=CLOSE
//*SETUP   DD   DSN=TCPIP.SETUP.RESOLVER,DISP=SHR,FREE=CLOSE
//*SETUP   DD   PATH='/etc/setup.resolver',PATHOPTS=(ORDONLY)
```

- Create a resolver setup file (MVS data set or HFS file)

  The setup file defines the location of the global TCPIP.DATA file (MVS data set or HFS file) and the default TCPIP.DATA name (MVS data set or HFS file). The following statements are supported:

  – comments (; or #)
  – COMMONSEARCH
  – DEFAULTIPNODES
  – DEFAULTTCPIPDATA

– GLOBALIPNODES
– GLOBALTCPIPDATA
– NOCOMMONSEARCH

The z/OS Communications Server provided sample setup file below can be found as member EZBRECNF (alias RESSETUP) in SEZAINST:

```
;
;   IBM Communications Server for z/OS
;   SMP/E distribution name: EZBRECNF
;
;   5694-A01 (C) Copyright IBM Corp. 2002, 2005,
;   Licensed Materials - Property of IBM
;
;   Function: Sample Resolver setup file
;
;
;  The following statement defines the final search location for
;   TCPIP.DATA statements. It will replace TCPIP.TCPIP.DATA
;  It may be an MVS data set or HFS file.
;
DEFAULTTCPIPDATA('TCPIP.TCPIP.DATA')
;
#  The following statement defines the first search location for
#   TCPIP.DATA statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; GLOBALTCPIPDATA('TCPCS.SYS.TCPPARMS(GLOBAL)')
;
; GLOBALTCPIPDATA(/etc/tcpipglobal.data)
;
#  The following statement defines the first search location for
#   IPNODES statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; GLOBALIPNODES('TCPCS.SYS.TCPPARMS(IPNODES)')
;
; GLOBALIPNODES('TCPCS.ETC.IPNODES')
;
; GLOBALIPNODES(/etc/ipnodes)
;
#  The following statement defines the final search location for
#   IPNODES statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; DEFAULTIPNODES('TCPCS.SYS.TCPPARMS(IPNODES)')
;
; DEFAULTIPNODES('TCPCS.ETC.IPNODES')
;
; DEFAULTIPNODES(/etc/ipnodes)
;
#  The following statement defines if the common search order
#   should be used or not.
;
NOCOMMONSEARCH
;
; COMMONSEARCH
;
```

If the resolver setup file is an MVS data set it must be either sequential (PS) or partitioned (PO) organization, fixed (F) or fixed block format (FB), a logical record length (LRECL) between 80 and 256, and have any valid blocksize (BLKSIZE) for fixed block. If the setup file may need to be modified, a member of an MVS partitioned data set is recommended.

If the file is an HFS file, it can reside in any directory. The maximum length of line supported is 256 characters. If the line is greater than 256 it will be truncated to 256 and processed.

The user ID assigned to the resolver address space needs read access (through RACF or equivalent) to SYS1.PARMLIB, the resolver setup file, the global TCPIP.DATA file, the default TCPIP.DATA file, the global IPNODES file, and the default IPNODES file. Likewise, any user IDs or jobs using TCPIP facilities will need read access to the global TCPIP.DATA file, the default TCPIP.DATA file, the global IPNODES file, and the default IPNODES file. For example, for the MVS data set RACF UACC=READ and for the HFS file, permission bits of 644 (Owner can read and write, Group can read, Other can read) could be used. For an HFS file, an OMVS segment or the default OMVS segment must be configured for the resolver user ID and any user IDs or jobs using TCPIP facilities.

Errors detected by a security product (for example, RACF) or OPEN services can generate messages that indicate that the information cannot be accessed. For example, IEC141I 013-C0 is generated if a file does not have the correct permission bit settings to allow it to be read. Also, RACF message ICH408I is issued if no OMVS segment is defined or if insufficient authorization is granted to read a data set.

- Update the z/OS UNIX BPXPRMxx parmlib member

  The resolver start procedure name should be specified as the *procname* in the BPXPRMxx parmlib member's RESOLVER_PROC(*procname*) statement. If for some reason the recommended method of using z/OS UNIX to start the resolver is not desired, use the MVS START command to start the resolver address space.

Managing the resolver address space

A BPXPRMxx statement, RESOLVER_PROC, is used to specify the procedure name, if any, to be used to start the resolver address space. If the RESOLVER_PROC statement is not in the BPXPRMxx parmlib member or is specified with a procedure name of DEFAULT, z/OS UNIX will start a resolver address space with the assigned name of RESOLVER. This name is used with the following MVS system commands to manage the resolver address space:

- Start (S)
- Stop (P)

  Stopping and restarting of the resolver should only be used if a new level of the resolver code has been installed.
- Force
- Modify (F)

  The MODIFY command should be used to dynamically change resolver setup statements, update the resolver's usage of TCPIP.DATA statements, or update the resolver's usage of local host and services tables. Dynamic changes are not supported by the resolver provided by the SMTP server, BIND 9 DNS and DNS V9 utilities.

If the resolver has IBM-provided software service (PTFs) to be applied, do the following steps. You do not need to stop and restart the TCPIP started task or any

application programs. However, for the short duration of time that the resolver is not running, any resolver requests will fail.

1. Use SMPE to apply the PTFs.
2. If you have LLA running, refresh it with the following system operator command:

   ```
   MODIFY LLA,REFRESH
   ```
3. Stop the resolver:
   - If you have not customized the resolver, issue the following system operator command:

     ```
     STOP RESOLVER
     ```
   - If you have customized the resolver, issue the following system operator command:

     ```
     STOP procname
     ```

     where *procname* is the name of your PROCLIB member.
4. Restart the resolver:
   - If you have not customized the resolver, issue the following system operator command:

     ```
     START IEESYSAS.RESOLVER,PROG=EZBREINI,SUB=MSTR
     ```
   - If you have customized the resolver, issue the following system operator command:

     ```
     START procname  or  START procname,SUB=MSTR
     ```

     where *procname* is the name of the PROCLIB member you created.

If the resolver has terminated for any reason, to get the resolver operational again, restart the resolver as described above in step 4.

Restriction: The resolver address space provides PC-entered services that must be accessible to all address spaces, so a system LX is obtained. This causes the Address Space Identifier (ASID) associated with the resolver address space to be nonreusable when the address space is stopped or restarted. If the resolver address space is terminated enough times, all available ASIDs could be exhausted, preventing the creation of a new address space on the system. In this case, an IPL is required. For more information on tuning parameters for the maximum number of ASIDs on a system, refer to the MAXUSER parameter in *z/OS MVS Initialization and Tuning Reference*.

Refer to *z/OS Communications Server: IP System Administrator's Commands* for command details.

The following MVS System commands can be used to control and display the status of the resolver CTRACE facilities:

- Trace CT
- Display Trace

Refer to the *z/OS Communications Server: IP Diagnosis Guide* for CTRACE usage and control information.

Understanding search orders of configuration information

It is important to understand the search order for configuration files used by TCP/IP functions, and when you can override the default search order with environment variables, JCL, or other variables you provide. This knowledge allows you to accommodate your local data set and HFS file naming standards, and it is helpful to know the configuration data set or HFS file in use when diagnosing problems.

It is important to note that the z/OS Communications Server environment consists of the TCP/IP address space, z/OS Communications Server applications, and the TCP/IP MVS applications. The TCP/IP address space functions are also referred to as the *stack*. The z/OS Communications Server applications refer to those applications using the z/OS UNIX socket API. The TCP/IP MVS applications refer to those applications written to the MVS APIs (for example, C, Sockets-Extended, CICS, IMS, and REXX). The TCP/IP stack and both sets of applications have some common (or global) configuration files, but they also use configuration files that are different.

Another important point to note is that when a search order is applied for any configuration file, the search ends with the first file found. Therefore, unexpected results are possible if you place configuration information in a file that never gets found, either because other files exist earlier in the search order, or because the file is not included in the search order chosen by the application.

Configuration data set naming conventions

When searching for configuration files, you can explicitly tell TCP/IP where most configuration files are by using DD statements in the JCL procedures or by setting environment variables. Otherwise, you can let TCP/IP dynamically determine the location of the configuration files, based on search orders shown in Table 1 on page 30.

For example, in Table 1 on page 30, for the FTP server application, if the installation did not code the //SYSFTPD DD statement, the FTP server would search for *jobname*.FTP.DATA, then file /etc/ftp.data, then data set SYS1.TCPPARMS(FTPDATA), and finally *hlq*.FTP.DATA.

Dynamic data set allocation

TCP/IP makes extensive use of dynamically allocated data sets using the MVS dynamic data set allocation function to search for configuration files. Multiple versions of a configuration data set can exist, each having a different high-level qualifier or middle-level qualifier. The search order for any configuration file will determine which data set is found and used.

High-level qualifier: High-level qualifiers (HLQ) permit you to associate an application's configuration data set with a particular jobname or TSO user ID, or permit you to use a default configuration data set for the application. The possible high-level qualifiers are:

- *userid*

  Userid is the TSO user ID which invoked the application.
- *jobname*

  Jobname is the application's batch JCL jobname or the name of the application's started procedure.
- *hlq*

TCP/IP is distributed with a default high-level qualifier (HLQ) of *TCPIP*. To override the default HLQ used by dynamic data set allocation, specify the DATASETPREFIX statement in the TCPIP.DATA configuration file. For most configuration files, the DATASETPREFIX value is used as the high-level qualifier of the data set name in the last step in the search order. Note that the DATASETPREFIX value is not used as the high-level qualifier of the data set name used as the last step in the search order for the PROFILE.TCPIP and TCPIP.DATA configuration files.

Middle-level qualifiers: Multiple middle-level qualifiers (MLQ) permit the isolation of certain profile and translation table data sets. Two of the possible middle-level qualifiers are:

- Node name

  Node name is an MLQ used in the search order for finding the configuration file PROFILE.TCPIP. Node name is determined by the parameters specified during VMCF initialization. For further information on initializing VMCF, refer to *z/OS Program Directory*.

- Function name

  The TCP/IP implementation of national language support (NLS) and double-byte character set (DBCS) support requires the use of multiple translation tables. To facilitate the concurrent use of multiple languages and code pages, TCP/IP uses a middle-level qualifier to designate which server or client uses a particular translation table. STANDARD, the default MLQ, is available for use if a single translation table can be used by multiple servers or clients. The TCP/IP Telnet client and FTP provide a TRANSLATE parameter that permits you to specify your chosen MLQ to replace the function name for that invocation of the command. For example, SRVRFTP is used as an MLQ by the File Transfer Protocol server.

Following are some of the data sets that are only dynamically allocated by TCP/IP in a configuration file search order (you cannot specify them with DD statements in JCL):

```
ETC.PROTO               ETC.RPC
HOSTS.ADDRINFO          HOSTS.SITEINFO
SRVRFTP.TCPCHBIN        SRVRFTP.TCPHGBIN
SRVRFTP.TCPKJBIN        SRVRFTP.TCPSCBIN
SRVRFTP.TCPXLBIN        STANDARD.TCPCHBIN
STANDARD.TCPHGBIN       STANDARD.TCPKJBIN
STANDARD.TCPSCBIN       STANDARD.TCPXLBIN
```

For each of these data sets, the fully qualified name is established by using one of the following values as the data set HLQ:

- User ID or job name
- DATASETPREFIX value

Naming conventions for dynamically allocated data sets: A data set that you allocate explicitly (with a DD statement in JCL) can have any valid MVS data set name or HFS file name. A data set that you create for the purpose of being allocated dynamically by TCP/IP must use the following naming conventions.

Note: In the examples below, xxxx indicates an appropriate high-level qualifier, yyyy indicates an appropriate middle-level qualifier, and zzzz indicates an appropriate low-level qualifier.

- *userid.yyyy.zzzz*

  userid is the user ID of the logged on TSO user.

- *TSOprefix.yyyy.zzzz*

  TSOprefix is the data set prefix established by the TSO PROFILE command. *userid* is the default value of *TSOprefix*.
- *jobname.yyyy.zzzz*

  jobname is the job name specified on the JOB statement for a job stream or the procedure name for a started procedure.
- *hlq.yyyy.zzzz*

  hlq is the TCP/IP HLQ distributed as the system default, which can be overridden by the value in the DATASETPREFIX statement.
- *xxxx.nodename.zzzz*

  nodename is an MLQ that is used to define the data set name for the TCP/IP stack profile data set.
- *xxxx.function_name.zzzz*

  function_name denotes an acronym specifying a particular TCP/IP server (for example SRVRFTP for the FTP server) and is used as an MLQ for the translation table data set for that application.
- *xxxx.private_name.zzzz*

  private_name is a user-specified private qualifier that can be specified as an option on some TCP/IP commands.
- SYS1.TCPPARMS(TCPDATA)

  The member of a system data set used to find the *configuration file* TCPIP.DATA.

TCP/IP configuration data sets

Table 1 lists the configuration data sets and z/OS UNIX HFS files used by the TCP/IP servers and functions. It includes the name of the sample of the data set or file that is provided by Communications Server, and the usage of the data set or file.

Table 1. TCP/IP configuration data sets

Name (search order)	Copied from	Usage
hlq.ETC.IPNODES	SEZAINST(EZBREIPN)	One of the local host files used for IPv6 name query, or IPv4 and IPv6 name query when COMMONSEARCH is specified in the resolver setup file.
/etc/mail/sendmail.cf	/usr/lpp/tcpip/samples/sendmail/cf/sample.cf	Provides configuration information for the sendmail daemon when being used as a message transfer agent (MTA). If /etc/mail/submit.cf does not exist, this data set also provides configuration information for the end-user sendmail application when being used as a mail user agent (MUA).

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
/etc/mail/submit.cf	/usr/lpp/tcpip/samples/sendmail/cf/submit.cf	Provides configuration information for the end-user sendmail application when being used as a mail user agent (MUA).
/etc/mail/zOS.cf	/usr/lpp/tcpip/samples/sendmail/cf/zOS.cf	Provides z/OS-specific information for the sendmail daemon when being used as a message transfer agent (MTA). Currently the file consists of Secure Sockets Layer (SSL) information only.
ETC.PROTO	usr/lpp/tcpip/samples/protocol	Used to map types of protocol to integer values to determine the availability of the specified protocol. Required by several z/OS Communications Server components. **Note:** The search order depends on the type of application (z/OS UNIX or native MVS).
ETC.RPC	SEZAINST(ETCRPC)	Defines RPC applications to the Portmapper function.
ETC.SERVICES	usr/lpp/tcpip/samples/services	Establishes port numbers for servers using TCP and UDP. Required for z/OS UNIX SNMP and OMPROUTE (if the RIP protocol is used). **Note:** The search order depends on the type of application (z/OS UNIX or native MVS).
FTP.DATA 1. //SYSFTPD 2. *userid/jobname*.FTP.DATA 3. /etc/ftp.data 4. SYS1.TCPPARMS(FTPDATA) 5. *hlq*.FTP.DATA	SEZAINST(FTCDATA) for the client and (FTPSDATA) for the server	Overrides default FTP client and server parameters for the FTP server. For more information about *hlq*, *jobname*, or *userid*, see Chapter 11, "Transferring files using FTP," on page 541.
HOSTS.LOCAL (or /etc/hosts)	SEZAINST(HOSTS)	Input data set to MAKESITE for generation of HOSTS.ADDRINFO and HOSTS.SITEINFO.
IKE daemon configuration 1. The name of an HFS file or MVS data set specified by the IKED_FILE environment variable 2. /etc/security/iked.conf	/usr/lpp/tcpip/samples/iked.conf	Contains IKE configuration statements.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
LBADV.CONF The data set or file specified on the CONFIG DD statement in the z/OS Load Balancing Advisor start procedure.	SEZAINST(LBADVCNF)	Contains z/OS Load Balancing Advisor configuration statements.
LBAGENT.CONF The data set or file specified on the CONFIG DD statement in the z/OS Load Balancing Agent start procedure.	SEZAINST(LBAGECNF)	Contains z/OS Load Balancing Agent configuration statements.
LPD.CONFIG	SEZAINST(LPDDATA)	Configures the Line Printer Daemon for the Remote Print Server.
LU62CFG	SEZAINST(LU62CFG)	Provides configuration parameters for the SNALINK LU6.2 interface.
MASTER.DATA	No sample provided	DNS database input required for authoritative name servers.
MIBS.DATA 1. The name of an HFS file or an MVS data set specified by the MIBS_DATA environment variable 2. /etc/mibs.data HFS file	No sample provided	Defines textual names for MIB objects for the z/OS UNIX snmp command.
NPSIDATE	SEZAINST(NPSIDATE)	Operates the TCP/IP X.25 NCP Packet Switching Interface.
NPSIGATE	SEZAINST(NPSIGATE)	Supports GATE MCHs for X.25 NCP Packet Switching Interface.
OMPROUTE configuration 1. The name of an HFS file or MVS data set specified by the OMPROUTE_FILE environment variable 2. /etc/omproute.conf 3. *hlq*.ETC.OMPROUTE.CONF	SEZAINST(EZAORCFG)	Contains OMPROUTE configuration statements.
OSNMP.CONF 1. /etc/osnmp.conf 2. /etc/snmpv2.conf	/usr/lpp/tcpip/samples/snmpv2.conf	Defines target host security parameters for the osnmp command.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
OSNMPD.DATA 1. The name of an HFS file or MVS data set specified by the OSNMPD_DATA environment variable 2. /etc/osnmpd.data HFS file 3. The data set specified on the OSNMPD DD statement in the agent procedure 4. *jobname*.OSNMPD.DATA, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(OSNMPD) 6. *hlq*.OSNMPD.DATA, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used **Note:** The first file found in the search order is used.	/usr/lpp/tcpip/samples/osnmpd.data	Used by SNMP for setting values for selected MIB objects.
PAGENT.CONF 1. File or data set specified with -c startup option 2. File or data set specified with PAGENT_CONFIG_FILE environment variable 3. /etc/pagent.conf	/usr/lpp/tcpip/samples/pagent.conf	Defines Policy Agent configuration parameters and optionally defines QoS service policies (rules and actions).
PROFILE.TCPIP 1. //PROFILE 2. *job_name.node_name*.TCPIP 3. *hlq.node_name*.TCPIP 4. *job_name*.PROFILE.TCPIP 5. *hlq*.PROFILE.TCPIP	SEZAINST(SAMPPROF)	Provides TCP/IP initialization parameters and specifications for network interfaces and routing.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
PW.SRC 1. The name of an HFS file or an MVS data set specified by the PW_SRC environment variable 2. /etc/pw.src HFS file 3. The data set specified on SYSPWSRC DD statement in the agent procedure 4. *jobname*.PW.SRC, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(PWSRC) 6. *hlq*.PW.SRC, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used **Note:** The first file found in the search order is used.	No sample provided	Defines a list of community names used when accessing objects on a destination SNMP agent.
Resolver Setup File	SEZAINST (RESSETUP)	Provides configuration statements for the resolver.
RSVPD.CONF 1. File or data set specified with -c startup option 2. File or data set specified with RSVPD_CONFIG_FILE environment variable 3. /etc/rsvpd.conf 4. *hlq*.RSVPD.CONF	/usr/lpp/tcpip/samples/rsvpd.conf	Defines RSVP Agent configuration parameters.
SNMPD.BOOTS 1. The name of an HFS file or an MVS data set specified by the SNMPD_BOOTS environment variable. 2. /etc/snmpd.boots **Note:** The first file found in the search order is used.	No sample provided	Defines the SNMP agent security and notification destinations. **Note:** If the SNMPD.BOOTS file is not provided, the SNMP agent creates the file. If multiple SNMPv3 agents are running on the same MVS image, use the environment variable to specify different SNMPD.BOOTS files for the different agents. For security reasons, ensure unique engine IDs are used for different SNMP agents.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
SNMPD.CONF 1. The name of an HFS file or an MVS data set specified by the SNMPD_CONF environment variable. 2. /etc/snmpd.conf **Note:** The first file found in the search order is used.	/usr/lpp/tcpip/samples/snmpd.conf	Defines the SNMP agent security and notification destinations. **Note:** If the SNMPD.CONF file is found, the PW.SRC file and the SNMPTRAP.DEST files are not used.
SNMPTRAP.DEST 1. The name of an HFS file or an MVS data set specified by the SNMPTRAP_DEST environment variable 2. /etc/snmptrap.dest HFS file 3. The data set specified on SNMPTRAP DD statement in the agent procedure 4. *jobname*.SNMPTRAP.DEST, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(SNMPTRAP) 6. *hlq*.SNMPTRAP.DEST, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used **Note:** The first file found in the search order is used.	No sample provided	Defines a list of managers to which the SNMP agent sends traps.
SMTPCONF	SEZAINST(SMTPCONF)	Provides configuration parameters for the Simple Mail Transfer Protocol.
SMTPNOTE	SEZAINST(SMTPNOTE)	Defines note parameters for Simple Mail Transfer Protocol.
TCPIP.DATA	SEZAINST(TCPDATA)	Provides parameters for TCP/IP client programs. **Note:** The search order depends on the type of application (z/OS UNIX or native MVS).
TNDBCSCN	SEZAINST(TNDBCSCN)	Provides configuration parameters for Telnet 3270 Transform support.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
TRAPFWD.CONF 1. An HFS file or an MVS data set specified by the TRAPFWD_CONF environment variable 2. /etc/trapfwd.conf **Note:** The first file found in the search order is used.	No sample provided	Defines addresses to which the Trap Forwarder Daemon forwards traps. **Note:** If the environment variable is set and if the file specified by the environment variable is not found, the Trap Forwarder daemon terminates.
VTAMLST	SEZAINST(VTAMLST)	Defines VTAM applications and their characteristics. Entries required for Telnet, SNALINK LU0, SNALINK LU6.2, and X.25 NPSI Server.
X25CONF	SEZAINST(X25CONF)	Provides configuration parameters for the X.25 NCP Packet Switching Interface.
X25VSVC	SEZAINST(X25VSVC)	Provides switched virtual circuit configuration for the X.25 NCP Packet Switching Interface.

Configuration files for the TCP/IP stack

Two configuration files are used by the TCP/IP stack, PROFILE.TCPIP and TCPIP.DATA. PROFILE.TCPIP is used only for the configuration of the TCP/IP stack. TCPIP.DATA is used during configuration of both the TCP/IP stack and applications; the search order used to find TCPIP.DATA is the same for both the TCP/IP stack and applications.

PROFILE.TCPIP search order

During initialization of the TCP/IP stack, system operation and configuration parameters for the TCP/IP stack are read from the configuration file PROFILE.TCPIP. As shown in Table 1 on page 30, the search order used by the TCP/IP stack to find PROFILE.TCPIP involves both explicit and dynamic data set allocation as follows:

- //PROFILE DD DSN=*aaa.bbb.ccc(anyname)*
- *jobname.nodename*.TCPIP
- *hlq.nodename*.TCPIP
- *jobname*.PROFILE.TCPIP
- TCPIP.PROFILE.TCPIP

Note: Explicitly specifying the PROFILE DD statement in the TCPIPROC JCL is the recommended way to specify PROFILE.TCPIP. If this DD statement is present, the data set it defines is explicitly allocated by MVS and no dynamic allocation is done. If this statement is not present, the search order continues to use dynamic allocation for the PROFILE.TCPIP.

Examples

The following examples show the search order used by TCP/IP to find the configuration file PROFILE.TCPIP. These examples use the sample TCP/IP started procedure, TCPIPROC, installed in the SEZAINST data set.

Example when DD cards are in your TCP/IP startup procedure: In this example, the PROFILE DD cards are specified as follows:

```
//TCPIP  PROC PARMS='CTRACE(CTIEZB00)'
//*
//* z/OS Communications Server
//* SMP/E Distribution Name: EZAEB01G
//*
//*     5694-A01 (C) Copr. IBM Corp. 1991,2001.
//*     All rights reserved.
//*     US Government Users Restricted Rights -
//*     Use, duplication or disclosure restricted
//*     by GSA ADP Schedule Contract with IBM Corp.
//*     See IBM Copyright Instructions
//*
//TCPIP    EXEC PGM=EZBTCPIP,
//             PARM='&PARMS',
//             REGION=0K,TIME=1440
//*
:
//PROFILE DD DISP=SHR,DSN=MVSA.PROD.PARMS(PROFILE)
:
```

Because the PROFILE DD is the first step in the search order, TCP/IP uses the data set MVSA.PROD.PARMS(PROFILE) as the PROFILE.TCPIP configuration file.

Example when no DD cards are in your TCP/IP startup procedure: In this example, the PROFILE DD statement is not specified:

```
//TCPIP  PROC PARMS='CTRACE(CTIEZB00)'
//*
//* z/OS Communications Server
//* SMP/E Distribution Name: EZAEB01G
//*
//*     5694-A01 (C) Copr. IBM Corp. 1991,2001.
//*     All rights reserved.
//*     US Government Users Restricted Rights -
//*     Use, duplication or disclosure restricted
//*     by GSA ADP Schedule Contract with IBM Corp.
//*     See IBM Copyright Instructions
//*
//TCPIP    EXEC PGM=EZBTCPIP,
//             PARM='&PARMS',
//             REGION=0K,TIME=1440
//*
:
```

For the configuration file PROFILE.TCPIP, the search order used is as follows:

1. PROFILE DD

   No PROFILE DD exists...search continues.
2. *jobname.nodename.*TCPIP

   If *jobname.nodename.*TCPIP is found, the search stops here.
3. *hlq.nodename.*TCPIP

   If *hlq.nodename.*TCPIP is found, the search stops here.
4. *jobname*.PROFILE.TCPIP

   If *jobname*.PROFILE.TCPIP is found, the search stops here.

5. TCPIP.PROFILE.TCPIP
   TCPIP.PROFILE.TCPIP is searched last if necessary.

TCPIP.DATA search order

TCPIP.DATA is used by the stack address space as follows:

- The TCP/IP stack's configuration component uses TCPIP.DATA during TCP/IP stack initialization to determine the stack's HOSTNAME. To get its value, the z/OS UNIX environment search order is used.
- The TCP/IP stack's TN3270 Telnet server component uses TCPIP.DATA statements to resolve a client's IP address to a name. To obtain the resolver-related statements for address resolution, the native MVS environment search order is used.

For details on the z/OS UNIX environment and native MVS environment search orders and the usage of z/OS UNIX environment variables, see "Resolver configuration files" on page 40.

Configuration files for TCP/IP applications

This section describes environment variables, the resolver configuration files that can be used by TCP/IP applications, and the search orders for those files. In addition to resolver files, an application can also have its own configuration files that are specific to that application. For more information about application-specific configuration files, see the descriptions of the individual applications in Part 2, "Server applications," on page 403.

Environment variables

Environment variables are named variables with assigned values that can be accessed by various processes in the Communications Server configuration. Applications use environment variables to define the characteristics of their specific environment.

Table 2 lists where to find more information about the environment variables that are explicitly set by or used by z/OS Communications Server and its applications. Language Environment and UNIX System Services also provide environment variables. For information regarding these other variables, refer to the section on Understanding Shell Variables in *z/OS UNIX System Services User's Guide*.

Table 2. Environment variables

Application	For environment variable information, refer to:
Bind 4.9.3 - DNS	Bind 4.9.3-based DNS environment variables in *z/OS Communications Server: IP Configuration Reference*
Bind 9 - DNS	Bind 9-based DNS environment variables in *z/OS Communications Server: IP Configuration Reference*
Digital certificate access server (DCAS)	DCAS environment variables in *z/OS Communications Server: IP Configuration Reference*
Dynamic host configuration protocol (DHCP)	DHCP environment variables in *z/OS Communications Server: IP Configuration Reference*

Table 2. Environment variables (continued)

Application	For environment variable information, refer to:
FTP client	Environment variables in *z/OS Communications Server: IP User's Guide and Commands*
FTP server	FTP server environment variables in *z/OS Communications Server: IP Configuration Reference*
IKE daemon	IKE environment variables in *z/OS Communications Server: IP Configuration Reference*
MIBDESC	MIBDESC environment variables in *z/OS Communications Server: IP Configuration Reference*
Motif	Motif environment variables in *z/OS Communications Server: IP Programmer's Guide and Reference*
nslapm2	Network SLAPM2 subagent environment variables in *z/OS Communications Server: IP Configuration Reference*
OMPROUTE	OMPROUTE environment variables in *z/OS Communications Server: IP Configuration Reference*
ORSHD	RSHD command (orshd) environment variables in *z/OS Communications Server: IP Configuration Reference*
OSNMP	OSNMP environment variables in *z/OS Communications Server: IP Configuration Reference*
Policy Agent and SLA subagent (pagtsnmp)	Policy Agent environment variables in *z/OS Communications Server: IP Configuration Reference*
Resolver	"Setting environment variables for configuration files" on page 43
SNMP agent	OSNMPD environment variables in *z/OS Communications Server: IP Configuration Reference*
SNMP DPI	SNMP DPI environment variables in *z/OS Communications Server: IP Programmer's Guide and Reference*
SNTPD	SNTPD environment variables in *z/OS Communications Server: IP Configuration Reference*
TRAPFWD daemon	TRAPFWD environment variables in *z/OS Communications Server: IP Configuration Reference*
UNIX rexecd	REXECD command environment variables in *z/OS Communications Server: IP Configuration Reference*
UNIX sendmail	"Environment variables" on page 1141
UNIX Telnet server (otelnetd)	"Environment variables" on page 532

Table 2. Environment variables (continued)

Application	For environment variable information, refer to:
X window system interface V11R4 and Motif version 1.1	Environment variables in *z/OS Communications Server: IP Programmer's Guide and Reference*
X window system	X Window System environment variables in *z/OS Communications Server: IP Programmer's Guide and Reference*

Resolver configuration files

Understanding the resolver search orders used in native MVS and z/OS UNIX environments is key to setting up your system properly.

As described in "Understanding resolvers" on page 20, the resolver can use available name servers, local definitions, or a combination of both, to process API resolver requests. Figure 2 shows how local definitions can be specified and searched for when needed.

Use the trace resolver facility to determine what TCPIP.DATA values are being used by the resolver and where they were read from. For information on dynamically starting the trace, refer to *z/OS Communications Server: IP Diagnosis Guide*. Once the trace is active, issue a TSO `NETSTAT HOME` command and a z/OS UNIX shell `netstat -h` command to display the values. Issuing a `Ping` of a host name from TSO and from the z/OS UNIX shell also shows activity to any DNS servers that might be configured.

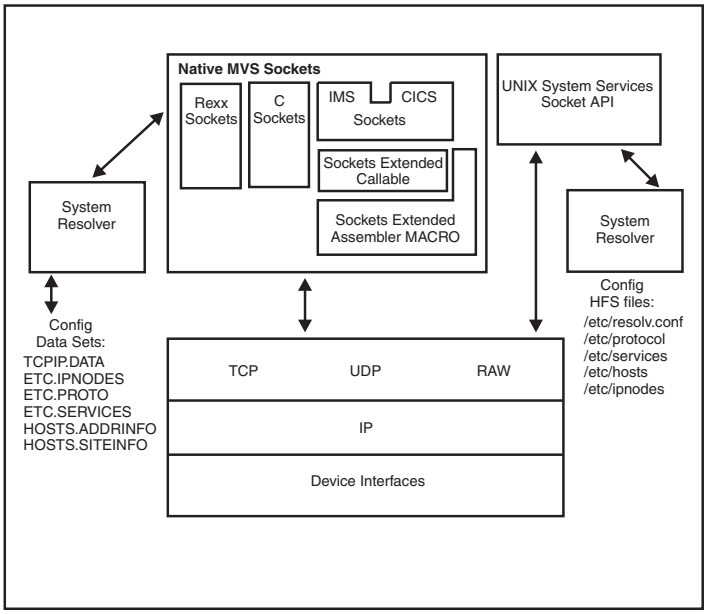


Figure 2. Resolver related configuration files in z/OS UNIX and native MVS environments

Table 3 on page 41 shows the complete set of local definition possibilities available to the resolver.

Table 3. Local definitions available to resolver

File type description	APIs affected	Candidate files
Base resolver configuration files	All APIs	1. GLOBALTCPIPDATA 2. RESOLVER_CONFIG environment variable 3. /etc/resolv.conf 4. SYSTCPD DD-name 5. *userid*.TCPIP.DATA 6. *jobname*.TCPIP.DATA 7. SYS1.TCPPARMS(TCPDATA) 8. DEFAULTTCPIPDATA 9. TCPIP.TCPIP.DATA
Translate tables	All APIs	1. X_XLATE environment variable 2. *userid*.STANDARD.TCPXLBIN 3. *jobname*.STANDARD.TCPXLBIN 4. *hlq*.STANDARD.TCPXLBIN 5. Resolver-provided translate table, member STANDARD in SEZATCPX
Local host tables	endhostent endnetent getaddrinfo gethostbyaddr gethostbyname gethostent GetHostNumber GetHostResol GetHostString getnameinfo getnetbyaddr getnetbyname getnetent IsLocalHost Resolve sethostent setnetent	1. X_SITE environment variable 2. X_ADDR environment variable 3. /etc/hosts 4. *userid*.HOSTS.xxxxINFO 5. *jobname*.HOSTS.xxxxINFO 6. *hlq*.HOSTS.xxxxINFO 7. GLOBALIPNODES 8. RESOLVER_IPNODES environment variable 9. *userid*.ETC.IPNODES 10. *jobname*.ETC.IPNODES 11. *hlq*.ETC.IPNODES 12. DEFAULTIPNODES 13. /etc/ipnodes
Protocol information	endprotoent getprotobyname getprotobynumber getprotoent setprotoent	1. /etc/protocol 2. *userid*.ETC.PROTO 3. *jobname*.ETC.PROTO 4. *hlq*.ETC.PROTO
Services information	endservent getaddrinfo getnameinfo getservbyname getservbyport getservent setservent	1. /etc/services 2. SERVICES DD-name 3. *userid*.ETC.SERVICES 4. *jobname*.ETC.SERVICES 5. *hlq*.ETC.SERVICES
Host alias table	getaddrinfo gethostbyname	HOSTALIASES environment variable

The actual search order of the candidate files varies depending on the type of API used and the resolver's setup. The search orders are explained in more detail in

"Search orders used in the z/OS UNIX environment" on page 43 and "Search orders used in the native MVS environment" on page 50.

Information about an application's search order can be obtained by using the trace resolver facility. Trace resolver output provides a caller API value that determines which search order is used. For information on dynamically starting the trace, refer to *z/OS Communications Server: IP Diagnosis Guide*.

The following caller API values indicate the z/OS UNIX environment search order is used:

- Language Environment C Sockets
- Unix System Services

The following caller API values indicate the native MVS environment search order is used:

- TCP/IP C Sockets
- TCP/IP Pascal Sockets
- TCP/IP Rexx Sockets
- TCP/IP Sockets Extended

Following are some examples of Communications Server TSO commands that use the native MVS search order:

- DIG
- FTP (batch)

  Note: The TSO FTP command uses the z/OS UNIX search order.
- LPR
- NETSTAT
- NSLOOKUP
- PING
- REXEC
- RPCINFO
- RSH
- TRACERTE

Following are some examples of Communications Server UNIX commands that use the z/OS UNIX search order:

- dig
- dnsdomainname
- domainname
- ftp
- host
- hostname
- netstat
- nslookup
- ping
- rexec
- rpcinfo
- sendmail

- snmp
- traceroute

Following are some examples of Communications Server applications that use the native MVS search order:

- CICS Listener
- LPD
- Miscellaneous server
- PORTMAP
- RSHD
- SMTP
- TN3270

Following are some examples of Communications Server applications that use the z/OS UNIX search order:

- FTP
- SNMP agent
- z/OS UNIX OPORTMAP
- z/OS UNIX OREXECD
- z/OS UNIX ORSHD

Search orders used in the z/OS UNIX environment

This section describes setting environment variables for configuration files, and the search orders used in the z/OS UNIX environment for the different file types shown in Table 3 on page 41. The z/OS UNIX socket functions utilize various types of TCP/IP data sets and HFS files. They include:

- Base resolver configuration files
- Translate tables
- Local host tables
- Protocol information
- Services information
- Host alias table

The particular file or table chosen can be either an MVS data set or an HFS file, depending on the resolver configuration settings and the presence of given files on the system.

Note: A program's first resolver service request initializes the resolver definitions that will be used for all resolver requests. For long running programs, the definitions can be modified by use of the MODIFY REFRESH operator command. For command usage and syntax, see *z/OS Communications Server: IP System Administrator's Commands*.

Setting environment variables for configuration files:

A z/OS C/C++ environment variable is an identifier used like a variable in a program. In Table 3 on page 41, the following environment variables appear:

HOSTALIASES
The host aliases data set, file, or DDNAME.

RESOLVER_CONFIG
The resolver configuration data set, file, or DDNAME. Used by TCP/IP to include the name of an MVS data set or HFS file in the search order for TCPIP.DATA.

RESOLVER_IPNODES
The IPNODES data set, file, or DDNAME.

X_SITE and X_ADDR
The HOSTS.SITEINFO and HOSTS.ADDRINFO data sets, files, or DDNAMEs created by the MAKESITE TSO command. X_SITE influences how gethostbyname() resolves the network address of the specified host name. X_ADDR is used by some TCP/IP functions, such as getnetbyaddr, to include the name of an MVS data set or HFS file in the search order for HOSTS.ADDRINFO.

X_XLATE
The ASCII-EBCDIC translate table data set, file, or DDNAME created by the CONVXLAT TSO command. Used by TCP/IP to include the name of an MVS data set or HFS file in the search order for STANDARD.TCPXLBIN.

Other environment variables that can be explicitly set by the resolver include the following:

LOCALDOMAIN
Defines the domain origin. Once this environment variable is set, it overrides any setting for DOMAIN, DOMAINORIGIN, or SEARCH found in TCPIP.DATA

RESOLVER_TRACE
Defines the data set, file, or DDNAME into which the resolver trace output is written.

MESSAGECASE
Determines whether messages are translated to all uppercase characters before being sent to the console.

Setting an environment variable so that a z/OS UNIX application is able to retrieve the value depends on whether the z/OS UNIX application is started from the z/OS shell or from JCL.

If the z/OS UNIX application is to be started from the z/OS shell, the export shell command can be used to set the environment variable. For example, to set the value of RESOLVER_CONFIG to the HFS file /etc/tcpa.data, you can code the following export command:

```
export RESOLVER_CONFIG=/etc/tcpa.data
```

If instead of an HFS file, you want to set RESOLVER_CONFIG to the data set MVSA.PROD.PARMS(TCPDATA), you can specify the following export command. Be sure to put the single quotation marks around the data set name. If you do not, your user ID will be added as a prefix to the data set name when the resolver tries to open the file.

```
export RESOLVER_CONFIG="//'MVSA.PROD.PARMS(TCPDATA)'"
```

If the z/OS UNIX application is to be started from JCL instead of from the z/OS shell, the environment variable needs to be passed as a parameter in the JCL of the application. For example, the following shows the RESOLVER_CONFIG variable set to pick up the TCPIP.DATA information from a file in the HFS:

```
//OSNMPD    PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)/',
//      'ENVAR("RESOLVER_CONFIG=/etc/tcpa.data")/-d 0')
⋮
```

The following example shows the RESOLVER_CONFIG variable set to pick up the TCPIP.DATA information from a partitioned data set:

```
//OSNMPD    PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)/',
//  'ENVAR("RESOLVER_CONFIG=//''TCPA.MYFILE(TCPDATA)''")/-d 0')
⋮
```

The following example shows the RESOLVER_CONFIG variable set to pick up the TCPIP.DATA information from a DD card:

```
//OSNMPD PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)/',
// 'ENVAR("RESOLVER_CONFIG=DD:TCPDATA")/-d 0')
//TCPDATA DD DSN=TCPA.MYFILE(TCPDATA),DISP=SHR
⋮
```

Tip: A DDNAME can also be specified as //DD:*ddname* as follows:

```
⋮
// 'ENVAR("RESOLVER_CONFIG=//DD:TCPDATA")/-d 0'))
```

The following example shows an alternate method of accessing environment variables:

```
//OSNMPD    PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)/',
// 'ENVAR("_CEE_ENVFILE=DD:STDENV")/-d 0')
//STDENV DD DSN=TCPA.MYFILE(TCPDATA),DISP=SHR
```

In this case, the environment variables will be read from the file specified on the STDENV DD statement. If this file is an MVS data set, the data set must be allocated with RECFM=V. RECFM=F is not recommended, because RECFM=F enables padding with blanks for the environment variables. See *z/OS XL C/C++ Programming Guide* for more information on specifying a list of environment variables using the _CEE_ENVFILE environment variable.

Regardless of whether the z/OS UNIX application is started from the z/OS shell or from JCL, the RESOLVER_CONFIG environment variable can also be set to indicate that a DDname should be used. The following directs the resolver to read its TCPIP.DATA statements from the DDname MYTCPIPD:

```
RESOLVER_CONFIG=DD:MYTCPIPD
```

For information on how to use a DDname when specifying what kind of file to use, refer to *z/OS XL C/C++ Programming Guide*.

Base resolver configuration files: The base resolver configuration file contains TCPIP.DATA statements. In addition to resolver directives, it is referenced to determine, among other things, the data set prefix (DATASETPREFIX statement's value) to be used when trying to access some of the configuration files specified in this section.

The search order used to access the base resolver configuration file is as follows:

1. GLOBALTCPIPDATA

   If defined, the resolver GLOBALTCPIPDATA setup statement value is used. For a description of the GLOBALTCPIPDATA statement, see "Understanding resolvers" on page 20.

   The search continues for an additional configuration file. The search ends with the next file found.
2. The value of the environment variable RESOLVER_CONFIG

   The value of the environment variable is used. This search will fail if the file does not exist or is allocated exclusively elsewhere.
3. /etc/resolv.conf
4. //SYSTCPD DD card

   The data set allocated to the DDname SYSTCPD is used. In the z/OS UNIX environment, a child process does not have access to the SYSTCPD DD. This is because the SYSTCPD allocation is not inherited from the parent process over the fork() or exec function calls.
5. *userid*.TCPIP.DATA

   userid is the user ID that is associated with the current security environment (address space or task/thread)
6. SYS1.TCPPARMS(TCPDATA)
7. DEFAULTTCPIPDATA

   If defined, the resolver DEFAULTTCPIPDATA setup statement value is used. For a description of the DEFAULTTCPIPDATA statement, see "Understanding resolvers" on page 20.
8. TCPIP.TCPIP.DATA

Any TCPIP.DATA statements that have not been found will have their default values, if any, assigned.

Translate tables: The translate tables (EBCDIC-to-ASCII and ASCII-to-EBCDIC) are referenced to determine the translate data sets to be used.

The search order used to access this configuration file is as follows. The search order ends at the first file found:

1. The value of the environment variable X_XLATE

   The value of the environment variable is the name of the translate table produced by the CONVXLAT TSO command.
2. *userid*.STANDARD.TCPXLBIN

   userid is the user ID that is associated with the current security environment (address space or task/thread).
3. *hlq*.STANDARD.TCPXLBIN

hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

4. If no table is found, the resolver uses a hardcoded default table that is identical to the STANDARD member in the SEZATCPX data set.

Tip: Preallocating the STANDARD.TCPXLBIN data set using a JCL DD statement stops the resolver from issuing dynamic allocations for the data set. This eliminates the dynamic allocation messages (for example, IEF237I and IEF285I) from being written to the job's output joblog.

Local host tables: By default, resolver first attempts to use any configured domain name servers for resolution requests. If the resolution request cannot be satisfied, local host tables are used. Resolver behavior is controlled by the following:

- TCPIP.DATA statements

  The TCPIP.DATA resolver statements define if and how domain name servers are to be used. The LOOKUP TCPIP.DATA statement can also be used to control how domain name servers and local host tables are used. For more information on TCPIP.DATA statements, refer to *z/OS Communications Server: IP Configuration Reference*.
- How your application is written and compiled

  If your application program uses the TCP/IP-provided C/C++ API and the RESOLVE_VIA_LOOKUP symbol was defined, only local host tables will be used. For information on the use of the RESOLVE_VIA_LOOKUP symbol, refer to *z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference*.

The local host table supplies sitename information for, as one example, resolving hostnames to host or network addresses. The local host table can also supply address information, for example, for resolving addresses to hostname or network names. There are different search orders used for selecting the local host table for these different purposes. The search order to use is based on certain resolver setup statements, the type of API invocation, and possibly the type of host address (IPv4 versus IPv6) being requested or being resolved.

IPv4-unique search order for sitename information: The resolver uses the IPv4-unique search order for sitename information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the:

- getaddrinfo API is attempting to locate an IPv4 address.
- gethostbyname, sethostent, gethostent, or endhostent API is invoked.

If the COMMONSEARCH statement is specified, see "IPv6/common search order" on page 48, where the resolver can use IPNODES to locate sitenames.

The resolver uses the IPv4-unique search order for sitename information unconditionally for getnetbyname API calls.

The IPv4-unique search order for sitename information is as follows. The search ends at the first file found:

1. The value of the environment variable X_SITE

   The value of the environment variable is the name of the sitename information file created by the TSO MAKESITE command.
2. /etc/hosts

3. *userid*.HOSTS.SITEINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.HOSTS.SITEINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv4-unique search order for address information: The resolver uses the IPv4-unique search order for address information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the getnameinfo API is attempting to resolve an IPv4 address or the gethostbyaddr API is invoked. If the COMMONSEARCH statement is specified, see "IPv6/common search order," where the resolver can use IPNODES to locate IPv4 and IPv6 addresses.

The resolver uses the IPv4-unique search order for address information unconditionally for the setnetent, getnetent, endnetent, or getnetbyaddr APIs.

The IPv4-unique search order for address information is as follows. The search ends at the first file found:

1. The value of the environment variable X_ADDR

   The value of the environment variable is the name of the address information file created by the TSO MAKESITE command.
2. /etc/hosts
3. *userid*.HOSTS.ADDRINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.HOSTS.ADDRINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv6/common search order: The resolver uses the IPv6/common search order when it determines that any of the following conditions exist:

- The resolver setup statement COMMONSEARCH is specified (to have the resolver use IPNODES to locate IPv4 addresses, IPv6 addresses, and sitenames), and the getaddrinfo, gethostbyname, getnameinfo, gethostbyaddr, sethostent, gethostent, or endhostent APIs are invoked.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getaddrinfo API is attempting to locate an IPv6 address.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getnameinfo API is attempting to resolve an IPv6 address.

Note: The IPv6/common search order is never used for the following API socket calls:

- getnetbyname
- getnetbyaddr
- setnetent
- getnetent
- endnetent

The IPv6/common search order is as follows. The search ends at the first file found:

1. GLOBALIPNODES value

   If defined, the resolver GLOBALIPNODES setup statement value is used. For a description of the GLOBALIPNODES statement, see "Understanding resolvers" on page 20.
2. The value of the environment variable RESOLVER_IPNODES
3. *userid*.ETC.IPNODES

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.ETC.IPNODES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.
5. DEFAULTIPNODES

   If defined, the resolver DEFAULTIPNODES setup statement value is used. For a description of the DEFAULTIPNODES statement, see "Understanding resolvers" on page 20.
6. /etc/ipnodes

Protocol information: The protocol information supplies protocol related information for the socket calls listed in Table 3 on page 41.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. /etc/protocol
2. *userid*.ETC.PROTO

   userid is the user ID that is associated with the current security environment (address space or task/thread).
3. *hlq*.ETC.PROTO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Services information: The services information supplies the service information for the socket calls listed in Table 3 on page 41.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. /etc/services
2. *userid*.ETC.SERVICES

   userid is the user ID that is associated with the current security environment (address space or task/thread).
3. *hlq*.ETC.SERVICES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Host alias table: The host alias table supplies hostname alias information for the socket calls listed in Table 3 on page 41. The format of the alias information is the alias name, followed by a space, followed by the fully qualified domain name that corresponds to the alias name. The domain name is written without a trailing dot, and the alias name cannot contain dots. The search order used to access this configuration file consists only of the value of the environment variable HOSTALIASES.

Search orders used in the native MVS environment

The native MVS environment socket functions utilize various type of TCP/IP data sets, including:

- Base resolver configuration files
- Translate tables
- Local host tables
- Protocol information
- Services information

The particular file or table chosen depends on the resolver configuration settings and the presence of given files on the system.

Note: A program's first resolver service request initializes the resolver definitions that will be used for all resolver requests. For long running programs, the definitions can be modified by use of the MODIFY REFRESH operator command. For command usage and syntax, see *z/OS Communications Server: IP System Administrator's Commands*.

Base resolver configuration files: The base resolver configuration file contains TCPIP.DATA statements. In addition to resolver directives, it is referenced to determine, among other things, the data set prefix (DATASETPREFIX statement's value) to be used when trying to access some of the configuration files specified in this section.

The search order used to access the base resolver configuration file is as follows:

1. GLOBALTCPIPDATA.

   If defined, the resolver GLOBALTCPIPDATA setup statement value is used. For a description of the GLOBALTCPIPDATA statement, see "Understanding resolvers" on page 20.

   The search continues for an additional configuration file. The search ends with the next file found.
2. //SYSTCPD DD card

   The data set allocated to the DDname SYSTCPD is used.

   Rule: Since TCPIP.DATA statements might need to be read and used multiple times by the resolver, the FREE=CLOSE JCL parameter should not be used when allocating SYSTCPD. To allow TCPIP.DATA statements to be changed while still allocated for long running programs, consider using a member of an MVS partitioned data set instead of an MVS sequential data set. For these long running applications, the resolver MODIFY REFRESH command should then be used to indicate that TCPIP.DATA statements have been changed.
3. *userid/jobname*.TCPIP.DATA

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
4. SYS1.TCPPARMS(TCPDATA)
5. DEFAULTTCPIPDATA

   If defined, the resolver DEFAULTTCPIPDATA setup statement value is used. For a description of the DEFAULTTCPIPDATA statement, see "Understanding resolvers" on page 20.
6. TCPIP.TCPIP.DATA

Translate tables: The translate tables are referenced to determine the translate data sets to be used.

The search order used to access this configuration file is as follows. The search order ends at the first file found:

1. *userid/jobname*.STANDARD.TCPXLBIN

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
2. *hlq*.STANDARD.TCPXLBIN

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.
3. If no table is found, the resolver uses a hardcoded default table that is identical to the STANDARD member in the SEZATCPX data set.

Tip: Preallocating the STANDARD.TCPXLBIN data set using a JCL DD statement stops the resolver from issuing dynamic allocations for the data set. This eliminates the dynamic allocation messages (for example, IEF237I and IEF285I) from being written to the job's output joblog.

Local host tables: By default, resolver first attempts to use any configured domain name servers for resolution requests. If the resolution request cannot be satisfied, local host tables are used. Resolver behavior is controlled by the following:

- TCPIP.DATA statements

  The TCPIP.DATA resolver statements define if and how domain name servers are to be used. The LOOKUP TCPIP.DATA statement can also be used to control how domain name servers and local host tables are used. For more information on TCPIP.DATA statements, refer to *z/OS Communications Server: IP Configuration Reference*.
- How your application is written and compiled

  If your application program uses the TCP/IP-provided C/C++ API and the RESOLVE_VIA_LOOKUP symbol was defined, only local host tables will be used. For information on the use of the RESOLVE_VIA_LOOKUP symbol, refer to *z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference*.

The local host table supplies sitename information for, as one example, resolving hostnames to host or network addresses. The local host table can also supply address information, for example, for resolving addresses to hostname or network names. There are different search orders used for selecting the local host table for these different purposes. The search order to use is based on certain resolver setup statements, the type of API invocation, and possibly the type of host address (IPv4 versus IPv6) being requested or being resolved.

IPv4-unique search order for sitename information: The resolver uses the IPv4-unique search order for sitename information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the:

- getaddrinfo API is attempting to locate an IPv4 address.
- gethostbyname, GetHostNumber, GetHostResol, IsLocalHost, Resolve, sethostent, gethostent, or endhostent API is invoked.

If the COMMONSEARCH statement is specified, see "IPv6/common search order," where the resolver can use IPNODES to locate sitenames.

The resolver uses the IPv4-unique search order for sitename information unconditionally for getnetbyname API calls.

The IPv4-unique search order for sitename information is as follows. The search ends at the first file found:

1. *userid/jobname*.HOSTS.SITEINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
2. *hlq*.HOSTS.SITEINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv4-unique search order for address information: The resolver uses the IPv4-unique search order for address information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the getnameinfo API is attempting to resolve an IPv4 address or the gethostbyaddr or GetHostString API is invoked. If the COMMONSEARCH statement is specified, see "IPv6/common search order," where the resolver can use IPNODES to locate IPv4 and IPv6 addresses.

The resolver uses the IPv4-unique search order for address information unconditionally for the setnetent, getnetent, endnetent, or getnetbyaddr APIs.

The IPv4-unique search order for address information is as follows. The search ends at the first file found:

1. *userid/jobname*.HOSTS.ADDRINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
2. *hlq*.HOSTS.ADDRINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv6/common search order: The resolver uses the IPv6/common search order when it determines that any of the following conditions exist:

- The resolver setup statement COMMONSEARCH is specified (to have the resolver use IPNODES to locate IPv4 addresses, IPv6 addresses, and sitenames), and the getaddrinfo, gethostbyname, getnameinfo, gethostbyaddr, GetHostNumber, GetHostResol, GetHostString, IsLocalHost, Resolve, sethostent, gethostent, or endhostent APIs are invoked.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getaddrinfo API is attempting to locate an IPv6 address.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getnameinfo or Resolve API is attempting to resolve an IPv6 address.

Note: The IPv6/common search order is never used for the following API socket calls:

- getnetbyname
- getnetbyaddr
- setnetent
- getnetent
- endnetent

The IPv6/common search order is as follows. The search ends at the first file found:

1. GLOBALIPNODES value

   If defined, the resolver GLOBALIPNODES setup statement value is used. For a description of the GLOBALIPNODES statement, see "Understanding resolvers" on page 20.
2. *userid/jobname*.ETC.IPNODES

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
3. *hlq*.ETC.IPNODES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.
4. DEFAULTIPNODES

   If defined, the resolver DEFAULTIPNODES setup statement value is used. For a description of the DEFAULTIPNODES statement, see "Understanding resolvers" on page 20.
5. /etc/ipnodes

Protocol information: The protocol information supplies protocol related information for the socket calls listed in Table 3 on page 41.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. *userid/jobname*.ETC.PROTO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
2. *hlq*.ETC.PROTO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Services information: The services information supplies service information for the socket calls listed in Table 3 on page 41.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. //SERVICES DD card

   The data set allocated to the DDname SERVICES is used.
2. *userid/jobname*.ETC.SERVICES

userid is the user ID that is associated with the current security environment (address space or task/thread).

jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

3. *hlq*.ETC.SERVICES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

MVS-related considerations

MVS system symbols

Use of MVS system symbols in the PROFILE.TCPIP data set, and data sets referenced by VARY TCPIP,,OBEYFILE commands, is automatically supported. This automatic support first tries to use hiperspace memory files to perform the symbol translation, but if an error occurs, then a temporary HFS file will be used. The temporary HFS file is created in either the directory specified by the TMPDIR environment variable or, if the TMPDIR environment variable is not defined, in the /tmp directory.

Use of MVS system symbols in the resolver setup file and the TCPIP.DATA file is also automatically supported. The resolver reads and processes the TCPIP.DATA file on behalf of TCP/IP applications that invoke resolver services. System symbols are resolved as file records are read.

Use of MVS system symbols in the values of resolver environment variables, like RESOLVER_CONFIG and RESOLVER_TRACE, is also supported.

For MVS system symbols in other configuration files, use the symbol translator utility, EZACFSM1, to translate the symbols before the files are read by TCP/IP. EZACFSM1 reads an input file and writes to an output file, translating any symbols in the process.

Note: The input file and output file can be MVS data sets or HFS files, but do not specify the same file for both the input and output files (this results in a return code of 45 and no translation is attempted).

For more information about the use of MVS system services, refer to *z/OS MVS Initialization and Tuning Guide*.

Following is the symbol translator JCL, found in SEZAINST(CONVSYM), which is used to start EZACFSM1:

```
//________ JOB (accounting,information),programmer.name,
//             MSGLEVEL=(1,1),MSGCLASS=A,CLASS=A
//*
//*   CS for z/OS
//*   SMP/E distribution name: EZACFCSY
//*
//*   5694-A01 (C) Copyright IBM Corp. 1998, 2005
//*   Licensed Materials - Property of IBM
//*
//*   Function: System Symbols Translator JCL
//*
//*   This JCL kicks off a utility that will read from
//*   an input file that contains MVS System Symbols
//*   and produce an output file which has those symbols
//*   replaced with their substitution text, as defined
```

```
//*   in the appropriate IEASYMxx PARMLIB data set; see MVS
//*   Initializaton and Tuning Reference for rules about symbols.
//*
//*   This JCL can be run against any of the TCP/IP configuration
//*   files that contain MVS System Symbols.  An example of how it
//*   could be used is this; a customer could have one base TCPIP.DATA
//*   file containing MVS System Symbols which they edit and maintain.
//*   They would run this utility against this one file the various
//*   MVS systems to produce the TCPIP.DATA file for each different
//*   system.
//*
//STEP1   EXEC  PGM=EZACFSM1,REGION=0K
//SYSIN   DD DSN=TCP.DATA.INPUT,DISP=SHR
//*SYSIN   DD    PATH='/tmp/tcp.data.input'
|//*          The input file can be either an MVS data set or an z/OS
|//*            UNIX file.
//*
//*
//SYSOUT  DD DSN=TCP.DATA.OUTPUT,DISP=SHR
//*SYSOUT  DD    PATH='/tmp/tcp.data.output',PATHOPTS=(OWRONLY,OCREAT),
//*              PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
|//*          The output file can be either an MVS data set or an z/OS
|//*            UNIX file.
//*
//*          The output file cannot be the same file as the input file-
//*          doing so will result in a return code of 45.
//*
//*          You can mix input and output file types (i.e., the input
|//*          can be an MVS data set with the output being a z/OS UNIX
|//*          file or visa versa).
//*          Note: Other pathmodes for sysout may be used if needed.
```

The symbol translator utility can be used on any of the TCP/IP configuration files, but because the PROFILE.TCPIP file is automatically translated during TCP/IP initialization, there is no need to run the utility against that file.

Automatic restart manager (ARM)

Automatic restart manager is an MVS component that can automatically restart the TCP/IP stack after an abnormal end (ABEND).

During initialization, TCP/IP automatically registers with the automatic restart manager, using the following options:

```
REQUEST=REGISTER
ELEMENT=EZAsysclonetcpname
```

where:

- *sysclone* is a 1– or 2–character shorthand notation for the name of the MVS system. Refer to *z/OS MVS Initialization and Tuning Guide* for a complete description of the SYSCLONE static system symbol.
- *tcpname* is a 1– to 8–character name of the TCP/IP stack which registers with the automatic restart manager. For example, if the SYSCLONE value is 02 and the TCP/IP stack name is TCPCS, the resulting ELEMENT value is EZA02TCPCS.

```
ELEMTYPE=SYSTCPIP
TERMTYPE=ELEMTERM
```

For more information about automatic restart manager, refer to *z/OS MVS Setting Up a Sysplex*.

Logging of system messages

Syslog daemon (syslogd) is a server process that must be started as one of the first processes in your z/OS UNIX environment. TCP/IP server applications and components use syslogd for logging purposes and can also send trace information to syslogd. Servers on the local system use AF_UNIX sockets to communicate with syslogd; remote servers use AF_INET sockets. z/OS Communications Server components use the **local1**, **daemon**, **mail**, **user**, and **auth** facilities names.

Note: Each application activates and deactivates traces in a slightly different manner. For details, refer to the chapter on the individual application in this document.

The syslog daemon reads and logs system messages to the MVS console, log files, SMF, other machines, or users as specified by the configuration file. If syslogd is not started, log data from some applications will be displayed on the MVS console. For more information on syslogd, refer to Chapter 5, "Customization," on page 175.

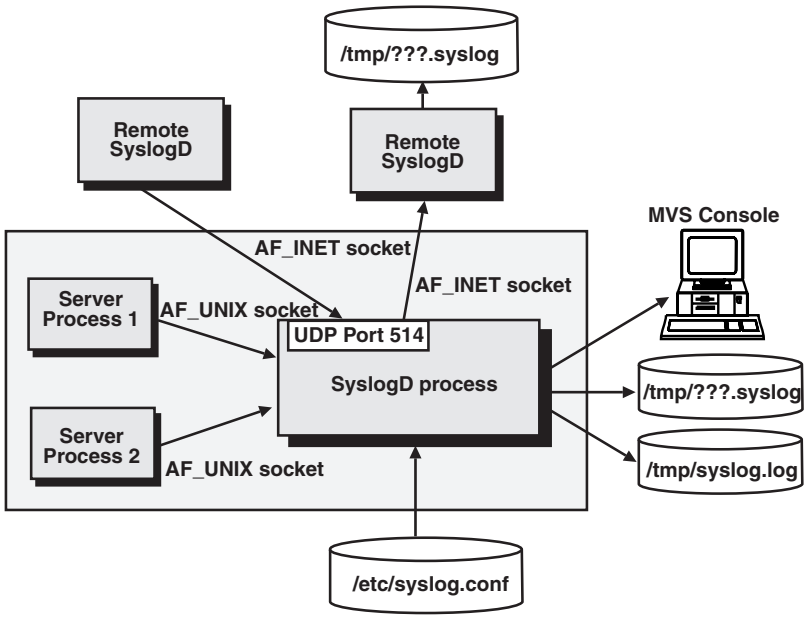


Figure 3. syslogd operation

Note: /tmp/???.syslog is the file specified in the syslogd.conf file.

The syslogd facility uses a common mechanism for segregating messages. Table 4 on page 57 shows the facilities used by z/OS Communications Server functions which write messages to syslogd. The Primary syslog facility column shows the syslog facility used for most messages logged by the application. Some applications use other facilities for certain messages. Table 4 on page 57 also shows any additional facilities.

Table 4. syslogd facilities

Application	syslogd record identifications	Primary syslog facility	Other syslog facility
Application Transparent Transport Layer Security (AT-TLS)	TTLS	daemon	auth
DHCP server	dhcpsd	user	None
FTP server	ftpd, ftps	daemon	None
IKE daemon	IKED	local4	None
NAMED	named	daemon	None
Network SLAPM2 subagent	NSLAPM2	daemon	None
OMPROUTE	omproute	user	None
OPORTMAP server	oportmap	daemon	None
OREXECD	rexecd	daemon	auth
ORSHD	rshd	daemon	auth
OTELNETD	telnetd	local1	auth
Policy Agent	Pagent	daemon	None
POPPER	popper	mail	None
PWCHANGE command	pwchange	daemon	None
PWTOKEY command	pwtokey	daemon	None
SENDMAIL	sendmail	mail	None
Service Level Agreement SNMP subagent	PASubA	daemon	None
Simple Network Time Protocol daemon	sntpd	daemon	None
SNMP agent (OSNMPD)	snmpagent	daemon	None
syslogd	syslogd	daemon	None
TCP/IP subagent	M2SubA	daemon	None
TFTP server	tftpd	user	None
TIMED daemon	timed	user	None
TN3270 Telnet subagent	TNSubA	daemon	None
Traffic Regulation Management Daemon (TRMD)	TRMD	daemon (used for IDS logging)	local4 (used for IPSEC logging)
Trap Forwarder daemon	trapfwd	daemon	None

Accounting - SMF records

Installations may use Systems Management Facilities (SMF) records for various purposes such as:

Performance management

Performance management includes the tasks that are related to verifying that defined service levels are met, and if not, identifying possible causes.

Aggregated information about delivered service, structured by organizational units (for which service levels have been defined) is needed to perform these tasks. These reports are typically time series with varying levels of time intervals, ranging from weeks through days to a time interval that matches the SMF interval. Some examples of potential reports related to performance management are:

- TCP connection elapsed time per server port number per time of day (potentially broken down on source IP address, or netmask)
- Number of TCP connections per server port number per time of day (potentially broken down on source IP address, or netmask)
- Number of inbound/outbound bytes transferred in TCP connections per time of day (potentially broken down in various ways: per destination or source port, per destination IP address, netmask, or in total, etc.)
- TCP retransmission activity per time of day (potentially broken down per destination IP address, or netmask)
- Number of outbound TCP connections per time of day (potentially broken down per destination IP address, or netmask)
- Number of inbound/outbound UDP datagrams per time of day (potentially broken down on server port number)
- Number of discards, error packets, and unknown protocol packets inbound and outbound per time of day (potentially broken down per interface)

Capacity planning

Capacity planning includes tasks that are related to forecasting capacity in terms of central processing power, memory, channel-based I/O subsystem, network attachments, and network bandwidth. Such planning tasks are based on analyzing trends for use of capacity during a preceding period (typically one to two years), and applying forecasting metrics, along with knowledge about planned launches of new applications or use of existing applications, to this trend in order to predict capacity needs during the next one to two year period. Some examples of potential reports related to capacity planning are:

- Total number of TCP connections per reserved server port number per day including analysis of average and variations around average during daily peak periods
- Total number of UDP inbound/outbound UDP datagrams per reserved server port number per day including average and variations around average during daily peak periods
- Number of bytes and/or packets transferred inbound and outbound per interface (LINK) per time of day (potentially broken down into unicasts, broadcasts, and multicasts)
- Size of queue length per interface per time of day

Auditing

Auditing involves tasks that are related to identifying and proving that individual events have taken place. Some examples of potential reports related to auditing are:

- Detailed information about specific TCP connections or UDP sockets, IP addresses, server/client identification, duration, number of bytes, etc.
- Details about activity that involves a specific client or server

- Details about a given application session based on server-specific SMF recording, such as individual TN3270 sessions or FTP sessions

Accounting

Accounting involves tasks that are related to calculating how much each individual user or organizational unit should be charged for use of the shared central IS resources. Input to such calculations vary, but is often based on CPU cycle use, data quantities, bandwidth usage, and memory use. For TCP/IP additional metrics may be defined, such as type of service used (FTP, LPD, Web server, etc), and TCP connection-related information (number of connections, duration, byte transfer counts, etc). Some examples of potential reports related to accounting are:

- Aggregated number of connections to a given server from a given source in terms of a specific client IP address, or netmask
- Accumulated connect time to a given server from a given source in terms of a specific client IP address, or netmask.
- Number of bytes transferred to or from a given source in terms of a specific client IP address, or netmask.
- Application-level accounting information specific to each individual server, for example:
  - For FTP: number of transfer operations and bytes retrieved or stored per user ID
  - For TN3270: number of sessions and session-type (TN3270/TN3270E/LINEMODE)

In general, SMF records are created for deferred processing and analysis. SMF recording is generally not used for real-time monitoring purposes. In a TCP/IP environment, real-time monitoring is implemented using the SNMP protocol and is based on internal variables that are maintained by SNMP subagents, but on z/OS a lot of the information that is written in SMF records is useful from a real-time monitoring perspective, too.

As can be seen from the above, all disciplines require detailed data as input. Depending on the discipline, certain levels of aggregation is performed on the raw detailed data in order to perform the tasks of that discipline. The objective of the TCP/IP product is to define and generate the lowest level of detail that is needed by all disciplines. How to aggregate and the actual aggregation is performed by other products, such as Performance Reporter for z/OS (PR), MVS Information Control System (MICS), or SAS-based tools or, in many cases, customer-written programs.

TCP/IP– produced SMF records should not be viewed isolated. Other components in MVS produce SMF records for the same purposes as those produced by TCP/IP. An installation is likely to combine information from a series of subsystems in performing detailed performance, or capacity planning. SMF records with information about use of CPU resources and memory resources per address space is, for example, produced by other components in MVS, and TCP/IP produced SMF records should not duplicate that information.

The events that trigger SMF records to be written and the information included in the SMF records must accommodate the intended purposes. There can be multiple purposes for given SMF records.

SMF records can be cut at multiple levels in the TCP/IP protocol stack, and the type of information that can be included depends on where the SMF record is created:

- At the IP and interface layer we know about ICMP activity, IP packet fragmentation and reassembly activity, IP checksum errors, IGMP activity, and ARP activity. At this layer, it is difficult to relate the information to specific users (remote clients, local socket address spaces, and so on), so from an accounting point of view, this information is not very interesting. From a performance and capacity planning point of view, this information is of interest because it allows the installation to aggregate network-layer activity to physical interfaces, which is an important aspect of both performance and capacity management.
- At the transport protocol layer, we know about IP addresses, port numbers, and host names. For TCP related work, we know about connections and information that is related to TCP connections, such as byte counts, connection times, reliability metrics, and performance metrics. For UDP related work, each UDP datagram is a separate entity, and the only way we can aggregate information for UDP is on a UDP socket-level, where we could cut SMF records every time a UDP socket is closed.
- At the application layer, we know more details about what goes on, but every application is different and it requires separate SMF record definitions and ability to write the SMF records to implement application-layer SMF recording. We currently do it for the stack Telnet server and the FTP server, but not for any other servers.

SMF accounting issues (Record type 118)

Many installations rely on the MVS component SMF for job accounting and for performance analysis. TCP/IP can create SMF type 118 records for certain events. If you are running multiple stacks, SMF does not always allow you to distinguish among them. Consider the following issues:

- There is no stack identity in SMF type 118 records. SMF records that are written by the system address space or by standard servers may be identified as belonging to one stack or another, based on address space naming conventions.
- SMF records written by client address spaces cannot be identified as belonging to a single stack based on the address space naming conventions used in standard servers.
- The only technique currently available to distinguish among records written by various client address spaces is to assign unique SMF type 118 record subtype intervals to each stack:
  - FTP server: One or nine subtypes in FTP.DATA
  - Telnet server: Two subtypes on TELNETPARMS
  - API: Two subtypes on SMFPARMS
  - FTP, Telnet client: One subtype on SMFPARMS

If you choose to assign subtypes, there will be an obvious impact on your local accounting programs. SMF type 118 subtype changes and additions must be coordinated with persons responsible for managing the use of SMF.

SMF type 118 records do not support IPv6 addresses. Thus, if you choose to exploit IPv6 in your environment, migrate your SMF processing to use the SMF type 119 records, which do support IPv6 addresses.

An external mapping (EZASMF76 macro) is available for customers to parse the SMF type 118 records that TCP/IP generates. EZASMF76 produces assembler level DSECTs for the Telnet (server and client), FTP (server and client), and API SMF records.

Note: If the BPX.SMF facility is defined and SMF records are to be written by syslogd, the user ID with which syslogd runs must be permitted to BPX.SMF.

To create the Telnet SMF Record layout, code:

```
EZASMF76 TELNET=YES
```

To create the FTP SMF Record layout, code:

```
EZASMF76 FTP=YES
```

To create the API SMF Record layout, code:

```
EZASMF76 API=YES
```

SMF accounting issues (Record type 119)

SMF type 119 records contain unique stack identification sections designed to eliminate the confusion of the type 118 records. They provide uniformity of date and time (UTC), common record format (self-defining section and TCP/IP identification section), and support for IPv6 addresses and expanded field sizes (64 bit versus 32 bit) for some counters. The kinds of SMF type 119 records available are:

- TCP connection initiation and termination
- UDP socket close
- TCP/IP, interface and server port statistics
- TCP/IP stack start/stop
- FTP server transfer completion
- FTP server logon failure
- FTP client transfer completion
- TN3270 server session initiation and termination
- Telnet client connection initiation and termination.

The SMF type 119 records utilize a common structure. Each record is organized as follows:

- SMF header
- Self-defining section containing pointers to:
  - TCP/IP identification section (identifies system, stack etc)
  - Sections containing the data for the record

An external mapping (EZASMF77 macro) is available for customers to parse the SMF type 119 records that TCP/IP generates.

For more detailed information refer to *z/OS Communications Server: IP Configuration Reference*.

For more information about SMF, refer to *z/OS MVS System Management Facilities (SMF)*.

Security considerations

Multilevel security is an enhanced security environment that can be configured on a z/OS system. In this environment, the security server and trusted resource managers enforce mandatory access control policies in addition to the usual discretionary access control policies. To participate in a multilevel secure environment, the user IDs associated with z/OS CS tasks and the resource profiles in the SERVAUTH class need to have security labels defined. For more information on the multilevel secure environment and configuring z/OS CS in that environment, see Chapter 4, "Preparing for TCP/IP networking in a multilevel secure environment," on page 145.

z/OS Communications Server relies on a System Authorization Facility (SAF) to protect several resources:

- Started tasks require access to a STARTED resource. This is documented in the server information in the *z/OS Communications Server: IP Configuration Reference*. Also, refer to SEZAINST(EZARACF) for SAF authorizations required for the TCP/IP stack and servers started tasks.
- Restricting access to a network, subnetwork or particular IP address in the network is provided by resources in the SERVAUTH class. Using NETACCESS statements, z/OS CS can map networks, subnetworks and IP addresses to SAF resource names. Users that are not permitted access to a particular SAF resource are not allowed to communicate with the corresponding network, subnetwork, or IP address. Refer to the NETACCESS statement in *z/OS Communications Server: IP Configuration Reference* or "Setting up SAF Server Access Authorization (SERVAUTH) (optional)" on page 207 for more information.

  Restricting users' ability to run applications that access specific TCP and UDP ports is also provided by resources in the SERVAUTH class. z/OS Communications Server provides a one-to-one mapping between port numbers and SAF resource names. Refer to the PORTACCESS statement in the *z/OS Communications Server: IP Configuration Reference* or "Setting up SAF Server Access Authorization (SERVAUTH) (optional)" on page 207 for more information.

  Also, similar to PORTACCESS, z/OS Communications Server ensures a user attempting to connect to a TN3270 secure port is allowed access to the port. This support is used in conjunction with TN3270 SSL client authentication support. Refer to the CLIENTAUTH statement in the *z/OS Communications Server: IP Configuration Reference* or "Setting up SAF Server Access Authorization (SERVAUTH) (optional)" on page 207 for more information.

  Restricting access to the TCPIP stack is also controlled under z/OS CS by defining a resource in the SERVAUTH class. Refer to "Setting up SAF Server Access Authorization (SERVAUTH) (optional)" on page 207 for more information.
- Restricting access to operator commands is provided through the OPERCMDS resource. z/OS Communications Server verifies that users have access to specific OPERCMDS resources before executing the operator command. Refer to the operator commands information in the *z/OS Communications Server: IP System Administrator's Commands* or "Setting up SAF Server Access Authorization (SERVAUTH) (optional)" on page 207 for more information about limiting access to z/OS Communications Server commands.
- Restricting access to the TSO and UNIX shell Netstat command is provided by SERVAUTH resources. z/OS Communications Server verifies that users have access to specific SERVAUTH resources before executing the Netstat command. Refer to the Netstat command information in the *z/OS Communications Server: IP System Administrator's Commands* for more information about limiting access to

Netstat command. The security product resource names in the SERVAUTH class do not apply to DISPLAY TCPIP,,NETSTAT command. If you wish to restrict access to DISPLAY TCPIP,,NETSTAT command, you can do so using standard operator command restriction facility, OPERCMDS class profiles. Refer to *z/OS MVS Planning: Operations* for more information.

Nonreusable ASIDs

The TCP/IP address space provides PC-entered services that must be accessible to all address spaces, so a system LX is obtained. This causes the Address Space Identifier (ASID) associated with the TCP/IP address space to be nonreusable when the address space is stopped or restarted. If the TCP/IP address space is terminated enough times, all available ASIDs could be exhausted, preventing the creation of a new address space on the system. In this case, an IPL is required. For more information on tuning parameters for the maximum number of ASIDs on a system, refer to the MAXUSER parameter in *z/OS MVS Initialization and Tuning Reference*.

TSO command authorization

If a command such as PING or TRACERTE is invoked using another method other than a TSO command, you might have to do additional authorization customization, such as also adding the program names to the AUTHPGM NAMES section of your IKJTSO*xx* SYS1.PARMLIB member.

UNIX System Services security considerations

This section describes some security considerations that have a product-wide effect. For descriptions of security considerations that affect specific servers or components, see the sections of this document that describe each server and component.

Requirement for an OMVS segment

Many TCP/IP Services components in z/OS Communications Server now exploit z/OS UNIX services in both the native MVS environment and in the z/OS UNIX environment. For example, all TCP/IP socket APIs and TCP/IP applications (whether they are provided by z/OS Communications Server, z/OS, other IBM and non-IBM products, or written by users) now make use of z/OS UNIX services.

Use of z/OS UNIX services requires a z/OS UNIX security context, referred to as an *OMVS segment*, for the user ID associated with any unit of work requesting these services. In other words, most user IDs requiring access to TCP/IP functions now require an OMVS segment to be defined in Resource Access Control Facility (RACF).

Note: The tasks, examples, and references in this section assume that you are using the z/OS Communications Server Security Server (RACF). If you are using a security product from another vendor, read the documentation for that product for instructions on task performance.

To satisfy the requirement for an OMVS segment in RACF, do one of the following:

- Identify all the users in your environment that use TCP/IP services and then define OMVS RACF segments for the associated user IDs.
- Use the default OMVS segment support provided by RACF and z/OS UNIX for users and groups.

The default OMVS segments reside in the USER profile and GROUP profile. The names of these profiles are identified by the installation, using the BPX.DEFAULT.USER facility class profile. The application data field in the class profile contains the user ID, or the user ID/group ID, that is used to access the default OMVS segments for users and groups, respectively.

Notes:

1. An HFS must be defined for the OMVS segment, and the home directory must exist.
2. If you use a trusted or privileged started task in ICHRIN03 or the STARTED class (especially a generic entry), be careful in assigning a default UID and GID with facility class BPX.DEFAULT.USER. Whenever trusted or privileged is specified, all default tasks have superuser authority.

To set up default OMVS segments, follow the steps in the Table 5.

Table 5. Setting up default of OMVS segment

Task	Details
Define a Group ID (GID) to the system to be used as an anchor for a default OMVS group segment.	Use the following command: `ADDGROUP DEFGRP OMVS(GID(777777))` Make the GID unique so that it is easily identifiable. The GID can be either very high or very low. The other fields related to the GID are not likely to be used for anything.

Define a user ID (UID) to be used as an anchor for the default OMVS user segment.	Use the following commands:
	`ADDUSER DEFUSR DFLTGRP(DEFGRP) NAME('DEFAULT USER')`
	`OMVS(UID(999999) HOME('/') PROGRAM('/bin/sh'))`
	
	Note: To avoid giving superuser authority, do not use 0 as the UID.
	
	When defining a UID, consider the following:
	• **UID** should be unique so that it is easily identifiable. The number can be very high or very low.
	• **HOME** — Use one of the following options when defining the home directory for the default user:
	– Define the HOME directory as the root (/). The users do not have write access. They do not need to update their home directory.
	– Define the HOME directory in the /tmp directory.
	– Define a directory as you would for any other user. This directory is then used concurrently by many users that do not have an OMVS segment. (*Not recommended*)
	• **PROGRAM** defines the default shell in this field.
	
	The other fields related to this UID are not likely to be used for anything.

Table 5. Setting up default of OMVS segment (continued)

Task	Details

Set up a default for the USER OMVS segment or set up a default UID and GID.	• To set up a default for the USER OMVS segment only, create a facility class profile named BPX.DEFAULT.USER, and then specify the default UID in the application data field. Use the following commands:
	`RDEFINE FACILITY BPX.DEFAULT.USER APPLDATA('DEFUSR')`
	`SETROPTS RACLIST(FACILITY) REFRESH`
	
	Note: You cannot set up a default GROUP OMVS segment alone.
	• To set up a default UID and GID, create a facility class profile named BPX.DEFAULT.USER, and then specify the default UID and GID in the application data field. Use the following commands:
	`RDEFINE FACILITY BPX.DEFAULT.USER APPLDATA('DEFUSR/DEFGRP')`
	`SETROPTS RACLIST(FACILITY) REFRESH`
	
	Be aware that the facility class must be activated. In addition, the USER profile of the default UID and the GROUP profile of the default GID must exist, and must contain OMVS segments with a UID and GID, respectively.
	Note: RACF does not check to ensure that the application data points to a valid UID or UID and GID, or that the USER and GROUP profiles contain OMVS segments with the required UID and GID.

The following process shows how the BPX.DEFAULT.USER facility class profile works:

1. A user requests a UNIX service, which is serviced by the kernel.
2. The kernel calls the security product to extract the UID, GID, HOME, and PROGRAM information.
3. The security product attempts to extract the OMVS segment associated with the user. If the user is not defined, the security product attempts to extract and use the OMVS segment for the default user that was listed in the BPX.DEFAULT.USER profile.

A similar process is followed to obtain a GID when the user default group does not have an OMVS segment.

Authorization of TCP/IP started task user ID

The TCP/IP address space operates as a transport provider for the INET physical file system. For this to occur, the TCP/IP system address space must connect to z/OS UNIX and become a z/OS UNIX process. Therefore, the started task UID that is assigned to the TCP/IP system address space must have a valid OMVS segment.

As a transport provider, the TCP/IP address space requires superuser privileges in z/OS UNIX. Define the TCP/IP system address space started task UID as UID=0, or define the TCP/IP system address space as a trusted environment in the RACF started class profile for the TCP/IP system address space. Use the following command to assign an OMVS segment to the TCP/IP started task user ID specified as UID=0:

```
ALU tcpip_userid OMVS(UID(0) HOME(/) PGM(/bin/sh))
```

Other user IDs requiring z/OS UNIX superuser authority

When a started procedure is used to start the following servers, daemons, and agents, the user must be a superuser [UID(0)] or permitted to BPX.SUPERUSER class profile.

- Domain Name System (DNS) BIND v4 server
- File Transfer Protocol (FTP) daemon
- Network Print Facility (NPF) queue manager
- OMPROUTE server
- Sendmail Mail Transfer Agent (MTA)
- SNMP agent (OSNMPD)
- TN3270 Telnet server

When the BIND v9 DNS server is started from either a start procedure or from the z/OS UNIX shell command line, the assigned user ID must be a superuser [UID(0)].

The following daemons are managed by the inetd server, and the user specified in file /etc/inetd.conf must be defined to RACF with UID(0). For details on inetd, refer to *z/OS UNIX System Services Planning*. For details on individual daemons, refer to the *z/OS Communications Server: IP Configuration Reference*.

- z/OS UNIX remote execution daemon (REXECD)
- z/OS UNIX remote shell daemon (RSHD)
- z/OS UNIX Telnet daemon

BPX.DAEMON facility class

Certain z/OS Communications Server TCP/IP Services servers need to change the security environment of the process in which they currently execute. For example, the FTPD daemon creates a new z/OS UNIX process for every FTP client connecting to it. After the new process is created, the daemon changes the security environment of the process so that it is associated with the security context of the logged-in user. The RACF facility class resource BPX.DAEMON is used for this purpose.

Table 6. BPX.DAEMON

Task	Details
Decide if you want to activate the BPX.DAEMON level of security by reviewing the section about BPX.DAEMON authority in *z/OS UNIX System Services Planning* to determine whether this level of security is appropriate for your installation.	This is not required. It is recommended, however, because it provides additional security in the z/OS UNIX environment. The following TCP/IP Services servers and daemons in z/OS Communications Server change the security environment of their processes: • z/OS UNIX TELNETD • z/OS UNIX RSHD • z/OS UNIX REXECD • FTPD
Plan the time at which you define BPX.DAEMON carefully.	As soon as you define the BPX.DAEMON resource, MVS will not let programs change the security environment unless the programs are retrieved from a program-controlled library and unless the UID under which the program executes has access to BPX.DAEMON.

Table 6. BPX.DAEMON (continued)

Task	Details
If you decide *not* to define the BPX.DAEMON facility class, assign UID(0) for the UIDs associated with these servers and daemons.	This is sufficient for processing. It is described in "Other user IDs requiring z/OS UNIX superuser authority" on page 66.
If you decide **to** define the BPX.DAEMON facility class, grant READ access to this profile for the UIDs associated with the listed daemons. Also, enable BPX.DAEMON security by defining the BPX.DAEMON facility class profile in RACF	To define the BPX.DAEMON facility class profile in RACF, use the following command: `RDEFINE FACILITY BPX.DAEMON UACC(NONE)` **Note:** You must specify the name BPX.DAEMON in this command. Substitutions for the name are not allowed.

If all the required conditions are not met, your server programs will fail as soon as you define BPX.DAEMON. If the server programs fail, delete BPX.DAEMON, and the setup reverts to its previous state. Check all your definitions, and make the required corrections before trying to define BPX.DAEMON again.

If this is the first facility class profile that your installation is using, activate the facility class using the following commands:

```
SETROPTS CLASSACT(FACILITY) GENERIC(FACILITY) AUDIT(FACILITY)
SETROPTS RACLIST(FACILITY)
```

If you start server programs using MVS start commands or from shell scripts that execute after startup of z/OS UNIX, you must allow the UIDs access to the BPX.DAEMON facility class resource. The following example shows the UID (ftpd_user_ID) with which you can start the FTPD daemon:

```
PERMIT BPX.DAEMON CLASS(FACILITY) ID(ftpd_user_ID) ACCESS(READ)
```

Authorization to change the user security environment is granted only if both of the following two conditions are true:

- The server program is executing under a UID that has READ permission to the BPX.DAEMON facility class profile and a UID=0.
- All programs running in the address space have been retrieved from a controlled library. Program control is discussed in the following section.

Program control

In a z/OS UNIX environment, there are additional security concerns related to the HFS and the loading of programs that are considered trusted. Program control facilities in RACF and z/OS UNIX provide a mechanism for ensuring that the z/OS UNIX program loading process has the same security features that APF authorization provides in the native MVS environment.

It is recommended that you enable program control in your installation. If you define the BPX.DAEMON facility class, then you *must* enable program control for certain z/OS Communications Server load libraries. Review the section on program control in *z/OS UNIX System Services Planning* to decide whether program control is appropriate for your installation.

To enable program control, follow the tasks in the following table.

Table 7. Program control

Task	Details
Activate program control.	Use the following command: `SETROPTS WHEN(PROGRAM)`

Set the universal access for public library data sets (those in LINKLST*xx*) to READ. This allows access to the controlled programs and any other program in those libraries. (MVS opens the LNKLST*xx* libraries during IPL and makes these programs public. However, users cannot make changes.)	Use the following commands to create RACF data set profiles:
	`ADDSD 'cee.version.SCEERUN' UACC(READ)`
	`ADDSD 'SYS1.LINKLIB' UACC(READ)`
	`ADDSD 'TCPIP.SEZALOAD' UACC(READ)`
	`ADDSD 'TCPIP.SEZATCP' UACC(READ)`

Ensure all load modules that are loaded by the BPX.DAEMON servers into an address space come from controlled libraries.	If the MVS contents supervisor loads a module from a noncontrolled library, the address space becomes *dirty* and loses its authorization. To prevent this from happening, define all the libraries from which load modules can be loaded as program controlled. At a minimum, this should include the C run-time library, the TCP/IP Services SEZALOAD and SEZATCP libraries, SYS1.LINKLIB, and any load libraries containing FTP security exits.
	
	Use the following commands:
	`RDEFINE PROGRAM * ADDMEM('SYS1.LINKLIB'/'volser'/NOPADCHK) UACC(READ)`
	`RDEFINE PROGRAM * ADDMEM('SYS1.SIEALNKE'/'volser'/NOPADCHK) UACC(READ)`
	`RALTER PROGRAM * ADDMEM('cee.version.SCEERUN'/'volser'/NOPADCHK) UACC(READ)`
	`RALTER PROGRAM * ADDMEM('TCPIP.SEZALOAD'/'volser'/NOPADCHK) UACC(READ)`
	`RALTER PROGRAM * ADDMEM('TCPIP.SEZATCP'/'volser'/NOPADCHK) UACC(READ)`
	`RALTER PROGRAM * ADDMEM('db2.DSNLOAD'/'volser'/NOPADCHK UACC(READ)`
	`RALTER PROGRAM * ADDMEM('db2.DSNEXIT'/'volser'/NPPADCHK UACC(READ)`
	`RALTER PROGRAM * ADDMEM('ftp.userexits'/'volser'/NOPADCHK UACC(READ)`
	
	Note: If you define the load libraries as controlled, do not specify a universal access of NONE for the PROGRAM resources. If you do so for your SYS1.LINKLIB programs, you cannot IPL your z/OS system. Be aware also that in z/OS, the volser specification is optional.

Activate RACF changes.	Use the following command: `SETROPTS WHEN(PROGRAM) REFRESH`

Defining TCP/IP as a UNIX System Services physical file system (PFS)

The TCP/IP services stack in z/OS Communications Server must be defined as a z/OS Communications Server UNIX System Services PFS before it can be started. This involves updating the BPXPRM*xx* parmlib member. The following sample definition in BPXPRM*xx* defines TCP/IP as a z/OS Communications Server UNIX System Services PFS, where the network layer is IP Version 4 (IPv4) and communication at the sockets layer is through the AF_INET family:

```
FILESYSTYPE TYPE(INET)  ENTRYPOINT(EZBPFINI)

NETWORK DOMAINNAME(AF_INET)
        DOMAINNUMBER(2)
        MAXSOCKETS(60000)
        TYPE(INET)
```

The sample definition above shows how to define a single TCP/IP stack as IPv4 only. To define a single TCPIP stack as both IPv4 and IPv6, add an additional NETWORK statement in the BPXPRM*xx* member. The following sample definition

in BPXPRM*xx* defines TCP/IP as a z/OS Communications Server UNIX System Services PFS, where the network layer is IP Version 6 (IPv6) and communication at the sockets layer is through the AF_INET6 family:

```
NETWORK DOMAINNAME(AF_INET6)
        DOMAINNUMBER(19)
        MAXSOCKETS(60000)
        TYPE(INET)
```

The BPXPRM*xx* member contains additional z/OS Communications Server UNIX System Services parameters that are crucial to the proper operation of TCP/IP. Carefully examine and specify these parameters:

- MAXPROCSYS — Specifies the maximum number of z/OS UNIX processes that the system allows.
- MAXPROCUSER — Specifies the maximum number of processes associated with a single z/OS Communications Server UNIX System Services user ID.
- MAXUIDS — Specifies the maximum number of z/OS UNIX user IDs that can operate concurrently.
- MAXFILEPROC — Specifies the maximum number of z/OS Communications Server UNIX System Services file descriptors a z/OS Communications Server UNIX System Services process can allocate. This includes access to both HFS files and z/OS Communications Server UNIX System Services socket descriptors. In z/OS Communications Server, most TCP/IP applications access z/OS Communications Server UNIX System Services sockets, either directly or indirectly, using the TCP/IP socket APIs. You should set the MAXFILEPROC value high enough to accommodate the largest number of sockets a single TCP/IP application (or z/OS Communications Server UNIX System Services process) can allocate.

  Be aware that the tn3270 Telnet server is exempt from the limit specified in this parameter. The tn3270 Telnet server can obtain the maximum number of socket connections for a single z/OS Communications Server UNIX System Services process.
- MAXPTYS — Specifies the maximum number of pseudo-terminals for the system.
- MAXTHREADTASKS — Specifies the maximum number of MVS tasks that a single process can have concurrently active.
- MAXTHREADS — Specifies the maximum number of threads that a single process can have concurrently active.
- MAXQUEUEDSIGS — The sum of MAXQUEUEDSIGS and MAXFILEPROC multiplied by 2 is the system wide maximum for the total number of asynchronous z/OS UNIX socket calls that can be outstanding. When specifying this number, consider the following:
  - For every TCP/IP connection that the TN3270 Telnet server has, there is an asynchronous z/OS UNIX socket call outstanding. This is true for both TN3270 and TN3270E clients.
  - Any TCP/IP application, IBM or vendor supplied, that uses either the z/OS UNIX assembler callable services asyncio call or the TCPIP provided Sockets Extended asynchronous API could have one or more outstanding asynchronous socket calls.

The MAXSOCKETS() parameter specifies the total number of z/OS Communications Server UNIX System Services sockets that can be active at any one time. You must ensure that this specification is large enough to accommodate your installation's workload. For example, each connection to your tn3270 Telnet server or FTP server requires one z/OS Communications Server UNIX System

Services socket. Once the maximum number of sockets is allocated, then no more Telnet sessions, FTP sessions, or other applications that require z/OS Communications Server UNIX System Services sockets can be started.

Note: If multiple NETWORK statements are defined, MAXSOCKETS can be specified for each NETWORK statement and will be enforced separately.

References

For details on the BPXPRM*xx* member, please refer to the following guides:

- *z/OS UNIX System Services Planning*
- *z/OS MVS Initialization and Tuning Reference*
- *z/OS UNIX System Services File System Interface Reference*

Performance considerations

Follow the guidelines found in the *z/OS MVS Initialization and Tuning Reference*. If your installation is running Workload Manager, follow the guidelines found in *z/OS MVS Planning: Workload Management*.

It is necessary that VTAM, TCP/IP, and some associated server applications be able to obtain cycles to maintain their network presences. In general, VTAM and TCP/IP should have a higher dispatching priority than the applications that use their services. Server applications such as OMPROUTE, TN3270 Telnet, IKED, and FTPD should be set at or just below TCPIP's value. If running WLM, these tasks should be assigned to the SYSSTC service class. Additionally, making these tasks non-swappable will assure that they will be available during periods of high CPU usage. The MVS default program property table sets Telnet non-swappable and privileged, which automatically assigns the task to the SYSSTC service class. Non-critical applications such as Policy Agent and TRMD should be set to a lower priority.

Fast path support

For applications that have extremely strict communications path-length requirements, an optional extension has been provided to further reduce overhead resulting from the z/OS UNIX-to-TCP/IP stack communications. This extension is only available to applications using the UNIX System Services Callable Services Socket API or the C/C++ socket API supported by the Language Environment. It is not available to applications using the native MVS socket APIs (such as C/C++, EZASMI macro, EZASOKET, REXX, or CICS socket APIs) provided by the Communications Server. Exploitation of this extension is entirely optional.

This feature can be activated for an entire z/OS UNIX process using the z/OS UNIX _BPXK_INET_FASTPATH environmental variable. The value of this variable determines whether a socket application is marked *fast path*. A C/C++ Language Environment application can set the variable by invoking the setenv() service, or you can export the variable to the z/OS UNIX shell environment before the socket application is invoked. An application using the z/OS UNIX Callable Services APIs can set this variable using the BPX1ENV service.

Note: z/OS UNIX environmental variables have a process-wide scope only—that is, they usually affect a single MVS address space only. It is possible, however, to have multiple UNIX processes within a single address space. In this scenario, the setting of this environmental variable might vary for each process within the address space. It is not a problem if some of your

applications exploit fast path services, while others do not. When a socket application is marked as fast path, the communications overhead is reduced on the following socket syscalls:

- - send()
- - recv()
- - sendto()
- - recvfrom()
- - sendmsg()
- - recvmsg()

Although applications are more efficient when using the environmental variable, they are not XPG compliant, and POSIX signals are not supported. Applications can be interrupted only with the SIGKILL terminating signal, and they cannot be debugged using the interactive z/OS UNIX dbx debugger. You can, however, develop and test an application using the dbx debugger without setting the environmental variable, and then execute the application in production with the environmental variable set. Also, note that applications using the z/OS UNIX asynchronous socket interface (BPX1AIO) to invoke synchronous socket operations (that is, setting the AioSync bit in the AIOCB) cannot use the BPX1AIO service to cancel outstanding synchronous calls on sockets that are marked as fast path. Doing so will cause the cancel operation to hang.

For environments that do not use common INET, the value of this variable should be set to the name specified on the FILESYSTYPE TYPE() parameter in the BPXPRMxx parmlib member.

For common INET environments, the value used to set the environmental variable depends on whether the application is using the TCP or UDP protocols. In a common INET environment, the variable should be set as follows:

- For UDP applications, it should be set to the name of the TCP/IP stack as specified on the SUBFILESYSTYPE NAME() parameter in the BPXPRMxx parmlib member. The socket application is explicitly associated with the TCP/IP stack named in the environmental variable (that is, the TCP/IP stack name). This means that the socket application can communicate with partners that are accessible only through the specific TCP/IP stack interfaces. For UDP, the environmental variable effectively overrides the support provided by common INET. You should take this contingency into account before activating fast path for a UDP-based application.

  Note that if the UDP application already establishes affinity to a specific TCP/IP stack using other means, such as setting the _BPXK_SETIBMOPT_TRANSPORT environment variable, using setibmopt(), BPX1PCT, and so on, the setting of the fast path variable is ignored. As a result, UDP applications that require fast path support and affinity to a specific TCP/IP stack must do so using the _BPXK_INET_FASTPATH environmental variable.
- For TCP applications, the variable can be set to an asterisk (*), indicating that any TCP/IP stack in the common INET configuration can be used. This allows all TCP/IP stacks that support the fast path model to obtain the fast path performance benefits automatically. TCP servers are not bound to a specific TCP/IP stack, even if they specify a specific TCP/IP stack name on the environmental variable; instead, they can listen for inbound connections across all TCP/IP stacks. When a connection arrives from the TCP/IP stack named in the environmental variable [at the time of the accept()], it is automatically marked as fast path. Connections that arrive from TCP/IP stacks that are not named by the current environmental variable value are not marked as fast path.

Note, however, that certain TCP/IP API functions, such as the resolver services [that is, gethostbyname(), gethostbyaddr(), getaddrinfo(), and getnameinfo()] and the network interface identification services [that is, if_nameindex(), if_nametoindex(), and if_indextoname()] use UDP sockets internally to perform their processing. Consequently, if a specific TCP/IP stack name is specified on the environmental variable, these hidden UDP sockets will only be associated with the named TCP/IP stack, which might have undesirable effects. For example, any resolver API queries resulting in communications with a domain name server will occur only over the specified TCP/IP stack. As a result, it is strongly recommended that TCP applications set the environmental variable to the special asterisk (*) value. If the application requires affinity to a specific TCP/IP stack, it should do so using any of the facilities that are provided by z/OS UNIX, such as setibmopt(), BPX1PCT, and so on. For more details on establishing affinity to a specific TCP/IP stack, refer to *z/OS UNIX System Services Planning*.

Applications can also enable fast path processing for a single socket by issuing the Iocc#FastPath IOCTL for the socket, using the w_ioctl() or the BPX1IOC APIs. Note that this IOCTL is only effective if it is issued against a socket that is already associated with a specific TCP/IP stack. Sockets are considered associated with a specific TCP/IP stack if they meet any of the following conditions:

- The application has explicit process affinity to a specific TCP/IP stack [that is, by setting the _BPXK_SETIBMOPT_TRANSPORT environmental variable, using setibmopt(), BPX1PCT, and so on].
- TCP/IP stack affinity has been explicitly established for this socket (that is, using the SIOCSETRTTD IOCTL).
- A bind() has already been issued for the socket using a specific IP address (that is, not INADDR_ANY).
- A TCP (that is, streams) socket that is connected. This includes TCP sockets that are returned as a result of accept() or sockets that a connect() was issued for.

The Iocc#FastPath constant is defined in the BPXYIOCC. Note that this IOCTL requires a 4-byte argument as input. This argument should be set to a nonzero value to activate fast path, or a zero value to disable fast path on the specified socket.

Considerations for multiple instances of TCP/IP

The z/OS Communications Server TCP/IP stack is a multiple-processor capable stack, which means that it can concurrently exploit all available processors on a system. Starting multiple stacks will not yield a significant increase in throughput.

In addition, running multiple z/OS Communications Server TCP/IP stacks requires additional system resources, such as storage, CPU cycles, and DASD. It also adds a significant level of complexity to the system administration tasks for TCP/IP.

For these reasons, it is suggested that in most cases you use the INET configuration, which supports a single TCP/IP stack. However, there are some special situations where running multiple stacks can provide a benefit. For example, you might want to run two separate stacks for intranet and Internet traffic, or AnyNet Sockets over SNA in conjunction with one or more TCPIP stacks.

Common INET physical file system (CINET PFS)

If you wish to run multiple z/OS Communications Server TCP/IP stacks concurrently, you must use the Common INET (CINET) configuration. In this configuration, up to a maximum of eight TCP/IP stacks can be active at any time.

When the CINET configuration is used, the CINET PFS is inserted between the LFS and the TCP/IP PFS for each stack. The CINET PFS maintains an internal copy of each TCP/IP stack's IP configuration, so that it can preroute a socket call to the correct TCP/IP stack. This allows most socket programs to run with multiple stack support with no change to the application. In addition, CINET supports IPv6, and is capable of supporting underlying TCP/IP stacks in IPv4/IPv6 dual mode or in IPv4-only mode.

You can specify your choice of INET (single stack) or CINET (multiple stack) support on the NETWORK, DOMAINNAME, FILESYSTYPE, and SUBFILESYSTYPE statements of SYS1.PARMLIB(BPXPRMxx). For more information about the BPXPRMxx statements, refer to "Specifying BPXPRMxx values for a CINET configuration" on page 82 and *z/OS UNIX System Services Planning*.

Port management overview

When there is a single transport provider, and the relationship of server to transport provider is 1:1, port management is relatively simple. Using the PORT statement, the port number can be reserved for the server in the PROFILE.TCPIP for that single transport provider.

Port management becomes more complex in a CINET environment where there are multiple transport providers (multiple instances of TCP/IP) and a potential for multiple combinations of the same server (for example, z/OS UNIX and TN3270/TN3270E Telnet).

In a multiple transport provider environment, the following questions need to be answered for each server in an installation:

- Is the server generic so that it can communicate with multiple TCP/IPs or does the server have an affinity for one instance of the transport providers and can only communicate with one TCP/IP?
- How can ports be reserved across multiple transport providers? When is the port reservation determined by MVS rather than by the job name, procedure name, or user ID?
- How can you synchronize between BPXPARMS and PORTRANGE for ephemeral port reservation?
- How can TCP/IP distinguish between two different instances of Telnet (z/OS UNIX Telnet and TN3270/TN3270E Telnet)?

Generic server versus server with affinity for a specific transport provider

The following sections describe the differences between generic servers and servers with affinities for specific transport providers.

Generic server: A generic server, a server without an affinity for a specific transport provider, provides service to any client on the network. (See Figure 4 on page 74.) FTP is an example of a generic server. The transport provider is merely a connection linking client and server. The service File Transfer is not related to the

internal functioning of the transport provider, and the server can communicate concurrently over any number of transport providers.

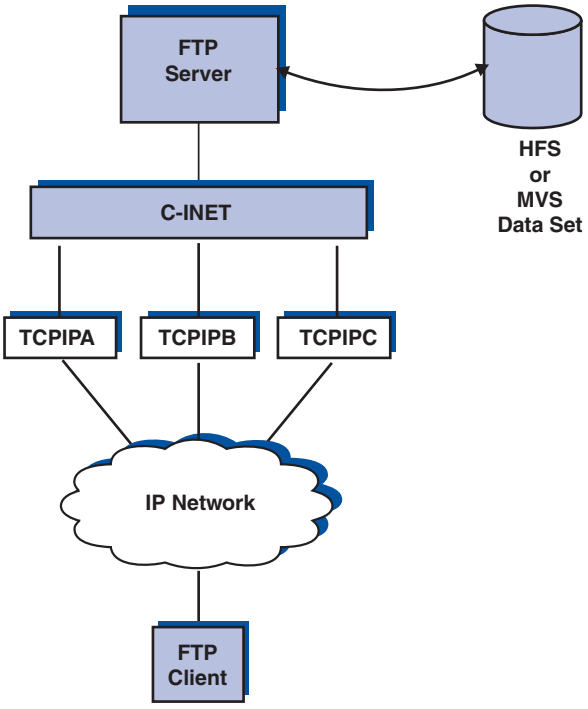


Figure 4. Generic server

Server with an affinity for a specific transport provider: When the service is related to the internal functioning of the transport provider (for example, Telnet, OMPROUTE, OSNMPD, and the Netstat command), there must be an explicit binding of the server application to the chosen transport provider. (See Figure 5 on page 75.) There must also be a way to specify the single transport to be chosen.

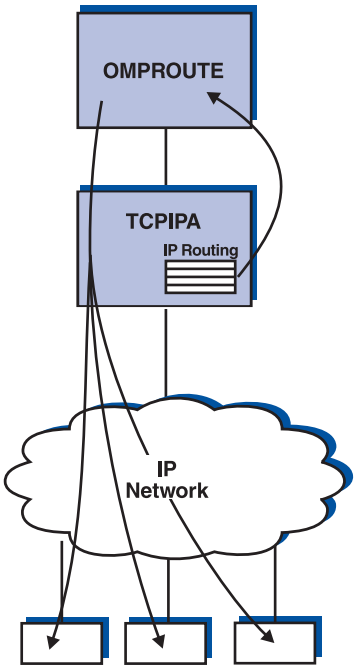


Figure 5. Server with affinity for a specific transport provider

With the exception of applications that use the socket API provided by TCP/IP, other IBM-supplied applications that use the z/OS UNIX socket API and that must bind to a specific transport provider use the z/OS UNIX socket call `setibmopt()` (refer to *z/OS XL C/C++ Run-Time Library Reference*) to specify which TCP they have chosen. A C function `__iptcpn()`, described in the *z/OS XL C/C++ Run-Time Library Reference*, enables the application to search the TCPIP.DATA file to find the name of the specific TCP/IP. (See Figure 6.) An application that uses the z/OS Language Environment runtime can also establish stack affinity by setting the environment variable _BPXK_SETIBMOPT_TRANSPORT.

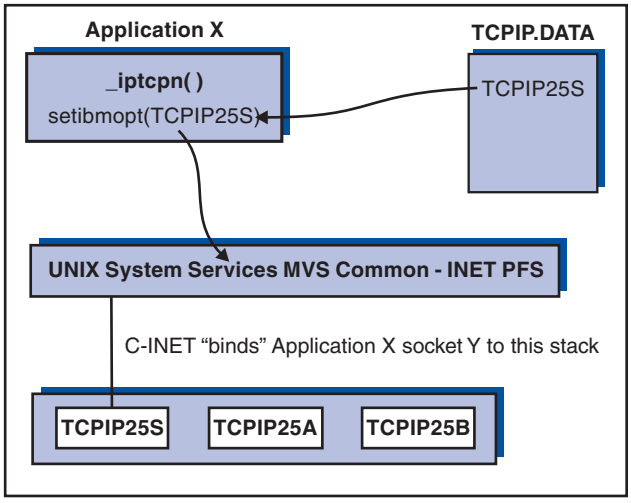


Figure 6. Example of binding an application to a specific transport provider

Generic servers in a CINET environment

In z/OS Communications Server, you can configure multiple TCP/IP stacks in a single MVS image using the CINET feature. In a CINET configuration, an application using the z/OS UNIX socket interface can get transparent access to all the TCP/IP protocol stacks configured under CINET. For example, when an application coded to z/OS UNIX sockets performs a SOCKET/BIND/LISTEN in a CINET environment, the request is propagated by CINET to all the TCP/IP stacks. This application can then service client requests that arrive into any of the configured TCP/IP stacks without having any awareness of this fact. This type of application is often referred to as a *generic server* or *daemon*.

The following servers or daemons shipped by z/OS Communications Server are generic:

- DCAS
- FTPD
- SNTPD
- syslogd
- Telnet in its own address space
- TFTPD
- TIMED
- z/OS UNIX POPPER
- z/OS UNIX Portmap
- z/OS UNIX REXECD
- z/OS UNIX RSHD
- z/OS UNIX SENDMAIL
- z/OS UNIX TELNETD

z/OS UNIX RSHD, REXECD and TELNETD are usually started by the INETD daemon, which is shipped as part of the z/OS UNIX. Because INETD is also a generic daemon, any server processes started by INETD inherently become generic servers as well.

If a server started by INETD (a generic server) requires affinity to a specific stack, this affinity can be accomplished by use of the _BPXK_SETIBMOPT_TRANSPORT environment variable. For more information about the _BPXK_SETIBMOPT_TRANSPORT environment variable refer to *z/OS UNIX System Services Planning*.

The _BPXK_SETIBMOPT_TRANSPORT environment variable, when set, has an effect similar to the setibmopt() function call provided by the XL C/C++ compiler and described in the *z/OS XL C/C++ Run-Time Library Reference*. This variable can be set in the JCL for a started procedure or batch job that executes a z/OS UNIX C/C++ program to indicate which TCP/IP stack instance the application should bind to. TCP/IP applications that require affinity to a specific TCP/IP stack, like OSNMPD and OMPROUTE, use the setibmopt() function call directly. The _BPXK_SETIBMOPT_TRANSPORT environment variable basically provides the ability to bind a generic server type of application to a specific stack.

For example, if you had two TCP/IP stacks configured under CINET, one named TCPIP and the other TCPIPOE, and you wanted to start an FTPD server instance that was associated with TCPIPOE, you could modify the FTPD procedure as follows:

```
//FTPD   PROC MODULE='FTPD',PARMS='TRACE'
//FTPD   EXEC PGM=&MODULE,REGION=7M,TIME=NOLIMIT,
//       PARM=('POSIX(ON) ALL31(ON)',
//      'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE")',
//      '/&PARMS')
//CEEDUMP  DD SYSOUT=*
//SYSFTSX DD DISP=SHR,DSN=TCPV34.STANDARD.TCPXLBIN
```

All the parameters specified prior to the slash (/) in the parameter statement are processed by the XL C/C++ run-time library. Parameters to be passed to the FTPD program must appear after the slash (/). Also note how the parameters were split over three lines in this example because they could not fit on a single line.

The following example uses JCL for the started procedure for INETD:

```
//INETD  PROC
//*************************************************************
//INETD  EXEC PGM=BPXBATCH,
//*      PARM='PGM /usr/sbin/inetd -d /etc/inetd.conf'
//       PARM='PGM /usr/sbin/inetd    //''USER1.INETD.CONF'''
//*
//STDERR  DD PATH='/tmp/inetd.debug.stderr',
//  PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//  PATHMODE=SIRWXU
//STDOUT  DD PATH='/tmp/inetd.debug.stdout',
//  PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//  PATHMODE=SIRWXU
//STDENV   DD  DISP=SHR,DSN=USER1.INETD.ENVIRON
```

The STDENV data set would contain the _BPX_SETIBMOPT_TRANSPORT variable as follows:

```
_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE
```

In the previous examples, INETD was also passed its configuration file as a parameter. In our examples, this file is an MVS data set rather than an HFS file; therefore, it requires the additional double slash (//) and quotes that the example shows.

Multiple instances of INETD are not allowed, even if each instance is bound to a different TCP/IP stack. This is an INETD restriction, not a TCP/IP restriction. Therefore, if you decide to make INETD have affinity to a specific stack, then that is the only INETD instance that you will be able to have running in that MVS image.

Notes:

1. The _BPXK_SETIBMOPT_TRANSPORT variable should be specified only for a generic server type of application.

   If specified for a non-generic server and/or non-z/OS UNIX application it will not have any effect.
2. The name specified for _BPXK_SETIBMOPT_TRANSPORT must match the job name associated with the TCP/IP stack.

   If the name specified does not match the job name of any TCP/IP stacks defined for CINET, the application will receive a z/OS UNIX return code of X'3F3' and a return value of X'005A' and may be accompanied by the following message:

   ```
   EDC8011I A name of a PFS was specified that either is
   not configured or is not a Sockets PFS.
   ```

If the name specified does not match the job name of any currently active TCP/IP stack defined under CINET, the application will receive a z/OS UNIX return code of X'70' and a return value of X'0296' and may be accompanied by the following message:

```
EDC5112I Resource temporarily unavailable.
```

3. For more detailed information about requesting transport affinity, refer to *z/OS UNIX System Services Planning*.

Port reservation across multiple transport providers

When there are multiple transport providers, be sure to synchronize the PORT statements in each of the PROFILE.TCPIP files to ensure that the port reservations for each stack match the port definitions for the servers that will be using that stack.

For more information about reserving ports with the PORT statement, see Chapter 5, "Customization," on page 175.

Ephemeral ports: When running with multiple transport providers, just as it is necessary to synchronize PORT reservations for specific applications across all stacks, it is required to synchronize reservations for port numbers that will be dynamically assigned across all stacks. These are the ephemeral ports above 1023, which are assigned by the stack when none is specified on the application `bind()`. To reserve a group of ports in the PROFILE.TCPIP, use PORTRANGE. For more information about PORTRANGE, see Chapter 5, "Customization," on page 175. Specify the same PORTRANGE for every stack. In addition, you need to let the z/OS UNIX CINET know which ports are guaranteed to be available on every stack. The following is an example of reserving ports 40000 to 41999 in the two required files:

- PROFILE.TCPIP
  - PORTRANGE 40000 2000 TCP OMVS ; Reserved for OMVS
  - PORTRANGE 40000 2000 UDP OMVS ; Reserved for OMVS
- BPXPRMxx parmlib member
  - NETWORK DOMAINNAME(AF_INET)
  - INADDRANYPORT(40000)
  - INADDRANYCOUNT(2000)

Note: When IPv6 is configured and there are two NETWORK statements, INADDRANYPORT and INADDRANYCOUNT only need to be specified for the NETWORK statement for AF_INET and not for AF_INET6. If they are specified for AF_INET6, they are ignored and the values from the NETWORK statement for AF_INET are used if provided. Otherwise, the default values are used.

Selecting a stack when running multiple instances of TCP/IP

Socket application programs in a multi-stack (CINET) environment must contend with the following:

- How the socket program selects which TCP/IP stack to use for its socket communication
- How the TCP/IP resolver code executing in the socket application address space decides which TCP/IP resolver configuration data sets to allocate

Note: If a resolver GLOBALTCPIPDATA setup file is used, a local TCPIP.DATA cannot override any explicit statements in the global file and cannot

override any resolver statements. Therefore, in a CINET environment, the TCPIPJOBNAME statement should not be specified in the GLOBALTCPIPDATA file. Also, using the GLOBALTCPIPDATA file with CINET requires that the resolver TCPIP.DATA statements are able to be used by all stacks. For example, the IP addresses specified by the NameServer statement must be accessible from all stacks. If they are not, then the GLOBALTCPIPDATA file should not be used and you should continue with multiple TCPIP.DATA data sets. For details, see "Understanding resolvers" on page 20.

To answer these questions, a distinction must be made between standard servers and clients (those that come with the z/OS Communications Server product), and other socket application programs, including those you might have written yourself.

Standard servers and clients

The anchor configuration data set is the TCPIP.DATA data set. This is the base resolver configuration data set with information on host name, domain origin, and so on. It holds the TCPIPJOBNAME statement, which identifies the TCP/IP stack to use, and the DATASETPREFIX statement, which is used by the resolver code and other services when allocating configuration data sets. For more information on these data sets, see "Configuration files for TCP/IP applications" on page 38.

The key to selecting both a specific stack and resolver configuration data sets is to control which TCPIP.DATA data set a standard server or client address space allocates. Applications that use the z/OS UNIX API can use Common INET to determine which stack an application will use. But, it is important to ensure that the search order and the contents of the resolver configuration data set are understood.

Native MVS servers and clients search for TCPIP.DATA in sequences as described in "Search orders used in the native MVS environment" on page 50.

z/OS UNIX servers and clients will search for TCPIP.DATA in sequences as described in "Search orders used in the z/OS UNIX environment" on page 43.

Nonstandard servers and clients

Nonstandard servers and clients (those that do not come with the z/OS CS product) also use TCPIP.DATA to decide which resolver configuration data sets to allocate. Depending on the socket API used, they might or might not use the TCPIPJOBNAME parameter to select a stack.

If you run sockets programs from other products or vendors, you may want to know which sockets API was used to develop the program, and which techniques, if any, the program uses to specify the name of the TCP/IP system address space. As long as application programs that use a TCP/IP socket library do not specify anything specific on calls `setibmopt()`, `Initialize`, or `INITAPI`, the TCPIPJOBNAME from a TCPIP.DATA data set will be used for finding a TCP/IP system address space name.

Table 8 on page 80 depicts the differences that prevail in stack selection depending on the TCP/IP socket API under which you are running the socket program.

Table 8. How your own socket programs select a stack

C sockets	Callable and Macro	Pascal sockets	REXX sockets
SETIBMOPT or TCPIPJOBNAME from TCPIP.DATA	TCPNAME on INITAPI or TCPIPJOBNAME from TCPIP.DATA	TCPIPJOBNAME from TCPIP.DATA	Service on Initialize or TCPIPJOBNAME from TCPIP.DATA
Callable and Macro programs might have a configuration option to specify the TCP/IP system address space name, or might interrogate the available stacks via the getibmopt() call.			

A Callable or Macro program does not have to call INITAPI. If INITAPI is not called, an implicit INITAPI is performed with the value taken from TCPIPJOBNAME in a TCPIP.DATA data set. If INITAPI is called with the TCPNAME parameter specified as a space, the TCP/IP system address space name results in the TCPIPJOBNAME keyword value.

In a z/OS UNIX INET (single stack) environment, the socket application program is always associated with the single TCP/IP stack. In the z/OS UNIX Common INET (CINET) environment, your application will be associated with multiple TCP/IP stacks unless the application specifically associates with a particular stack using the z/OS UNIX socket call setibmopt(). For other ways of requesting stack affinity in a CINET environment, refer to *z/OS UNIX System Services Planning*.

TCP/IP TSO clients

TSO client functions can be directed against any of a number of TCP/IP stacks. Obviously, the client function must be able to find the TCPIP.DATA appropriate to the stack of interest at any one time. Some TSO client commands provide a parameter to specify the stack to be used. For those that do not, the following methods are available for finding the relevant TCPIP.DATA:

- Add a SYSTCPD DD statement to your TSO logon procedure. The issue with this approach is that a separate TSO logon procedure per stack is required, and users have to log off TSO and log on again using another TSO logon procedure in order to switch from one stack to another.
- Use one common TSO logon procedure without a SYSTCPD DD statement. Before a TSO user starts any TCP/IP client programs, the user has to issue a TSO ALLOC command wherein the user allocates a TCPIP.DATA data set to DD name SYSTCPD. To switch from one stack to another, the user simply has to deallocate the current SYSTCPD allocation (for example, TSO FREE command) and allocate another TCPIP.DATA data set.
- Combine the first and second methods. Use one logon procedure to specify a SYSTCPD DD for a default stack. To switch stacks, issue TSO ALLOC to allocate a new SYSTCPD. To switch back, issue TSO ALLOC again with the name that was on the SYSTCPD DD in the logon procedure. The disadvantage to this approach is that the name that was on the SYSTCPD DD is hidden in the logon procedure and needs to be retrieved or remembered.

The last method can be implemented by creating a small REXX program for every TCP/IP stack on your MVS system. For each stack create a REXX program with the name of the stack (for example, T18A or T18B). Whenever TSO users want to use the T18A stack, they run the T18A REXX program. Any TCP/IP functions invoked thereafter will use the T18A stack for socket communication. If users want to switch to the T18B stack, they run the T18B REXX program. See Figure 7 on page 81 for an example.

```
/* REXX  "T18B" */
/*********************************************************************/
/*                                                                   */
/* Switch TSO Address Space to use the T18B Stack.                   */
/* Subsequent NETSTAT command will be directed toward                 */
/*  the T18BTCP stack.                                               */
/*                                                                   */
/*********************************************************************/
Say 'Switching to T18BTCP stack'

msgstat = msg()
z = msg("OFF")
"FREE FI(SYSTCPD)"
"ALLOC FI(SYSTCPD) DA('TCPIP.T18B.TCPPARMS(TCPDATA)') SHR"
z = msg(msgstat)

exit(0)
```

Figure 7. REXX program to switch TSO user to another TCP/IP stack

Selecting configuration data sets

The resolver code and other services that execute as part of the socket program address space to service calls such as `gethostbyname()`, `getservbyname()` and `getprotobyname()` allocate one or more resolver configuration files to service these calls. All socket programs, including standard servers and clients and homegrown socket programs, need access to resolver configuration files. For information on how the resolver configuration files are found and used, see "Configuration files for TCP/IP applications" on page 38.

Sharing resolver configuration data sets

The general recommendation is to use separate DATASETPREFIX values for each stack and create separate copies of the required configuration data sets; at the very least, create separate copies of the resolver configuration data sets. For a test and a production stack, however, you would probably use different DATASETPREFIX values. However, if the stacks are functionally identical, you may share the same DATASETPREFIX values and many of the same configuration data sets. You need separate TCPIP.DATA data sets because of the two different TCPIPJOBNAMEs. On the other hand, you may choose to share the resolver configuration data sets between the stacks by using the same DATASETPREFIX value in each TCPIP.DATA data set.

In addition to separate TCPIP.DATA data sets, separate /etc/resolv.conf files might also be necessary. If this is the case, use the environment variable RESOLVER_CONFIG to point to the appropriate resolver information.

Exercise caution if servers use DATASETPREFIX to allocate server-specific configuration data sets. Try to use explicit allocation as far as possible in your server JCL procedures. Most servers allow you to explicitly allocate their configuration data sets using DD statements.

Some servers may use DATASETPREFIX to create new data sets. Servers that do create new data sets allow you to specify an alternate data set prefix for the data sets that are created. NPF creates new sequential data sets with captured print data. NPF has a special keyword in NPF.DATA for this purpose; it is called

NPFPRINTPREFIX. If this keyword is specified, NPF will use that as the high-level qualifier for newly created print data sets instead of taking the DATASETPREFIX value from TCPIP.DATA. Another example of a server that creates new data sets is the SMTP server.

Specifying BPXPRMxx values for a CINET configuration

For a detailed description of parameters in SYS1.PARMLIB(BPXPRMxx), refer to *z/OS UNIX System Services Planning* and *z/OS MVS Initialization and Tuning Guide*.

```
/* AF_INET file system for sockets      */
/* CINET support - BPXTCINT  */

FILESYSTYPE TYPE(CINET)
            ENTRYPOINT(BPXTCINT)  1
NETWORK DOMAINNAME(AF_INET)
        DOMAINNUMBER(2)
        MAXSOCKETS(10000)  2          TYPE(CINET)
        INADDRANYPORT(40000)  3
        INADDRANYCOUNT(2000)
NETWORK DOMAINNAME(AF_INET6) 4
        DOMAINNUMBER(19)
        MAXSOCKETS(10000)  2          TYPE(CINET)
SUBFILESYSTYPE NAME(TCPIP1A)  5
               TYPE(CINET)
               ENTRYPOINT(EZBPFINI)  6
               DEFAULT  7
SUBFILESYSTYPE NAME(TCPIP1B)  5
               TYPE(CINET)
               ENTRYPOINT(EZBPFINI)
```

Figure 8. SYS1.PARMLIB(BPXPRMxx) for CINET

1 CINET and BPXTCINT specify the use of CINET.

2 The MAXSOCKETS operand specifies the maximum number of sockets that can be obtained for the given file system type. It should be large enough for the number of sockets needed for applications using z/OS Communications Server. MAXSOCKETS is enforced independently for AF_INET (IPv4 sockets) and AF_INET6 (IPv6 sockets).

3 INADDRANYPORT and INADDRANYCOUNT specify the first ephemeral port number and the range of ports for z/OS UNIX CINET usage. The starting port number should be set at least as high as the value for MAXSOCKETS. Setting the high port number should help ensure that the ephemeral port is available to CINET and is less likely to be used by any z/OS UNIX application. These values have to match the PORTRANGE definitions in your PROFILE data sets for both TCP/IP stacks. INADDRANYPORT and INADDRANYCOUNT should not be specified in the NETWORK statement for AF_INET6. If these parameters are specifed in the NETWORK statement for AF_INET6, they are ignored. The INADDRANYPORT and INADDRANYCOUNT values for AF_INET6 are set to the same values specifed for AF_INET.

4 This additional NETWORK statement is required if you want a TCP/IP stack to also support IPv6. Omit this statement if you do not want the stack to support IPv6 (that is, the stack will support IPv4 only).

5 A transport provider stack for CINET is specified with a SUBFILESYSTYPE statement. The NAME field must match the address space name for the TCP/IP

started task as well as the TCPIPJOBNAME parameter in TCPIP.DATA. In our example, the name of the first stack is TCPIP1A and the name of the second stack is TCPIP1B.

6 EZBPFINI identifies a z/OS Communications Server TCP/IP stack. For a z/OS Communications Server TCP/IP stack, this is the only valid value.

7 Keyword DEFAULT specifies which transport provider stack is to be used as the default stack for z/OS UNIX. If DEFAULT is not specified, the first active stack will be used as the default stack. The sequence of SUBFILESYSTYPE statements is arbitrary if one stack is identified with the keyword DEFAULT. TCPIP1A is the default stack in Figure 8 on page 82.

Considerations for Enterprise Extender

The Enterprise Extender (EE) network connection is a simple set of extensions to the existing open high-performance routing (HPR) technology. It performs an efficient integration of the HPR frames using UDP/IP packets. To the HPR network, the IP backbone is a logical link. To the IP network, the SNA traffic is UDP datagrams that are routed without any hardware or software changes to the IP backbone. Unlike gateways, there is no protocol transformation and unlike common tunneling mechanisms, the integration is performed at the routing layers without the overhead of additional transport functions. The advanced technology enables efficient use of the intranet infrastructure for support of IP-based client accessing SNA-based data (for example, TN3270 emulators or Web browsers using services such as IBM's Host On-Demand) as well as SNA clients using any of the SNA LU types.

Enterprise Extender seamlessly routes packets through the network protocol *edges*, eliminating the need to perform costly protocol translation and the store-and-forward associated with transport-layer functions. Unlike Data Link Switching (DLSw), for example, there are no TCP retransmit buffers and timers and no congestion control logic in the router because it uses connectionless UDP and the congestion control is provided end system to end system. Because of these savings, the *edge* routers have less work to do and can perform the job they do best, which is forwarding packets instead of incurring protocol translation overhead and maintaining many TCP connections. Data center routers can handle larger networks and larger volumes of network traffic, thus providing more capacity.

Enterprise Extender supports both the IPv4 and IPv6 addressing models. For more information on EE, refer to *z/OS Communications Server: SNA Network Implementation Guide*, the EE information in *Migrating Subarea Networks to an IP Infrastructure Using Enterprise Extender* (SG24–5957–00, an IBM Redbook), or the following Web site:

`http://www.ibm.com/servers/eserver/zseries/library/techpapers/enterprise_extender.html`

Considerations for VIPA

The Internet Protocol (IP) is a connectionless protocol. IP packets are routed from the originator through a network of routers to the destination. All physical adapter devices in such a network, including those for client and server hosts, are identified by an IP Address which is unique within the network. The important point about IP is that a failure of an intermediate router node or adapter will not prevent a packet from moving from source to destination, as long as there is an alternate path through the network.

TCP sets up a connection between two endpoints, identified by the respective IP addresses and a port number on each. Unlike failures of an adapter in an intermediate node, if one of the endpoint adapters (or the link leading to it) fails, all connections through that adapter fail and must be reestablished. If the failure is on a client workstation host, only the relatively few client connections are disrupted and usually only one person is inconvenienced. However, an adapter failure on a server means that hundreds or thousands of connections may be disrupted. On an S/390 or zSeries server with large capacity, the number may run to tens of thousands.

A Virtual IP Address, or VIPA in TCP/IP for z/OS , alleviates this situation. A VIPA is configured in the same way as a normal IP address for a physical adapter, except that it is not associated with any particular device. To an attached router, the TCP on z/OS simply looks like another router. When the TCP receives a packet destined for one of its VIPAs, the inbound IP function of the stack notes that the IP address of the packet is in the stack's Home list and passes the packet up the stack. Assuming the stack has multiple adapters or paths to it (including XCF from other TCP stacks in a sysplex), if a particular physical adapter fails, the attached routing network will simply route VIPA-targeted packets to the stack via an alternate route.

While this removes hardware and associated transmission media as a single point of failure for large numbers of connections, the connectivity of a server can still be lost through a failure of a single stack or an MVS image. The VIPA can be configured on another stack with a manual process, but this requires the presence of an operator or programmed automation.

Dynamic VIPA Takeover enables Dynamic VIPAs to be moved without human intervention or programmed automation to allow new connections to a server at the same IP address as soon as possible. This can reduce downtime significantly. With Dynamic VIPA Takeover you can configure one or more TCP/IP stacks to be backups (VIPABACKUP statement) for a particular Dynamic VIPA. If the stack or MVS image where the Dynamic VIPA is active is terminated, one of the backup stacks automatically activates that Dynamic VIPA. The existing connections will be terminated but can be quickly reestablished on the stack that is taking over.

Notes:

1. Because a VIPA is associated with a z/OS TCP/IP stack and is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex, or even to a z/OS TCP/IP stack not in the sysplex as long as the address fits into the installation's network configuration.
2. If using VIPA along with an intelligent bridge or switch, ensure that 'Port fast mode' (Cisco) is enabled. This helps to decrease the amount of time the VIPA is unreachable in scenarios where there is dynamic movement of VIPA (dynamic or static). For more information, see your bridge or switch manual.

You may also associate a particular Dynamic VIPA address with an application using the SIOCSVIPA or SIOCSVIPA6 ioctl command or by BINDing explicitly to the Dynamic VIPA address. If the Dynamic VIPA address is within the VIPARANGE profile statement, then this Dynamic VIPA address will be created dynamically. This type of configuration enables a Dynamic VIPA to become an address of an application in a sysplex.

With Sysplex Distributor you can spread connection requests destined for Dynamic VIPAs to other stacks in the sysplex. You can use the VIPADISTRIBUTE profile

statement to designate up to 32 stacks and 64 ports where connections for a particular DVIPA can be distributed, including the stack where the DVIPA is defined.

The distributing stack (the stack where the VIPADISTRIBUTE statement was coded) might use either WLM or a combination of WLM and Quality of Service (QoS) performance information to determine where to forward new connection requests. If the distributing stack/MVS image fails, connections forwarded to target stacks can be preserved by having the Dynamic VIPA address backed up on another stack.

Similarly, a stack can immediately take back a Dynamic VIPA address from another stack. If the original stack used an address specified with VIPADEFINE with the keyword MOVEABLE IMMEDIATE (the default), then the Dynamic VIPA is moved as soon as the second stack requests ownership. The second stack assumes responsibility for forwarding packets for existing connections to the appropriate stack. If MOVEABLE WHENIDLE was specified, ownership does not pass until all existing connections on the current stack are closed.

For detailed information about VIPA, see Chapter 7, "Virtual IP Addressing," on page 297.

Considerations for Fast Response Cache Accelerator

Fast Response Cache Accelerator (FRCA) is a Communications Server function that significantly improves the performance of the WebSphere® HTTP server on z/OS. Web pages are cached within the operating system kernel and requests are handled without traversing the entire kernel or entering the user space. For more information on configuring the WebSphere HTTP server's use of the FRCA function, refer to *z/OS HTTP Server Planning, Installing, and Using*.

FRCA also provides the ability to perform content-based quality-of-service (QoS) classification by selecting an appropriate QoS policy for each individual URI in the HTTP request. For more information on specifying individual URIs, refer to the Policy Agent and policy applications chapter of *z/OS Communications Server: IP Configuration Reference*.

Use the Netstat CACHINFO/-C commands to display information about FRCA statistics. Statistics are displayed for each listening socket configured for FRCA support. For more information on the Netstat CACHINFO/-C reports, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Considerations for networking hardware attachment

This section provides general networking hardware attachment information and considerations associated with the IBM zSeries platform. Most of the information included here is associated with the IBM Open Systems Adapter (OSA) and the QDIO system architecture.

Virtual LAN (VLAN)

A local area network (LAN) is a broadcast domain. Nodes on a LAN can communicate with each other without a router, and nodes on different LANs need a router to communicate. A virtual LAN (VLAN) is a configured logical grouping

of nodes using switches. Nodes on a VLAN can communicate with each other as if they were on the same LAN, and nodes on different VLANs need a router to communicate.

OSA VLAN

The IBM Open Systems Adapter provides support for IEEE standards 802.1p/q, which describes priority tagging and VLAN identifier tagging. Deploying VLAN IDs allows a physical LAN to be partitioned or subdivided into discrete virtual LANs. This support is provided by the z/OS TCP/IP stack and OSA-Express in QDIO mode. The OSA registration process of a VLAN ID supports a single VLAN ID per IPv4 or IPv6 connection.

Primary router

OSA also provides primary (PRIRouter) and secondary (SECRouter) router support. This function allows a single TCP/IP stack, on a per protocol (IPv4 and IPv6) basis, to register and act as a router stack on a per OSA basis. Secondary routers can also be configured to provide for conditions in which the primary router becomes unavailable and the secondary router takes over for the primary router. The primary router stack is the only stack that OSA forwards packets to when the destination IP address has not been previously registered with OSA. Note that if PRIRouter or SECRouter is not specified, then the default value is NONRouter.

Relationship of VLAN and primary router

The OSA primary router support takes into consideration and interacts with the VLAN ID support (VLAN ID registration and tagging). OSA supports a primary and secondary router on a per VLAN basis (per registered VLAN ID). Therefore, if TCP/IP is configured with a specific VLAN ID and also configured as a primary or secondary router, that stack serves as a router for just that specific VLAN. This allows each OSA (CHPID) to have a primary router per VLAN. Configuring multiple primary routers (one per VLAN) has many advantages and preserves traffic isolation for each VLAN.

This support becomes more important when a single OSA is shared by multiple stacks. In this type of configuration, when each stack was configured with a unique VLAN ID, each stack could also be configured as a primary router for its respective VLANs.

OSA also continues to support a primary and secondary router that is not associated with a specific VLAN. This primary router is referred to as the default primary router. It continues to function as the router for inbound packets that are not VLAN ID tagged, or packets that are VLAN ID tagged with a VLAN ID that is not registered to OSA. This is the same primary router support that existed prior to introduction of the VLAN ID support. Therefore, multiple specific VLAN primary routers and a single default primary router can concurrently activate and share a single OSA.

Each VLAN-specific primary and secondary router is subject to the same OSA rules (that is, supporting a single primary router and allowing multiple secondary routers) as the default primary router.

Figure 9 on page 87 shows a configuration where multiple TCP/IP stacks are sharing a single OSA, and multiple VLANs with primary routers are configured.

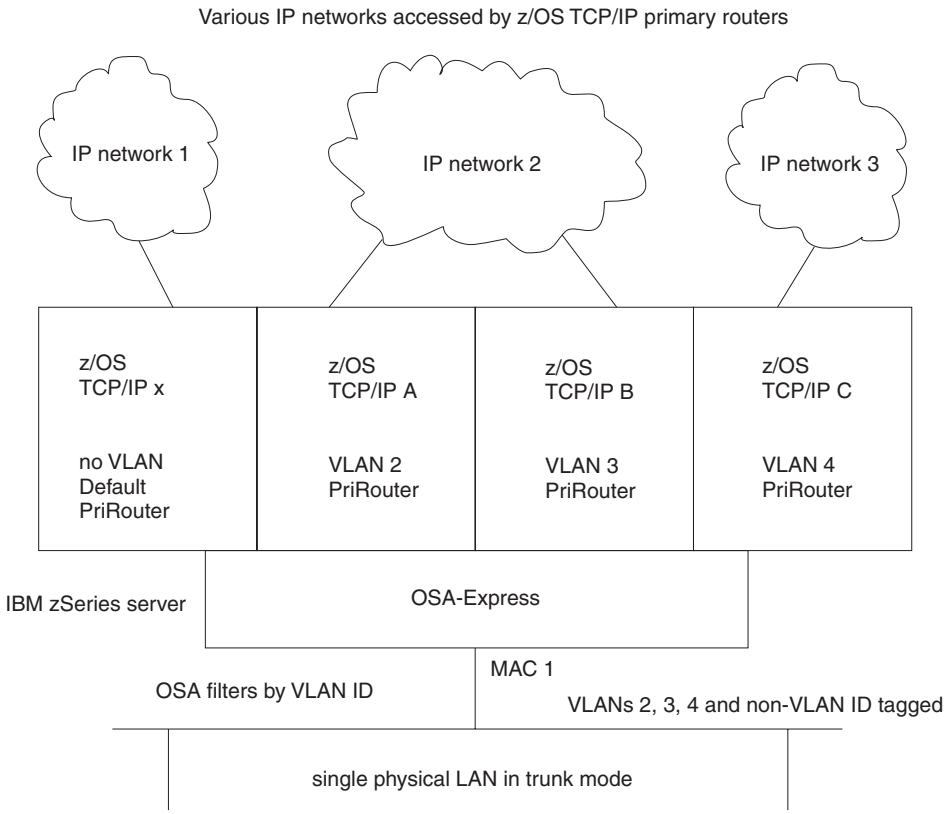


Figure 9. Primary router per VLAN (shared OSA with multiple primary routers)

In this example, TCP/IP x serves as the default primary router (PRIRouter without a VLANID configured). The other three TCP/IP stacks serve as a PRIRouter for just their specific VLANs.

For additional information regarding the details and syntax for configuring a VLAN Identifier (VLANID), and how to configure a TCP/IP stack as a primary or secondary router, refer to *z/OS Communications Server: IP Configuration Reference*. For IPv4 information on the PRIRouter, SECRouter, NONRouter, and VLANID parameters, refer to the section entitled DEVICE and LINK – MPCIPA devices. For IPv6 information regarding these parameters, refer to the section entitled INTERFACE - IPAQENET6 interfaces.

Network configuration strategy with VLAN

The IBM OSA-Express VLAN support allows a TCP/IP stack to register a specific VLAN identifier for both IPv4 and IPv6. Note that the VLAN ID for IPv4 can be different than the VLAN ID for IPv6. When a VLAN ID is configured, the following occurs:

1. TCP/IP becomes VLAN aware or enabled, and this TCP/IP (IPv4 or IPv6 connection) is considered to be part of the configured VLAN.
2. During activation, TCP/IP registers the configured VLAN ID value to OSA.
3. A VLAN ID tag is added to all outbound packets.
4. OSA filters inbound VLAN ID tagged packets based on the configured VLAN ID.
5. If this stack is also configured as a router (for example, PRIRouter), the OSA primary router support is extended to this stack, allowing it to serve as the primary router for the configured VLAN ID.

When configuring a z/OS TCP/IP stack with a VLAN ID, consideration must also be given to how the LAN is partitioned and how the VLAN aware switches are configured.

VLAN switch concepts

In conjunction with the IEEE standards, most VLAN aware switches recognize and support at least two modes, referred to as trunk and access modes. This support is provided on a switch port basis. The general concepts of the two modes are as follows:

Trunk mode
: Indicates that the switch should allow all VLAN ID tagged packets to pass through the switch port without altering the VLAN ID. Trunk mode is intended for servers that are VLAN capable, and filters and processes all VLAN ID tagged packets. In trunk mode, the switch expects to see VLAN ID tagged packets inbound to the switch port.

Access mode
: Indicates that the switch should filter on specific VLAN IDs and only allow packets that match the configured VLAN IDs to pass through the switch port. The VLAN ID is then removed from the packet before it is sent to the server (that is, VLAN ID filtering is controlled by the switch). In access mode, the switch expects to see packets without VLAN ID tags inbound to the switch port.

For the specific details regarding how to configure a VLAN switch, consult the specific product documentation.

VLAN configuration recommendations

When deploying the z/OS TCP/IP VLAN ID support in conjunction with IBM OSA-Express in QDIO mode, it is recommended that deployment be symmetrical with the configuration of the corresponding VLAN switch. Specific recommendations are as follows:

- When using a VLAN ID, configure the switch port in trunk mode.

  When a VLAN ID is configured in any z/OS TCP/IP stack that is sharing an OSA, the corresponding switch port associated with the OSA should be configured in trunk mode. In this mode, OSA performs VLAN ID filtering.

  Conversely, access mode should not be configured on the switch port if a VLAN ID is configured on any stack sharing this OSA.
- When not using a VLAN ID, configure the switch port in access mode.

  When a VLAN ID is not configured on any z/OS TCP/IP stack that is sharing an OSA, access mode should be configured at the switch (if VLAN filtering is desired and therefore required at the switch).

  Conversely, trunk mode should not be configured on the switch port if a VLAN ID is not configured on any stack sharing this OSA.
- Multiple OSAs on the same physical LAN

  When a z/OS TCP/IP stack has access to multiple OSAs that are on the same physical LAN, and a VLAN ID is configured on any of the OSAs, it is recommended that this stack configure a VLAN ID for all OSAs on the same physical LAN. That is, do not mix VLAN and no-VLAN on the same physical network when a stack has access to the same LAN through multiple OSAs.
- VLAN ID 1 considerations

  Some switch vendors use VLAN ID 1 as the default value when a VLAN ID value is not explicitly configured. It is recommended that you avoid the value of 1 when configuring a VLAN ID value.

Figure 10, Figure 11 on page 90, and Figure 12 on page 90 illustrate the preceding recommendations.

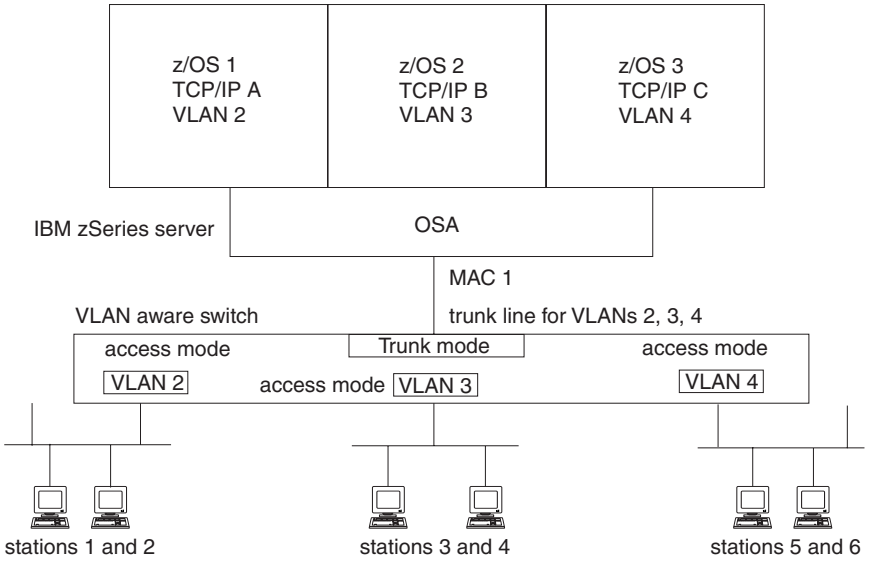


Figure 10. Single OSA and VLAN switch configuration

Figure 10 shows the recommended VLAN switch port configuration when a VLAN ID is configured in the TCP/IP stack. A single physical LAN is divided into three separate virtual LANs (2, 3, and 4), the OSA port is configured as a trunk line, and the other ports on the switch are configured in access mode for their specific VLAN.

In Figure 10 there are three virtual LANs deployed through the same shared OSA, where each TCP/IP stack appears to have a unique and isolated physical network as follows:

- VLAN 2 - TCP/IP A and stations 1 and 2
- VLAN 3 - TCP/IP B and stations 3 and 4
- VLAN 4 - TCP/IP C and stations 5 and 6

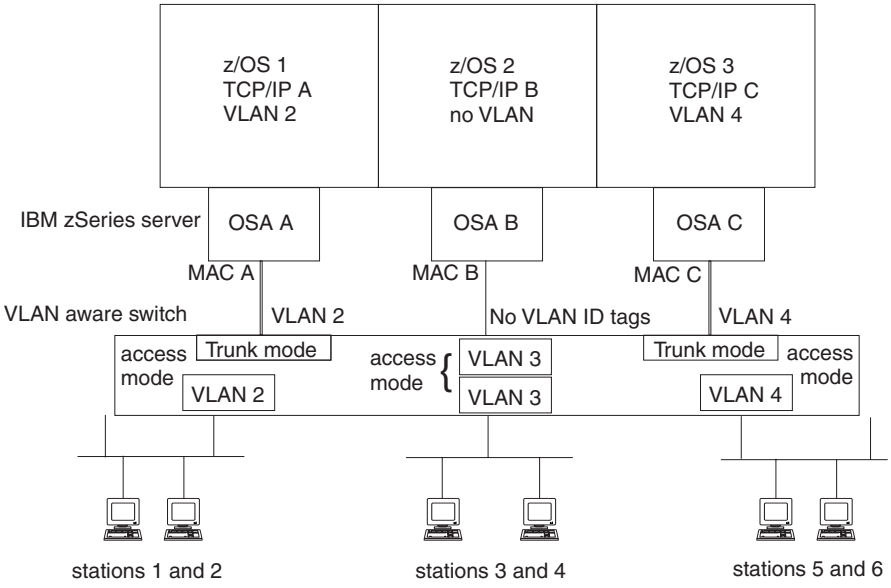


Figure 11. Matching VLAN switch configuration to multiple OSAs (VLAN configuration)

Figure 11 illustrates using multiple OSAs and TCP/IP stacks. Three unique VLANs are created. However, TCP/IP stack B will not deploy a VLAN ID, and the corresponding switch port is configured in access mode. No VLAN ID tags will flow to this OSA port.

In Figure 11 there are also three virtual LANs deployed. Access to each VLAN is provided through separate OSAs, yet the functionality of having three physical networks is still provided. TCP/IP B is not configured with a VLAN ID, and therefore stack B is unaware of the existence of VLAN 3 (although stations 3 and 4 on VLAN 3 have access to stack B through OSA B). Note that the switch port for OSA B is configured in access mode, while the other two switch ports are configured in trunk mode.

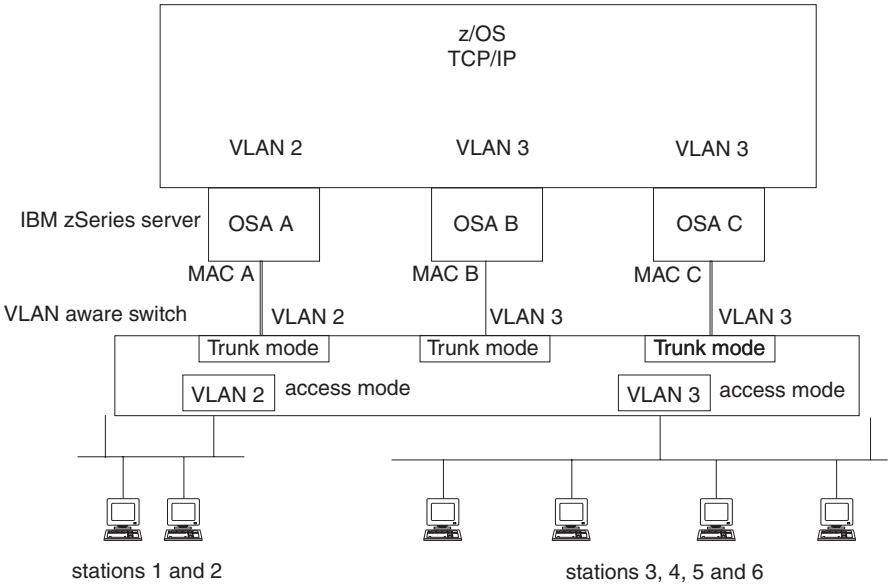


Figure 12. Single stack using multiple OSAs on the same physical network

Figure 12 on page 90 illustrates a single TCP/IP stack using multiple OSAs that are on the same physical network. There are two VLANs deployed, where OSA A is on VLAN 2, and OSA B and OSA C are on VLAN 3.

Configuring OSA B and OSA C with the same VLAN ID has significance for failure or takeover scenarios. The interface takeover (ARP takeover) function, with redundant connectivity onto a LAN, applies within the VLAN. Therefore, if OSA B becomes unavailable, OSA C can take over. Similarly, OSA B can take over if OSA C becomes unavailable. However, OSA A cannot take over for either OSA B or OSA C, because OSA A is on a different VLAN.

In Figure 12 on page 90, a single TCP/IP stack has access to two VLANs through three OSAs, which provides the following network isolation:

- VLAN 2 - through OSA A to stations 1 and 2
- VLAN 3 - through OSA B and OSA C to stations 3, 4, 5 and 6

Checksum offload

When sending or receiving packets over OSA-Express in QDIO mode with checksum offload support, TCP/IP offloads most IPv4 (outbound and inbound) checksum processing (IP header, TCP, and UDP checksums) to the OSA. The TCP/IP stack performs checksum processing in the following cases where checksum cannot be offloaded:

- Packets that go to another stack sharing the same OSA
- IPSec-encapsulated packets
- Fragmented and reassembled packets
- Outbound multicast and broadcast packets
- Outbound TCP packets that contain only a TCP header
- When multipath is in effect (unless all interfaces in the multipath group support checksum offload)

TCP segmentation offload

When sending packets over OSA-Express in QDIO mode with TCP segmentation offload support, TCP/IP offloads most IPv4 outbound TCP segmentation processing to the OSA. The TCP/IP stack performs TCP segmentation processing in the following cases where segmentation cannot be offloaded:

- Packets that go to another stack sharing the same OSA
- IPSec-encapsulated packets
- When multipath is in effect (unless all interfaces in the multipath group support segmentation offload)

Tip: Applications that use large TCP send buffers will obtain the most benefit from TCP segmentation offload. The size of the TCP receive buffer on the other side of the TCP connection also affects the negotiated buffer size. You can control the size of these buffers using the following mechanisms:

- The TCPSENDBFRSIZE parameter on the TCPCONFIG statement sets the default TCP send buffer size for all applications.
- An application can use the SO_SNDBUF socket option to override the default TCP send buffer size.
- The TCPRCVBUFRSIZE parameter on the TCPCONFIG statement sets the default TCP receive buffer size for all applications.

- An application can use the SO_RCVBUF socket option to override the default TCP receive buffer size.

HiperSockets concepts and connectivity

HiperSockets is a zSeries hardware feature that provides high performance internal communications between LPARs within the same central processor complex (CPC), without the use of any additional or external hardware equipment (for example, channel adapters, LANs, and so on). When the processor supports HiperSockets and the CHPIDs have been configured in HCD (IOCP), TCP/IP connectivity can occur for two reasons:

- DYNAMICXCF is configured on the IPCONFIG (IPv4) or the IPCONFIG6 (IPv6) statements.
- A user-defined HiperSockets (MPCIPA) DEVICE and LINK (IPv4) or INTERFACE (IPv6) is configured and started.

Therefore, there are two types of HiperSockets devices:

- DYNAMICXCF HiperSockets device or interface (TRLE ″IUTIQDIO″ and an MPC group of subchannel devices). The PORTNAME will be IUTIQDxx, where xx = the IQD CHPID that VTAM uses (for example, IUTIQDFD when using IQD CHPID x'FD').
- A user-defined HiperSockets device or interface (TRLE ″IUTIQDxx″ and an MPC group of subchannel devices). The PORTNAME is not applicable for this TRLE.

In both cases, the TRLE is dynamically built by VTAM. For additional details regarding how to configure a user-defined HiperSockets device or interface, refer to "HiperSockets" on page 375 and to *z/OS Communications Server: IP Configuration Reference*.

Concepts and considerations for the IQD CHPID

The HiperSockets hardware device is represented by the IQD CHPID and its associated subchannel devices. All LPARs that are configured (HCD) to use the same IQD CHPID have internal connectivity and therefore have the capability to communicate using HiperSockets. If the system supports multiple channel subsystems, and if HiperSockets connectivity is required across multiple channel subsystems, the IQD CHPID must also be configured (HCD) to span the applicable channel subsystems. The IQD CHPID can be viewed as a logical LAN within the CPC. The HiperSockets hardware allows up to four (16 with systems that support multiple channel subsystems) separate IQD CHPIDs to be defined per CPC, creating the capability of having four (16 with systems that support multiple channel subsystems) separate logical LANs within the same CPC. Figure 13 on page 93 and Figure 14 on page 93 illustrate this concept:

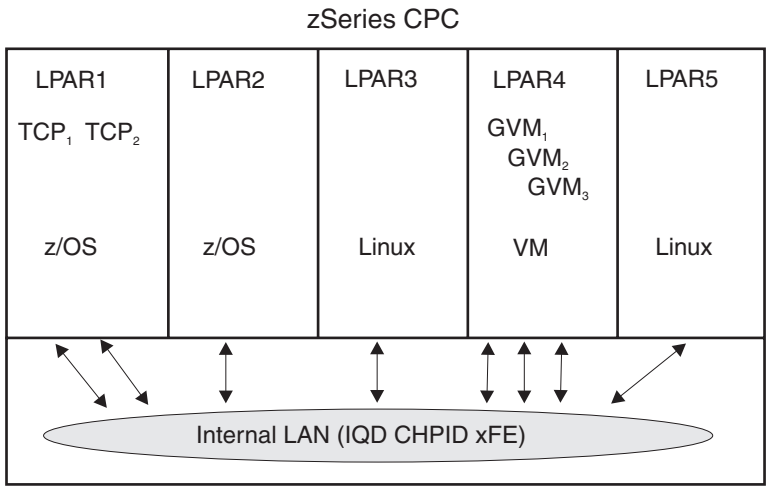


Figure 13. HiperSockets internal LAN

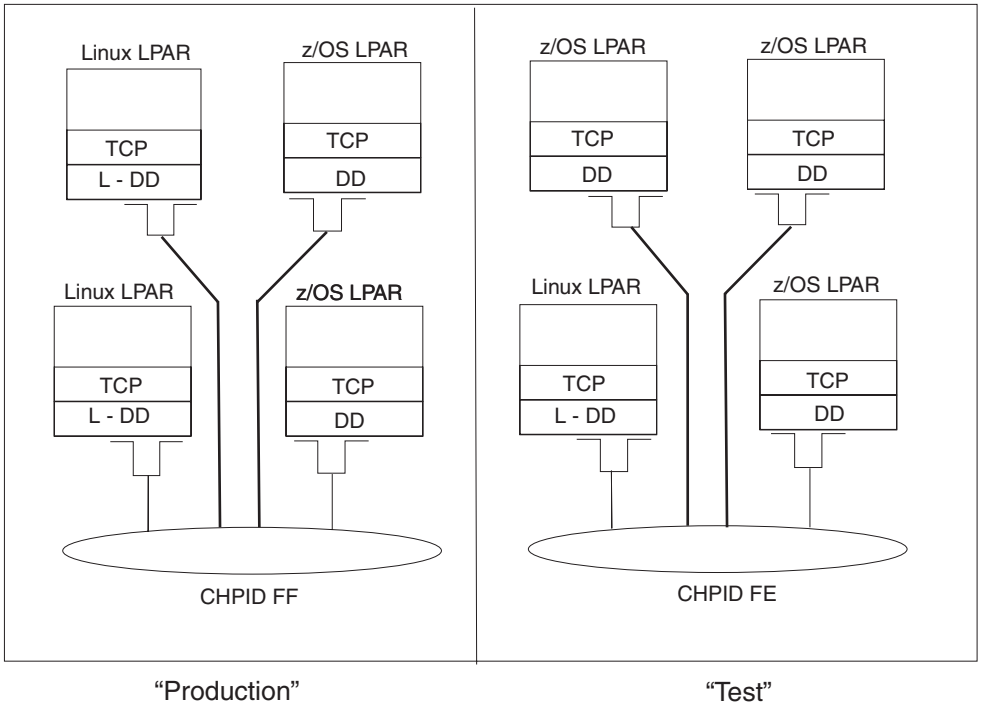


Figure 14. HiperSockets multiple internal LANs

With this capability, the system administrator can logically separate (or control) the internal connectivity. This is accomplished by controlling which specific LPARs are allowed to internally connect using HiperSockets. Further examples of this are as follows:

- SYSPLEX 'A' LPARs running on LPs 1 through 4 could use IQD CHPID x'FC'.
- SYSPLEX 'B' LPARs running on LPs 5 through 8 could use IQD CHPID x'FD'.
- A VM LPAR runs in LP 9 running various second level systems (Linux and z/OS) which use IQD CHPID x'FE'.
- combinations of the examples above could be:

– Another set of LPARs on LPs 10 through 12 which are not using DYNAMICXCF (non SYSPLEX) are connected to IQD CHPIDs x'FE' and x'FF'.
– Subsets of LPARs 1 through 8 are using both the DYNAMICXCF IQD CHPIDs and a non-DYNAMICXCF IQD CHPIDs.
– Some LPARs are connected to all four IQD CHPIDs.

Planning for IQD CHPID spanning

Figure 15 illustrates how an IQD CHPID can span both logical channel subsystem 0 and logical channel subsystem 1.

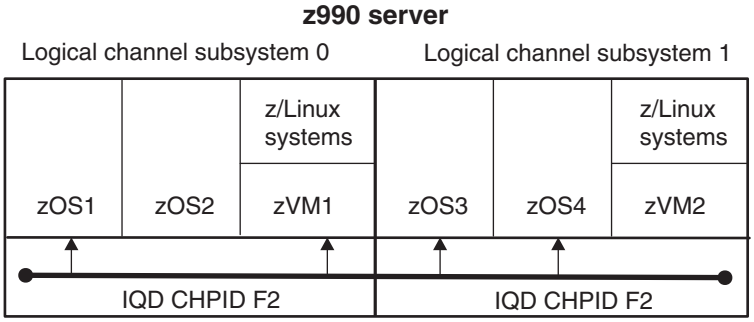


Figure 15. Spanned IQD (HiperSockets) CHPID

In Figure 15, note the following:

- CHPID spanning is applicable to servers that support multiple channel subsystems, such as the IBM z990 server.
- IQD CHPIDs and their spanning attributes are configured using HCD.
- Logical channel subsystem 0 and logical channel subsystem 1 are both using IQD CHPID F2, which was configured (in HCD) to span.
- All images can exploit IQD CHPID spanning.
- Using HiperSockets support along with IQD spanning support, logical partitions in both logical channel subsystem 0 and 1 have internal connectivity.
- Specific logical partitions can be configured to have access to the spanned IQD CHPID. In this example:
  – IQD CHPID F2 is configured to span both logical channel subsystem 0 and logical channel subsystem 1.
  – In logical channel subsystem 0, zOS1 and zVM1 have connectivity (through IQD CHPID F2) to zOS3 and zOS4 in logical channel subsystem 1.
  – zOS2 and zVM2 do not have access to IQD CHPID F2.
- Up to 16 IQD CHPIDs can be configured in HCD.
- There are no TCP/IP or VTAM configuration changes required (spanning support is transparent to z/OS Communications Server).

  Tip: TCP/IP Dynamic XCF support uses HiperSockets connectivity when the remote system is reachable through HiperSockets. This is dynamically determined, and this internal logic also handles a spanned IQD CHPID.

To configure a spanned IQD CHPID, refer to *z/OS HCD User's Guide*.

The HiperSockets MPC group

VTAM will build a single HiperSockets MPC group, using the subchannel devices associated with a single IQD CHPID. VTAM will use two subchannel devices for

the read and write control devices, and 1 to 8 devices for data devices. Each TCP/IP stack will be assigned a single data device.

Therefore, in order to build the MPC group, there must be a minimum of 3 subchannel devices defined (within HCD) and associated with the same IQD CHPID. The maximum number of subchannel devices that VTAM will use is 10 (supporting 8 data devices or 8 TCP/IP stacks) per LPAR or MVS image. The subchannel devices must be configured for the LPAR and online prior to when the TCP/IP stack is initialized. Generally, the number of HiperSockets subchannel devices you should configure per LPAR is:

```
     2   (read / write control devices)
    + N  (where N = number of TCP/IP stacks)
    -----------------
     N+2  (total subchannel devices per LPAR)
Example (LPAR 1 starts two TCP/IP stacks and both stacks use HiperSockets):

   -  define 4 subchannel devices on the same IQD CHPID
   -  where 2 are used for read / write control and 2 data devices are available
```

The first TCP/IP stack within the LPAR to initialize DYNAMICXCF will cause the HiperSockets MPC group (IUTIQDIO) to be dynamically created. Each TCP/IP stack can then start the IUTIQDIO device, and each stack will be assigned a unique (dedicated) subchannel data device from the IUTIQDIO MPC group.

Recommendation: Configure the IQD CHPIDs using CHPIDs x'F0' through x'FF' (but any valid CHPID value x'00' through x'FF' can be configured as TYPE = IQD). Refer to *z/OS HCD Planning* and Appendix D, "Using HCD," on page 1203 for additional details.

HiperSockets maximum frame size

The HiperSockets hardware supports four different frame sizes referred to as the HiperSockets MFS (maximum frame size). Using HCD (or IOCP), the HiperSockets MFS is configured on the IQD CHPID using the 'OS=' parameter. All LPARs communicating over the same IQD CHPID will then use the same IQD MFS. The MFS affects the largest packet that TCP/IP can transmit. TCP/IP will adjust the MTU (Maximum Transmission Unit) based on the MFS, which is discovered during activation.

The following table depicts the four possible TCP/IP MTU sizes resulting from the HiperSockets frame sizes:

OS=*value*	**HiperSockets frame size**	**TCP/IP MTU size**
00 (default)	16 KB	8 KB
40	24 KB	16 KB
80	40 KB	32 KB
C0	64 KB	56 KB

The default HiperSockets MFS is 16 KB. However, in cases in which increased bandwidth is required (such as large file transfers, file backup, and so on), a larger MFS could be used. In most workload environments the default size will result in better storage and CPU utilization.

```
************************************************************
* OS  values are '00'=16K, '40'=24K, '80'=40K  and 'C0'=64K.  *
*                                                          *
* Need at least 3 addresses per z/OS, maximum of 10:       *
*   - 2 addresses for control                              *
```

```
*   - 1 address for data for each TCP stack (between 1 and 8) *
*************************************************************
 ID SYSTEM=(2064,1)
*
 CHPID PATH=FC,TYPE=IQD,SHARED,OS=00
 CHPID PATH=FD,TYPE=IQD,SHARED,OS=40
 CHPID PATH=FE,TYPE=IQD,SHARED,OS=80
 CHPID PATH=FF,TYPE=IQD,SHARED,OS=C0
*
 CNTLUNIT  CUNUMBR=FC00,PATH=FC,UNIT=IQD
 IODEVICE  ADDRESS=(2C00,16),CUNUMBR=FC00,UNIT=IQD
*
 CNTLUNIT  CUNUMBR=FD00,PATH=FD,UNIT=IQD
 IODEVICE  ADDRESS=(2C10,16),CUNUMBR=FD00,UNIT=IQD
*
 CNTLUNIT  CUNUMBR=FE00,PATH=FE,UNIT=IQD
 IODEVICE  ADDRESS=(2C20,16),CUNUMBR=FE00,UNIT=IQD
*
 CNTLUNIT  CUNUMBR=FF00,PATH=FF,UNIT=IQD
 IODEVICE  ADDRESS=(2C30,16),CUNUMBR=FF00,UNIT=IQD
```

Refer to *z/OS HCD Planning* and Appendix D, "Using HCD," on page 1203 for additional details.

Modifying HiperSockets connectivity [TCP/IP device and link and the VTAM HiperSockets MPC group (IUTIQDIO)]

Certain modifications can be made to the HiperSockets device (MPC group) without disrupting an active TCP/IP stack.

z/OS supports dynamic I/O for the HiperSockets CHPID and subchannel devices, allowing subchannel devices to be added or removed to or from an LPAR which has already been IPLed.

TCP/IP supports the STOP and START command for the DYNAMICXCF HiperSockets device (IUTIQDIO). However, the commands are only supported when the (internal) start (activation) was successful during stack initialization. TCP/IP also supports the STOP and START command for the user-defined HiperSockets devices (IUTIQDxx). Since a user-defined HiperSockets device is supported as an MPCIPA device, STOP and START function just as they would for other MPCIPA devices.

VTAM supports a MODIFY IQDCHPID command, which allows the user to change the initial setting of the IQDCHPID start option.

Therefore, it is possible to make certain changes to the DYNAMICXCF HiperSockets MPC group (IUTIQDIO) without restarting VTAM or an active TCP/IP stack. Examples of changes that can be made are (STOP/START device required):

- Alter which specific IQD CHPID is used for DYNAMICXCF (for example, move from the x'FC' CHPID to the x'FD' CHPID).
- Add or remove subchannel devices (for example, from the current IQD CHPID).
- Alter the IQD MFS which alters the TCP/IP MTU (for example, increase the current IQD CHPID from 16k to 64k).

Although VTAM supports modifications to the start option IQDCHPID (and the modification will be immediately displayed), the effects will vary depending on what the current usage was and the change (from or to) that was made. For example:

- When MODIFIED from ANY (or CHPID) to NONE, there is no effect on current usage but blocks subsequent activations of the DYNAMICXCF HiperSockets device.
- When MODIFIED from NONE to ANY (or CHPID), there is no effect on current usage but allows subsequent activations.
- When MODIFIED from CHPID_X to CHPID_Y, there is no effect on current usage.

  Note: VTAM only uses the CHPID value when building the IUTIQDIO MPC group.

  To change CHPIDs for an active MPC group the following must be done:

  1. TCP/IP IUTIQDIO devices and IQDIOINTF6 interfaces that are changing must be stopped.
  2. Make any necessary HCD/IOCDS changes.
  3. Verify new subchannel devices are varied online.
  4. Verify the MPC group has deactivated (with no usage it times out after approximately 2 minutes).
  5. Modify IQDCHPID = CHPID (to new CHPID).
  6. Restart the TCP/IP IUTIQDIO devices and IQDIOINTF6 interfaces.

To use HiperSockets communications, the processor must have the necessary hardware support. If the processor does not support HiperSockets communications, modifications to this start option will not be accepted, and the IQDCHPID option will not be displayed (displayed as ***NA***) .

HiperSockets connectivity and routing

For each pair of stacks within a sysplex that are not on the same MVS image, if all of the following conditions are true, the stacks will use HiperSockets DYNAMICXCF connectivity (versus standard XCF connectivity):

- The two stacks must be on the same CPC.
- For the DYNAMICXCF HiperSockets device (IUTIQDIO):
  - The two stacks must be using the same IQD CHPID.
  - If running on a system that supports multiple channel subsystems and the two stacks are also in different channel subsystems, the IQD CHPID must be configured (HCD) to span.
- Both stacks must be configured (HCD) to use HiperSockets.
- For IPv4 HiperSockets connectivity, both stacks must be at the z/OS V1R2 level, or a later release. For IPv6 HiperSockets connectivity, both stacks must be at the z/OS V1R7 level.
- The initial HiperSockets activation must complete successfully.

When an IPv4 DYNAMICXCF HiperSockets device and link are created and successfully activated, a subnet route is created across the HiperSockets link. The subnet is created by using the DYNAMICXCF IP address and mask. This allows any LPAR within the same CPC to be reached, even ones that are not within the sysplex. For example, an LPAR that is running z/Linux or z/VM that does not support joining the sysplex can still be reached. The z/Linux or z/VM LPAR must define at least one IP address for the HiperSockets endpoint that is within the subnet defined by the DYNAMICXCF IP address and mask.

Similarly, when an IPv6 DYNAMICXCF HiperSockets interface is created and successfully activated, a prefix route is created across the HiperSockets interface (if prefix_route_len is specified on DYNAMICXCF). This allows any LPAR within the

same CPC to be reached, even ones that are not within the sysplex. For example, an LPAR that is running z/Linux or z/VM that does not support joining the sysplex can still be reached. The z/Linux or z/VM LPAR must define at least one IP address for the HiperSockets endpoint that uses the same prefix as the DYNAMICXCF IP address.

Therefore, TCP/IP can communicate with other LPARs within the CPC over the HiperSockets connectivity created by DYNAMICXCF even when the TCP/IP in the other LPAR is not part (joins or supports) of the sysplex. You can also elect to manually configure a HiperSockets device for non-sysplex communications.

When multiple stacks reside within the same LPAR that supports HiperSockets, both IUTSAMEH and HiperSockets links or interfaces will coexist. In this case, it is possible to transfer data across either link or interface. Because IUTSAMEH links or interfaces have better performance, it is better to always use them for intra-stack communication. A host route will be created by DYNAMICXCF processing across the IUTSAMEH link or interface but not across the HiperSockets link or interface. To avoid using the HiperSockets link or interface for communication within the same host, the following rules should be observed:

- Specify DYNAMICXCF IP addresses in a separate subnet than that of VIPA addresses (IPv4) or using a separate prefix than that of VIPA addresses (IPv6).
- Do not specify static IUTSAMEH links or interfaces.

It is also possible with multiple stacks in the same LPAR to end up with both XCF and HiperSockets links or interfaces. This occurs when the availability of the (preferable) HiperSockets link or interface changes as each TCP stack (within the same LPAR) is started. For example, stack A is started with HiperSockets available and later stack B is started with HiperSockets unavailable. This type of configuration should be avoided.

Efficient routing using HiperSockets Accelerator

Communications Server leverages the technological advances and high performing nature of the I/O processing offered by HiperSockets with the IBM zSeries servers and the IBM OSA-Express using QDIO architecture by optimizing IP packet forwarding processing that occurs across these two types of links. This function is referred to as HiperSockets Accelerator. It is a configurable option, and activated by configuring the IQDIORouting option on the IPCONFIG statement.

Restriction: HiperSockets Accelerator is IPv4 only.

When configured, it allows unicast IPv4 packets that are received over a HiperSockets link and are to be forwarded over a QDIO link (or received over QDIO and are to be forwarded over HiperSockets) to be forwarded by the z/OS Communications Server HiperSockets device driver. That is, the IP forwarding function is pushed down as close to the hardware [or to the lowest software DLC (Data Link Control)] layer as possible so that these packets do not have to be processed by the TCP/IP stack or address space. Therefore, valuable TCP/IP resources (storage and machine cycles) are not expended for purposes of routing and forwarding packets. The following figure illustrates a configuration before the utilization of HiperSockets Accelerator.

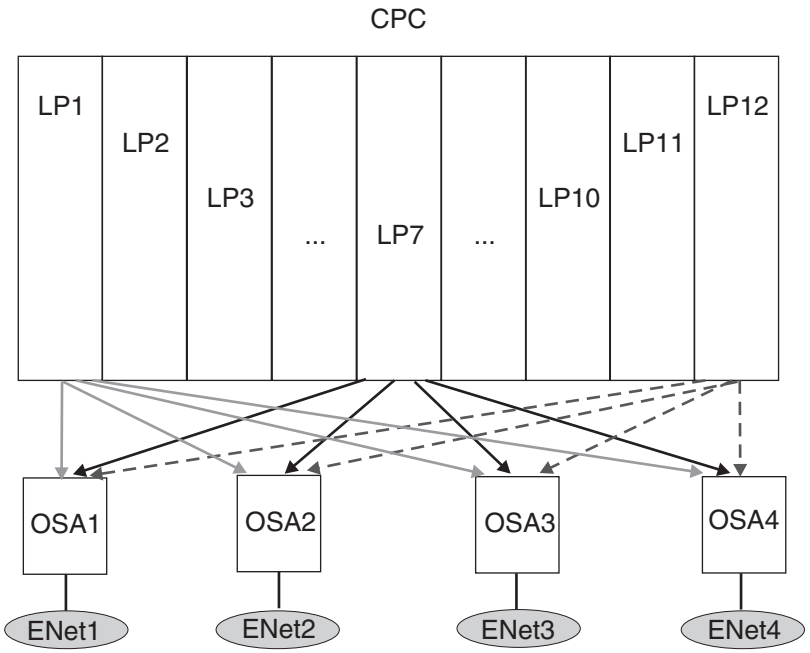


Figure 16. Candidate configuration for HiperSockets Accelerator

HiperSockets Accelerator presents a different configuration and approach to obtain full connectivity as shown in the figure below.

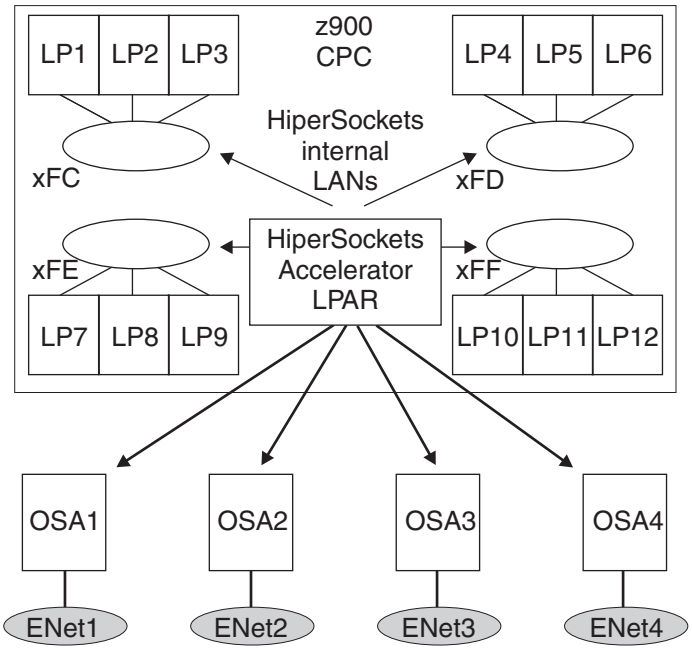


Figure 17. HiperSockets Accelerator configuration

This function allows a user to position a specific or single TCP/IP stack which has direct physical connectivity to the OSAs LANs as the HiperSockets router. This stack can then connect to all remaining TCP/IP stacks in other images (LPARs) within the same CPC that require connectivity to the same OSA LANs using HiperSockets connectivity.

This approach becomes more beneficial as the number of LPARs within a given CPC increase. Instead of attempting to directly attach each LPAR to each physical network attachment using an OSA LAN, a smaller number of OSAs could be concentrated through a single z/OS LPAR. From a performance perspective, HiperSockets Accelerator attempts to make the intermediate (or router) TCP/IP stack appear as if it did not exist in the path. Instead, each LPAR will appear as if each were directly attached to the physical network (for example, packets are forwarded without traversing the router TCP/IP stack). There are no additional routing configuration tasks required by the user. The prerouting occurs automatically. The TCP/IP stack automatically detects IP packet forwarding is occurring across a HiperSockets eligible route (QDIO/HiperSockets or HiperSockets/QDIO), and dynamically creates a HiperSockets Accelerator route entry. All subsequent packets will then take the optimized device driver path, and will not traverse the TCP/IP stack.

The dynamically created HiperSockets Accelerator routing entries can be displayed by the Netstat ROUTE/-r report option with the IQDIO modifier. VTAM tuning statistics are provided to allow the user to monitor or measure prerouting activity.

QDIOPriority (IQDIORouting option) is an optional choice that allows the user to specify which of the four priority queues should be used when prerouting packets from a HiperSockets link outbound to a QDIO link. The default is 1 (highest priority), and in most cases should be sufficient.

For additional details regarding the IQDIORouting configuration option, refer to the IPCONFIG statement in the *z/OS Communications Server: IP Configuration Reference*.

Maximum Transmission Unit (MTU) considerations

TCP/IP uses the MTU to determine the largest size frame to send. The MTU in effect for a given outbound send depends on several factors:

interface_MTU
This is either a hardcoded size based on the physical device or a value learned from the device during activation. For information about the interface MTU values for the various network interface types supported by TCP/IP, refer to the Summary of DEVICE and LINK statements in *z/OS Communications Server: IP Configuration Reference*. For an IPAQENET6 interface, you can configure a lower interface_MTU using the MTU keyword on the INTERFACE statement.

For an active link or interface, TCP/IP reports the interface_MTU in the ActMtu field of the Netstat DEVLINKS/-d command.

configured_route_MTU
This is the MTU size configured for a route.

For a static route, specify configured_route_MTU on either a ROUTE statement in a BEGINROUTES block or on a GATEWAY statement in the TCP/IP profile.

For each IPv4 dynamic route added by OMPROUTE over an interface, the configured_route_MTU comes from the value of the MTU keyword specified on the RIP_INTERFACE, OSPF_INTERFACE or INTERFACE statement for that interface in the OMPROUTE configuration file. If you do not specify an MTU for an interface, OMPROUTE uses 576. For IPv6,

OMPROUTE learns the interface_MTU value from TCP/IP, and you cannot specify a configured_route_MTU in the OMPROUTE configuration file.

actual_route_MTU
This is the minimum of the interface_MTU and the configured_route_MTU. When path MTU discovery is not in effect, TCP/IP uses the actual_route_MTU for sending outbound packets.

path_MTU
This is the value determined by the path MTU discovery function. When path MTU discovery is in effect, TCP/IP uses path_MTU for sending outbound packets.

You can enable path MTU discovery for IPv4 using IPCONFIG PATHMTUDISCOVERY. TCP/IP automatically enables path MTU discovery for IPv6. Path MTU discovery starts out by setting path_MTU to the actual_route_MTU of the route. If packets would require fragmentation to get to the final destination, path MTU discovery determines path_MTU by repeatedly decreasing the value until it can send packets to the final destination without fragmentation.

These factors comprise a general set of rules for how TCP/IP determines the MTU, but there are some exceptions. For example, if an application uses the IPV6_USE_MIN_MTU socket option, TCP/IP sends outbound packets using the IPv6 minimum MTU of 1280.

Guidelines:

- Enable path MTU discovery in configurations where traffic originating in the z/OS TCP/IP stack will traverse multiple hops with different MTU sizes.
- When using OSA-Express Gigabit Ethernet (which supports an interface MTU of 8992), be aware that not all routers and switches support a value this large. Either ensure that all routers and switches in your configuration support 8992 or specify a lower configured_route_MTU.
- When using OMPROUTE, specify the MTU keyword for each IPv4 interface.
- When using OMPROUTE, configure all nodes on a LAN to use the same MTU value. Otherwise, you might encounter problems, such as OSPF adjacency errors.

Considerations for multiple servers sharing a TCP port

For a TCP server application to support a large number of client connections on a single system while providing good performance for those connections, it might be necessary to run more than one instance of the server application to service the connection requests. For all instances of the server application to receive client connection requests without changing the client applications, the servers must all bind to the same server IP address and port. To enable this in the TCP/IP stack, you must add the SHAREPORT or SHAREPORTWLM keyword to the PORT profile statement that reserves the TCP port for the server instances. For more detailed information on the PORT statement and its keywords, refer to *z/OS Communications Server: IP Configuration Reference*.

The set of server instances sharing the same TCP port on the same TCP/IP stack is called a shareport group. As incoming client connections arrive for this port and IP address, TCP/IP distributes them across the servers in the shareport group.

If the SHAREPORT keyword was specified, the connections are distributed across the available servers using a weighted round-robin distribution based on the Servers' accept Efficiency Fractions (SEFs).

If the SHAREPORTWLM keyword was specified, the connections are distributed across the available servers using a weighted round-robin distribution based on the WLM server-specific recommendations, modified by the SEFs.

The SEF is a measure, calculated at intervals of approximately 1 minute, of the efficiency of the server application in accepting new connection requests and managing its backlog queue. You can use the Netstat ALL/-A command to obtain the current SEF value for a server. For more detailed information about the Netstat command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Configuring a shareport group can also be used to improve the availability of key applications. For example, should an application server instance become inactive (encounters a failure or is stopped for planned maintenance), the other applications in the shareport group can continue to process client requests.

Applications that might be good candidates for being placed in a shareport group must also satisfy the following requirements:

- Multiple instances of the application can be started within a single system.
- Each application instance must provide the same functional capabilities. That is, each must be capable of processing any TCP connection requests to the shared port.

In addition, each TCP connection directed to a shareport group must be eligible for load balancing (it can not have an affinity to a specific server). If specific client affinities exist to a specific server, you should consider using Sysplex Distributor, which provides for load balancing of connections while maintaining client affinities to a specific server instance over a period of time (Timed Affinity). While the primary focus of the Sysplex Distributor is on load balancing across servers on multiple target systems, it also supports multiple servers in a shareport group on the target systems. For more information on Sysplex Distributor configuration, see Chapter 8, "TCP/IP in a sysplex," on page 363.

Considerations for Common Information Model (CIM) providers

The Common Information Model (CIM) provides a model for describing and accessing data across an enterprise. CIM is a standard developed by a consortium of major hardware and software vendors, including IBM, called the Distributed Management Task Force (DMTF). CIM data is defined by standardized CIM classes. Class definitions define properties for each class. For example, some properties of the CIM_EthernetPort (Ethernet interface) class are interface type, enabled state, and speed. Platforms can extend the standardized CIM classes by defining their own platform-specific classes and their properties.

CIM data instrumentation is supplied by CIM components called providers. The providers gather data on a system in support of the CIM classes. Clients can retrieve the data through the Common Information Model Object Manager. On z/OS, this function is provided by the z/OS CIM server.

Communications Server supplies CIM providers for the CIM classes shown in Table 9 on page 103. Only IPv4 data is currently supported. To support data from

eight TCP/IP stacks on an MVS image, the TCP/IP stack name has been added to the IBMzOS_EthernetPort and IBMzOS_IPProtocolEndpoint classes as a platform-specific property.

For more information about CIM, the z/OS CIM server, the properties support for the Communications Server CIM classes, and other z/OS CIM provider support, refer to *z/OS Common Information Model User's Guide*.

Table 9. Communications Server CIM providers

CIM base class name	z/OS class name	Description
CIM_EthernetPort	IBMzOS_EthernetPort	IPv4 Ethernet interfaces defined to the TCP/IP stack
CIM_IPProtocolEndpoint	IBMzOS_IPProtocolEndpoint	IPv4 IP addresses defined to the TCP/IP stack
CIM_PortImplementsEndpoint	IBMzOS_NetworkPortImplementsIPEndpoint	IPv4 Ethernet interfaces and their associated IP addresses
CIM_SystemDevice	IBMzOS_CSNetworkPort	MVS system and its associated IPv4 Ethernet interfaces

This CIM provider function resides in the /usr/lpp/tcpip/lib HFS directory. There is no configuration necessary to activate this CIM provider support. The z/OS CIM server must be configured and activated for the data supported by the Communications Server CIM providers to be available to clients. The z/OS Communications Server CIM classes are shipped with the z/OS CIM server. The files that define these classes and any platform specific properties are also installed in the /usr/lpp/tcpip/mof HFS directory.

There are security configuration considerations that govern the ability of providers to gather CIM data. For more information about security configuration, see "CIM provider access control" on page 130.

Required steps before starting TCP/IP

The following sections describe the steps you must complete before starting TCP/IP.

Planning your installation and migration

It will be to your advantage to have thoroughly studied the following documentation prior to the installation and customization of z/OS Communications Server:

- Program Directory for z/OS for CBPDO Installation and ServerPac Reference, Program Number 5694-A01
- Preventive Service Planning (PSP) bucket
- *z/OS Communications Server: New Function Summary*
- *z/OS UNIX System Services Planning*
- OS390CKL, an IBM MKTTOOLS document for the z/OS UNIX System Services implementer

It is also recommended that you attend a z/OS UNIX System Services concepts class and a class in using z/OS UNIX System Services prior to migrating to z/OS Communications Server. If this is not possible, then you will want to ensure that

the z/OS UNIX System Services implementer and the RACF administrator work together with you during the installation and customization process.

Planning for and installing z/OS Communications Server requires MVS, UNIX, and networking skill. If your background is in traditional MVS programming or systems programming, the z/OS UNIX System Services terminology might at first seem to be somewhat confusing. If your background is in the UNIX environment, the terms should be familiar to you.

In the past, MVS TCP/IP system programmers have needed a working knowledge of the MVS or z/OS system. These programmers have been accustomed to working closely with the RACF administrator and z/OS system programmer for authorizations; the VTAM and NCP system programmers for SNALINK and NCP connections; the IP address administrator for basic name and address assignments; and the administrators of the router network and channel-attached peripherals for connection definition and problem determination.

With the introduction of z/OS Communications Server, the TCP/IP system programmer needs to develop an additional alliance with the z/OS UNIX System Services system programmer. The TSO interfaces that have been traditionally available in the host-based TCP/IP still stand at the system programmer's disposal and additional MVS console commands simplify some TCP/IP operations. However, another user interface provided by the UNIX shell environment, either with the OMVS shell or the ISPF SHELL, is a useful and sometimes necessary tool that the TCP/IP system programmer will need to work with. Additionally, the tight coupling of z/OS Communications Server with z/OS UNIX System Services means that the TCP/IP system programmer needs more than a passing knowledge of UNIX conventions, commands, and Hierarchical File System (HFS) concepts. Even if the system programmer is familiar with other UNIX environments, work with the UNIX shell requires more than basic familiarity.

In the first version of a full TCP/IP stack based on native MVS and on z/OS UNIX System Services, few have all the requisite skills to successfully implement z/OS Communications Server on their own. As more and more systems programmers acquire skills in UNIX System Services and in TCP/IP, this will become less and less the case. Working with the z/OS UNIX System Services implementer when implementing z/OS Communications Server provides the most effective solution to establishing a working z/OS Communications Server environment.

Additional assistance is available to the z/OS UNIX System Services implementor at the z/OS wizards website, http://www.ibm.com/eserver/zseries/zos/wizards/. Wizards are interactive assistants that simplify tasks such as installation planning, as well as configuration and customization.

If you are migrating to z/OS Communications Server, establish a migration process to move all your existing applications, and after this, consider the use of new and enhanced functions based on *z/OS Communications Server: New Function Summary*. z/OS Communications Server allows multiple copies of the TCP/IP protocol stack to execute on the same MVS image. However, with all the performance enhancements introduced in z/OS Communications Server, it is probably not necessary to implement a multi-stack system for production purposes unless one is considering building a system programming test stack.

You are now ready to move on to the following steps.

Step 1: Install z/OS Communications Server

Before you begin the installation:

- Read *z/OS and z/OS.e Planning for Installation* to help you plan the installation and migration of z/OS Communications Server.
- Be sure you understand the data set naming conventions used in TCP/IP. You can find this information in "Configuration data set naming conventions" on page 28.
- Consult the *z/OS Program Directory* (Customization considerations for Wave 1D) for current information about the material, procedures, and storage estimates of the MVS image.

Install z/OS Communications Server with other elements of z/OS. If you use the ServerPac method of installation, see *z/OS Installing Your Order*; if you use the CBPDO method of installation, refer to *z/OS Program Directory*. When appropriate, those two documents will direct you back to this document to customize the TCP/IP data sets and procedures and verify their configuration.

Verifying the initial installation

Both the *z/OS Program Directory* and *z/OS Installing Your Order* contain step-by-step instructions that can be used to set up and verify a basic TCP/IP configuration with only the loopback address and a few key servers. For more information regarding these instructions, refer to the information about Wave 1D customizations in the *z/OS Program Directory* or the information about verifying your installation in *z/OS Installing Your Order*.

Step 2: Customize z/OS Communications Server

To customize TCP/IP you need to update the cataloged procedures and configuration data sets for the TCP/IP address space, its clients, and servers.

z/OS Communications Server runs as a started task in its own address space. Each of the servers runs in its own address space and is started with its own procedure. The TCP/IP address space requires:

- A procedure in a system or recognized PROCLIB.
- A data set that provides configuration definitions for the TCP/IP address space and includes statements affecting many of the servers. This data set is referred to as PROFILE.TCPIP.
- A data set to provide the parameters that are common across all clients. This data set is referred to as TCPIP.DATA.

Many of the servers also require other data sets for their specific functions.

Making SYS1.PARMLIB changes

You need to make certain changes to SYS1.PARMLIB. These changes depend on which of the following installation methods you use:

ServerPac method
: After the file system is restored (through the RESTFS job), you will see that ServerPac has changed some of the parmlib members. Follow the instructions to change the BPXPRMxx member of parmlib.

CBPDO method
: Change the parmlib members according to the instructions listed in the chapters that describe installation instructions for Wave 1. Tables describing changes to parmlib and changes to the BPXPRMxx member are included.

Note: z/OS Communications Server exploits z/OS UNIX services even for traditional MVS environments and applications. Prior to utilizing TCP/IP services, therefore, a full-function mode z/OS UNIX environment—including a Data Facility Storage Management Subsystem (DFSMSdfp), a Hierarchical File System (HFS), and a security product (such as Resource Access Control Facility (RACF))—needs to be defined and active before z/OS Communications Server can be started successfully.

Additional information about required TCP/IP definitions for the UNIX environment can be found in "Defining TCP/IP as a UNIX System Services physical file system (PFS)" on page 68 and "UNIX System Services security considerations" on page 63.

Common z/OS UNIX configuration problems: Following are some explanations and possible solutions for common problems that you may encounter when configuring the z/OS UNIX environment.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIP-BPX1SOC,
     00000003,FFFFFFFF,00000070,112B00B6
```

  These messages usually indicate that both INET and CINET FILESYSTYPE have been specified. Only one should be specified; refer to the FILESYSTYPE section in *z/OS UNIX System Services Planning* for additional information.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIP-BPX1SOC,
     00000003,FFFFFFFF,0000006F,112B00B0
```

  These messages indicate that the requester of the service is not privileged. The service requested requires a privileged user. Check the documentation for the service to understand what privilege is required.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR
     FOR TCPIPA-BPX1IOC,8008C981,FFFFFFFF,0000009E,12B2005A

EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED
```

  These messages usually indicate that an incorrect jobname was specified in the SUBFILESYSTYPE NAME() definition in the BPXPRM*xx* member for a common INET environment. In this scenario, the NAME() must match TCPIPA.

- TCP/IP initialization fails with the following messages:

```
IEA8481 DUMP SUPPRESSED - ABDUMP MAY NOT DUMP STORAG FOR KEY 0-7 JOB TCPV34A
IEF4501 TCPIPA TCPIPA - ABEND=SEC6 U0000 REASON=0F01C008
```

  These messages are usually an indicator that an OMVS RACF segment has not been defined for the user ID associated with the TCP/IP started procedure. Define an OMVS segment with a UID of 0 for the user ID associated with the TCP/IP started procedure.

- TCP/IP initialization fails with the following messages:

```
IEF4031 TCPIPA - STARTED - TIME=16.01.25
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIPA-BPX11OC,
     8008139A,FFFFFFFF,00000079,12D2025E
EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED.

==> The 0079 value is EINVAL - The parameter is incorrect
==> The 025E value is JRSocketCallParmError - A socket syscall
    contains incorrect parameters
```

These messages usually indicate that an incorrect entry point name has been specified in the SUBFILESYSTYPE ENTRYPOINT() definition. The correct value is ENTRYPOINT(EZBPFINI).

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIPA-BPX1SOC,
     00000003,FFFFFFFF,0000045A,112B0000
EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED.

==> The 045A value is EAFNOSUPPORT - The address family is not supported
```

  These messages indicate that AF_INET was not defined or did not initialize properly. Check for any earlier z/OS UNIX messages and verify that the z/OS UNIX NETWORK DOMAINNAME(AF_INET) statement is in your BPXPRM*xx* member.

- After issuing a NETSTAT command from TSO, the following message is displayed:

```
netstat
CEE5101C During initialization, the z/OS UNIX callable service
     BPX1MSS failed. The system return code was 0000000156,
     the reason code was 0507014D. The application will be
     terminated.
NETSTAT ENDED DUE TO ERROR+
READY
?
USER ABEND CODE 4093 REASON CODE 00000090
READY

==> The 0156 value is EMVSINITIAL - Process initialization error
==> The 014D value is JRFsFailChdir - The dub failed, due to
    an error with the initial home directory
```

  These messages indicate that the user ID issuing the NETSTAT command does not have an OMVS RACF segment defined for it. Define an OMVS segment for this user ID or activate the default OMVS segment support. For details, see "UNIX System Services security considerations" on page 63.

- Socket applications using the z/OS Communications Server TCP/IP Services APIs fail with an ERRNO of 156.

  ERRNO 156 indicates a z/OS UNIX process initialization failure. This is usually an indication that a proper OMVS RACF segment is not defined for the user ID associated with the application. The RACF OMVS segment may not be defined or may contain errors such as an improper HOME() directory specification. If the OMVS segment is not defined, you may also receive the following message:

```
ICH4081 USER(USER8   ) GROUP(SYS1   )   NAME(TSO USERID USER8  )
   CL(PROCESS )
   OMVS SEGMENT NOT DEFINED
```

  In this example, USER8 is the user ID associated with the failing application. To correct this problem, define a proper OMVS segment for the user ID associated with the failing application. For details, see "UNIX System Services security considerations" on page 63.

Completion of these steps ensures that the applications and resources on the target system will function correctly at the new level.

The subsequent chapters in this document show you how to:

- Configure the TCP/IP address space by updating the samples provided in SEZAINST(SAMPPROF) and SEZAINST(TCPIPROC).

- Configure the universal client parameters provided in SEZAINST(TCPDATA).
- Configure the site table, defined in *hlq*.HOSTS.LOCAL or *hlq*.ETC.IPNODES, to identify the Internet names and addresses of your TCP/IP host.
- Customize the TCP/IP Component Trace parameters by updating the CTRACE parameter in the PARM= field of the EXEC JCL statement in the TCP/IP started procedure.

  You can find a description of the MVS Component Trace support in the *z/OS Communications Server: IP Diagnosis Guide*.
- Specify the ENVAR parameter on the PARM=keyword to override the resolver file. For more information on setting the environment variable RESOLVER_CONFIG using the ENVAR parameter, see "Considerations for multiple instances of TCP/IP" on page 72.
- Configure each of the servers you want to run. This might require:
  - Modifying sample procedures and adding them in your PROCLIB
  - Modifying the configuration data set, PROFILE.TCPIP
  - Adding port numbers to *hlq*.ETC.SERVICES
  - Modifying other data sets containing server-specific parameters

You can find the sample procedures and data sets in SEZAINST or the HFS. Table 1 on page 30 provides additional reference information you can use as you configure and customize each server.

You can find general information about starting, stopping, and dynamically controlling the servers in *z/OS Communications Server: IP System Administrator's Commands*.

Step 3: Configure VMCF and TNF

The Pascal socket interface makes use of the IUCV/VMCF services for a limited set of inter-address space communication flows. As a result, if you are using any applications (provided by IBM or others) that use the Pascal socket API, you must insure that the VMCF and TNF subsystems are active before the applications are started. TCP/IP provides several applications and commands that exploit these interfaces, such as the SMTP and LPD servers; therefore, some installations will require setting up VMCF and TNF.

The restartable VMCF must be started before TCP/IP if you want the VMCF node name used as a default host name during TCP/IP initialization (in cases where no other host name can be located).

Note: Host name is the value normally specified on the TCPIP.DATA HOSTNAME statement.

Also note that the VMCF node name is used as a system name qualifier when processing the TCPIP.DATA file and by the SMTP server as the NJE node name. It is recommended that the MVS system name is used for the VMCF node name specification and that the NJE node name is specified explicitly by using the NJENODENAME statement in the SMTP configuration data set.

You can configure Virtual Machine Communication Facility (VMCF) and Termination Notification Facility (TNF) in two different ways: as restartable subsystems or as non-restartable subsystems.

Restartable subsystems

Configuring VMCF and TNF as restartable subsystems has the following advantages:

- Error detection is provided when the subsystems do not seem to be initializing properly.
- You can change the system name on the restart.
- Commands are available to remove users from internal tables, display current users and to terminate the subsystem.

In summary, a restartable VMCF and TNF configuration provides better availability and is therefore recommended.

If you choose to use restartable VMCF and TNF, follow these steps:

1. Update your IEFSSN*xx* member in SYS1.PARMLIB with the TNF and VMCF subsystem statements required by TCP/IP. The specification can be in either the IBM recommended keyword parameter form or the positional parameter form of IEFSSNxx. For example:

   ```
   * The keyword parameter form is:

   SUBSYS SUBNAME(TNF)
   SUBSYS SUBNAME(VMCF)

   * The positional parameter form is:
   TNF
   VMCF
   ```

2. Add procedure EZAZSSI to your system PROCLIB. A sample of this procedure is located in the SEZAINST library.

   ```
   //EZAZSSI PROC P=&SYSNAME.
   //STARTVT EXEC PGM=EZAZSSI,PARM=&P,TIME=1440
   ```

3. Start VMCF and TNF using the procedure EZAZSSI before starting TCP/IP. If your nodename is the same as the MVS system symbolic &SYSNAME, then you can start VMCF and TNF with the following command:

   ```
   S EZAZSSI
   ```

   If your nodename is different than the MVS system symbolic &SYSNAME, start VMCF and TNF as follows:

   ```
   S EZAZSSI,P=nodename
   ```

   Replace nodename with the SYSTEM NAME of your MVS system.

Non-restartable subsystems

If you will not be using restartable VMCF and TNF, you should update your IEFSSN*xx* member in SYS1.PARMLIB with the following subsystem statements required by TCP/IP. The specification can be in either the IBM recommended keyword parameter form or the positional parameter form of IEFSSNxx. For example:

```
* The keyword parameter form is:

SUBSYS SUBNAME(TNF) INITRTN(MVPTSSI)
SUBSYS SUBNAME(VMCF) INITRTN(MVPXSSI) INITPARM(nodename)


* The positional parameter form is:

TNF,MVPTSSI
VMCF,MVPXSSI,nodename
```

Replace node name on the VMCF line with the NJE node name of your MVS system.

VMCF commands

If you will be using restartable VMCF, the following VMCF commands let you display the names of the current users of VMCF and TNF, and if necessary, remove names from the name lists.

Note: Removing names from the name lists and stopping either subsystem can have undesired results, if done hastily. Use the REMOVE and stop (P) commands carefully and only as a last resort.

If you remove a user, the application is not canceled, nor is the connection severed. In other words, the *removed* application may remain active in the system, and may subsequently abend 0D6/0D4/0C4, or cause TCP/IP to hang. A user that is removed from VMCF may still be a user of TNF and even TCP/IP, and vice versa.

To terminate users and stop VMCF or TNF properly, follow these steps:

1. Display the current users of the subsystems, using one of the following:

   ```
   F VMCF,DISPLAY,NAME=*
   F TNF,DISPLAY,NAME=*
   ```

2. Terminate those users. If termination fails, use the REMOVE command as a last resort to force them from the name list.
3. Stop the subsystem, using one of the following commands:

   ```
   P VMCF
   P TNF
   ```

   If the P command fails, use one of the following commands:

   ```
   FORCE ARM VMCF
   FORCE ARM TNF
   ```

Following are descriptions of the commands:

F TNF,DISPLAY,NAME=[name|*]
Displays the named user [or all (*) users] of TNF, sorted by ASID.

F TNF,REMOVE,NAME=[name|*]
Removes either the named user [or all (*) users] from the TNF internal tables.

P TNF Requests TNF to terminate.

F VMCF,DISPLAY,NAME=[name|*]
Displays the named user [or all (*) users] of VMCF, sorted by name.

F VMCF,REMOVE,NAME=[name|*]
Removes either the named user [or all (*) users] from the VMCF internal tables.

P VMCF
Requests VMCF to terminate

Following are sample commands:

```
F TNF,DISPLAY,NAME=TCPV3
F VMCF,DISPLAY,NAME=*
F TNF,REMOVE,NAME=FTPSERV
F VMCF,REMOVE,NAME=*
P TNF
```

Common VMCF problems

Following are some common VMCF problems:

- VMCF or TNF fail to initialize with an 0C4 abend.

  This is probably an installation problem; check the program properties table (PPT) entries for errors. Some levels of MVS do not flag PPT syntax errors properly.

- Abends 0D5 and 0D6 after REMOVEing a user.

  This is probably because the application is still running and using VMCF. It is not recommended that users be removed from VMCF or TNF without first terminating the affected user.

- VMCF or TNF do not respond to commands.

  This is probably because one or both of the non-restartable versions of VMCF or TNF are still active. To get them to respond to commands, stop all VMCF/TNF users, FORCE ARM VMCF and TNF, then use EZAZSSI to restart.

- VMCF or TNF cannot be stopped.

  This is probably because users still exist in the VMCF and TNF lists. Use the *F VMCF,DISPLAY,NAME=** and *F TNF,DISPLAY,NAME=** commands to identify those users who are still active. Then either cancel those users or remove them from the lists using the *F VMCF,REMOVE* and *F TNF,REMOVE* commands.

- Address Space Identifiers (ASIDs) become nonreusable when VMCF and TNF address spaces are stopped or restarted.

  Because VMCF and TNF address spaces provide PC-entered services that must be accessible to all address spaces, they each obtain a system LX. This causes the ASIDs associated with these address spaces to be nonreusable when these address spaces are terminated. If VMCF and TNF are terminated enough times all available ASIDs could be exhausted, preventing the creation of new address spaces on the system. In this case, an IPL will be required. For more information on tuning parameters for the maximum number of ASIDs in a system, refer to the MAXUSER parameter in *z/OS MVS Initialization and Tuning Reference*.

IUCV/VMCF considerations

The IUCV/VMCF inter-address space communication API enables applications running in the same MVS image to communicate with each other without requiring the services of the TCP/IP protocol stack. The VMCF/TNF subsystems provide these services, which are still available in z/OS Communications Server. Several components of TCP/IP in z/OS Communications Server continue to make some use of these services for the purpose of inter-address space communications. These include:

- The AF_IUCV domain sockets for the TCP/IP C socket interface. The AF_IUCV domain enables applications executing in the same z/OS image and using the TCP/IP C socket interface to communicate with each other using a socket API, but without requiring the services of the TCP/IP protocol stack, as no network flows result in these communications. This is quite different from the more common AF_INET domain that enables socket communication over a TCP/IP network. AF_IUCV sockets continue to be supported in z/OS Communications Server.

  An example of a TCP/IP-provided application that exploits AF_IUCV sockets is the SNMP Query Engine component (SQESERVE). The z/OS UNIX socket

library provides a similar functionality to the AF_IUCV domain sockets with its AF_UNIX domain. Users creating new applications should consider using AF_UNIX domain sockets.

- The Pascal socket interface also makes use of the IUCV/VMCF services for a limited set of inter-address space communication flows. As a result, any applications (provided by IBM or others) that use the Pascal socket API also still have a requirement for the VMCF/TNF subsystems. TCP/IP provides several applications and commands that exploit these interfaces, such as the SMTP and LPD servers, and the TSO TELNET, HOMETEST, TESTSITE, RSH, REXEC, and LPR commands. IUCV/VMCF services require the usage of an address space name of SYSTEM. This means a TSO user cannot have the user ID name of SYSTEM.

Therefore, in z/OS Communications Server you must continue to configure and start the VMCF and TNF subsystems as you did in TCP/IP V3R2. However, because the VMCF/TNF subsystems are no longer used to communicate directly with the TCP/IP protocol stack in z/OS Communications Server, the amount of CPU they will consume will be significantly lower than in the TCP/IP V3R2 environment.

Step 4: Update the VTAM application definitions

You must update the VTAM definitions for TN3270 Telnet and any other of these applications that you configure on your system. You can find example VTAM definitions for each of these applications in their respective chapters.

- SNALINK
- SNALINK LU6.2
- TN3270 Telnet
- X.25 NPSI Server

SEZAINST(VTAMLST) contains a sample of the VTAM definitions for TN3270 Telnet applications. You should copy this member, update it, and add it to the ATCCON*xx* member of VTAMLST. This will ensure that the TN3270 Telnet applications are activated when VTAM is started.

Because the TCP/IP LU code cannot handle multiple concurrent sessions, you must code SESSLIM=YES for each TN3270 Telnet LU defined to VTAM. Otherwise, if SESSLIM=NO, menu or session manager applications that use return session processing might cause session termination.

Step 5: Verify that the required address spaces are active

TCP/IP uses services from other address spaces in its processing. While TCP/IP waits for availability of those services, it is recommended that you verify that the OMVS, resolver, and VTAM address spaces have completed initialization before starting TCP/IP. If using automation to start TCP/IP, have it look for the BPXI004I, EZZ9291I, and IST020I messages, respectively, before issuing the START command.

For information on how the resolver can be started, see “Understanding resolvers” on page 20. You can use the resolver’s MODIFY DISPLAY command to check that the resolver is active and what resolver setup statements are being used. For the syntax and usage of the command, see *z/OS Communications Server: IP System Administrator’s Commands*.

Step 6: Start the TCP/IP address space

Enter the MVS START command from the operator's console to start TCP/IP, specifying the member name of your cataloged procedure. This will start the TCP/IP address space and any of the servers you have defined in the AUTOLOG statement in PROFILE.TCPIP. For example, if the procedure to start the TCP/IP address space was called TCP1 in your PROCLIB, you would enter:

```
START TCP1
```

For information on updating the TCPIP cataloged procedure or configuration statements used to configure the TCPIP address space, refer to *z/OS Communications Server: IP Configuration Reference*.

Step 7: Set up cataloged procedures and configuration data sets

At this point in the configuration process, you can choose to either set up procedures or you can do each one individually when you set up the appropriate application, function, or server.

See the remaining chapters in this document for more information about setting up the appropriate application, function, or server.

Step 8: Customize TCP/IP messages

The messages for every TCP/IP server program are compiled and linked with the program and reside in an internal message repository. Some of the server programs that are written in the C language also have their messages in external data sets. You can edit these external message data sets to translate the messages to another language or customize them to suit your installation.

How to access the message data sets

The procedures for these servers have a special DD statement that point to the external message data set. If you are going to override the internal messages and use external customized messages, you need to remove the comment from the appropriate DD statement and ensure it points to the correct data set.

The following table shows the servers that have external messages, the DD statement used, and the name of the message data set delivered with the system:

Server	DD statement	Data set
NCPROUTE	//MESSAGE	SEZAINST(EZBNRMSG)
SNMP Query Engine	//MSSNMPMS	SEZAINST(MSSNMP)
MISC Server	//MSMISCSR	SEZAINST(MSMISCSR)

Message text

The message text might include special characters for the variable fields that are converted when the message is printed or displayed and control characters that affect the message format. The conversion characters start with a percent sign (%) and the control characters start with a backslash (\). These are all standard notations for the C language print function. The messages might also contain comments which start with `/*` and end with `*/`.

In the following simulated message, the control character `\n` forces a new line to print and the string variables, represented by `%s`, are converted in the order they are passed from the program.

```
29999  I  Command %s received from user %s\n
```

Message format

The following diagram explains the syntax for TCP/IP message IDs on the host:

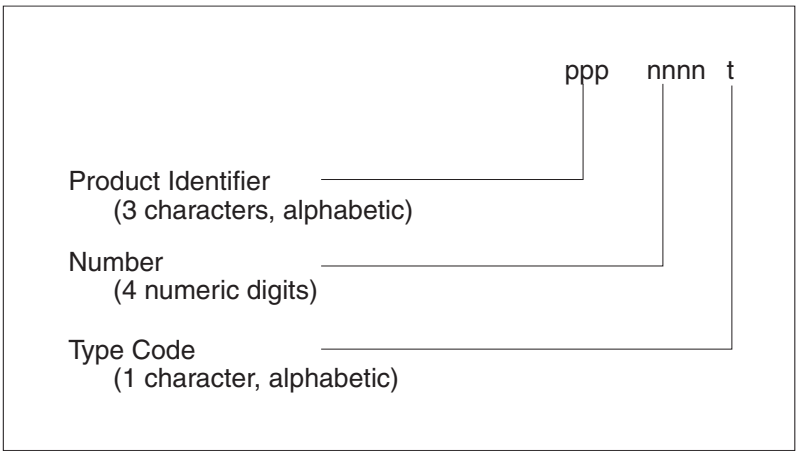


Figure 18. Syntax for TCP/IP message IDs

The ***product identifiers*** (ppp) for TCP/IP are EZA, EZB, EZY, and EZZ. The ***number*** (nnnn) indicates a unique 4-digit numeric value assigned to the message by product. The ***type*** (t) indicates the severity assigned to the message.

Rules for customizing the messages

The general rule for customizing or translating messages is to only change the text portion of the message.

- Do not change the MARGIN, PRODUCT, and COMPONENT definitions at the top of the data set. These are required definitions for the program. For example, these entries at the top of the MISC server message data set should not be changed:

  ```
  MARGINS(1,72)
  PRODUCT EZA
  COMPONENT MSC
  ```

- Do not change the message numbers and the severity code. These parts of the message have specific meaning; if you change them the program may not work correctly.
- Do not change the conversion characters. These indicate that the program is passing data, the type of data it is passing, and the appropriate way to display or print this data. For example, do not change or delete %s and %d in the following message:

  ```
  4858 W   "Route from %s in unsupported address family %d\n"
  ```

- You can reorder the variables that are passed in the message. For example, you can reverse the order of the two string variables that are passed when translating a message by specifying the new order of the arguments in parentheses following the message text:

  Before: `29999I  Command %s received from user %s\n`

  After: `29999I  Utilisador %s envio instrucion %s\n (2, 1)`

  The result would be `EZY9999I Utilisador MANNY envio instrucion FTP` instead of `EZY9999I Command FTP received from user MANNY`.

- Watch for any program parameters or keywords that might be in the message text. In most cases, you should not translate them.

  For example, in the following message, `'active'` is a keyword used in the gateway definition and should not be translated:

  ```
  4851 E   "First two elements must be 'active' for active gateway\n"
  ```

Chapter 3. Security

The z/OS Communications Server, along with other elements of z/OS, provide numerous enterprise-strength security services to protect your mission-critical data. This chapter provides an overview of these technologies and how they can be used for a safe and secure z/OS TCP/IP deployment.

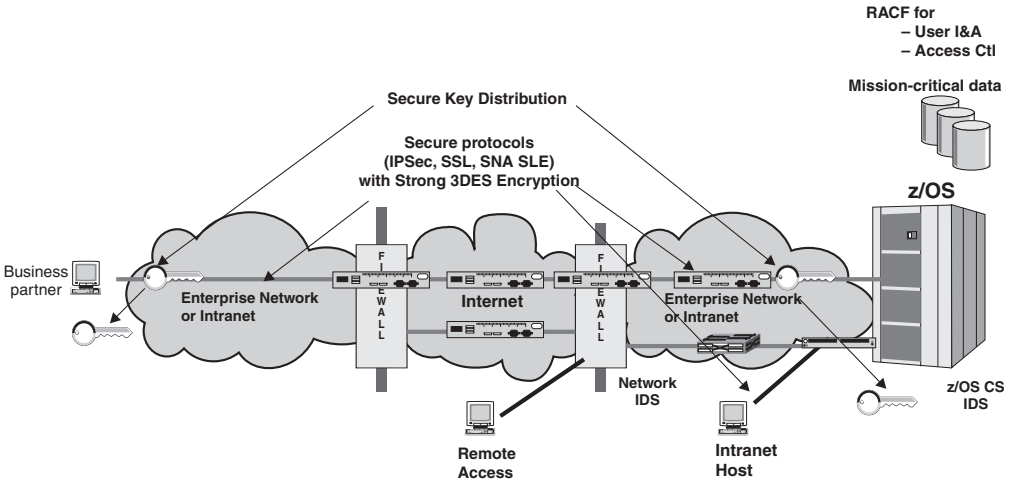


Figure 19. Elements of a secure TCP/IP deployment

Tip: Many of the tasks, examples, and references in this chapter assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

The Communications Server protects data and other resources on the system. Communications Server applications use RACF services to ensure that users requesting application access are identified and authenticated, and to protect data and other system resources from unauthorized access. The Communications Server safeguards the availability of the system by protecting against denial of service attacks from the network.

The Communications Server protects data in the network by supporting a variety of cryptographic-based network security protocols such as IPSec, SSL, and SNA Session Level Encryption. These security protocols ensure that data received is originated by the claimed sender (data origin authentication), that contents were unchanged in transit (message integrity), and that sensitive data is concealed using encryption (data privacy).

The Communications Server provides security event reporting to record potential security violations. These services may help you identify potential sources of subsequent attacks, respond more quickly to network attacks, and manage system resources during periods of high network traffic for key applications.

Note: Some of the security features described in this chapter have not yet been implemented for IPv6. To determine which functions are supported for IPv6, refer to the IPv6 support tables in *z/OS Communications Server: IPv6 Network and Application Design Guide*.

System resource protection

Application security

The Communications Server protects data and other system resources accessed by applications included in the Communications Server element. This protection requires verification of the identity of the end user requesting access. This process is called identification and authentication. In addition, access to resources must be limited to those users with permission. This process is called access control. Communications Server applications use RACF for identification and authentication, and access control decisions. Authenticated users are granted access to RACF resources only for which they have permission

Some applications allow anonymous access. Applications that allow anonymous access include anonymous FTP, Remote Execution, and Trivial File Transfer Program (TFTP). The Communications Server ensures that all anonymous access can be controlled by the installation. If anonymous access is allowed, the resources accessed can be limited in several ways:

- The application can be configured to limit resources for which access will be attempted.
- The application can be configured to use a RACF user ID to represent the anonymous user. In this case, access is allowed for those resources specifically permitted for the anonymous RACF user ID and for those resources that are universally accessible.

Most Communications Server applications must be configured specifically to allow anonymous access. One exception is TFTP. TFTP can be configured to control those directories that contain files that can be transferred.

The following chart depicts a representative set of Communications Server applications, whether end user identification is required, and the security credentials under which resource access is made. For more information on specific application considerations, refer to the individual chapters for each application.

Server	End User Identification	Resource Access
FTP	Optional (1)	End user ID or configured anonymous user ID (2)
LPD	Optional (1)	Server ID or end user ID
TFTP	No	Server user ID (2)
MVS REXECD	Required	End user ID
MVS RSHD	Required (password optional) (1)	Surrogate user ID or end user ID
UNIX REXECD	Required	End user ID
UNIX RSHD	Required (password optional) (1)	End user ID or Server user ID(exit routine to verify request)
UNIX SHELL (telnet/rlogin)	Required	End user ID

Figure 20. User identification, authentication, and access control for z/OS Communications Server applications

(1) All optionals are installation controlled and can all be configured to require full end user identification.

(2) Files accessible can be configured on a server basis to limit access.

TCP/IP resource protection

The Communications Server uses the System Authorization Facility (SAF) to protect TCP/IP resources from unauthorized access. These resources are represented by resource profiles defined in the SERVAUTH class. The use of SERVAUTH is optional. The installation can choose to use any combination of the protections provided by SERVAUTH.

In addition to the use of SERVAUTH protection, further resource protection is provided through such functions as Intrusion Detection Services (IDS), syslogd isolation, and IP filtering, or through controlling access to the VARY TCPIP command. These and other topics are discussed in more detail later in the chapter.

Local user access control to TCP/IP resources using the SAF

The SAF can control the ability of users executing on z/OS to access select TCP/IP resources, protecting against unauthorized user access to these resources.

With this solution, the administrator defines the above TCP/IP resources as SAF resources. The resource profiles are defined as part of the SERVAUTH class. The Communications Server allows the local user access to these resources based on the user or group permissions associated with the SAF resource.

Table 10 on page 120 summarizes the SERVAUTH profiles used by TCP/IP. These profiles are described in this chapter and elsewhere in this document.

Table 10. SERVAUTH profiles used by TCP/IP

Function	Description	SERVAUTH profile
TN3270 server access control	Controls ability to access TN3270 server based on SAF user ID associated with TLS-authenticated X.509 client certificate	EZB.TN3270.*sysname*.*tcpname*.PORT*xxxxx*
FTP server access control	Controls ability to access FTP server based on SAF user ID associated with TLS-authenticated X.509 client certificate	EZB.FTP.*sysname*.*ftpdaemonname*.PORT*xxxxx*
DCAS server access control	Controls ability to access DCAS server based on SAF user ID associated with TLS-authenticated X.509 client certificate	EZB.DCAS.*cvtsysname*
TCP stack access control	Controls user ability to open a socket and get host name or host ID	EZB.STACKACCESS.*sysname*.*tcpname*
TCP local port access control	Controls user ability to bind to a non-ephemeral TCP or UDP port	EZB.PORTACCESS.*sysname*.*tcpname*.*port_safname*
TCP netaccess access control	Controls local user inbound and outbound access to network resources, and local user access to local IP address when explicitly binding to local interface (or using job-specific source IP addresses)	EZB.NETACCESS.*sysname*.*tcpname*.*security_zonename*
Netstat command access control	Provides ability to restrict Netstat usage	EZB.NETSTAT.*sysname*.*tcpname*.*netstat_option*
Policy Agent command control	Provides ability to restrict **pasearch** command usage by policy type	EZB.PAGENT.*sysname*.*tcpname*.*policy_type*
FTP SITE command control	Provides ability to restrict usage of SITE DUMP and DEBUG commands (commands generate large amount of output)	EZB.FTP.*sysname*.*ftpdname*.SITE.DUMP EZB.FTP.*sysname*.*ftpdname*.SITE.DEBUG
SNMP agent control	Provides ability to control usage of SNMP subagents that connect to the TCP/IP SNMP agent	EZB.SNMPAGENT.*sysname*.*tcpname*
MODDVIPA utility program control	Provides ability to restrict usage of MODDVIPA utility program (creates new DVIPA on system)	EZB.MODDVIPA.*sysname*.*tcpname*
Fast Response Cache Accelerator (FRCA) Access Control	Provides ability of user to create FRCA cache (FRCA used by Web servers for caching static Web pages in the stack)	EZB.FRCAACCESS.*sysname*.*tcpname*
TCP connection information service access control	Provides ability to restrict access to the TCP connection information using TCP connection information service; intended for network management applications	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPCN
Real-time SMF information service access control	Provides ability to restrict access to select real-time SMF records accessible using the SMF information service; intended for network management applications	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPSM

Table 10. SERVAUTH profiles used by TCP/IP (continued)

Function	Description	SERVAUTH profile
TCP/IP packet trace service access control	Provides ability to restrict access to select real-time packet trace records accessible using the TCP/IP packet trace service; intended for network management applications	EZB.NETMGMT.*sysname.tcpname*.SYSTCPDA
FTP HFS access control	Provides ability to generally restrict FTP user access to HFS	EZB.FTP.*sysname.ftpdaemonname*.ACCESS.HFS
Broadcast access control	Provides ability to control whether an application is permitted to set the SO_BROADCAST socket option needed to send broadcast datagrams	EZB.SOCKOPT.*sysname.tcpname*.SO_BROADCAST
IPv6 Advanced Socket API access control	Provides ability to control whether an application is permitted to set IPv6 advanced socket API options: IPv6_NEXTHOP IPv6_TCLASS IPv6_RTHDR IPV6_HOPOPTS IPV6_DSPOPTS IPV6_RTHDRDSTOPT IPV6_PKTINFO IPV6_HOPLIMIT	EZB.SOCKOPT.*sysname.tcpname*.IPV6_NEXTHOP EZB.SOCKOPT.*sysname.tcpname*.IPV6_TCLASS EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDR EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPOPTS EZB.SOCKOPT.*sysname.tcpname*.IPV6_DSTOPTS EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDRDSTOPTS EZB.SOCKOPT.*sysname.tcpname*.IPV6_PKTINFO EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPLIMIT
TCP/IP stack initialization access control	Controls ability of applications to open a socket before AT-TLS policy is loaded into the TCP/IP stack	EZB.INITSTACK.*sysname.tcpname*
CIM provider access control	Provides ability to restrict access to CIM data	EZB.CIMPROV.*sysname.tcpname*
ipsec command access control	Provides ability to control **ipsec** command usage	EZB.IPSECCMD.*sysname.tcpname.command_type*

Stack access control

Stack access control allows control of access to a TCP/IP stack using an SAF security server. It provides a way to generally allow or disallow users or groups of users access to a TCP/IP stack. The function controls the ability of a user to open an AF_INET or AF_INET6 socket, and to get the host ID or host name. The TCP/IP stack to be protected is represented by the resource name *EZB.STACKACCESS.sysname.tcpname*. Access to the stack is allowed if the user is permitted to the security profile in the SERVAUTH class covering this resource or if the security server indicates there is no profile covering this resource. There are no new TCP definitions required.

Guideline: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. If your product does not make this distinction, you must define the stack access resource profile and permit users to it whenever the SERVAUTH class is active.

The following example provides an overview of stack access control. *sysname* refers to the MVS system variable *sysname*. *tcpname* refers to the TCP/IP job name. As shown in the example below, user Tom has permission to access both Stack1 and Stack2, Joe does not have permission to access any stack, and Bob has permission to access Stack2 but not Stack1.

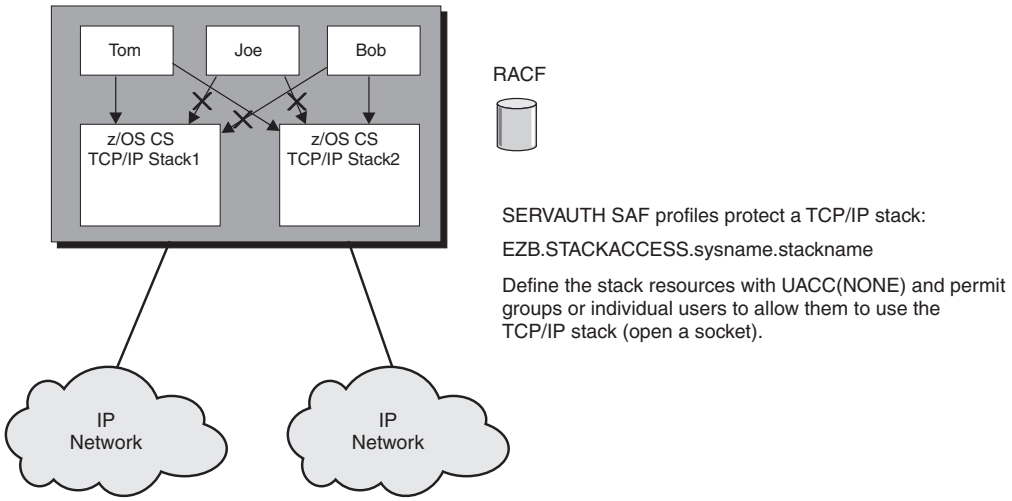


Figure 21. Stack access control overview

Port access control

Port access control uses the PORT and PORTRANGE statements to protect against unauthorized use of non-ephemeral ports. It allows control of an application's ability to bind to specific TCP and UDP ports or port ranges using the SAF. The port access support is enabled if the keyword, SAF, is specified on the PORT or PORTRANGE statement. The SAF keyword value specifies a portion of the resource name that represents the port. The user ID associated with the application at the time of the bind request must be permitted to the resource before the application is allowed to bind to the port. The port is represented by a SERVAUTH profile name of *EZB.PORTACCESS.sysname.tcpname.SAFkeyword*. *SAFkeyword* is the value specified on the SAF keyword on the PORT and PORTRANGE statement.

The following example provides an overview of port access control. As shown in the example below, z/OS user WEBSERV is permitted to bind to port 80. User Bob is not permitted to bind to port 80.

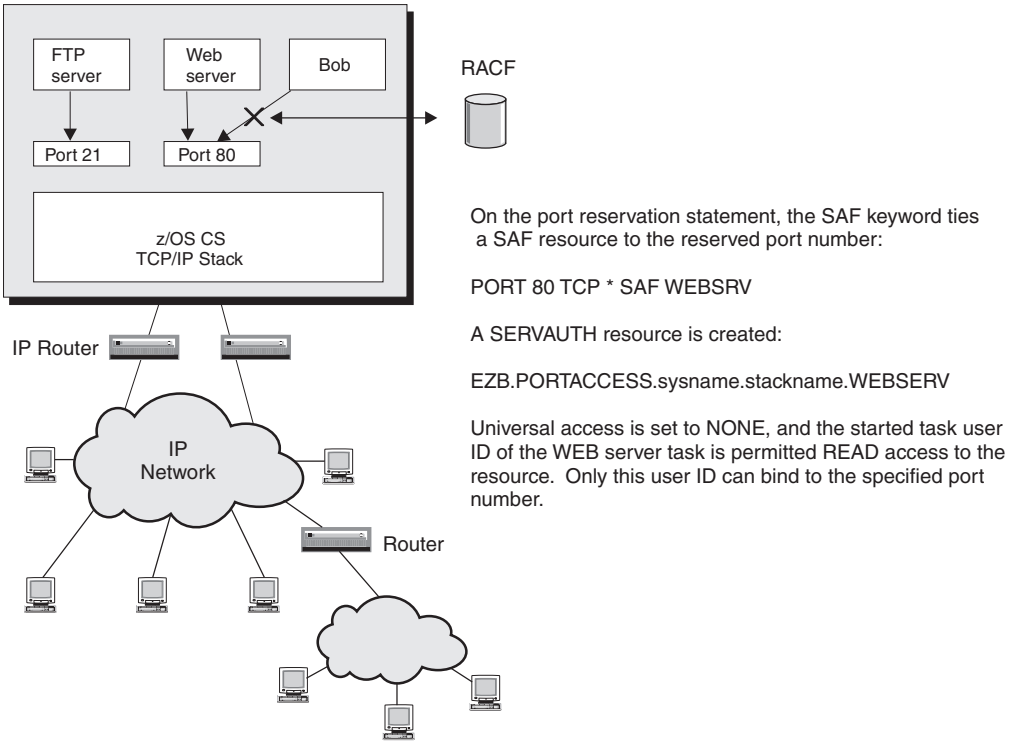


Figure 22. Port access control overview

Port access control also augments the job name reservation method. The PORT or PORTRANGE can be reserved with a job name, a wildcard job name (*), a partial wildcard job name (0-7 characters, followed by *), or the special job name of RESERVED. If job name is specified, the port is reserved for an application with the specified job name. If the wildcard job name is specified, the port is not reserved for any particular job name. For both of these cases, the SAF keyword, if specified, still verifies that the userid associated with the application at the time of the bind to the port is permitted access to the port. The RESERVED job name shuts down the use of a port or range of ports for any application.

The IPCONFIG, UDPCONFIG, and TCPCONFIG RESTRICTLOWPORTS statements specify that all applications binding to a low port (1–1024) must be APF-authorized or superuser, unless the SAF keyword is specified and the user ID binding to the port is permitted to the SAF resource. z/OS Communications Server client applications that need to bind to a low port are shipped as APF-authorized.

Network access control

Network access control gives system administrators the ability to assign permission for z/OS users to access certain networks and hosts. With this function, the ability of users to send or receive data between z/OS and certain networks can be controlled through z/OS. Network access control provides an additional layer of security to any authentication and authorization security that is used in the network or at the peer system by disallowing the unauthorized user to communicate with the peer network resource.

Essential elements of this function are as follows:

- The IP network is considered the resource to be protected.
- IP addresses are classified into *security zones*, in which each zone has a certain level of security sensitivity. A default security zone exists for interfaces that are

not explicitly associated with a specific security zone. Security zones consist of one or more, perhaps discontiguous, IP address ranges that have the same security sensitivity and are identified by a specific zone name.

- The SAF is used to check permission of users or groups of users to access the security zone.
- The installation defines a network access resource for each security zone and permits users or groups of users access to the resource. The security zone is represented by an SAF SERVAUTH profile name of *EZB.NETACCESS.sysname.tcpname.zonename*.
- TCP/IP keeps a mapping of network resources by IP address to security zones. This mapping is consulted on certain inbound and outbound operations to determine the corresponding resource zone name for the most specific network defined. Then the current user's access to that resource is queried using the SAF, and the operation will be allowed or denied completion accordingly. This mapping is also consulted when the security ioctl is issued to extract the port of entry zone name of a socket's current peer.
- Network access control is used to control z/OS user access to a peer address in an IP network through a sockets application. Resource access checks occur at connection setup or acceptance time for TCP, peer identification time for UDP and RAW, and on the first and potentially subsequent sends or receives (TCP, UDP, or RAW) to a particular destination in a socket's lifetime.
- Network access control is used to control z/OS user access to local addresses when a socket is bound to a local address. Resource access checks occur when an application explicitly binds a socket to a local address, including the addresses INADDR_ANY (0.0.0.0/32) or IN6ADDR_ANY (::/128). Job-specific source IP addresses are handled as if the application did an explicit bind to the configured address.
- Network access control security checks are made at the transport layer (TCP, UDP, and RAW). Other IP-specific packets generated by the stack are not covered under this function (such as ICMP echo replies, for example). Additionally, there is no user concept when dealing with packets that are being forwarded through the stack, and hence no checks are made.
- Network access control for outbound and inbound can be individually enabled or disabled.
- TCP/IP caches security information following network access control checks. Access is automatically rechecked on the next use of the cache whenever the RACF commands SETROPTS RACLIST, SETROPTS RACLIST REFRESH, or SETROPTS NORACLIST are issued for the SERVAUTH class. If you are using a security product other than RACF, the NetAccess zone table in the TCPIP PROFILE must be rebuilt to cause TCP/IP to recognize changes to the SERVAUTH class profiles for existing sockets.
- The socket calls bind, connect, and those that send data will fail with errno EACCES if the specified IP address is not permitted to be used by the application user. Inbound UDP and RAW datagrams that are not permitted under the current network access control policy are normally filtered out before they get to socket calls that receive data. It is possible for a datagram to arrive under a network access control policy that would allow it to be read, but then be received after a policy change that does not allow it. If the application (or common INET) has issued a select or poll on the socket, the receive call returns an EACCES errno to avoid blocking the application. Inbound TCP connections that are not permitted under the current network access control policy are also normally reset and discarded before they get to the socket backlog. In those cases where the only available new connections are not allowed and a select has

been issued, accept processing returns a new socket and the next attempt to send or receive data returns an ECONNRESET error.

- NetAccess inbound filtering needs to predict whether a future receive or accept will succeed. When there are multiple processes or threads operating on the same socket, to achieve consistent results you must ensure that certain calls are done under the same identity, or identities with equivalent network access control policies. For UDP and RAW sockets, the select, receive, and send calls must be done under equivalent policies. For TCP listening sockets, the select and accept calls must be done under equivalent policies.

The following example provides an overview of network access control. As shown in the example below, z/OS user Bob is permitted access to Security Zone A but not Security Zone B. An outbound connect from Bob is permitted to Security Zone A, but not Security Zone B. Bob is permitted to accept connections from Security Zone A but not Security Zone B.

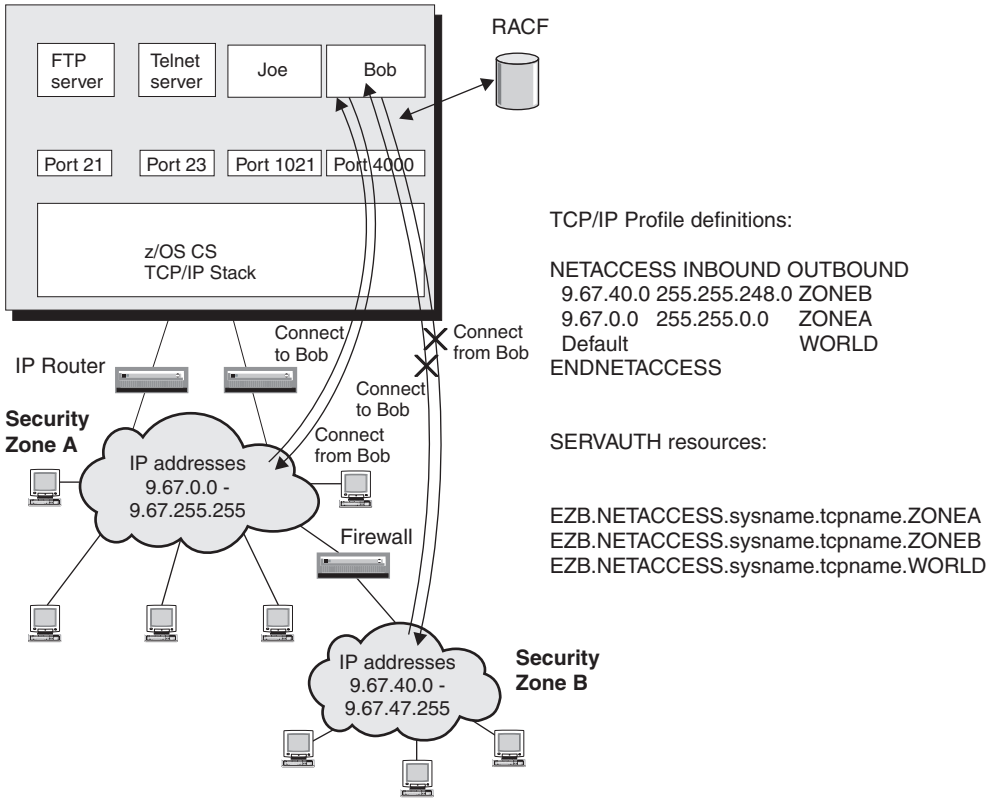


Figure 23. Network access control example

Socket option access control

Socket option access control gives system administrators the ability to assign permission for z/OS users to set selected socket options using a SAF-compliant security server. Access control is provided for the SOL_SOCKET level, SO_BROADCAST option, and the IPv6 advanced socket API options.

SO_BROADCAST socket option: IPv4 IP addresses are divided into a network portion, an optional subnetwork portion, and a host portion. Normally, UDP and RAW datagrams are delivered to the single peer system identified by the destination IP address. However, when the destination address is a subnetwork or network base address (host portion is all zeros), or a subnetwork or network broadcast address (host portion is all ones), the datagram is delivered to every peer

system in that subnetwork or network. Socket semantics require that an application set the SO_BROADCAST option on before attempting to send a datagram to a base or broadcast address. This protects the application from accidentally sending a datagram to many systems. Network access control does not check to see if there are other security zones defined within the scope of a destination subnetwork or network address or whether the user is permitted to send a datagram to all of these security zones.

Control over setting this socket option allows the system administrator to restrict use of subnetwork or network addresses to those users or programs that require them. This support is enforced at the PFS layer and applies to all z/OS Communications Server socket APIs.

TCP/IP programs known to set the SO_BROADCAST socket option include:

- OMPROUTE
- DHCP
- binlsd
- sntpd, when invoked with the -b option

Additionally, any programs that use the clnt_broadcast() service in the SUN RPC libraries, or the send_pkt(sock, pkt, addr, broadcast) service in the NCS RPC library with the broadcast parameter set, require permission to the SO_BROADCAST socket option. The following TCP/IP programs use RPC services that require permission to broadcast:

- rpcinfo, when invoked with the -b option
- orpcinfo, when invoked with the -b option

The socket option to be protected is represented by the resource name EZB.SOCKOPT.*sysname*.*tcpname*.SO_BROADCAST. When this profile is defined, users of any program setting this option require READ permission. Access to the option is also allowed if the security server indicates there is no profile covering this resource. Conditional access lists, such as PERMIT WHEN(PROGRAM(...)), are supported for profiles covering socket option access control resources. There are no new TCP definitions required.

Guideline: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. If your product does not make this distinction, you must define the socket option resource profile and permit users to it whenever the SERVAUTH class is active.

Following is an example of the definitions:

```
RDEFINE SERVAUTH EZB.SOCKOPT.*.*.SO_BROADCAST UACC(NONE)
PERMIT EZB.SOCKOPT.*.*.SO_BROADCAST CLASS(SERVAUTH) ACCESS(READ) ID(OMPROUT)
PERMIT EZB.SOCKOPT.*.*.SO_BROADCAST CLASS(SERVAUTH) ACCESS(READ) ID(*)
WHEN(PROGRAM(ORPCINFO))
```

The program name listed in the conditional access list must be the name the program is invoked by. Most TCP/IP applications are invoked by an alias name rather than the module name. Table 11 lists TCP/IP applications that send broadcast datagrams:

Table 11. TCP/IP application load module and alias names

Load module	Alias
EZAORRTE	OMPROUTE

Table 11. TCP/IP application load module and alias names (continued)

Load module	Alias
EZATDHSD	DHCPSD
EZATDLSD	BINLSD
EZASNTPD	SNTPD
EZARPCIN	RPCINFO
EZARORNP	ORPCINFO

Tip: In the UNIX System Services environment, both /bin/rpcinfo and /bin/orpcinfo are externally linked to ORPCINFO. Either command executes the EZARORNP program.

To use program names in conditional access lists, the program must be loaded into a controlled environment from a program controlled data set. TCP/IP applications are distributed in the SEZALOAD load library. To program control this data set, you must add it to the ** profile in the PROGRAM class as follows:

```
RALTER PROGRAM ** ADDMEMBER('TCPIP.SEZALOAD'//NOPADCHK)
```

For more information on program control, refer to *z/OS Security Server RACF Security Administrator's Guide*.

IPv6 advanced socket API options: You can control access for the IPv6 advanced socket API options that influence outbound packets.

For the IPV6_NEXTHOP, IPV6_TCLASS, IPV6_RTHDR, IPV6_HOPOPTS, IPV6_DSTOPTS, IPV6_RTHDRDSTOPTS, and IPV6_PKTINFO socket options, to set the socket option on setsockopt() or to use the corresponding ancillary data item on sendmsg(), an application must meet one of the following criteria:

- Be APF authorized.
- Have superuser authority.
- The corresponding SERVAUTH resource name in Table 12 is defined, and the application has at least READ access to the resource.

Table 12. Socket option resource names

Socket option/Ancillary data item	Resource name
IPV6_NEXTHOP	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_NEXTHOP
IPV6_TCLASS	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_TCLASS
IPV6_RTHDR	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDR
IPV6_HOPOPTS	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPOPTS
IPV6_DSTOPTS	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_DSTOPTS
IPV6_RTHDRDSTOPTS	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDRDSTOPTS
IPV6_PKTINFO	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_PKTINFO
IPV6_HOPLIMIT	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPLIMIT

For the IPV6_HOPLIMIT socket option, to set a hop limit greater than the default using either the IPV6_UNICAST_HOPS or IPV6_MULTICAST_HOPS socket option on setsockopt() or the IPV6_HOPLIMIT ancillary data item on sendmsg(), an application must meet one of the following criteria:

- Be APF authorized.
- Have superuser authority.
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPLIMIT is defined, and the application has at least READ access to this resource.

TCP/IP applications that set IPv6 advanced socket API options: Table 13 lists the TCP/IP applications that set IPv6 advanced socket API options. In a multilevel secure environment, APF and superuser authority are not sufficient; the security product resource name for the socket options appearing in Table 13 must be defined for every stack. Usage must be explicitly permitted by user ID, or conditional access lists, such as PERMIT WHEN(PROGRAM(...)) must be defined for the load module and alias names listed in the table. For more information about applications in a multilevel secure environment, see "Required configuration in a multilevel secure environment" on page 149.

Table 13. TCP/IP applications that set IPv6 advanced socket API options

Load module	Alias	Socket options set
EZACDOPN	OPING	IPV6_PKTINFO
EZACDTPN	PING	IPV6_PKTINFO
EZACDTRT	OTRACERT	IPV6_HOPLIMIT, IPV6_PKTINFO
EZACDTTR	TRACERTE	IPV6_HOPLIMIT, IPV6_PKTINFO
EZADNSVR	NAMED	IPV6_PKTINFO
EZAORRTE	OMPROUTE	IPV6_HOPLIMIT, IPV6_PKTINFO
EZASNTPD	SNTPD	IPV6_PKTINFO

For more details on these IPv6 advanced socket API options, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

For examples of the security product definitions, refer to the EZARACF sample in data set SEZAINST.

Netstat access control

Netstat access control allows control of access to Netstat command output from the TSO or UNIX System Services shell environments using an SAF security server. The Netstat command output is considered the resource to be protected and is represented by the resource name *EZB.NETSTAT.sysname.tcpname.netstatoption*. Access to the Netstat output is allowed if the user is permitted to the security profile in the SERVAUTH class covering this resource or if the security server indicates there is no profile covering this resource. There are no new TCP definitions required.

Guideline: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. If your product does not make this distinction, you must define the netstat resource profiles and permit users to them whenever the SERVAUTH class is active.

An installation can implement a security policy that indicates which users have authorization to selected Netstat options. The level of granularity for this security policy can be either by individual or all Netstat options.

Fast Response Cache Accelerator access control

Fast Response Cache Accelerator access control allows control of application access to Fast Response Cache Accelerator (FRCA) services. For more information on FRCA, see "Considerations for Fast Response Cache Accelerator" on page 85.

FRCA services are represented with a resource profile in the SERVAUTH class named EZB.FRCAACCESS.*sysname.tcpname*. Access to FRCA services is allowed if the Web server user is permitted to this resource, or if the resource is not defined. There are no new TCP definitions required. This function is enabled if the SERVAUTH class is active and the FRCA resource profile is defined. If it is not defined, the check is not made.

Guideline: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. If your product does not make this distinction, you must define the FRCA resource profile and permit users to it whenever the SERVAUTH class is active.

TCP/IP stack initialization access control

TCP/IP stack initialization access control enables control of application access to a TCP/IP stack before required policies have been installed. Access checking is performed only if Application Transparent Transport Layer Security (AT-TLS) has been configured in the initial PROFILE.TCPIP. If AT-TLS is active, the resource name EZB.INITSTACK.*sysname.tcpname* must be defined to allow access to the stack. Access checking is done during the period of time after the stack is active and before AT-TLS has been initiated from policy. During this period, access is restricted based on the RACF profile. If no RACF profile is specified, all socket requests are blocked. If a RACF profile is specified, access to the stack is prohibited, except by user IDs permitted to the resource. Checking ceases the first time that the Policy Agent completes the processing of the AT-TLS policy, or when a profile change deactivates AT-TLS.

Guideline: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. However, if AT-TLS is enabled, a profile that is not defined has the same result as a user that is not permitted, regardless of the security product you are using. If AT-TLS is enabled, you must activate the SERVAUTH class, and define the INITSTACK resource profile and permit users to it.

TCP/IP packet trace service access control

The TCP/IP packet trace service provides an interface for network management applications to obtain packet trace data. The information provided through the service is considered the resource to be protected. Access to this information can be controlled through an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPDA.

Access to the information is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR PKTTRCService statement in PROFILE.TCPIP. For details, refer to *z/OS Communications Server: IP Configuration Reference*.

If the resource profile is not defined, the service allows access to the packet trace information only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

TCP connection information service access control

The TCP connection information service allows network management applications to obtain information about TCP connection activity. Access to this information can be controlled by an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPCN.

Access to the TCP connection information is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR TCPCONNService statement in PROFILE.TCPIP. For details, refer to *z/OS Communications Server: IP Configuration Reference*.

If the resource profile is not defined, the service allows access to the TCP connection information only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

Real-time SMF information service access control

The SMF information service allows network management applications to obtain selected TCP/IP SMF records, such as SMF records supported by FTP and Telnet, in a real-time fashion. Access to this information can be controlled through an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPSM.

Access to these SMF records is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR SMFService statement in PROFILE.TCPIP. For details, refer to *z/OS Communications Server: IP Configuration Reference*.

If the resource profile is not defined, the service allows access to the SMF data only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

CIM provider access control

CIM provider access control permits the Communications Server CIM providers to gather CIM data, when the user ID associated with the client of the z/OS CIM server is not defined as a superuser. For more information on the CIM providers, see "Considerations for Common Information Model (CIM) providers" on page 102.

Access can be controlled by an external security manager product, such as RACF, by defining the resource profile name EZB.CIMPROV.*sysname*.*tcpname* in the SERVAUTH class. For examples of the security product commands needed to create this resource profile name and grant users access to it, see member EZARACF in sample data set SEZAINST.

Access is granted if the user ID associated with the client of the z/OS CIM server is permitted (has read access) to this resource profile.

Tip: Some security products do not distinguish between a resource profile not defined and a user not permitted to that resource. If your product does not make this distinction, you must define the CIM provider resource profile and permit the client user ID to it whenever the SERVAUTH class is active, if you want the Communications Server CIM providers to be able to gather CIM data.

Syslogd isolation

Syslogd isolation provides a capability for the installation to control which user IDs and job names can write syslogd records to specified syslogd facilities. This function enables the installation to segregate system and application syslogd records, and to segregate syslogd records from different applications. This function prevents an application level process from flooding a syslogd facility intended for system use, possibly causing system syslogd records to be lost. This function is enabled when user ID and/or job name are specified as additional criteria along with existing facility and priority criteria to select a syslogd repository.

In addition, the user ID and job name associated with the syslogd record writer can optionally be stored in a syslogd record based on a syslogd command-line parameter. This capability is useful when syslogd records for multiple jobs or users are recording in the same syslogd facility. This function enables positive identification of the creator of the syslogd records and ensures that the syslogd record, if spoofed, can be identified.

Syslogd isolation also provides a capability to disable reception of syslogd messages from other hosts in the network. This capability is provided by a syslogd command-line parameter. This parameter disables reception of syslogd messages from all hosts. If an installation wants to allow certain hosts in the network access to syslogd, IP Filtering can be used instead to specify which hosts are permitted to access the syslogd UDP port.

IP filtering

The IP security function can configure the Communications Server to perform packet filtering at the IP layer. Integrated Security Services Firewall Technologies can also provide similar functionality.

IP filters are rules defined to either discard or permit packets. IP filtering matches a filter rule to data traffic based on any combination of IP source or destination address (or masked address), protocol, source or destination port, direction of flow, or time. IP filtering can control traffic being routed, or control access at the host that has the communication endpoint. Even when an external firewall is providing filtering protection for the host, Communications Server IP filtering can provide a secondary line of defense.

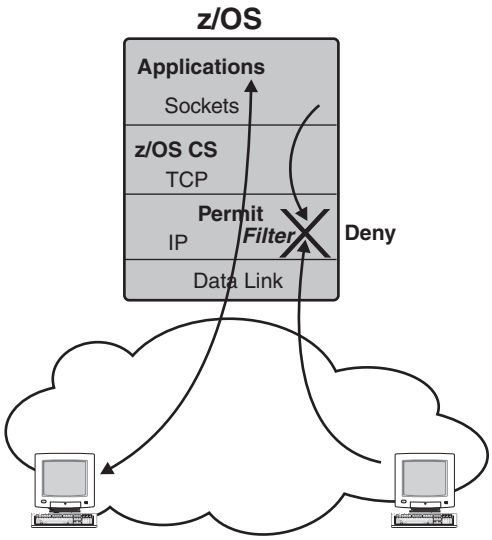


Figure 24. IP filtering at the z/OS communication endpoint

For more information about IP filtering, see Chapter 17, "IP security," on page 819, or refer to *z/OS Integrated Security Services Firewall Technologies*.

Security considerations for the VARY command

You can restrict access to the VARY TCPIP command by defining RACF profiles under the OPERCMDS class and specifying the list of users that are authorized to issue the VARY TCPIP command. You can decide on the level of control that is appropriate for your installation. For example, you might want to allow a user to be able to start or stop a TCP/IP device using the VARY TCPIP command, but you do not want the user to be able to modify the TCP/IP configuration. For further information on restricting access to the VARY TCPIP command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Multilevel security

Multilevel security is an enhanced security environment that can be configured on a z/OS system. In this environment, the security server and trusted resource managers enforce mandatory access control policies in addition to the usual discretionary access control policies. To participate in a multilevel-secure environment, the user IDs associated with tasks trying to access z/OS CS resources and those resource profiles in the SERVAUTH class need to have security labels defined. For more information on the multilevel-secure environment and configuring z/OS CS in that environment, see Chapter 4, "Preparing for TCP/IP networking in a multilevel secure environment," on page 145.

Protecting data in the network

Network security principles

Cryptography: The foundation of good security

The foundation of good security methods begins with cryptography. Cryptography keeps your data and communications secure using techniques such as encryption, authentication, and data integrity. Encryption services protect sensitive data from being read by other than the intended receiver. Cryptographic authentication and data integrity services allow communicating hosts to detect if data is altered in transit. Public key cryptography can identify and authenticate hosts or users.

Public key cryptography can also be used in the secure creation of symmetric session keys for both security endpoints. Once a secure session is created, successful data authentication and decryption occur only if both hosts have the correct session keys.

End to end security

Cryptographic security solutions can be applied to a portion of the data path or end to end, whichever is appropriate for your security policy. Generally, the greatest degree of security is provided when cryptographic methods are used end to end. However, if only portions of the data path are considered untrusted by an enterprise (such as the Internet) it may be adequate to protect only the untrusted portion with cryptography. z/OS offers security protocols that can be configured to protect portions of the data path or the entire data path.

Workload-based security deployment

In making a security protocol selection, an important consideration is the application workload to be protected. In order to illustrate this concept, it is helpful to understand where various protocols are implemented from a protocol layering perspective.

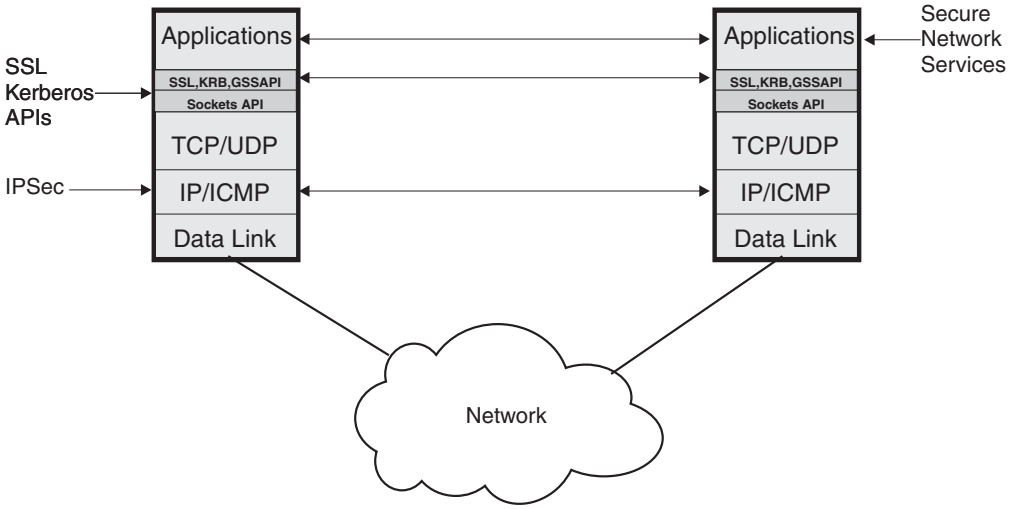


Figure 25. Security protocols from a protocol layering perspective

Existing workload: The network layer is the lowest layer in the protocol stack where end to end security over multiple hops can be applied. Network layer security protocols provide blanket protection for upper-layer application data without requiring modification to the application. IPSec is implemented at the network layer and provides authentication, integrity, and data privacy between any two IP entities. IPSec can protect a segment of the data path (e.g., between two routers), or it can secure the data path end to end. Because IPSec is applied at the IP layer, it is a connectionless security protocol and is applied on a per packet basis.

Secure Sockets Layer (SSL) is another popular security protocol implemented above the transport layer at the application interface layer. TCP applications can be modified to use SSL. Many existing socket applications might be able to use Application Transparent Transport Layer Security (AT-TLS) services provided within the TCP/IP TCP layer. For details, see Chapter 18, "Application Transparent Transport Layer Security (AT-TLS) data protection," on page 987.

SSL requires a reliable transport layer and is therefore not used for UDP applications. SSL provides authentication, integrity, and data privacy. SSL, originally used to secure traffic between a Web browser and Web server, can also secure other applications. SSL is a connection-oriented security protocol and protects all data on a connection or session.

The Communications Server has an SSL-enabled TN3270 server, thus allowing secure access to existing SNA applications being accessed over an IP network. Serving as a protocol gateway between the IP network and the SNA network, the SSL-enabled TN3270 server protects the data path in the IP network from the TN3270 client all the way to the z/OS TN3270 server. If the TN3270 Server resides on a different host from the target SNA application, SNA Session Level Encryption can be used to secure the SNA portion of the data path. SNA application data can be protected without modification to the SNA applications.

New workload: For new applications, security can be built-in. One method of building security into the application on z/OS is to use z/OS System SSL and Kerberos. Another method is to use the Communication Server's policy-driven Application Transparent Transport Layer Security (AT-TLS), which requires no change to the application unless the application must control certain portions of the AT-TLS support.

Newer versions of network services such as SNMPv3, Secure DNS, and z/OS UNIX sendmail, which are supported by the Communications Server, have security built into the application protocol using standards-based specifications for secure interoperability.

Network security protocols

IPSec and VPNs

IPSec is defined by the IPSec Working Group of the IETF. It provides authentication, integrity, and data privacy between any two IP entities. Management of cryptographic keys and security associations can be done manually or dynamically using an IETF-defined key management protocol called Internet Key Exchange (IKE).

With IPSec, you can create virtual private networks (VPN). A VPN enables an enterprise to extend its private network across a public network, such as the Internet, through a secure tunnel called a security association. IPSec VPNs enable the secure transfer of data over the public Internet for same-business and business-to-business communications, and protect sensitive data within the enterprise's internal network.

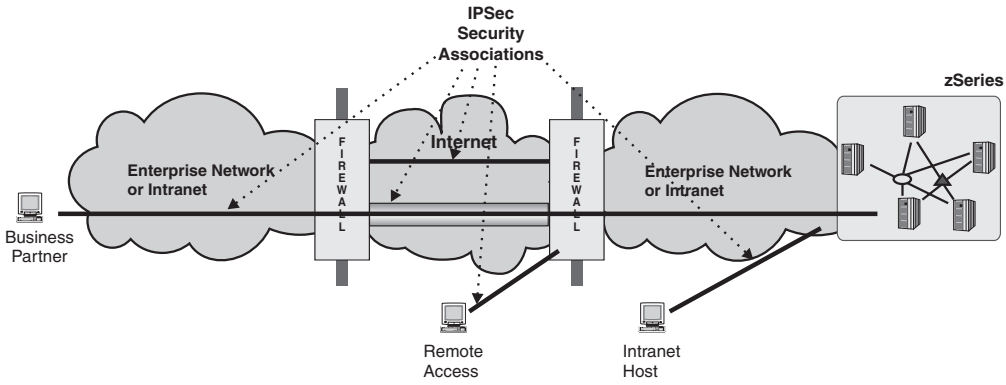


Figure 26. e-business scenarios with virtual private networks

z/OS provides support for IKE and IPSec VPNs, including the following:

- AH and ESP protocols
- Triple DES for strong encryption
- IPSec transport and tunnel mode encapsulation
- Aggressive and main mode negotiations
- Pre-shared key and RSA signature methods of authentication
- NAT traversal

IBM CP Assist for Cryptographic Functions, where available, or a cryptographic coprocessor, provide support for IPSec encryption and decryption. Where hardware support for encryption is not available, encryption and decryption are performed through software.

For more information about configuring IPSec and VPNs, see Chapter 17, "IP security," on page 819.

For more information on using IPSec with Dynamic VIPAs, see "Sysplex-wide security associations" on page 328.

SSL and TLS

The SSL protocol provides data encryption, data origin authentication, and message integrity. It also provides server and client authentication using X.509 certificates. SSL begins with a handshake during which the server is authenticated to the client using X.509 certificates. Also, the client can optionally be authenticated to the server. During the handshake, security session parameters, such as cryptographic algorithms, are negotiated and session keys are created. After the handshake, the data is protected during transmission with data origin authentication and optional encryption using the session keys.

The cryptographic algorithms that are used for the SSL session are based on the algorithms the server and client are willing to use. During the SSL handshake, the client and server exchange a list of algorithms. The algorithm selected is based on the best match between the client's list and the server's list. The selectable algorithms can be limited by configuring a subset of allowable algorithms at the server. Servers can support encryption using Triple DES as well as other encryption algorithms (RC2, RC4, and DES). A hardware crypto coprocessor, if available, is used for DES and Triple DES encryption.

SSL requires a server X.509 certificate, which is stored in its certificate key ring. The certificate is used as part of the SSL handshake server authentication process.

The client validates the server certificate. SSL optionally uses a client X.509 certificate that is used as part of the SSL handshake client authentication process. In order to use client authentication, the client must have a client X.509 certificate. Successful client authentication requires that the Certificate Authority (CA) that signed the client certificate be considered trusted by the server. To be considered trusted, the certificate of the CA must be in the key ring of the server.

Refer to "Transport layer security" on page 461 for detailed information on obtaining certificates.

SSL is not defined by the IETF. TLS is based on SSL and is defined by the IETF as RFC 2246.

TN3270 SSL: The Communications Server provides an SSL-enabled TN3270 server that protects the data path in the IP network to the z/OS TN3270 server using the SSL protocol. IBM Host On Demand and PCOMM provides a TN3270 client that is enabled for SSL.

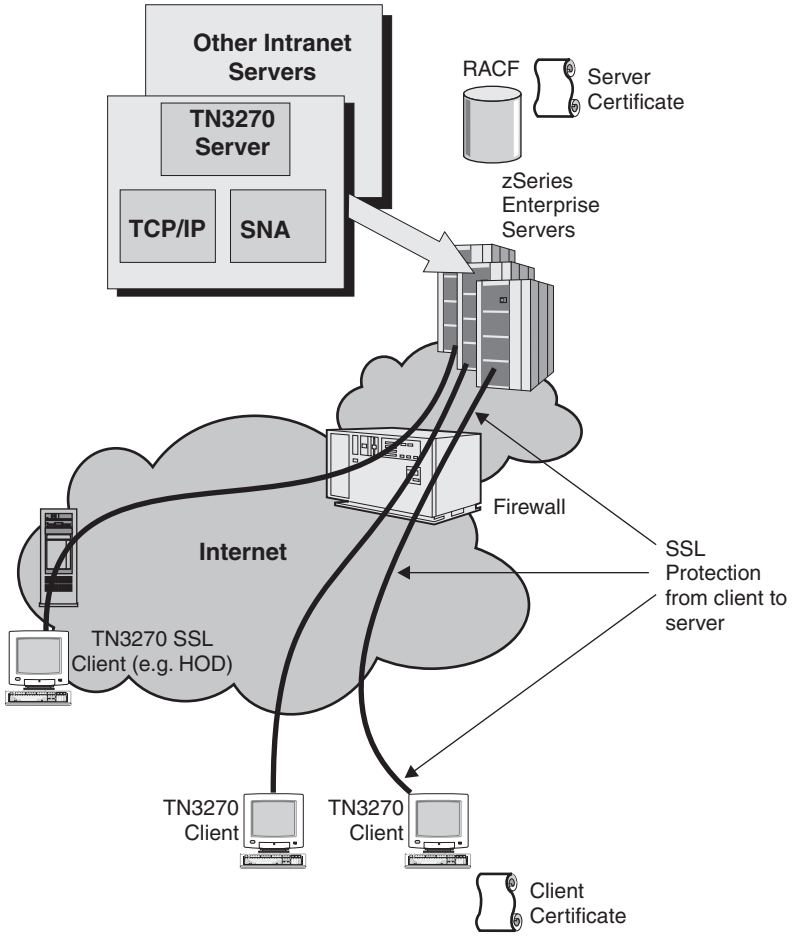


Figure 27. TN3270 SSL overview

The Communications Server TN3270 SSL support provides several extensions for RACF-based access control to the TN3270 server. These extensions prevent a client from seeing the USSMSG (log on screen) unless the client is authorized. In order to use this support, the client certificate must be defined to RACF using RACF digital certificate services. The first level of authorization checking verifies that the RACF userid represented by the client certificate is defined to RACF. The next level of

authorization requires that this RACF userid be permitted to access the TN3270 server port. In this case, the TN3270 server port is represented as a RACF resource using the SERVAUTH class.

Multiple port support: One method of enabling a mix of SSL and non-SSL traffic is to use TN3270 multiple port support. Using the multiple port support, separate ports can be defined with one port being dedicated to non-SSL traffic and another port dedicated to SSL traffic. Ports designated as SECUREPORT are capable of using SSL. The following diagram illustrates the use of multiple ports. In this case, intranet clients are not required to use SSL. These clients connect to the BASIC port (port 23 in this example). All clients connecting from the Internet are required to use SSL. These clients use the SECUREPORT (port 1023 in this example). Packet filtering is used at the firewall that separates the intranet and the Internet to control access to the TN3270 ports. In order to prevent Internet access to the BASIC port, port 23 is blocked at the firewall. The SECUREPORT, port 1023, is permitted at the firewall. In this scenario, the best security is achieved when SSL client authentication with the TN3270 RACF extensions is used. This support ensures that the client has authority to attempt to log on to SNA applications through TN3270. Regardless of the method of authentication used, the SNA application should identify and authenticate the end user using RACF before any application access is granted. SSL encryption services, if used, would encrypt the user ID and password.

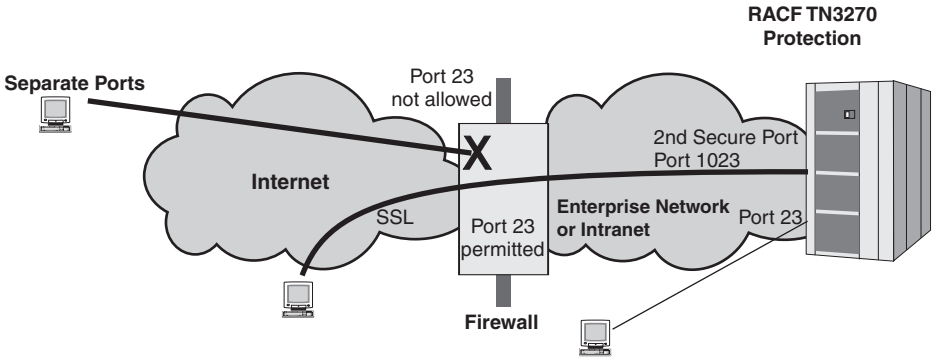


Figure 28. Using multiple TN3270 ports to separate SSL and non-SSL traffic

This next diagram illustrates how IPSec and SSL can be combined to provide a more secure remote access from the Internet to SNA applications than is depicted in the previous diagram. In this scenario, IPSec AH protocol is used between the user's PC and the firewall for authentication. The firewall is open for port 1023 for traffic that is authenticated with IPSec only. The firewall would discard traffic for port 1023 that cannot be authenticated by IPSec. The additional security provided by IPSec protects the zSeries from unauthorized access attempts and denial of service attacks by hosts outside the VPN.

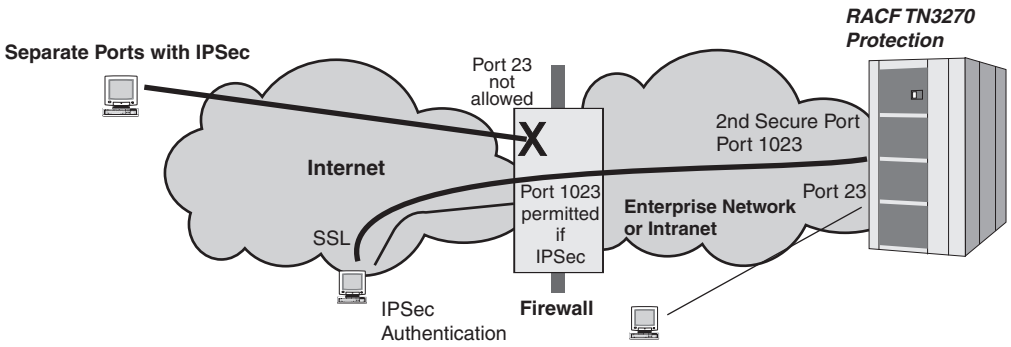


Figure 29. Combining TN3270 SSL with IPSec client-to-firewall authentication

TN3270 use of single port for SSL and non-SSL connections: A single port can be used to support a mix of SSL and non-SSL traffic. In this case the port is designated as SECUREPORT. In order to support the configuration of various SSL security policies for a single port, the SECUREPORT designation defines the port to be capable of using SSL, rather than the port must use SSL. The PARMSGROUP and BEGINVTAM blocks are used to specify the connection type (CONNTYPE) associated with a subset of the port's connections. A PARMSMAP statement is used to associate the PARMSGROUP information with specific IP address, hostname, or linkname. CONNTYPE specifies the SSL policy for the connections that are associated with it.

The TN3270 server supports both negotiated and non-negotiated SSL. TN3270 negotiated SSL is an IETF defined extension to the TN3270 protocol. With TN3270 negotiated SSL, the decision to use SSL for a connection is based on the outcome of a negotiation between the TN3270 client and server using TN3270 protocols. This negotiation is performed after the TN3270 connection is established, and if SSL is negotiated, the SSL handshake is performed. With non-negotiated SSL, an SSL handshake is required immediately after connection establishment. Concurrent use of both TN3270 negotiated and non-negotiated SSL connections are allowed for a single port.

The following diagram illustrates the use of a single TN3270 port that allows a mix of SSL and non-SSL traffic. In this case, intranet clients are not required to use SSL. All clients connecting from the Internet are required to use SSL. Both intranet and Internet clients connect to the SECUREPORT (port 23 in this example). In this scenario, IPSec AH protocol is used between the user's PC and the firewall for authentication. The firewall is open for port 23 for traffic that is authenticated with IPSec only. The firewall would discard traffic for port 23 that IPSec cannot authenticate. In this scenario, packet filtering without IPSec cannot be used at the firewall that separates the intranet and the Internet to control access on the basis of port since only one port is used. Without IPSec AH, all access control checks are deferred to the TN3270 Server. The additional security provided by IPSec at the firewall protects the zSeries from unauthorized access attempts and denial of service attacks by hosts outside the VPN.

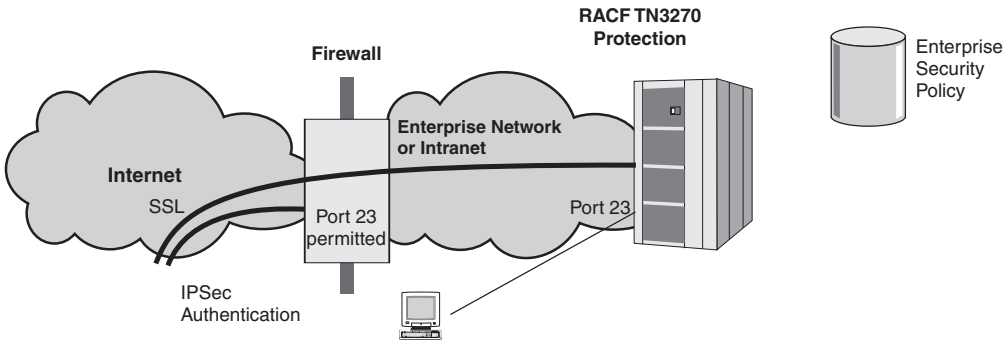


Figure 30. TN3270 SSL and non-SSL traffic using a single TN3270 port

Express Logon Feature (ELF): With emulator products, the traditional method of authenticating the user is through user ID and password which is kept in sync with the host access control facility (RACF, ACF/2, AS/400® user management, etc.). The Express Logon Feature simplifies user ID and password administration for users signing on to SNA applications using TN3270. ELF allows an end user to use an SSL-authenticated X.509 certificate for authentication to the SNA application instead of using a user ID and password. ELF requires IBM Host Integration software. The Host Integration requirements depend on the configuration.

There are two network designs available; a two-tier or a three-tier approach. Both are discussed in Appendix C, "Express Logon Feature (ELF)," on page 1195.

TLS-enabled FTP: The Communications Server FTP server and client support Transport Layer Security (TLS). This support enables secure file transfer by providing data privacy, message authentication, and message integrity services for data sent and received using the FTP control and data connections.

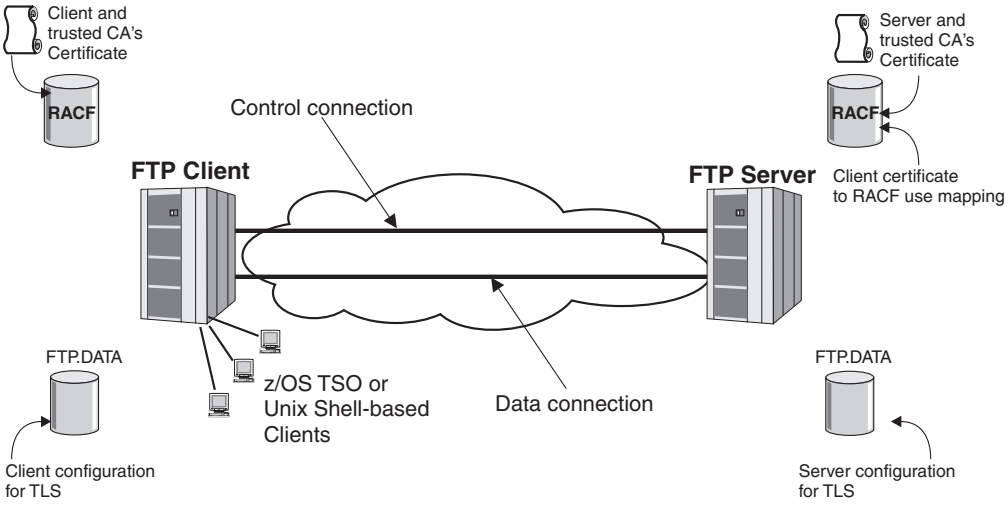


Figure 31. FTP client and server TLS overview

The TLS-enabled FTP server can be configured to run in two modes. Conditional mode allows an installation to use a single port for both TLS and non-TLS FTP control connections. In conditional mode, the FTP client and server negotiate the use of TLS based on a subset of the FTP security negotiation functions documented in RFC 2228. Once the use of TLS is negotiated, the TLS handshake is performed which establishes the TLS session and negotiates security parameters and session keys. Unconditional mode allows an installation to use a separate port for all TLS

traffic. Port 990 is the port designated for control connections for unconditional TLS mode. With unconditional mode, it is assumed that TLS is required, and after the FTP control connection is made, the TLS handshake is performed.

TLS secures the control connection and optionally the data connection. TLS for the data connection requires a TLS session for the control connection. FTP server configuration controls whether the FTP server requires TLS for the control and data connections. This TLS protection by connection type is negotiated during the FTP RFC 2228 negotiation that precedes the TLS handshake. During the lifetime of the control connection, the use of TLS or no TLS for the data connection can be requested by the FTP client using the FTP RFC 2228 commands.

FTP TLS optionally authenticates the client during the TLS handshake using a client X.509 certificate. FTP server configuration specifies whether TLS client authentication is required and what type of validation of the certificate is required. For example, the FTP server can be configured to map the client certificate to a RACF userid and then verify that the userid associated with the certificate matches the userid entered by the end user.

Configuration to control TLS capabilities and options for both FTP client and server TLS are stored in the FTP.DATA data set.

Application Transparent Transport Layer Security (AT-TLS)

Communications Server provides for invocation of System SSL in the TCP transport layer of the stack. Application Transparent Transport Layer Security (AT-TLS) support is controlled by the TTLS or NOTTLS parameter on the TCPCONFIG statement in the TCP/IP profile. When AT-TLS is enabled, AT-TLS statements in Policy Agent define the security attributes for connections that match AT-TLS rules. This policy-driven support can be deployed transparently underneath many existing sockets, leaving the application unaware of the encryption and decryption being done on its behalf. Support is also provided for applications that need to negotiate TLS or need to participate in client authentication. These applications must be aware of AT-TLS support and use ioctl support provided by AT-TLS. AT-TLS supports the TLS, SSLv3, and SSLv2 protocols. For more details, see Chapter 18, “Application Transparent Transport Layer Security (AT-TLS) data protection,” on page 987.

Kerberos

Kerberos is a network authentication protocol that is designed to provide strong authentication for client/server applications using secret-key cryptography. The Kerberos network authentication protocol assumes that services and workstations communicate over an insecure network. It allows clients and servers to do either one way, or two way (mutual) authentication. It allows for data encryption and prevents passwords from having to be retyped to access networked services and also prevents their transmission in plain text over the network. This feature can help reduce the need to manage multiple passwords.

z/OS Communications Server no longer ships Kerberos Version 4. z/OS Integrated Security Services ships Kerberos Version 5. Because Integrated Security Services Kerberos does not require DCE login and eliminates the need for multiple registries, it is recommended that new applications be written to Kerberos Version 5 and use z/OS Integrated Security Services.

The following Communications Server IP applications now include support for Kerberos Version 5 security protocol:

- The UNIX System Services Telnet Server now allows clients supporting Kerberos Version 5 (as described in RFC 1416) to log in to the shell environment using Kerberos as the authentication protocol.
- The FTP client and Server now support connections to or from other clients and servers supporting Kerberos Version 5 authentication for the FTP protocol (as described in RFC 2228).
- The UNIX System Services RSH server can now also be configured to support client authentication using Kerberos from RSH clients supporting Kerberos Version 5.

OSPF authentication

Communications Server OSPF (Open Shortest Path First) dynamic routing protocol supports message authentication and message integrity of OSPF routing messages through the use of the OSPF MD5 Authentication security protocol as defined by RFC 2328. OSPF MD5 Authentication ensures that an unauthorized IP resource cannot inject OSPF routing messages into the network without detection, thus ensuring the integrity of the routing tables in the OSPF routing network.

OMPROUTE computes a secure MAC for the routing message using the MD5 algorithm. This MAC is sent with the routing message so that the message can be authenticated by the receiver.

Secure DNS

The Communications Server supports DNS at the Version 9.1 of BIND. This level of DNS has built-in security features, DNSSEC and TSIG.

DNSSEC: DNSSEC ensures that DNS query results are not spoofed and in fact originate from a trusted DNS. DNSSEC defines extensions to DNS that provide data integrity and authentication to security aware resolvers and applications through the use of cryptographic digital signatures. DNSSEC is defined by the IETF in RFC 2535.

TSIG: TSIG is a protocol for Secret Key Transaction Signatures for DNS. This protocol allows for transaction level authentication using shared secrets and one way hashing. It authenticates dynamic updates as coming from an approved client, or responses as coming from an approved recursive name server.

SNMPv3

z/OS Communications Server SNMP supports SNMPv3. The legacy community-based protocols SNMPv1 and SNMPv2 are also supported. SNMPv3, defined in RFCs 2570 through 2575 is the standards-based solution for SNMP security. It is categorized as a User-based Security Model (USM) which provides different levels of security based on the user accessing the managed information. To support this security level, the SNMPv3 framework defines several security functions, such as USM for authentication and privacy, and view-based access control model (VACM) which provides the ability to limit access to different MIB objects on a per-user basis, and the use of authentication and data encryption for privacy. However, SNMP is not just enhanced security. It defines an architecture for SNMP management frameworks, with the intent that pieces of the architecture can advance over time without requiring the entire structure to be rewritten. For that reason, three major subsystems are defined:

- Message processing subsystem
- Security subsystem
- Access control subsystem

The framework is structured so that multiple models can be supported concurrently and replaced over time. For example, although there is a new message format for SNMPv3, messages created with the SNMPv1 and SNMPv2 formats can still be supported. Similarly, the user-based security model can be supported concurrently with the community-based security models previously used.

Security Event Reporting

Integrated Intrusion Detection Services (IDS)

Intrusion is a broad term encompassing many undesirable activities. The objective of an intrusion may be to acquire information that a person is not authorized to have (*information theft*). It may be to cause business harm by rendering a network, system or application unusable (*denial of service*). Or it may be to gain unauthorized use of a system as a stepping stone for further intrusions elsewhere. Most intrusions follow a pattern of information gathering, attempted access and then destructive attacks. Some attacks can be detected and neutralized by the target system. Other attacks cannot be effectively neutralized by the target system. Many of the attacks also make use of spoofed packets which are not easily traceable to their true origin. Many attacks now make use of unwitting accomplices - machines or networks that are used without authorization to hide the identity of the attacker. For these reasons, detecting information gathering, access attempts and attack accomplice behaviors is a vital part of intrusion detection.

Attacks can be initiated from outside the internal network or from inside the internal network. Particularly vulnerable is an open system such as a public Web server or any machine that is placed in service to serve those outside the internal network. A firewall can provide some level of protection against attacks from outside. However, it cannot prevent attacks once the firewall has authorized an external host to communicate with hosts in the internal network, nor can it provide protection in the case where the attack is initiated from inside the network. In addition, end to end encryption limits the types of attacks that can be detected by an intermediate device such as a firewall.

An Intrusion Detection System can provide detection of some types of attacks. Common intrusion detection system types currently deployed are network sniffers or sensors and vulnerability scanners. Sniffers, placed at strategic points in the network (in front or behind a firewall, in the network, or in front of a host), operate in promiscuous mode, examining traffic real-time that passes through on the local network. Sniffers use *pattern matching* to try to match a packet against a known attack which is expressed as an *attack signature*. Sniffers work best against single packet attacks. Limitations are that they cannot deflect the attacking packet, and they cannot evaluate against encrypted data. Scanners do not detect intrusions in real-time. They examine a system periodically looking for vulnerabilities or evidence of intrusion. Some scanners evaluate historical data and can identify behavioral anomalies and patterns associated with intrusions.

The z/OS Communications Server provides Intrusion Detection Services (IDS) which enable the detection of attacks and the application of defensive mechanisms on the z/OS server. The focus of IDS is self-protection. IDS can be used alone or in combination with an external network-based Intrusion Detection System. The IDS is integrated into the z/OS Communications Server stack and can provide the following functions unavailable from an external Intrusion Detection System.

- z/OS CS IDS evaluates data that has been encrypted by IPSec end to end after decryption on the target server system.

- z/OS CS IDS avoids the overhead of per packet examination against a table of signatures for many known attacks. This is accomplished by integrating the attack detection probes into existing error detection logic. This detection is done in real-time. IDS policy is examined when an attack is detected to determine the action to be taken.
- z/OS CS IDS detects statistical anomalies real-time. Real-time detection is achieved since it is easier for the target system to keep stateful data/internal thresholds and counters.
- z/OS CS IDS implements prevention type of policies that are executed on the system that is the target of the attack. Prevention policies include packet discard and connection limiting.

The IDS is policy driven and the policies are kept in LDAP. These policies determine what actions to take for various IDS events. IDS events detected include scans, single packet attacks against the TCP/IP stack, and flooding. Actions include packet discard, connection limiting, and reporting. IDS events can be recorded in syslog files and/or the console. IDS statistics can be recorded in syslog. Packet traces can be taken to document suspicious activities. The TRMDSTAT command provides summary and detailed reporting of IDS events and statistics.

The following figure shows the z/OS Communications Server IDS architecture.

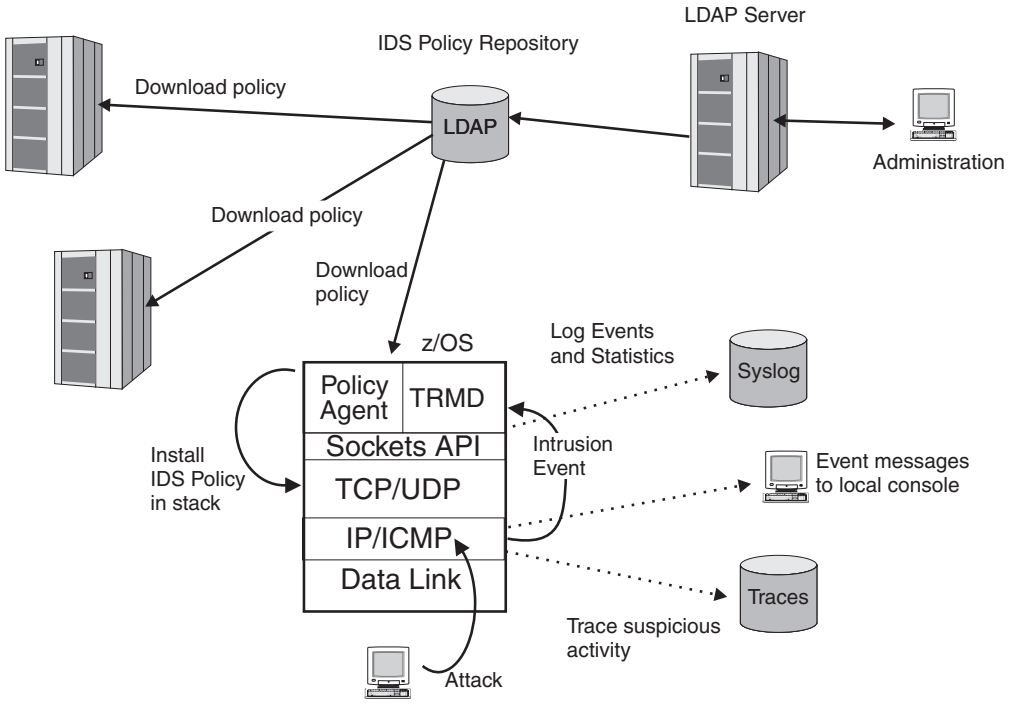


Figure 32. Intrusion Detection Services overview

For more information on IDS, see Chapter 16, "Intrusion Detection Services (IDS)," on page 789.

Chapter 4. Preparing for TCP/IP networking in a multilevel secure environment

Use this chapter to configure a z/OS Communications Server stack and applications in a multilevel secure environment.

Tip: Many of the tasks, examples, and references in this chapter assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

Understanding multilevel security concepts

IBM's multilevel security for z/OS has access control implications for your entire system. z/OS Communications Server TCP/IP is only one element of a multilevel secure z/OS system environment. Before planning your multilevel secure TCP/IP network, you should be familiar with the concepts and terminology presented in *z/OS Planning for Multilevel Security and the Common Criteria*.

Multilevel secure networking

The security administrator is responsible for defining the security levels and categories required in a multilevel secure environment. These become part of the set of security labels used to enforce mandatory access control policies when applications access resources on behalf of users. All of the systems enforcing mandatory access control policies in a multilevel secure network must have equivalent definitions of these security labels and the systems in the network to which they apply.

In the networking environment, the information that is being protected is the data being read and written through sockets. Sockets are opened and used by applications running under user IDs. In a z/OS multilevel secure environment:

- Each user ID is permitted to use one or more security labels.
- Every job or login session is associated with a user ID.
- A user ID can use only one security label for each job or login session.
- The security label used is limited by the port of entry (source type and location) of the job or login session.

Applications can have read access to information from many sources that can have various security labels. This information might be commingled in the buffers used to write information to the network. TCP/IP treats all socket data buffers as having the security label of the writing task. All sockets are inherently read/write, so TCP/IP requires communicating partners to have equivalent security labels in a multilevel secure environment.

Nonsecure systems

Most systems do not support mandatory access control processing. If these systems are not physically managed, they normally should not participate in a multilevel secure network. Some installations might need to permit them to participate. In

these cases, it is recommended that they be assigned a security label with the lowest security level and a single security category that is not common with any other security label.

Managed systems

Systems that do not support mandatory access control processing can participate in a multilevel secure network, if they are physically managed to guarantee that all information on the system has the same single security label and all users of the system are permitted to that security label. These systems are referred to as single-level security or managed systems in this document. This management requires both physical control of the systems and careful management of the network. Managed systems must be prevented from communicating with other managed systems that do not have equivalent security labels.

Systems that support mandatory access control and are configured to implicitly associate the correct security label with each managed system can also communicate with managed systems. The systems that perform mandatory access control are responsible for ensuring that only information from applications with an equivalent security label is sent to a managed system, and that information received from a managed system is given only to applications with an equivalent security label.

Multilevel secure systems

Some systems in the network provide multilevel secure environments. These systems have mechanisms to associate security labels with information accessed through the system and with users logged into the system. The system enforces mandatory access control policies to ensure proper separation of information.

Other systems cannot normally associate a single security label with IP addresses owned by a multilevel secure system. The packets being sent from a single IP address on the multilevel secure system might have originated from applications running under different security labels.

Applications on a multilevel secure system can securely communicate with applications on a managed system. Mandatory access control enforcement occurs only on the multilevel secure system. The multilevel secure system is responsible for ensuring that it sends only information from an application with an equivalent security label to any managed system. It also is responsible for ensuring that information received from a managed system is delivered only to an application with an equivalent security label.

When two applications on multilevel secure systems communicate, the security label of the sending application must be communicated to the receiving system so that the receiving system can enforce mandatory access control prior to delivering the information to an application. One mechanism to accomplish this is for the sending system to pass the security label with each packet. The two multilevel secure systems must share a common definition and representation of the security labels that are passed.

z/OS Communications Server TCP/IP stacks on z/OS multilevel secure systems

A z/OS CS TCP/IP stack running in a z/OS multilevel secure environment can optionally be configured as either a restricted stack or an unrestricted stack. A restricted stack is configured with a user ID that is defined with a security label

other than SYSMULTI. An unrestricted stack is configured with a user ID that is defined with a security label of SYSMULTI. A single z/OS system can concurrently run up to eight z/OS CS TCP/IP stacks, which can be any mix of restricted and unrestricted stacks.

In either mode of operation, appropriate mandatory access control processing is performed at the transport layer. z/OS Communications Server stacks can be host systems on trusted subnetworks. z/OS Communications Server stacks do not perform mandatory access control processing at the link or network layers, so security labels are not considered in packet forwarding with the exception of Sysplex Distributor, as discussed later. It is important to note that packets containing security labels are not forwarded by a restricted stack. These packets are discarded by the restricted stack.

Restricted stacks

In this mode of operation, the stack ensures that all sockets are opened by applications running with a security label that is equivalent to the security label of that stack. This guarantees that all information sent by the stack can be implicitly associated with that stack's security label. The stack also ensures that all information received from the network and delivered to an application is equivalent to the stack's security label. A restricted stack can be viewed by other multilevel secure systems as if it were a managed system.

It is important to note that even though a restricted stack is running under a specific security label, it still receives packets from the network with information covered by different security labels. A restricted stack discards this information rather than deliver it to any local applications. However, this information can appear in storage dumps, logic traces, and packet traces. Diagnostic information should always be protected under the highest security label (SYSHIGH).

Unrestricted stacks

In this mode of operation, the stack allows sockets to be opened by applications with any security label. The stack supports mandatory access control processing that allows its applications to communicate securely with any other managed system or restricted stack.

For applications on unrestricted stacks to communicate securely with each other, Communications Server must be able to determine the security label of the sending application. Unrestricted stacks are permitted to define VIPAs in network security zones with security labels other than SYSMULTI. When one of these is used as either the source or destination IP address of a packet, it implicitly identifies the security label of the information. When both IP addresses in a packet are in security zones with the SYSMULTI security label, the application security label must be explicitly transmitted in the packet. This is known as packet tagging.

Communications Server implements a proprietary form of packet tagging to pass the security label of the sending application to the receiving stack for mandatory access control enforcement against the receiving application. Because of this proprietary format, communications between applications on unrestricted stacks that require packet tagging are supported only when both of those stacks are on the same z/OS system and are communicating over an IUTSAMEHOST link, or are members of the same z/OS sysplex and are communicating over an XCF link.

Stack recognition of a multilevel secure environment

You can activate the SAF SECLABEL class and define security labels on SERVAUTH profiles. This causes the security server to enforce mandatory access

control policies for those resources without fully activating a multilevel secure environment. The z/OS Communications Server stack does not perform its extra mandatory access control policy enforcement until you issue the RACF command SETROPTS MLACTIVE.

When a NetAccess statement is encountered in TCPIP profile processing and MLACTIVE has been set, the stack activates extra mandatory access control policy enforcement in both restricted and unrestricted stacks as follows:

- New sockets are allowed only if a STACKACCESS profile covers this stack.
- Network access is allowed only to IP addresses that are mapped into network security zones covered by NETACCESS profiles.
- Restricted stacks do not normally allow SYSMULTI tasks to have network access to security zones with security labels that are not equivalent to the stack's security label. For more information, see "Exempting certain users of certain programs from full Network Access Control" on page 152.
- Unrestricted stacks transmit packet labels both internally and externally to enable an extra mandatory access control check, between the sending task's security label and the receiving task's security label, when both IP addresses are in security zones with a SYSMULTI security label.
- Distributing stacks consider security labels in choosing target applications.
- In-stack TN3270E servers consider security labels in mapping connections to LU names.
- Internal configuration consistency checks are performed whenever PROFILE.TCPIP or certain SERVAUTH class profile changes are made.

Common INET in a multilevel secure environment

When you start several TCP/IP stacks under OMVS, you are using the Common INET (CINET) PFS. Users and jobs can optionally establish affinity to a single stack, or they can allow CINET to choose a stack. If stack affinity is not set, CINET replicates the socket() command to all stacks attached to it. If a job or user does not have READ access to any of the attached stacks, RACF might generate audit failure messages for those stacks. As long as at least one stack accepts the socket() command, CINET will return success to the application. CINET then routes subsequent commands on that socket to one or more of the stacks that accepted the socket() call. For further information on CINET, refer to *z/OS UNIX System Services Planning*.

Network security zones

A network security zone is an administrative name for a collection of systems that require the same access control policy. IP addresses are used to map systems into security zones. This requires that the IP addresses used in your multilevel secure network be predictably associated with a single system or group of systems with the same access control policy. A network security zone can contain a single IP address or any combination of IP addresses and subnetworks. All of the IP addresses in a security zone must have the same security label, though all IP addresses with the same security label do not have to be in the same security zone.

Where your z/OS systems fit in your network

z/OS systems that are not configured with RACF SETROPTS MLACTIVE must be physically managed, as any other managed system.

z/OS systems at V1R5 or later that are configured with RACF SETROPTS MLACTIVE, have appropriate TCP/IP configuration, and have appropriate RACF

SERVAUTH class profiles defined, can be placed in trusted subnetworks. Firewalls can allow any managed subnetworks to communicate with these trusted subnetworks. The trusted subnetworks will often be defined as a SYSHIGH security zone and will likely contain several individual IP addresses in other security zones, depending on the mix of restricted and unrestricted stacks within the trusted subnetwork.

Planning stacks on your z/OS systems

Before you begin, determine what z/OS systems need to participate in your multilevel secure network. For each z/OS system, determine what release level it will be running, what sysplex it is a member of, which other multilevel secure systems it needs to communicate with, and what security labels will be used.

Further information on the following subsections is located elsewhere in this book, as well as in *z/OS Communications Server: IP Configuration Reference* and *z/OS Security Server RACF Security Administrator's Guide*.

Required configuration in a multilevel secure environment

Some configuration statements that are optional in a discretionary security environment are required in a multilevel secure environment. The default behavior of the stack in a discretionary security environment is to permit most applications when these statements are not defined. The default behavior of the stack in a multilevel secure environment is to fail every application when these statements are not defined. Every stack must have an EZB.STACKACCESS profile in the SAF SERVAUTH class. All referenced IP addresses must be mapped into security zones by NetAccess statements in the TCPIP profile. The EZB.SOCKOPT profile must be defined for the following options in the SAF SERVAUTH class:

- IPV6_DSTOPTS
- IPV6_HOPOPTS
- IPV6_NEXTHOP
- IPV6_PKTINFO
- IPV6_RTHDR
- IPV6_RTHDRDSTOPTS
- IPV6_TCLASS
- SO_BROADCAST

In addition, if setting a hop limit greater than the default hop limit, the EZB.SOCKOPT profile must also be defined for the following options in the SAF SERVAUTH class:

- IPV6_HOPLIMIT
- IPV6_MULTICAST_HOPS
- IPV6_UNICAST_HOPS

For more information, see "TCP/IP resource protection" on page 119.

Considerations for IPv6-enabled stacks

IPv6 architecture provides for a plug-and-play environment by automatically generating IP addresses when they are not manually configured. To guarantee predictable IP address association with specific systems, you must disable this capability on z/OS Communications Server stacks through manual configuration.

The stack automatically generates link local IP addresses using the link local prefix and the interface ID. To make these addresses predictable, you must manually configure the INTFID parameter on all IPv6 INTERFACE statements. If you are enabling dynamic XCF, you must also manually configure the DYNAMICXCF INTFID parameter on the IPCONFIG6 statement.

Some IPv6 interface types support router autogeneration. When this is enabled, the responsible router informs the stack about prefixes that are supported and the stack generates IP addresses from those prefixes and the interface ID. You must disable this capability on all interfaces that support it, by manually configuring at least one prefix or address using the IPADDR parameter on the INTERFACE statement.

Deciding whether to use restricted or unrestricted stacks

Many installations will find it easier to run a single unrestricted stack on each z/OS system. The following situations might require you to run one or more restricted stacks on a given system:

- Your installation has administrative requirements that prohibit the use of stacks that allow sockets with different security labels.
- You must allow applications to communicate with applications on a multilevel secure system that is not a z/OS system.
- You must allow applications to communicate with applications on a z/OS multilevel secure system that is not z/OS V1R5 or later.
- Your stacks are not on the same system and are not members of the same sysplex.

Configuring a restricted stack

To configure a restricted stack, do the following:

1. Determine the appropriate security label for this restricted stack. Associate a user ID with the stack job, and permit the user ID to the intended stack security label. Make that security label the default security label in the user ID profile.
2. Define a STACKACCESS profile for this stack in the SERVAUTH class with the same security label. This profile might often be UACC(READ).
3. Determine the interface and VIPA addresses needed for this stack.
4. Define one or more security zone names for this stack.

   If you have discretionary access control policies that require different treatment for some of the IP addresses on this stack, they will need to be in different security zones.
5. Define NETACCESS profiles for this stack in the SERVAUTH class with the same security label.

   If your access control policy requires only mandatory access control enforcement, this profile can be generic with respect to the z/OS system name and the TCP stack job name. It might often be UACC(READ).
6. Define a NETACCESS statement that maps this system's external IP addresses into security zone names. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.
7. If you define a SOURCEVIPA address for this stack, it must be an IP address in a security zone covered by a NETACCESS profile with the stack's security label.

Configuring an unrestricted stack

To configure an unrestricted stack, do the following:

1. Unrestricted stacks must run with the SYSMULTI security label. Associate a user ID with the stack job, and permit the user ID to the SYSMULTI security label. Make SYSMULTI the default security label in the user ID profile.
2. Define a STACKACCESS profile for this stack in the SERVAUTH class with the SYSMULTI security label. This profile might often be UACC(READ).
3. Determine the interface and VIPA addresses needed for this stack.

   If you are running server applications that require multiple instances running under different security labels, you will need at least one VIPA for each security label.
4. Define security zone names for this stack.

   If you have discretionary access control policies that require different treatment for some of the IP addresses on this stack, they will need to be in different security zones.

   VIPAs defined for use with specific security labels will need to be in separate security zones.
5. Define NETACCESS profiles for this stack in the SERVAUTH class.

   Profiles created for VIPAs used with specific security labels are defined with the appropriate security label. If your access control policy requires only mandatory access control enforcement, these profiles can be generic with respect to the z/OS system name and TCP stack job name. They might often be UACC(READ).

   Profiles for other zones on this stack are defined with the SYSMULTI security label. If there are z/OS systems sharing the RACF database that are not members of the same sysplex, or that are not z/OS V1R5 or later, any generic profile should have UACC(NONE). A fully qualified profile might often be UACC(READ).
6. Define a NETACCESS statement that maps this system's IP addresses into security zone names. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.

   Tip: If you have multiple unrestricted stacks and you use the OMPROUTE routing daemon, for more information on local address security zones, see "OMPROUTE" on page 158.
7. If you enable SOURCEVIPA on IPCONFIG and IPCONFIG6 statements, in many circumstances you might be able to avoid explicit packet tagging. Place IPv4 source VIPAs in the HOME list preceding the physical interface IP addresses that can use them. VIPAs in SYSMULTI security zones should precede those in other zones. Place IPv6 VIPAs on VIRTUAL6 INTERFACE statements. Use the SOURCEVIPAINTERFACE parameter on other INTERFACE statements to associate the set of source VIPAs that can be used with each IPv6 interface.
8. If you define TCPSTACKSOURCEVIPA for this stack, to avoid application failures, it must be in a security zone with a SYSMULTI security label.

Configuring global definitions for all stacks

To configure global definitions for all stacks, do the following:

- Define a security zone name for the INADDRANY and LOOPBACK addresses.

Define a NETACCESS profile for this zone in the SERVAUTH class for each stack. This profile should be specific with respect to the z/OS system name and TCP stack job name, and should have the same security label as the stack job. It might often be UACC(READ).

Define a NETACCESS statement that maps any system's INADDRANY and LOOPBACK IP addresses into this security zone name. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.

- Define one security label that has the lowest security level and one category that is not used in any other security labels. This security label can then be used for all unknown systems. Mandatory access control access under this security label will be more restrictive than under SYSLOW.

  A task using this security label will have R/O access to resources with SYSLOW, W/O access to resources with SYSHIGH, and R/W access to resources with this security label and SYSMULTI. It will have no access to resources with any other security labels because they will be disjoint.

  Any resources created under this security label will only be readable by tasks running under this security label, SYSHIGH, and SYSMULTI. This significantly reduces the risk from unintended, publicly readable or executable, SYSLOW resources.
- Define a security zone name for all unknown systems in the multilevel secure network.

  Define a NETACCESS profile for this zone in the SERVAUTH class. This profile can be generic with respect to the z/OS system name and TCP stack job name. If your installation supports communications with unknown systems on all z/OS systems, make this profile UACC(READ). Otherwise, make it UACC(NONE).

  Define a NETACCESS DEFAULT statement that maps all unspecified IP addresses into this security zone name. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.

Exempting certain users of certain programs from full Network Access Control

There are certain network administration programs that, to be fully functional, need to be exempted from some aspects of Network Access Control. For instance, the ping and traceroute functions test the network path to a destination system. They often need to traverse routers or firewalls that are at IP addresses mapped into security zones that are not normally mandatory access control accessible from a particular restricted stack. ICMP error messages from these systems will not be delivered to the function, if they are not exempted from the Network Access Control check that limits all traffic to security labels equivalent to the stack security label. Also, routing protocol daemons, such as OMPROUTE, frequently need to exchange routing table information with adjacent nodes that are in security zones that might not be mandatory access control accessible from a particular restricted stack. To operate correctly, these programs must also be exempted from some aspects of Network Access Control.

A SYSMULTI user with UPDATE authority to the EZB.STACKACCESS profile will be exempt from the Network Access Control restriction that all traffic must be with partners that are in security zones with security labels that are equivalent to the stack's security label or the security label associated with the local IP address. It is recommended that this authority be limited to their usage of the programs that must be exempted. This can be accomplished by first specifying UACC(READ)

when defining the STACKACCESS profiles, and then granting conditional update access to each by specifying the following:

```
PERMIT stackaccess_profile_name CLASS(SERVAUTH) ID(*) ACCESS(UPDATE) -
WHEN(PROGRAM(ping,oping,tracerte,otracert,omproute))
```

Note: The WHEN(PROGRAM()) conditional access parameter is not supported on profiles in the SERVAUTH class, except where explicitly stated. PERMITs with WHEN(PROGRAM()) on other profiles might be ignored.

Configuring stack sysplex features in a multilevel secure environment

The following apply to stack sysplex features in a multilevel secure environment:

- If TCPSTACKSOURCEVIPA is configured on a stack, the specified VIPA must be in a NetAccess security zone with a security label that is identical to the stack security label.
- If job-specific source IP addressing is used, the specified IP address must be in a NetAccess security zone with a security label that is permitted on the stack and is equivalent to the specified job. If an interface name is used, at least one of the IP addresses configured on that interface must be in a network security zone with a security label that is either SYSMULTI or equal to the specified job.
- For Sysplex Distributor, the distributing stack must either be an unrestricted stack or a restricted stack with a security label that is the same as all target stacks. The distributing stack will use the security label of the source security zone and the security labels of the active target applications when selecting a target. The distributing stack will also honor SECLBYSYSTEM when the target application is running under SYSMULTI on an unrestricted stack. In an environment using SECLBYSYSTEM, a distributing stack must be on a system where all security labels are active.
- VIPA takeover must be configured only between stacks with the same security label.
- Distribution of connections that require packet tagging are restricted to flowing over XCF or IUTSAMEHOST links. This restriction applies to the route from the client to the distributor, from the distributor to the target server, and from the target server back to the client.

Defining security labels on other profiles in the SERVAUTH class

A z/OS system with RACF SETROPTS MLACTIVE requires all resource profiles defined in the SERVAUTH class to have a security label. All z/OS Communications Server profiles in the SERVAUTH class have EZA, EZB, or IST as the first qualifier. Resource profiles that require meaningful security labels, such as EZB.STACKACCESS and EZB.NETACCESS as described previously in this chapter, are explicitly identified in the appropriate section of this document. The following resource profiles can be defined with the SYSNONE security label:

- EZB.SOCKOPT.*sysname.tcpname*.IPV6_DSTOPTS
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPLIMIT
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPOPTS
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_NEXTHOP
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_PKTINFO
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDR
- EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDRDSTOPTS

- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_TCLASS
- EZB.SOCKOPT.*sysname*.*tcpname*.SO_BROADCAST

However, some installations might want to define SO_BROADCAST with the SYSLOW security label to further reduce the exposure of data write_down by restricting datagram broadcast to users running with SYSLOW or SYSMULTI. All other z/OS Communications Server resource profiles can be defined with the SYSNONE security label.

Planning your multilevel secure network

Separate your network into security zones. Each subnetwork of physically managed systems should be defined as a single security zone. Several subnetworks with identical security labels and discretionary access control policy requirements can be assigned the same security zone name. Each trusted subnetwork of self-managed multilevel secure systems likely requires several security zones. The trusted subnetwork can also contain physically managed resources, such as routers and network administrator workstations. The trusted subnetwork security zone is likely to require a SYSHIGH security label. Multilevel secure stacks within the trusted subnetwork must have their interface addresses in security zones with the security label of the stack. VIPAs are usually placed in separate subnetworks dedicated to VIPAs and containing no real interface addresses. Multicast addresses, LOOPBACK addresses, and INADDR_ANY addresses require security zones as well.

The security administrator does the following:

- Defines security labels in the z/OS security server.
- Creates user IDs in the security server with appropriate security labels.
- Defines common discretionary access control policies.
- Defines the security zone name for each required combination of security label and discretionary access control policy.
- Identifies groups of managed machines that belong in each security zone.
- Provides physical security for each group of machines to ensure only users with appropriate security clearance can use them.
- Configures managed machines so that only the network administrator can set IP addresses.

The network administrator does the following:

- Isolates each group of machines into a subnetwork.
- Configures IP addresses on the machines.
- Configures a firewall to limit each group to only communicate with other subnetworks that have the same security label.
- Assigns network security zone names to subnetworks.
- Defines a NETACCESS statement that maps subnetworks and systems into network security zones. This can be placed in a shared data set that can be included in the PROFILE.TCPIP of all z/OS CS stacks in the network.

Planning for interactive UNIX System Services users in a multilevel secure environment

Interactive users of z/OS UNIX System Services that are permitted to log on with more than one security label must have a separate home directory for each security label. The approach recommended in *z/OS Planning for Multilevel Security and the Common Criteria* is similar to the following.

Steps for creating a separate home directory for each security label

Perform the following steps to create a separate home directory for each security label:

1. For each supported security label:
   a. Log on to an administrative user ID with that security label.
   b. Create a directory with the name of that security label under the /u directory.
   c. For each user permitted to that security label, create a home directory under that security label directory.

2. Create a symbolic link in the /u directory using the special value "$SYSSECR/", perhaps named symsecr as follows:

   ```
   ln -s "$SYSSECR/" /u/symsecr
   ```

   Tip: When issuing this command from the shell, use double quotation marks around the $SYSSECR/ string so that the shell does not attempt variable substitution before passing it to the ln command.

3. Define all users' home directories to be '/u/symsecr/*user*' as follows:

   ```
   ALTUSER user OMVS(HOME('/u/symsecr/user'))
   ```

This approach is useful in many other situations where a different configuration is required for different security labels.

Steps for setting stack affinity by security label

Installations that start several TCP/IP stacks under UNIX System Services common INET (CINET) often find it useful to set stack affinity for most interactive users. This can most easily be done by setting the environment variable _BPXK_SETIBMOPT_TRANSPORT to the name of the stack that users should use during login profile processing, as follows:

1. Create a directory for each security label under the /etc directory.

2. Create a profile script file in each directory containing the following:

   ```
   export _BPXK_SETIBMOPT_TRANSPORT=stackname
   ```

3. Create a "$SYSSECR/" symbolic link in the /etc directory, perhaps named seclbl as follows:

   ```
   ln -s "$SYSSECR/" /etc/seclbl
   ```

4. Edit the /etc/profile script file and add the following:

```
if  test  –f   /etc/seclbl/profile
then
.  /etc/seclbl/profile
fi
```

Host and domain name by security label

Installations that start several TCP/IP stacks under UNIX System Services CINET often find it necessary to define a different host name or domain name for interactive users based on their security label. The first decision that needs to be made is what host name and domain name each TCP/IP stack will have. Two common approaches are to:

- Assign each of the stacks a different host name in the same domain name.
- Assign each of the stacks on the same z/OS image the same host name, but define a different domain name for each security label.

These names are then implemented by creating a separate resolver configuration file for each security label.

Steps for creating a separate resolver configuration file for each security label

As described previously, this can be accomplished using the same directory structure under the /etc directory as follows:

1. Log in as a file system administrator with each security label.

---

2. Create a copy of your current resolver configuration file in each security label directory using the following command:

```
cp  /etc/resolv.conf  /etc/seclbl/rsv.cfg
```

---

3. Edit these new files as follows:
   a. Change the TCPIPJOBNAME statement to the appropriate stack job name.
   b. Change the HOSTNAME statement to the appropriate host name.
   c. Change the DOMAINORIGIN or first SEARCH statement to the appropriate domain for this security label.

      Tip: For information about interactions between the DOMAINORIGIN and SEARCH statements, refer to the TCPIP.DATA configuration statements chapter of *z/OS Communications Server: IP Configuration Reference*.

---

4. Replace your current resolver configuration file with a symbolic link as follows:

```
ln  -s  "$SYSSECR/rsv.cfg" /etc/resolv.conf
```

Planning for applications in a multilevel secure environment

In planning for applications in a multilevel secure environment, you must first understand the multilevel security programming rules that apply to socket applications.

Trusted network administration server applications
Applications that are part of the infrastructure, such as DNS and routing. The delivered information is not sensitive and needs to be accessible across many security labels.

Trusted multilevel secure server applications
Applications with a login process using port of entry, that do all resource access under the client's login identity, and that maintain separation of information accessed by different client tasks, such as FTP and otelnet.

Process applications that change identity must do the following:

- Issue _poe() prior to identity change (can be done by parent, for example, INETD).
- Change identity in parent process prior to fork, spawn, or exec to ensure interprocess communication (IPC) resources are properly labeled.
- Access any user resources after the identity change.
- Close unnecessary parent resources prior to exec.

Threaded applications that set identity on a thread must do the following:

- Issue _poe() on handling thread prior to _pthread_security_np().
- Ensure all related processing occurs on threads that have the same identity.
- Change identity on thread prior to spawn, fork, or exec.
- Access any user resources after identity change on the thread.
- Not access any other thread's resources.
- Close all user resources before removing identity from thread.
- Remove user identity from thread before acquiring new work.

All SYSMULTI applications must ensure that they do not put user data on server resources or other user resources. Pay special attention to debugging, logging, or tracing output. Many servers include user level data in output.

Trusted single-level secure server applications
Applications that deliver sensitive information accessed under the server's identity, such as TFTP and HTTP, and applications with a login process that do not use port of entry nor access client information under the server's identity, such as SMTP and MVSHRD.

Network administration commands and client applications
Local commands, requiring special controls and privileges, that are used to query, configure, or diagnose the network infrastructure.

General user commands and client applications
Local commands that access resources (including network resources) under the invoking user's security environment.

Unsupported applications
Applications that have not been inspected, or have been inspected and are not trusted in a multilevel secure environment.

Configuring z/OS CS applications in a multilevel secure environment

z/OS CS applications can be divided into the application categories previously described.

Trusted network administration server applications

The following are trusted network administration server applications:

- DNS name server (BIND 9)
- OMPROUTE
- Resolver
- SNTPD
- TIMED
- TRMD
- z/OS syslog daemon (syslogd)
- z/OS UNIX Policy Agent

They can run under a user ID with the SYSMULTI security label in a multilevel secure environment, provided that they adhere to the following configuration instructions.

DNS name server (BIND 9): You can run as many instances of DNS as appropriate for your installation. Name resolution is not considered sensitive information. You can run one instance of the DNS server per z/OS image under a user identity with a SYSMULTI security label.

OMPROUTE: You should run one instance of OMPROUTE for each stack that is using dynamic route configuration. Each instance of OMPROUTE must run under a user ID with SYSMULTI. OMPROUTE communicates with multicast IP addresses. These addresses must be configured into NetAccess security zones. If OMPROUTE must communicate with adjacent nodes that are not in network security zones with security labels equivalent to the security label of the restricted stack or the security label associated with the local IP address, OMPROUTE must run under a user ID that is SYSMULTI and has update authority to the EZB.STACKACCESS resource profile. A SYSMULTI user with UPDATE authority to the EZB.STACKACCESS resource profile is exempt from the restriction that all traffic must be with partners that are in security zones with security labels that are equivalent to the stack's security label or the security label associated with the local IP address. You should carefully protect your routing configuration files to maintain network security. You should consider using any application level security supported by the routing protocol you use. OMPROUTE must be run with stack affinity for the stack that it is servicing.

OMPROUTE uses multicast UDP datagrams to discover adjacent OSPF routing daemons on common subnetworks. Adjacent OMPROUTE instances then establish TCP connections with each other. When two unrestricted stacks running OMPROUTE are attached to a common subnetwork that is neither XCF nor IUTSAMEHOST, adjacency errors will occur. The multicast UDP datagram is successfully transmitted, but the subsequent TCP connection between the two SYSMULTI interface addresses fails because it requires packet tagging. These adjacency failures can be avoided by preventing OMPROUTE from receiving multicast datagrams from partners with which it cannot communicate.

Steps for avoiding adjacency failures: Perform the following steps to prevent OMPROUTE from receiving multicast datagrams from partners with which it cannot communicate.

1. Create a network security zone named URXCF for all interface addresses in XCF or IUTSAMEHOST networks on unrestricted stacks:
   a. Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

```
RDEFINE SERVAUTH EZB.NETACCESS.*.*.URXCF UACC(READ) SECLABEL(SYSMULTI)
```

b. Modify the common NETACCESS profile to define the addresses in this zone:

```
NETACCESS
192.168.10.0/24 URXCF ; xcf subnet perhaps
10.254.254.0/24 URXCF ; IUTSAMEHOST subnet perhaps
ENDNETACCESS
```

---

2. Create a network security zone named UROTHER for all interface addresses in other network types on other unrestricted stacks:
   a. Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

```
RDEFINE SERVAUTH EZB.NETACCESS.*.*.UROTHER UACC(READ) SECLABEL(SYSMULTI)
```

   b. Prevent the OMPROUTE running for each unrestricted stack from receiving datagrams from this zone with the following RACF command:

```
PERMIT EZB.NETACCESS.*.*.UROTHER CLASS(SERVAUTH) ID(ompurid) ACCESS(NONE)
```

   c. Modify the common NETACCESS profile to define the addresses in this zone:

```
NETACCESS
10.254.1.0/24 UROTHER ; ethernet subnet perhaps
ENDNETACCESS
```

---

3. Create a network security zone named URLOCAL for all interface addresses in other network types on each specific unrestricted stack. OMPROUTE is permitted to use this local interface to connect to adjacent OMPROUTE daemons on adjacent restricted stacks.
   a. Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

```
RDEFINE SERVAUTH EZB.NETACCESS.*.*.URLOCAL UACC(READ) SECLABEL(SYSMULTI)
```

   b. Modify the local NETACCESS profile for each stack to define the local addresses in this zone:

```
NETACCESS
10.254.1.17/32 URLOCAL ; local address in ethernet subnet perhaps
ENDNETACCESS
```

Resolver: The resolver task is started by OMVS and runs under the same identity as OMVS. This identity will normally have a security label of SYSMULTI. The resolver processes its configuration, system console commands, and its CTRACE under this identity.

The resolver is configured to use a set of files, data sets, and name servers in a given sequence when asked to resolve a name or IP address. For information on configuring the resolver search sequence, refer to *z/OS Communications Server: IP Configuration Reference*.

The name resolution process is performed on the requesting thread of execution under the user identity associated with that thread. Each user must have READ access to the files and data sets configured. This is best accomplished by making these UACC(READ) with SECLABEL(SYSLOW). For the resolver to contact a name server on their behalf, each user must also have appropriate STACKACCESS permission and NETACCESS permission to the security zone of the name server.

OMVS might be configured in a CINET environment with multiple AF_INET Physical File Systems (TCP/IP stacks). In this environment, users and jobs can optionally have affinity for a single stack or allow CINET to choose a stack for them. When stack affinity is not set, CINET replicates some AF_INET calls to all attached stacks. Other calls are routed by CINET to a single stack based on routing information that CINET has extracted from those stacks. The socket() call is one of the calls that is routed to all connected stacks. This might produce RACF Failure Audit messages to any stacks that the resolver user is not permitted to. These messages can be eliminated by setting stack affinity prior to using resolver functions.

SNTPD: You can run as many instances of SNTPD as appropriate for your installation. Time is not considered sensitive information. You can run any SNTPD server under a user identity with a SYSMULTI security label.

TIMED: You can run as many instances of TIMED as appropriate for your installation. Time is not considered sensitive information. You can run any TIMED server under a user identity with a SYSMULTI security label.

TRMD: You should run one instance of TRMD for each stack that has IDS functions configured. Each instance of TRMD can run under a user ID with the same security label as the stack it is servicing, or with SYSMULTI. TRMD must be run with stack affinity for the stack that it is servicing.

z/OS syslog daemon (syslogd): You should run one instance of syslogd per z/OS image under a user ID with the SYSMULTI security label. The AF_UNIX socket (default name /dev/log) created by syslogd must have a security label of SYSMULTI, so that applications running under various security labels can log to it.

Syslogd routes application log messages according to filters configured in /etc/syslog.conf. Log messages from applications running under different security labels can be mixed into an output log file by a filter rule. Each output log file created by syslogd should be configured in a directory with a SYSHIGH security label.

z/OS UNIX Policy Agent: You should run one instance of Policy Agent per z/OS image under a user ID with the SYSMULTI security label. The AF_UNIX socket /tmp/unix.str created by Policy Agent must have a security label of SYSMULTI so that applications running under various security labels can connect to it. The user ID that Policy Agent is running under should have READ access to the EZB.STACKACCESS resource profiles of all stacks on the system.

Trusted multilevel secure server applications

The following are trusted multilevel secure server applications:

- TN3270 and TN3270E Telnet server running as part of the TCP/IP stack
- z/OS UNIX FTP server
- z/OS UNIX REXEC server
- z/OS UNIX RSH server
- z/OS UNIX Telnet server

They can run under a user ID with the SYSMULTI security label in a multilevel secure environment, provided that they adhere to the following configuration instructions.

TN3270 and TN3270E Telnet server: You can run the TN3270 server on as many stacks as appropriate for your installation. Configure NACUSERID to enable Network Access Control for your TN3270 server. In a multilevel secure environment, all TN3270 ports must have a NACUSERID configured to ensure correct LU mapping.

The NACUSERID can be the same user ID used to run the stack. On a restricted stack, the security label associated with the NACUSERID must be equivalent to the security label of the stack user ID.

Any TN3270 server can listen on multiple ports. On an unrestricted stack, a TN3270 server listening on a port can run under a user ID with a security label other than SYSMULTI. In this case, you must have a VIPA defined in a security zone with the same security label. Use the BIND parameter on the PORT reservation statement in the TCPIP profile to override the bind of this TN3270 port to that VIPA.

The TN3270 server maps TCP connections to LU names configured in LU groups or supplied by an LU mapping exit. All of the LU names in a single LU group or provided by a single exit must have the same security label. The TN3270 server ensures that the mapped LU name has an equivalent security label to the NetAccess profile for the network security zone containing the client. Note that if the security label for the NetAccess profile is SYSMULTI, the TN3270 server ensures that the mapped LU name also has a security label of SYSMULTI. When performing user login authentication, the SNA application should use this LU name as the port of entry TERMID. The LU name profile in the TERMINAL class must have equivalent security label definitions on both the TN3270 system and the SNA application system.

You can also run Telnet as its own procedure. In this case, the NACUSERID is optional. Ports without NACUSERID are controlled under the server's user ID. Running as its own procedure, Telnet is not multilevel security compliant. Telnet must run as part of the TCP/IP stack to be compliant.

z/OS UNIX FTP server: You can run as many instances of the FTP daemon as appropriate for your installation. You can run any FTPD under a user identity with a SYSMULTI security label. A single FTPD can span a mix of restricted and unrestricted stacks in a CINET environment.

To use NetAccess profiles for IPv4 clients, configure PortOfEntry4 SERVAUTH in the FTP.DATA file of the server. If you do not, you must configure profiles in the TERMINAL class covering all IPv4 addresses in your multilevel secure network, with the same security label defined on your corresponding NetAccess profiles.

z/OS UNIX INET daemon: The z/OS UNIX INET daemon (inetd) can be configured to listen for requests for many different services. You can run inetd under a user identity with a SYSMULTI security label. Each connection to one of its ports causes it to issue the _poe() service and fork a new instance of the server configured to handle that service. The new connection is passed to the server. If a user ID for the forked server is configured in the inetd configuration file, that user ID must be authorized to the SERVAUTH NETACCESS profile for the network security zone containing the client IP address. If the forked server performs additional user login authentication, that user ID must also be permitted to the NetAccess profile.

The following z/OS CS servers are designed to be forked by inetd. They each perform additional user authentication.

- z/OS UNIX REXEC server
- z/OS UNIX RSH server
- z/OS UNIX Telnet server

Trusted single-level secure server applications

The following are trusted single-level secure server applications:

- SMTP server (SMTPPROC)
- TFTP
- TSO REXEC and RSH servers
- z/OS UNIX sendmail

They are not supported in a multilevel secure environment when they are run under a user ID with a security label of SYSMULTI. The following configuration instructions show how to run separate instances of these servers for each required security label.

SMTP server (SMTPPROC): The SMTP server receives mail files from TCP clients or the JES spool, and forwards mail files to other SMTP servers or to the JES spool. User identities and mail security labels are not processed.

You must run a separate instance of SMTP for each security label you need to support. Run each one under a different job name assigned to a user ID with the appropriate security label. You can run multiple SMTP servers on the same unrestricted stack. Define a VIPA in a network security zone with the appropriate security label for each server on that stack. Use the PORT reservation statement in the TCPIP profile or the LISTENONADDRESS keyword in the SMTP configuration file to override the bind address of each job to the appropriate VIPA. You can also run multiple SMTP servers with one server per restricted stack using stack affinity. In this case, it is not necessary to override the INADDR_ANY bind address.

Spool files routed to an instance of SMTP must have an equivalent security label.

Spool files created by SMTP carry the SMTP security label. SMTP servers are only permitted to connect to other SMTP servers with an equivalent security label.

TFTP: The TFTP server delivers files to any requester without user authentication. It should be configured to limit the files it attempts to access. All file access is done under its own user identity. Clients can get files that are publicly readable and have security labels that the TFTP server dominates. Clients can put existing files that are publicly writable and have a security label that dominates the TFTP server's security label.

Requirement: You must run a separate instance of TFTP for each security label you need to support.

Steps for running a separate instance of TFTP for each security label: Before you begin, plan to run each instance of TFTP under a different job name assigned to a user ID with the appropriate security label. Understand that you can run multiple TFTP servers on the same unrestricted stack.

Perform the following steps to run a separate instance of TFTP for each security label:

1. Define a VIPA in a network security zone with the appropriate security label for each server on that stack.

---

2. Perform one of the following:
   - With a unique port number for each entry, use the PORT reservation statement in the TCPIP profile to override the bind address of each job name to the appropriate VIPA and port number.
   - Instead of using the PORT reservation statement and specifying the port using -p in the TFTP start procedure, use the -b TFTP start option to specify the IP address to which this instance of TFTP should bind. If the -b start option is used, each instance of TFTP can use the same well-known port (69).

---

3. Ensure the procedure for each instance of TFTP specifies on which port or IP address it will run.

You know you are done when you have established an environment where separate instances of TFTP can use the same well-known port (if the -b start option was used.)

In a CINET environment, you should establish stack affinity to the intended server stack prior to starting each instance of TFTP.

Guideline: Ensure that files and directories have appropriate security labels prior to using the TFTP server. Be especially careful not to have publicly writable files with a SYSMULTI security label, to eliminate the possibility of two users with different security labels (including two TFTP servers) passing data through the file.

TSO REXEC and RSH servers: The MVRSHD server listens for both rexec and rsh client connections. Requests are submitted as TSO batch jobs and spool output is returned to the client. MVRSHD does perform client authentication and does isolate user data for each connection. However, it does not request port of entry processing during client authentication, and job submission and spool access are performed under the identity of the MVRSHD job rather than the client identity.

RSH client requests are run under the identity and security label of the MVRSHD job. REXEC client requests are run under the user ID specified on the request.

MVRSHD supports an optional startup parameter to control whether or not it inserts the SECLABEL parameter on the generated job card. When SECLABEL=Y is specified at startup, MVRSHD passes its own security label in the SECLABEL job parameter. The user ID specified on the request must be permitted to the MVRSHD server's security label. When the parameter is not specified, or is specified as SECLABEL=N, the default security label of the user ID specified on the request must be identical to the MVRSHD server's security label.

You must run a separate instance of MVRSHD for each security label you need to support. Run each one under a different job name assigned to a user ID with the appropriate security label. You can run multiple MVRSHD servers on the same unrestricted stack. Define a VIPA in a network security zone with the appropriate security label for each server on that stack. Use the PORT reservation statement in the TCPIP profile to override the bind address of each job to the appropriate VIPA.

z/OS UNIX sendmail: The objective of this section is to guide you through the steps required to set up sendmail in a multilevel secure environment. The contents of this section are based on the assumption that you understand sendmail concepts and terminology. If you are not familiar with sendmail, you should first read "Configuring z/OS UNIX sendmail and popper" on page 1122.

Considerations for sendmail daemons: Mail must be configured so that it can only be exchanged among equivalent security labels. Essentially, multiple independent mail networks must be set up. Mail support does not need to be configured for every security label supported on a multilevel secure system. On z/OS systems, users must not be configured for sendmail when they log on with the SYSMULTI security label.

The most straightforward way to accomplish multiple independent mail networks is to define a different domain for each security label supported. Single-level security systems have their host name and IP address defined in the domain intended for their security label. Multilevel secure systems can have the same host name defined in each domain name intended for one of the security labels they support. An appropriate IP address on a restricted stack, or a VIPA on an unrestricted stack, is used in each domain for the multilevel secure system. When users log on to a multilevel secure system, their mail address becomes their user ID at that host name within the security label-specific domain. They use only the sendmail daemon on their system that supports that domain. When a user directs mail to a user ID at another multilevel secure host name, it defaults to the sendmail daemon on that host that is supporting the same security label-specific domain.

On a multilevel secure system in a single domain environment, each security label with mail support has a different host name. When users log on, their mail address becomes their user ID at the security label-specific host name in the common domain. They use only the sendmail daemon on their system that supports that host name. Users must know which host names are located in network security zones with equivalent security labels.

Of course, a user can address mail to another user at any host and domain name. However, their sendmail daemon will only be able to connect to other mail servers at IP addresses in network security zones with an equivalent security label. Mail sent to hosts that reside in security zones with security labels that are not equivalent will time out. Mail received by the z/OS sendmail daemon, addressed to a local user ID that is defined but is not permitted to the security label of the sendmail daemon, is returned with the unknown user error message.

The sendmail daemon receives mail files from TCP clients or other mail servers. It forwards these mail files to other mail servers, queues mail for later transmission to other mail servers, or passes mail to tsmail (or another local delivery agent) to complete local delivery. In a multilevel secure environment, you must run a separate instance of the sendmail daemon for each security label. The sendmail daemon must not be run under the SYSMULTI security label. In some cases, the sendmail daemon queues mail for delivery. There are several different configuration options that allow the sendmail daemon to process the queue. Each mail queue must have a security label that matches the security label of the sendmail daemon.

Run each sendmail daemon under a different job name assigned to a user ID with the appropriate security label. In a multilevel secure environment, there are special

configuration considerations and changes needed to support multiple sendmail daemons running under different security labels. These considerations and changes are as follows:

- You can run multiple sendmail daemons on the same unrestricted stack (using multiple network security zones), or you can run one sendmail daemon on each restricted stack.
- If CINET is active and multiple restricted stacks are used, each sendmail daemon can establish stack affinity through the _BPX_SETIBMOPT_TRANSPORT environment variable and bind to INADDR_ANY. Otherwise, you must define a VIPA in a network security zone with the appropriate security label for each sendmail daemon.

CINET or INET	**Restricted stack**	**Unrestricted stack**
CINET with single stack	VIPA is optional.	VIPA with same SECLABEL is required.
CINET with multiple stacks	Use VIPA or stack affinity.	VIPA with same SECLABEL is required.
INET	VIPA is optional.	VIPA with same SECLABEL is required.

- Sendmail clients are permitted to connect only to sendmail daemons with an equivalent security label. NETACCESS configuration permits sendmail daemons (mail transmission agents, or MTAs) to connect outbound only to other MTAs with an equivalent security label and accept inbound connections only from clients and other MTAs with an equivalent security label. Sendmail configuration should direct sendmail clients to attempt connections that will succeed.
- If the mail submission agent (MSA) or MTA is bound to a specific IP address, a shared sendmail.cf can be used if you are using the same host name in security label-specific domains. If you are using different host names, you must provide a separate sendmail.cf for each security label.
- Whenever a sendmail daemon (MTA) is bound to a specific IP address, the use of job-specific source IP addressing specifying the same IP address is suggested. This ensures that name resolution by peer MTAs results in the intended host name and domain name.

Considerations for sendmail clients and sendmail MSP: If the mail submission program (MSP) feature is used, the feature(`msp') statement in submit.mc directs the sendmail client to connect to localhost by default, which is typically LOOPBACK or LOOPBACK6. In a CINET environment, these addresses can be ambiguous if you are not using stack affinity. This default approach works when you are running only one restricted stack, or the client is running with stack affinity to a restricted stack. The sendmail daemon must be running on the same restricted stack and must not bind the MSA socket to a specific IP address. In all other cases, the feature(`msp') statement in submit.mc must include the host name of the sendmail daemon. If you are using the same host name for your sendmail daemons with security label-specific domain names, you can use a shared submit.cf that specifies the host name without the domain (that is, not fully qualified). If you are using different host names for your sendmail daemons, you must provide a separate submit.cf for each security label.

The sendmail client and MSP are permitted to connect only to a sendmail MSA or MTA with an equivalent security label.

Other considerations: If DNS is to be used for mail server searches, each VIPA with a unique name should be added to the name server as an MX record. If DNS is not being used for mail server searches, DNS searching can be turned off by adding the /etc/mail/service.switch file containing the following line:

```
hosts files
```

Sendmail.cf has /etc/mail/service.switch as the default location for this file.

If you use a shared aliases file, the user specified as postmaster in /etc/mail/aliases must be permitted to all security labels supported for sendmail. If you need to have different aliases, you must provide a separate aliases file for each security label. The user specified as postmaster in each security label-specific aliases file must be permitted to that security label.

Steps for setting up and running sendmail in a multiple security label environment:
Before you begin: Read and understand "Steps for configuring z/OS UNIX sendmail" on page 1127. The objective of this section is to guide you through the steps required to set up sendmail in a multilevel secure environment. The contents of this section are based on the assumption that you understand your sendmail configuration and your network configuration.

Perform the following steps to set up and run sendmail in a multiple security label environment:

1. Create host and domain names as follows:
   a. Decide whether you are using security label-specific domain names or separate host names.
   b. Define a VIPA in a network security zone with the appropriate security label for each sendmail server to which you will bind.

      Tip: This step is optional if stack affinity is used with restricted stacks.
   c. Define the domains and host names in your DNS, or /etc/hosts and /etc/ipnodes files. For example, if your system supported the security labels SYSHIGH, SYSLOW, ORANGES and APPLES, the following are some sample /etc/hosts definitions:

      ```
      ;  DIFFERENT HOST NAMES IN THE SAME DOMAIN
      10.10.10.1  Z10HIGH.MYCORP.COM
      10.10.10.2  Z10LOW.MYCORP.COM
      10.10.10.3  Z10ORNGE.MYCORP.COM
      10.10.10.4  Z10APPLE.MYCORP.COM
      ;  SAME HOST NAMES IN DIFFERENT DOMAINS
      192.168.10.41 ZOS10.SYSHIGH.MYCORP.COM
      192.168.10.42 ZOS10.SYSLOW.MYCORP.COM
      192.169.10.43 Z0S10.ORANGE.MYCORP.COM
      192.168.10.44 ZOS10.APPLES.MYCORP.COM
      ```
   d. Set up stack affinity and resolver configuration, as described in "Planning for interactive UNIX System Services users in a multilevel secure environment" on page 155.
   e. If starting the sendmail servers using started procedures, do one of the following:
      - Add a STDENV DD line to the JCL, specifying a file that sets the environment variable RESOLVER_CONFIG to the appropriate file based on the security label for this sendmail server. For example:

        ```
        //STDENV DD PATH='/etc/seclbl/sendmail.env',
        //       PATHOPTS=(ORDONLY)
        ```

- Set the RESOLVER_CONFIG environment variable to the appropriate file (based on the security label of the sendmail server) using ENVAR on the PARM keyword of the EXEC statement. For example:

```
//SENDMAIL EXEC PGM=BPXBATCH,REGION=4096K,TIME=NOLIMIT,
// PARM=('PGM /usr/sbin/sendmail -bd -q1h -L sndmail1',
//      'ENVAR("RESOLVER_CONFIG=/etc/resolv.conf")'
```

---

2. Create security label-specific mail queue directories as follows:

   a. If necessary, create the parent SYSMULTI directories. To create these directories, you must log on as a superuser with the SYSMULTI security label and issue the following UNIX System Services commands:

      ```
      mkdir /var
      mkdir /var/spool
      ```

   b. Create new mail queue directories for each security label to be supported for sendmail. For each security label to be supported for servers, repeat the following steps:

      1) Log on to a TSO user ID with that security label.

      2) Issue the following UNIX System Services commands, replacing *seclbl* with the security label name and *sndmuser* with the appropriate sendmail user ID:

         ```
         mkdir /var/spool/seclbl
         mkdir /var/spool/seclbl/mqueue
         chown sndmuser:sndmgrp /var/spool/seclbl/mqueue
         mkdir /var/spool/seclbl/clientmqueue
         chown smmsp:smmspgrp /var/spool/seclbl/clientmqueue
         ```

         Tip: Do not create /var/spool/*seclbl* or its subdirectories for SYSMULTI or any other security label that is not to be supported for sendmail. There will not be a server to use them.

      3) Create a symbolic link for these mail directories. The following UNIX System Services command issued by a superuser creates the symbolic link and directs the mail to the appropriate queue:

         ```
         ln -s '$SYSSECR/'  /var/spool/secsymr
         ```

---

3. Change statements in the sendmail.mc configuration file as follows:

   a. Change the location of the daemon pid file so that there is a separate one for each security label:

      ```
      define(`confPID_FILE', `/var/spool/secsymr/sendmail.pid')dnl
      ```

   b. Change the location of the local host names file so that there is a separate one for each security label:

      ```
      define(`confCW_FILE', `/etc/secsymr/local-host-names')dnl
      ```

   c. Change the location of the queue directories:

      ```
      define(`MSP_QUEUE_DIR', `/var/spool/secsymr/clientmqueue')dnl
      define(`QUEUE_DIR', `/var/spool/secsymr/mqueue')dnl
      ```

   d. Define configuration file variables for the local host name and domain name:

      ```
      define(`MLS_hostname', esyscmd(`hostname -s'))dnl
      define(`MLS_domain', esyscmd(`domainname'))dnl
      ```

   e. Change the DAEMON_OPTIONS statements. Code the ADDR parameter, specifying the unqualified local host name. For each sendmail daemon you start, this name will be resolved within the domain specific to the security label that daemon is running under. For example:

```
FEATURE(`no_default_msa')dnl
DAEMON_OPTIONS(`Name=MTA, Addr='MLS_hostname`, Family=inet ')dnl
DAEMON_OPTIONS(`Name=MSA, Port=587,  Addr='MLS_hostname`, Family=inet ')dnl
```

---

4. Change statements in the submit.mc configuration file as follows:
   a. Change the location of the client queue directory:
      ```
      define(`MSP_QUEUE_DIR', `/var/spool/secsymr/clientmqueue')dnl
      ```
   b. Define configuration file variables for the local host name and domain name:
      ```
      define(`MLS_hostname', esyscmd(`hostname -s'))dnl
      define(`MLS_domain', esyscmd(`domainname'))dnl
      ```
   c. Change the FEATURE(`msp') statement. Code the unqualified local host name for this system. For each user that invokes /bin/sendmail, this name will be resolved within the domain specific to the security label that user is running under. For example:
      ```
      FEATURE(`msp', MLS_hostname)
      ```
      Tip: If MX records in DNS should not be searched, brackets [] must be placed around the name or address.

---

5. Use the m4 compiler to create sendmail.cf and submit.cf files as follows:
   a. Create symbolic links for the submit.cf and sendmail.cf files:
      ```
      ln  -s  /etc/seclbl/mail/submit.cf  /etc/mail/submit.cf
      ln  -s  /etc/seclbl/mail/sendmail.cf  /etc/mail/sendmail.cf
      ```
   b. Log on with each security label to be supported for mail and create the .cf files:
      ```
      /etc/m4 /etc/mail/sendmail.mc > /etc/mail/sendmail.cf
      /etc/m4 /etc/mail/submit.mc > /etc/mail/submit.cf
      ```

---

6. Create new system mailbox directories for each supported security label. When sendmail is configured to use /usr/lib/tsmail, you must configure /usr/mail to be a symbolic link to a security label-specific directory. The following approach creates a set of security label-specific mount points for mail file systems:
   a. Log on as a superuser with the SYSMULTI security label and issue the following UNIX System Services commands:
      ```
      mkdir  /mailmnt
      ln  -s  '$SYSSECR/mail'  /usr/mail
      ```
      Tip: If another program other than /usr/lib/tsmail is used, the above commands might need to be adjusted accordingly.
   b. Perform the following steps for each security label supported for mail on the system:
      1) Log on to a superuser ID using the security label.
      2) Set an environment variable to your current security label as follows:
         ```
         export SL=$(id -M)
         ```
      3) Issue the following UNIX System Services command:
         ```
         mkdir /mailmnt/$SL
         ```

---

7. Start a separate instance of the sendmail daemon for each security label you need to support. For instructions on setting up the appropriate user IDs and groups for sendmail, see “Steps for configuring z/OS UNIX sendmail” on page 1127

page 1127. You can start sendmail daemons either from UNIX System Services or as started procedures. Refer to the EZARACF sample for examples of defining user IDs and STARTED class profiles for sendmail.

Repeat either of the following steps to start a sendmail daemon for each supported security label:

- From UNIX System Services:
  a. Log on to UNIX System Services with the security label that you want to support.
  b. Start the sendmail daemon from the command line.
- From the MVS console:
  a. Add a user ID and permit the user ID to the appropriate security label.
  b. Use the started class facility to assign the user ID with the appropriate security label to the procedure.
  c. Start the procedure from the MVS console or using the AUTOLOG statement in PROFILE.TCPIP.

Network administration client applications

The following are network administration client applications:

- DNS utilities:
  - nsupdate
  - rndc
- DNS utilities used for DNS security (DNSSEC):
  - dnsmigrate
  - dnssec-keygen
  - dnssec-makekeyset
  - dnssec-signkey
  - dnssec-signzone
  - rndc-confgen
- netstat
- pasearch
- ping, oping
- traceroute, otracert, tracerte
- trmdstat

For security implications on the use of these commands, see the following information on configuring these applications for use in your multilevel secure environment.

nsupdate: The nsupdate command is used to update the name server database. Because the name server is often running with a SYSMULTI security label and listening on IP addresses in SYSMULTI network security zones, mandatory access control policy does not limit a general user's ability to use the nsupdate command to contact the server. File system access control lists and security labels should be used to limit execution of this command to intended network administrators. Application level security options should also be considered.

rndc: The remote name daemon control (rndc) command allows the system administrator to control the operation of a name server. Because the name server is often running with a SYSMULTI security label and listening on IP addresses in SYSMULTI network security zones, mandatory access control policy does not limit

a general user's ability to use the rndc command to contact the server. File system access control lists and keys should be used to limit execution of this command to intended network administrators.

DNS utilities used for DNS security (DNSSEC): The dnssec-keygen program is used to generate keys. The dnssec-makekeyset program is used to create a key set from one or more keys. The dnssec-signkey program is used to sign one child's key set. The dnssec-signzone program is used to sign a zone. The rndc-confgen program is used to create a rndc.conf file.

netstat: The netstat command can be used to display many stack settings, information about open ports and established connections. You might want to restrict the availability of some netstat options to network administrators. For information on restricting netstat options, see "Netstat access control" on page 128.

pasearch: The pasearch command is provided to interrogate the stack policies currently in place on a system. It requires that the user have superuser privileges. Execution should be limited to those network and security administrators that need to know policy settings. This can be done by setting the file mode access permission bits using the CHMOD command.

ping: The ping command is used to verify the network path to a given destination. It sends ICMP Echo Request datagrams to the destination and listens for ICMP Echo Reply responses. Normal Network Access Control limits a user's ability to send and receive these datagrams. On a restricted stack, all users are limited to sending and receiving datagrams to destinations in network security zones with security labels equivalent to the stack. On an unrestricted stack, users are limited to destinations equivalent to their own security label. Users with a SYSMULTI security label on an unrestricted stack are limited to those security zones with which they are authorized to communicate.

You can permit SYSMULTI users to ping addresses in security zones that are not equivalent to that stack by PERMITing them with UPDATE access to the STACKACCESS profile for that stack. This PERMIT can be limited to only apply when using certain programs, such as ping, by using the WHEN(PROGRAM(oping,ping)) clause on the PERMIT.

traceroute: The traceroute function is used to verify the network path to a given destination and identify each intermediate system. It sends UDP datagrams to the destination with increasing hop count values, and listens for ICMP TIME EXCEEDED INTRANSIT and PORT UNREACHABLE responses. Normal Network Access Control limits a user's ability to send and receive these datagrams. On a restricted stack, all users are limited to sending datagrams to destinations in network security zones with security labels equivalent to the stack, and receiving datagrams from intermediate systems in equivalent security zones. On an unrestricted stack, users are limited to destinations equivalent to their own security label. Users with a SYSMULTI security label on an unrestricted stack are limited to those security zones with which they are authorized to communicate.

You can permit SYSMULTI users to trace the route to addresses in security zones that are not equivalent to that stack by PERMITing them with UPDATE access to the STACKACCESS profile for that stack. This PERMIT can be limited to only apply when using certain programs, such as traceroute, by using the WHEN(PROGRAM(tracerte,otracert)) clause on the PERMIT.

trmdstat: The trmdstat command is provided to process syslogd output files and generate IDS reports. The network or security administrator using this command needs to be authorized to read the syslogd output files. This often requires a user security label of SYSHIGH or SYSMULTI.

General user client applications

The following are general user client applications:

- dig
- dnsdomainname, domainname
- host
- hostname
- MISCSERV (echo, discard, character generator)
- NSLookup
- TSO REXEC client
- TSO RSH client
- TSO Telnet client
- z/OS UNIX FTP client
- z/OS UNIX REXEC client
- z/OS UNIX RSH client

These commands are supported for use by general users in a multilevel secure environment. Mandatory access control policy enforcement limits the stacks that can be used, the servers that can be contacted, and the local resources that can be used. No extra, application specific, multilevel secure environment configuration is required. For discretionary security environment considerations, see the appropriate sections of this book and refer to *z/OS Communications Server: IP Configuration Reference*.

Unsupported applications

The following applications are not supported in a multilevel secure environment:

- BINL
- DHCP PXE
- LPD
- LPQ
- LPR
- LPRM
- LPRSET
- NCPROUTE
- NPF
- Portmapper
- SNMP NetView client
- z/OS UNIX DNS name server (BIND 4)
- z/OS UNIX Network SLAPM2 subagent
- z/OS UNIX OMPROUTE SNMP subagent
- z/OS UNIX popper
- z/OS UNIX RSVP agent
- z/OS UNIX SLA subagent
- z/OS UNIX SNMP client
- z/OS UNIX SNMP server and agent

- z/OS UNIX Trap Forwarder Daemon

Any other commands or applications not explicitly mentioned in this section or prior sections have not been inspected and are not supported in a multilevel secure environment.

Changing your multilevel secure networking environment

Changes to certain parts of your multilevel secure configuration should be well planned. Changes to the NETACCESS statement in PROFILE.TCPIP, the security label associated with the TCPIP started task, the security label associated with the EZB.STACKACCESS or EZB.NETACCESS profiles in the SERVAUTH class, or the definition of profiles in the SECLABEL class can result in an IP address being associated with a different security label. These changes imply that corresponding changes have been implemented that affect the security management of the affected systems. For systems that support mandatory access control policy enforcement, those policies must be updated at the same time. For systems that do not support mandatory access control policy enforcement, the physical user access procedures, system data content, firewall configurations, and router configurations must be updated at the same time.

The safest method of controlling mandatory access control policy changes is to use the RACF options MLSTABLE and MLQUIET. When the RACF options are set to MLACTIVE, MLSTABLE, and NOMLQUIET, TCP/IP does not permit an existing NETACCESS configuration to be changed with a VARY TCPIP,,OBEYFILE command. When the RACF option MLQUIET is set, RACF requires the TCP/IP job user ID to be RACF SPECIAL to open data sets referenced by VARY TCPIP,,OBEYFILE commands. For more information on setting and using these options, refer to *z/OS Security Server RACF Security Administrator's Guide*.

In an MLSTABLE environment, all user and application access to data should be halted before entering the MLQUIET environment. All network access can be halted by stopping all TCP/IP stacks. If security administrators must access the system through TN3270, consider running a restricted stack with only a TN3270 server for the security administrators. Permit the TN3270 server NACUSERID to the EZB.STACKACCESS profile for this stack. Stop all other TCP/IP stacks. After the local policy changes and all coordinated changes on other systems are complete, set NOMLQUIET and restart your production TCP/IP stacks and network servers.

Every stack running on a system with the RACF option MLACTIVE does an internal consistency check on several PROFILE.TCPIP statements and their associated SERVAUTH profiles. This consistency checking occurs at the end of initial profile processing, after the VARY TCPIP,,OBEYFILE command modifies the profile, and whenever RACLIST is issued for the SERVAUTH or SECLABEL classes. Some applications, such as OMPROUTE, also cause the consistency checking to occur because they internally issue an equivalent of the VARY TCPIP,,OBEYFILE command. TCP/IP writes a message to the job log for each inconsistency it finds that could compromise the security of information flowing through the stack. If inconsistencies are found, a final message (EZD1217I) summarizing the number of problems found is written to the system console.

By default, the stack will continue running when inconsistencies are found. It is recommended that you override this default by specifying GLOBALCONFIG MLSCHKTERMINATE in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, before starting production workloads. Before making

security related configuration changes, it is recommended that you first stop all production workloads. You can then specify GLOBALCONFIG NOMLSCHKTERMINATE in PROFILE.TCPIP or in the data set referenced by a VARY TCPIP,,OBEYFILE command. This parameter can only be changed from MLSCHKTERMINATE to NOMLSCHKTERMINATE when the RACF option NOMLSTABLE is set or when both MLSTABLE and MLQUIET are set.

The stack performs the following consistency checks:

- The stack does not currently have a NETACCESS statement configured.
- The TCP/IP job security label must be currently active on this system.
- A STACKACCESS profile for this stack must exist.
- The STACKACCESS profile must have the same security label as the TCP/IP job.
- The NETACCESS statement must have both INBOUND and OUTBOUND turned on.
- A NETACCESS profile must exist for each defined NETACCESS security zone.
- If NETACCESS DEFAULTHOME is configured, it must have a zone with the same security label as the TCP/IP job.
- If NETACCESS TcpStackSourceVIPA is configured, it must be in a zone with the same security label as the TCP/IP job.
- The special address INADDR_ANY (0.0.0.0) must be in a NETACCESS security zone.
- The special address INADDR_ANY (0.0.0.0) must be in a zone with the same security label as the TCP/IP job.
- On an IPv6-enabled stack, the special address IN6ADDR_ANY (::) must be in a NETACCESS security zone.
- On an IPv6-enabled stack, the special address IN6ADDR_ANY (::) must be in a zone with the same security label as the TCP/IP job.
- On an IPv6-enabled stack, every INTERFACE statement must have a manually configured INTFID.
- On an IPv6-enabled stack, every INTERFACE that supports autoconfiguration must have at least one manually configured IPADDR.
- On an IPv6-enabled stack with dynamic XCF, DYNAMICXCF INTFID must be configured on the IPCONFIG6 statement.
- Every home address on the stack must be in a NETACCESS security zone.
- Every home address that is not a VIPA must be in a zone with the same security label as the TCP/IP job.
- Every home address that is a VIPA must be in a zone with an equivalent security label as the TCP/IP job.
- On a restricted stack, VIPAs must not be in a security zone with the SYSMULTI security label.

There are several consistency checks that are not performed by the stack. These remain the responsibility of the system security administrator:

- The NetAccess zone definitions must agree across stacks. It is recommended that common definitions from a shared data set be included in all stack profiles.
- Subnetworks or networks are entirely contained in a single NetAccess security zone. NetAccess checks for authorization to a specific destination address, but broadcast packets are delivered to all addresses within a subnetwork or network.

Addresses owned by multilevel secure stacks are the only addresses that can be safely configured into a network security zone different than their subnetwork or network.

- RACF definitions must be consistent across RACF data sets. It is recommended that installations use RACF facilities to manage this consistency. SECDATA and SECLABEL class profile consistency is critical to preserve the meaning of SECLABEL profile names. It is suggested that installations define SERVAUTH profiles for all zones that are generic across all systems and stacks, except those containing INADDR_ANY and LOOPBACK.

Continue making changes to either PROFILE.TCPIP statements or RACF profiles until no consistency errors are reported. Then, specify GLOBALCONFIG MLSCHKTERMINATE in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command. At that point, it is safe to set NOMLQUIET and restart production workloads.

Chapter 5. Customization

Before you begin customizing, it is assumed that you know what configuration data sets are used by the TCP/IP address space, their search order, and considerations for what type of TCP/IP stack you will be running in your environment (for example, Enterprise Extender (EE) and multiple stacks). See Chapter 2, "Configuration overview," on page 11 for this information.

After reading this chapter, you will know how to configure and start syslogd and the TCP/IP stack. You should understand the relationships of TCP/IP configuration files as they apply to the TCP/IP address space. The four main configuration files that you will be working with are:

- TCPIP.DATA
- PROFILE.TCPIP
- HOSTS.LOCAL
- ETC.IPNODES

You should be able to use the following commands to verify customization:

TSO PING, z/OS UNIX oping, and z/OS UNIX ping
: Sends IP datagrams to a specified destination host, requesting a reply, and measures the round trip time. This helps you to verify the interfaces defined to the TCP/IP address space.

TSO NETSTAT, z/OS UNIX onetstat, and z/OS UNIX netstat
: Queries TCP/IP about the network status of the local host. With NETSTAT, you can verify most TCP/IP customization values that can be set from the PROFILE.TCPIP.

TSO HOMETEST
: Verifies your host name and address configuration.

TSO TRACERTE, z/OS UNIX otracert, and z/OS UNIX traceroute
: Displays the route that a packet takes to reach a requested destination.

Configuring the syslog daemon (syslogd)

Configuration statements

The syslogd processing is controlled by a configuration file called /etc/syslog.conf, in which you define logging rules and output destinations for error messages, authorization violation messages, and trace data. Logging rules are defined using a *facility* name, a *priority* code, and the user ID and job name of the program that generated the message. The *facility* name and *priority* code are passed on the logging request from an application when it wants to log a message. The user ID and job name are provided by the system. Refer to *z/OS Communications Server: IP Configuration Reference* for more information about logging rules.

As shown in the following sample /etc/syslog.conf file, comments can be added to the configuration file by placing the number (#) character in column one of the comment line. Everything following the number (#) character is treated as a comment. This sample is available in /usr/lpp/tcpip/samples/syslog.conf in the HFS.

```
#  Licensed Materials - Property of IBM
#  5694-A01
#  (C) Copyright IBM Corp. 1992,  2003
#  Status = CSV1R5
#
# /etc/syslog.conf - control output of syslogd
#
# The # sign begins a comment which extends to the end of the line.
#
# Blank lines are ignored.
#
# Rules in this file specify types of messages which syslogd will
# store, and where syslogd will store it.
#
# See IP Configuration Reference for detailed information about
# the syntax.  These comments are meant to provide only a general
# overview.
#
# Four criteria can be used to select messages for processing:
#
# 1) user ID associated with application generating the message
#
#    * can be specified for the user ID if the user ID is not
#    important.
#
# 2) job name of application generating the message
#
#    * can be specified for the job name if the job name is not
#    important.
#
# 3) facility of the message, as specified by the application
#
#    This is user, mail, news, uucp, daemon, auth, cron, lpr, or
#    local0-local7.  Consult the documentation for the application
#    to determine which facility the application specifies.
#
#    A special facility, mark, specifies that syslogd should log
#    mark messages on a regular basis.  These can be used to verify
#    that syslogd was operational during a specific time interval.
#
# 4) priority of the message, as specified by the application
#
#    This is emerg, panic, alert, crit, err, error, warn, warning,
#    notice, info, or debug.
#
#    A special priority, none, specifies that messages with the
#    specified user ID, job name, or facility should not be
#    selected.
#
# These criteria are specified together as
#
#    userid.jobname.facility.priority
#
# or, if user ID and job name are both *, as
#
#    facility.priority
#
# This can be combined in a series as
#
#   userid.jobname.facility.priority;userid.jobname.facility.priority
#
# The criteria for selecting messages for processing are combined
# with a destination, which tells syslogd what to do with selected
# messages.
```

```
#
#   criteria    destination
#
# The destination can be a file, one or more user IDs, SMF, syslogd
# at a remote host, or all logged-in users.
#
# The following example stores messages with facility daemon or
# local1 in the file /directory/logfile.
#
#   daemon.*;local1.*   /directory/logfile
#
# The directory structure used in this sample configuration is
# expected to be created automatically by syslogd, with a new
# directory of log files for each day.  This requires two types
# of configurations outside of the scope of this configuration
# file:
#
# 1) syslogd command-line option
#
#    The syslogd -c command-line option should be enabled, causing
#    syslogd to create log files and directories if they do not
#    already exist.
#
# 2) cron job
#
#    A cron job should be utilized to wake up syslogd at the
#    beginning of each day to switch to new log files in a new
#    directory.  Here is the cron job definition:
#
#      1 0 * * * kill -HUP `cat /etc/syslog.pid`
#
#    This job should be defined for a user ID with UID zero so that
#    it has permissions to send the signal to syslogd.
#
#    See UNIX System Services Planning and UNIX System Services
#    Command Reference for more information about cron.
#
# A sample shell script is provided for removing log files which are
# a specifed number of days old.  It assumes the same directory
# structure which is used in this sample configuration.
#
# All example rules except for the last one are commented-out.  Some
# or all of the example rules will need to be changed for your
# environment.  Each example rule contains an explanation of changes
# which may be required.
#
######################################################################
#
# Write all messages with priority crit or higher to the MVS operator
# console.  See the UNIX System Services Planning manual for more
# information about the /dev/console special file.
#
# *.crit             /dev/console
#
######################################################################
#
# Write all messages from syslogd itself to the file
# /var/log/YYYY/MM/DD/syslogd.log and to the system console.
#
# Notes:
#
# a) If syslogd is invoked as a started task or from a shell script
#    (e.g., /etc/rc) with job name SYSLOGD, the name of the
#    long-running syslogd job is SYSLOGD followed by a digit.
```

```
#
#    If syslogd runs with a different job name on your system, the
#    rule will have to be changed accordingly.
#
# b) During initialization, syslogd writes messages to
#    /dev/console.  These rules cover messages during steady-
#    state.
#
# *.SYSLOGD*.*.*      /var/log/%Y/%m/%d/syslogd
# *.SYSLOGD*.*.*      /dev/console
#
######################################################################
#
# Write all messages from inetd to the log file inetd and to the
# console.
#
# Notes:
#
# a) If inetd is invoked as a started task or from a shell script
#    (e.g., /etc/rc) with job name INETD, the name of the
#    long-running inetd job is INETD followed by a digit.
#
#    If inetd runs with a different job name on your system, the rule
#    will have to be changed accordingly.
#
# *.INETD*.*.*       /var/log/%Y/%m/%d/inetd
# *.INETD*.*.*       /dev/console
#
######################################################################
#
# Write all messages with priority err or higher from applications
# which specify facility "daemon" to the log file daemon.
# Because we chose to log messages from syslogd and inetd separately,
# we'll filter out those messages from this rule using special
# priority none.
#
# Notes:
#
# a) In this example, SYSLOGD followed by some other character is the
#    job name of syslogd.  If it is different on your system, change
#    the rule.
# b) In this example, INETD followed by some other character is the
#    job name of inetd.  If it is different on your system, change the
#    rule.
#
# daemon.err;*.SYSLOGD*.*.none;*.INETD*.*.none /var/log/%Y/%m/%d/daemon
#
######################################################################
#
# Write all messages from applications which specify facility "auth"
# to the log file auth.
#
# auth.* /var/log/%Y/%m/%d/auth
#
######################################################################
#
# Write all messages from applications which specify facility "mail"
# to the log file mail.
#
# mail.* /var/log/%Y/%m/%d/mail
#
######################################################################
#
# Write all messages with priority err and higher from otelnetd and
```

```
# other applications which specify facility "local1" to the log file
# local1.
#
# local1.err       /var/log/%Y/%m/%d/local1
#
#####################################################################
#
# Write all messages from otelnetd and other applications which
# specify facility "local1" when running as user SMITH to the log file
# local1.smith.  This could be useful if, for example, otelnetd traces
# need to be collected for a problem which user SMITH is experiencing
# and you do not wish to collect otelnetd traces from all user IDs.
#
# SmITh.*.local1.*  /var/log/%Y/%m/%d/local1.smith
#
#####################################################################
#
# Write all messages with priority err and higher to SMF.  These will
# be stored in SMF record type 109.  SMF must be active and
# configured to accept record type 109.  The user ID associated with
# syslogd must have read access to BPX.SMF.  See UNIX System Services
# Planning for more information about BPX.SMF.
#
# *.err        $SMF
#
#####################################################################
#
# Write all messages with priority crit and higher to the syslogd on
# host 192.168.1.9. The host may be specified by IPv4 address, by IPv6
# address, or by a name that resolves to an IPv4 or IPv6 address.
#
# *.crit      @192.168.1.9
#
#####################################################################
#
# Write all messages with priority err and higher to log file errors.
#
# THIS EXAMPLE STATEMENT IS UNCOMMENTED.
#
*.err            /var/log/%Y/%m/%d/errors
#
```

Starting and stopping syslogd

Syslogd can be started only by a task or user with superuser authority.

If there is no TCP/IP transport active when syslogd starts or if TCP/IP is recycled, syslogd will establish or reestablish communication with TCP/IP when it becomes available.

Following is the syntax for the syslogd command:

syslogd [−f *conffile***] [−i][−u[−c[−d][−m** *markinterval***] [−p** *logpath***]**

syslogd recognizes the following options:

-f Configuration file name.

-i Do not receive messages from the IP network.

-u For records received over the AF_UNIX socket (most messages generated on the local system), include the user ID and job name in the record. In this case, a forward slash, the user ID, and the job name will follow the

local host name for messages received over the AF_UNIX socket. The forward slash, which immediately follows the local host name, can be used to determine whether or not the user ID and job name is being recorded. If not recorded, a blank immediately follows the local host name. When user ID or job name is not available, *N/A* will be written in the corresponding field.

-c Create log files and directories automatically.

-d Run syslogd in debugging mode (see "Diagnosing syslogd configuration problems" on page 183 for more information).

-m Number of minutes between mark messages. The default value is 20 minutes. The following rule must be coded for each logfile that you want a mark record recorded in: mark.info.

-p Path name of z/OS UNIX character device for the datagram socket. The default value is /dev/log.

Note: This option is not used frequently. If you selected the -p option, syslogd will not function properly.

To specify the job name and pass the appropriate environment variables to the syslogd process, start syslogd using a shell script such as the following:

```
#
# Start the syslog daemon
#

export _BPX_JOBNAME='syslogd'
/usr/sbin/syslogd -f /etc/syslog.conf &
```

You can execute this shell script directly from the /etc/rc file to start syslogd at z/OS UNIX initialization.

If an incorrect argument or number of arguments is entered, syslogd exits and the return code is 1. In all other situations in which syslogd exits, the return code is 0.

To terminate syslogd, send a SIGTERM signal.

```
kill -s SIGTERM <PID>
```

To force syslogd to reread its configuration file and activate any modified parameters without stopping, send a SIGHUP signal. syslogd will continue to append log messages to the files you specify in /etc/syslog.conf.

```
kill -s SIGHUP <PID>
```

The syslog daemon stores its process ID in the /etc/syslog.pid file so that it may be used to terminate or reconfigure the daemon. For syslogd to successfully create this file, you must define the syslogd user ID as UID=0.

Note: If the BPX.SMF facility is defined and SMF records are to be written by syslogd, the user ID with which syslogd runs must also be permitted to SAF resource BPX.SMF. See SEZAINST(EZARACF) for more information.

Messages are read from the UNIX domain datagram socket and, unless the -i command line option is specified, the IPv4 (AF_INET) or IPv6 (AF_INET6) Internet domain datagram socket. Kernel messages are not logged by syslogd in z/OS UNIX.

Note: For more information about the facilities used by z/OS Communications Server functions, see Table 4 on page 57.

Offloading log files

z/OS Communications Server includes a syslogd configuration file in /usr/lpp/tcpip/samples/syslog.conf, a REXX program for removing old log files in /usr/lpp/tcpip/samples/rmoldlogs, and a JCL procedure for starting syslogd in SEZAINST(syslogd). These are intended to be used together, though each may need to be customized for your installation.

The sample syslogd configuration file is installed in /usr/lpp/tcpip/samples/syslog.conf. It can be copied to /etc/syslog.conf after customization. If it is copied somewhere else, the syslogd -f command-line option must be used to tell syslogd where to find the configuration file.

The sample REXX program for removing old log files is installed in /usr/lpp/tcpip/samples/ezaslrol. It can be copied to an installation-defined directory after customization. The sample JCL procedure can be copied to an installation-defined library after customization.

The sample configuration uses date stamps in the names of directories of log files to organize log files by year (%Y), month (%m), and day (%d) as follows:

```
*.err      /var/log/%Y/%m/%d/errors
```

Log files for February 14, 2001, for example, would be stored in directory /var/log/2001/02/14. Variable substitution occurs using the Language Environment C function strftime(). Variables are case sensitive. For more information and a complete list of variables, refer to *z/OS XL C/C++ Run-Time Library Reference*.

A cron job should be used to send the SIGHUP signal to syslogd every day at midnight so that it switches to a new set of files. The cron job should be created for a user ID with UID 0. The definition of the cron job is:

```
0 0 * * * kill -HUP `cat /etc/syslog.pid`
```

Tip: An accent mark (`) is used in the definition above, not a single quotation mark.

The log file names vary based on the day, so sending SIGHUP to syslogd after the day changes causes syslogd to create new files.

Because some messages sent just after midnight may be logged by syslogd before it processes the SIGHUP signal, it is possible that a few messages sent after midnight will be stored in the log files for the previous day.

The sample REXX program can be run daily to remove all log files older than the number of days specified in the program. Comments in the REXX program describe how to configure the number of days. The definition of a cron job to run the REXX program every day at 1:00 A.M. is:

```
0 1 * * * localdir/ezaslrol
```

localdir is the name of the installation-defined directory where the customized version of /usr/lpp/tcpip/sample/ezaslrol was copied.

Using syslogd for z/OS UNIX application programs

You can use the logging facilities of the syslogd server with your z/OS UNIX application programs. Include the syslog.h header file with C programs so that they can open a log facility, send log messages to syslogd, and close the facility:

```
#include <syslog.h>

1 openlog("oec", LOG_PID, LOG_LOCAL0);
2 syslog(LOG_INFO, "Hello from oec");
3 closelog();
```

1 Open a log facility with the name of local0. Prefix each line in the log file with the program name (oec) and the process ID.

2 Log an info priority message with the specified content.

3 Close the log facility name.

The preceding statements created the following line in the log file:

```
May 26 11:27:51 mvs18oe oec[3014660]: Hello from oec
```

For more information about the syslog function, refer to *Advanced Programming in the UNIX Environment*, published by Addison-Wesley or *z/OS XL C/C++ Run-Time Library Reference*.

Usage notes

- syslogd can run swappable or nonswappable. When an application makes an address space nonswappable, it might convert additional real storage in the system to preferred storage. Because preferred storage cannot be configured offline, allowing syslogd to run in a nonswappable state can reduce the installation's ability to reconfigure storage in the future. Use the following guidelines to set the desired state:
  - If the FACILITY class resource BPX.STOR.SWAP is not defined to the system:
    - syslogd will run nonswappable.
    - syslogd cannot be prevented from running nonswappable.
  - If the FACILITY class BPX.STOR.SWAP resource is defined to the system with UACC(NONE):
    - syslogd will run swappable by default (no access to BPX.STOR.SWAP).
    - syslogd can run nonswappable (given at least READ access to BPX.STOR.SWAP).
  - To define the FACILITY class resource BPX.STOR.SWAP issue the following commands:

    ```
    RDEFINE FACILITY BPX.STOR.SWAP UACC(NONE)
    SETROPTS RACLIST(FACILITY)REFRESH
    ```
- If you want syslogd to receive log data from remote syslogd servers, ensure that syslogd can bind to UDP port 514 by reserving that port for the syslogd job in your PROFILE.TCPIP data set. Ensure that the syslog service is defined in your services file or data set (for example, /etc/services). The following example port reservation in PROFILE.TCPIP assumes that syslogd runs as job syslogd1:

  ```
  PORT
  ...
  514 UDP syslogd1     ;syslogd daemon
  ...
  ```

  The following example shows the services file or data set file entry:

```
syslog     514/udp
```

- Configuration file errors are written to the operator console because initialization is not complete until the entire configuration file has been read.
- Facility mark is not affected by the *.priority usage. Mark messages are written only to the destinations of rules that specify mark.info.
- If a mark interval of zero minutes is specified, mark messages will be written every thirty seconds.

Diagnosing syslogd configuration problems

syslogd supports a debug mode, which is selected using the -d command-line option. In this debug mode, syslogd does not run as a daemon, but instead runs in the foreground and writes a large number of trace messages to STDOUT. These messages can be used to diagnose problems in the syslogd configuration or to collect documentation when reporting a syslogd problem to IBM support.

Note: Do not use the -d option for normal operations.

If you are running syslogd in batch with -d, debug output is written to SYSPRINT, SYSTERM, or SYSERR, whichever is found first. The sample syslogd procedure SEZAINST(syslogd) defines SYSPRINT so that debug messages are stored in the job output.

Use caution using -d when syslogd is started from /etc/rc. If -d is used in this way, the shell background character (&) must be used to run syslogd in the background. Otherwise, /etc/rc does not end and UNIX System Services initialization does not complete.

Also, use caution when using -d along with a port reservation statement for the syslogd port (UDP port 514) in the TCP/IP profile. The job name of syslogd might differ based on whether or not the -d option was specified. If a port reservation statement is coded based on the job name that syslogd uses without the -d option, syslogd might not be able to bind to the port when run with -d. When using debug mode with a port reservation statement for the wrong job name, the bind() error can be ignored or the -i command-line option can be specified along with -d so that syslogd will not get a UDP socket.

Configuring TCPIP.DATA

The TCPIP.DATA configuration data set is the anchor configuration data set for the TCP/IP stack and all TCP/IP servers and clients running in z/OS. With a z/OS TCP/IP stack, you can define the TCPIP.DATA parameters in an HFS file or in an MVS data set. The TCPIP.DATA configuration data set is read during initialization of all TCP/IP server and client functions. All functions must access this data set in order to find basic configuration information, such as the name of the TCP/IP address space, the TCP/IP host name, and the data set prefix to use when searching for other configuration data sets. The TCPIP.DATA file contains the following major groups of configuration parameters:

- Application operating characteristics
- Resolver operating characteristics
- Socket library diagnostic data statements
- Resolver diagnostic data statements

Use of TCPIP.DATA and /etc/resolv.conf

The TCPIP.DATA data set is also known as one of the resolver configuration data sets. In fact, this name is now more commonly used to refer to this important file in the UNIX System Services environment because the socket library contains a component called the resolver. In a UNIX system, you use the /etc/resolv.conf file for the same purpose as you use TCPIP.DATA in your MVS system.

TCPIP.DATA specifies the name of the TCP/IP address space. Because the data set search order can vary, your installation will determine which data set you can use. See Chapter 2, "Configuration overview," on page 11 for search order, data set, and file retrieval information.

If you use TCPIP.DATA, it can be shared between multiple systems with a system name. But, if TCPIP.DATA is allocated via SYSTCPD DD and an application forks, any allocations from the parent of SYSTCPD are lost to the child process.

In z/OS UNIX System Services, each application can have its own environment variable, RESOLVER_CONFIG='xxx'. There are no concerns for forked child processes; however, this means that you cannot share the same data set or file among multiple systems.

Creating TCPIP.DATA

Create a TCPIP.DATA file by copying the sample provided in SEZAINST(TCPDATA) and modifying it to suit your local conditions.

Allocate this data set with either sequential (PS) or partitioned (PO) organization, a fixed (F) or fixed block format (FB), a logical record length (LRECL) between 80 and 256, and any valid block size for a fixed block. This file can also be the HFS file /etc/resolv.conf, or an HFS file that is pointed to by either the environment variable RESOLVER_CONFIG or the SYSTCPD DD in a JCL procedure. If you have an HFS file, the maximum line length can be 256. The environment variable RESOLVER_CONFIG can also point to an MVS data set or PDS.

You can use any name for the TCPIP.DATA data set if you access it using the //SYSTCPD DD statement, or use ENVAR to set RESOLVER_CONFIG, in the JCL for all the servers, logon procedures, and batch jobs that execute TCP/IP functions. If you are not using the //SYSTCPD DD statement, the environment variable, or /etc/resolv.conf, then the data set name must conform to the conventions described in "Configuration files for the TCP/IP stack" on page 36. Another alternative is to use the well-known data set name SYS1.TCPPARMS(TCPDATA). You will issue the HOMETEST command with TRACE RESOLVER activated to verify the actual data set name the system finds for TCPIP.DATA later in this chapter. However, because HOMETEST is an MVS sockets application, it does not use RESOLVER_CONFIG or /etc/resolv.conf in its search order. For this reason, it is recommended that /etc/resolv.conf and TCPIP.DATA contain exactly the same information or consider using the resolver GLOBALTCPIPDATA setup statement.

Rule: Since TCPIP.DATA statements might need to be read and used multiple times by the resolver, the FREE=CLOSE JCL parameter should not be used when allocating SYSTCPD. To allow TCPIP.DATA statements to be changed while still allocated for long running programs, consider using a member of an MVS partitioned data set instead of an MVS sequential data set. For these long running applications, the resolver MODIFY REFRESH command should then be used to indicate that TCPIP.DATA statements have been changed.

TCPIP.DATA statements

Each configuration statement can be preceded by an optional *system_name*. This permits configuration information for multiple systems to be specified in a single *hlq*.TCPIP.DATA data set. The *system_name* is matched against the name of the system on which you are running. The name of the system is taken from the name in the IEFSSNxx member of parmlib, which is the third parameter of the VMCF line.

The statements are processed in the order they appear in the data set. The following rules apply to this processing:

- If the *system_name* does not match the name of the system, the configuration statement is ignored.
- If *system_name* is blank, the configuration statement is in effect on every system.
- If the *system_name* matches the host's name, the configuration statement that follows it is in effect.
- The ***last*** statement that matches is effective.

For example, if you have the following three TCPIPJOBNAME statements, MVS6 would look for a TCP/IP cataloged procedure named TCPBTA2, MVSA would look for TCPV3, and all other systems would look for TCPMCWN.

```
      TCPIPJOBNAME TCPMCWN
MVS6: TCPIPJOBNAME TCPBTA2
MVSA: TCPIPJOBNAME TCPV3
```

But if you reversed the order, all systems would try to find the procedure named TCPMCWN.

```
MVS6: TCPIPJOBNAME TCPBTA2
MVSA: TCPIPJOBNAME TCPV3
      TCPIPJOBNAME TCPMCWN
```

A sample TCPIP.DATA data set (TCPDATA) can be found in SEZAINST. For detailed information on each of the statements, refer to the *z/OS Communications Server: IP Configuration Reference*.

Using MVS system symbols in TCPIP.DATA

For ease of management when configuring a complex environment, use MVS system symbols such as &SYSCLONE, &SYSNAME, and &SYSPLEX. The resolver translates these symbols as it reads TCPIP.DATA, reducing the number of TCPIP.DATA files that must be maintained in a multisystem environment. For detailed information about symbols and how to define them, refer to *z/OS MVS Initialization and Tuning Reference*.

Configuring PROFILE.TCPIP

During TCP/IP address space initialization, a configuration profile data set (PROFILE.TCPIP) is read that contains system operation and configuration parameters. A sample data set, SEZAINST(SAMPPROF), can be copied and modified for use as your default configuration profile.

If you are not familiar with the search order for this data set, see "PROFILE.TCPIP search order" on page 36 for information about understanding data set search orders. Refer to *z/OS Communications Server: IP Configuration Reference* for the complete statement syntax and descriptions of the configuration statements.

For ease of management when configuring a complex environment, you can use one of the following PROFILE.TCPIP data set features:

- Group related statements into separate files and use the INCLUDE statement in PROFILE.TCPIP to include them in your configuration.
- Use MVS system symbols (such as &SYSCLONE, &SYSNAME, and &SYSPLEX). Because TCP/IP translates these symbols as it reads this file, this feature reduces the number of PROFILE.TCPIP data sets that must be maintained in a multi-TCP/IP environment.

Note: For detailed information about symbols and how to define them, refer to *z/OS MVS Initialization and Tuning Reference*.

The PROFILE data set contains the following major groups of configuration parameters:

- TCP/IP operating characteristics
- TCP/IP physical characteristics
- TCP/IP reserved port number definitions (application configuration)
- TCP/IP network routing definitions
- TCP/IP diagnostic data statements

This chapter discusses the first three areas of configuration. For routing configuration information, see Chapter 6, "Routing," on page 217. For information about configuring diagnostic statements, see *z/OS Communications Server: IP Diagnosis Guide*.

Changing configuration information

If you want to change the TCP/IP configuration without stopping and starting the TCP/IP address space, you can dynamically change many of the TCP/IP configuration options established by the PROFILE.TCPIP data set. To do this, put the changed configuration statements in a separate data set and process it with the VARY TCPIP,,OBEYFILE command.

For more information about VARY TCPIP, refer to *z/OS Communications Server: IP System Administrator's Commands*. Also, see the Modifying section in each configuration statement in *z/OS Communications Server: IP Configuration Reference* for a description of how to dynamically change the information for that configuration statement.

Note: If you attempt to edit PROFILE.TCPIP while TCPIP is active, and PROFILE.TCPIP is defined in the TCPIP PROC as a sequential data set (for example, //PROFILE DD DISP=SHR,DSNAME=TCPIP.PROFILE.TCPIP), the *Dataset in use* message might be displayed. To avoid this, specify FREE=CLOSE, as follows:

```
//PROFILE  DD DISP=SHR,DSNAME=TCPIP.PROFILE.TCPIP,FREE=CLOSE
```

This allows you to edit the profile while TCP/IP is active. Typically, when TCP/IP starts, it keeps the PROFILE allocated and does not release the allocation until the end of the step (in this case, the end of the job). If you specify FREE=CLOSE, the release occurs once the data set is read. MVS releases the enqueue on the PROFILE, which allows you to edit it.

If the PROFILE is a member of a PDS, [for example, SYS1.TCPPARMS(PROFILE)], FREE=CLOSE is not needed.

Setting up TCP/IP operating characteristics in PROFILE.TCPIP

Figure 33 on page 188 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. This sample can be copied from SEZAINST(SAMPPROF). Figure 33 on page 188 includes the portion of the sample that shows how to set up TCP/IP operating characteristics. Descriptions for the statements follow Figure 33 on page 188. For more information about any of these statements, refer to *z/OS Communications Server: IP Configuration Reference*. For information specific to IPv6 support, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

```
; ====================================================================
; General TCP/IP address space configuration
; ====================================================================
;
; ARPAGE: Specifies the number of minutes between creation or
;   revalidation of an LCS ARP table entry and the deletion of the
;   entry.
;
ARPAGE 20
;
;
; GLOBALCONFIG: Provides settings for the entire TCP/IP stack
;
GLOBALCONFIG NOTCPIPSTATISTICS
;
; IPCONFIG: Provides settings for the IPv4 IP layer of TCP/IP.
;
; Example IPCONFIG for single stack/single system:
;
| IPCONFIG DATAGRAMFWD SYSPLEXROUTING
;
; Example IPCONFIG for automatic activation of inter-stack dynamic XCF
;   and Same Host (IUTSAMEH) links
;
; IPCONFIG DYNAMICXCF 201.1.10.10 255.255.255.0 2
| ;
| ; Example IPCONFIG for IPSECURITY support:
| ;
| ; IPCONFIG IPSECURITY
| ;
;
; IPCONFIG6: Provides settings for the IPv6 IP layer of TCP/IP.
;
; IPCONFIG6 DATAGRAMFWD SOURCEVIPA
;
; Example IPCONFIG6 for automatic activation of inter-stack dynamic XCF
;   and Same Host (IUTSAMEH) links
;
; IPCONFIG6 DYNAMICXCF 2001::151:0000
;
; SOMAXCONN: Specifies maximum length for the connection request queue
;   created by the socket call listen().
;
SOMAXCONN 10
;
;
; TCPCONFIG: Provides settings for the TCP layer of TCP/IP.
;            RESTRICTLOWPORTS limits access to ports below 1024
;            to authorized applications.  Applications can be
;            authorized to low ports in three ways:
;             - via PORT or PORTRANGE with the appropriate jobname
;                or wildcard jobname
;             - APF authorized
;             - superuser
;
```

```
TCPCONFIG TCPSENDBFRSIZE 16K TCPRCVBUFRSIZE 16K SENDGARBAGE FALSE
TCPCONFIG RESTRICTLOWPORTS
;
;
; UDPCONFIG: Provides settings for the UDP layer of TCP/IP
;            RESTRICTLOWPORTS limits access to ports below 1024
;            to authorized applications.  Applications can be
;            authorized to low ports in three ways:
;             - via PORT or PORTRANGE with the appropriate jobname
;                or wildcard jobname
;             - APF authorized
;             - superuser
;
UDPCONFIG RESTRICTLOWPORTS
;
; SRCIP allows the substitution of the source IP address on a
; jobname specific basis for client applications which specify
; inaddr_any for the source IP address. This may be done when
; an application issues an explicit bind() call with inaddr_any
; or when it bypasses issuing an explicit bind() call and
; issues a connect().
;
;SRCIP
; JOBNAME USER15  9.43.242.5
; JOBNAME USER*   9.43.242.4
; JOBNAME USER15  2001::092B:F203
; JOBNAME JOB*    ETHER1
; JOBNAME *       9.43.242.3
;ENDSRCIP
```

Figure 33. Example of TCP/IP operating characteristics in PROFILE.TCPIP

The following section explains the grouping of statements shown in Figure 33.

ARPAGE

Use ARPAGE to set the number of minutes between a revalidation and deletion of ARP table entries for LCS devices. An installation that wants to describe this value in seconds versus minutes should use the IPCONFIG ARPTO statement.

Note: The ATM ARP requests are controlled via the ATMLIS statement, and the MPCIPA and MPCOSA ARP requests are not controlled by the TCP/IP address space.

GLOBALCONFIG

Use GLOBALCONFIG to print out several counters in text format. These counters include number of TCP retransmissions and total number of TCP segments sent from the TCPIP system. Most installations will use the SMF facility of MVS to collect these counters in a more standard way. Use the ECSALIMIT parameter on the GLOBALCONFIG statement to limit TCP/IP's use of common storage. The POOLLIMIT parameter can be used to limit TCP/IP's use of private storage pools.

IPCONFIG

Use IPCONFIG to configure various settings of the IP layer of TCP/IP. Use ARPTO to specify the ARP time out value in seconds for LCS devices. See page 188 for more information.

Use CLAWUSEDOUBLENOP on vendor devices that document the need for double NOPs on each CCW.

Use DATAGRAMFWD if this TCP/IP is to be a router and needs to forward datagrams to other routers. Use IGNOREREDIRECT when a dynamic routing program is used and ICMP redirect packets are to be ignored by the TCP/IP address space. MULTIPATH is used to inform TCP/IP how to distribute traffic across equal cost routes.

Use FIREWALL to restrict this host to be a network firewall. To make IPSEC tunnels associated with Dynamic VIPA addresses eligible for distribution, if the VIPA addresses are being distributed and are eligible to be moved during VIPA takeover or giveback, add the DVIPSEC keyword to FIREWALL.

SOURCEVIPA enables interface fault tolerance for z/OS clients that establish outbound connections. When SOURCEVIPA is set, outbound datagrams use the corresponding virtual IP address (VIPA) in the HOME list instead of the physical interfaces IP address. SOURCEVIPA has no effect on RIP servers such as NCPROUTE or OMPROUTE.

TCPSTACKSOURCEVIPA allows z/OS clients to specify a sysplex wide source IP address for TCP connections. When TCPSTACKSOURCEVIPA is set, outbound TCP datagrams use the IP address specified in the TCPSTACKSOURCEVIPA statement instead of static VIPA addresses or physical interface addresses.

Use SYSPLEXRouting to communicate interface changes within a sysplex domain to the workload manager (WLM). DYNAMICXCF allows the cross communication facility within a sysplex to dynamically generate connections within a sysplex domain. If DYNAMICXCF is used with a dynamic routing program like OMPROUTE, the BSDROUTINGPARMS and the OMPROUTE configuration files need to be updated with subnet mask and cost information. For more information on additional configuration parameters required, see the usage notes related to the DYNAMICXCF parameter under the IPCONFIG statement in *z/OS Communications Server: IP Configuration Reference*.

Use REASSEMBLYTIMEOUT to specify the TCP/IP reassemble timeout value in seconds, and the TTL specifies the TCP/IP time to live or hop count value.

Use PATHMTUDISCOVERY to indicate to TCP/IP that it is to dynamically discover the path MTU, which is the minimum of MTUs of each hop in the path.

Use STOPONCLAWERROR to indicate to the TCP/IP stack to stop channel programs (HALTIO and HALTSIO) when a device error is detected.

IPCONFIG6

Use IPCONFIG6 to update the IP layer of TCP/IP with information that pertains to IPv6.

Use DATAGRAMFWD to enable the transfer of data between networks.

Use DYNAMICXCF to enable Dynamic XCF support for IPv6.

SOMAXCONN

Use SOMAXCON to specify the maximum number of sockets queued on a listener.

SRCIP

Use of the SRCIP and ENDSRCIP profile statement block allows a job to use a unique IP address as its source address for outbound TCP connections. When a source IP address has been designated for a job and

the source IP address exists at the time the outbound TCP connection is initiated, this source IP address is used, overriding any other source IP address selection method. This source address selection is performed for jobs that issue a connect() and have not previously bound the socket to an IP address, or for those that bind() to inaddr_any before issuing the connect() call.

TCPCONFIG

Use TCPCONFIG to configure various settings of the TCP protocol layer. If a keep-alive value other than 120 minutes is needed by an installation, use the INTERVAL statement to change the default keep-alive value. FINWAIT2TIME can be used to specify a different timeout value for a TCP Connection which is in a FINWAIT2 state. SENDGARBAGE will cause the keep-alive packet to contain one byte of random data and an incorrect sequence number, assuring that the data is not accepted by the remote TCP. The TCPTIMESTAMP option can be used to choose whether or not to participate in timestamp negotiation.

The behavior of acknowledgments and delaying their transmission can be altered by using the DELAYACKS statement.

If RESTRICTLOWPORTS is specified, only applications that meet at least one of the following criteria are allowed to bind to low ports (1–1023):

- The port is reserved for the application via the PORT or PORTRANGE statement.
- The application runs with APF authorization.
- The application runs with effective POSIX UID zero.

If an installation wants to control TCP buffering (to limit storage usage or to manage large bandwidth devices), use the TCPSENDBFRSIZE, TCPRCVBUFRSIZE, and TCPMAXRCVBUFRSIZE parameters.

UDPCONFIG

Use UDPCONFIG to configure various settings of the UDP protocol layer. NOUDPCHKSUM can be used to eliminate check summing overhead for IPv4 UDP packets. This option is ignored for UDP datagrams flowing over an IPv6 network, as UDP Checksum is a required function on an IPv6 network.

If RESTRICTLOWPORTS is specified, only applications that meet at least one of the following criteria are allowed to bind to low ports (1–1023):

- The port is reserved for the application via the PORT or PORTRANGE statement.
- The application runs with APF authorization.
- The application runs with effective POSIX UID zero.

If an installation wants to control UDP buffering (to limit storage usage or to manage large bandwidth devices), use the UDPSENDBFRSIZE and UDPRCVBUFRSIZE parameters. UDPQUEUELIMIT can be used to set a queue limit for UDP. This is useful for installations that want to limit the size of the queue of UDP datagrams that an application can have waiting before the TCP/IP address space starts discarding them.

Setting up physical characteristics in PROFILE.TCPIP

Figure 34 on page 196 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. Figure 34 on page 196 includes the portion

of the sample that shows how to set up physical characteristics. This sample can be copied from SEZAINST(SAMPPROF). Following Figure 34 on page 196, several of the statements that are used to set up physical characteristics in PROFILE.TCPIP are described. For more information about any of these statements, or information on statements not described, refer to *z/OS Communications Server: IP Configuration Reference*. For information specific to IPv6 support, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

```
; ====================================================================
; Hardware definitions
; ====================================================================
;
; DEVICE: Defines name (and sometimes device number) for various types
;   of network devices for IPv4 only
; LINK: Defines a network interface to be associated with a particular
;   device.  For IPv4 only.
; INTERFACE: Defines an IPv6 interface.
;
;
; DEVICE and LINK for CTC devices
;
;DEVICE  CTC1    CTC  D00  AUTORESTART
;LINK    CTCD00  CTC  0  CTC1
;
; DEVICE and LINK for HYPERchannel A220 devices:
;
;DEVICE  HCH1    HCH  E00  AUTORESTART
;LINK    HCHE00  HCH  1  HCH1
;
; DEVICE and LINK for LAN Channel Station and OSA devices:
;   DEVICE: Defines name and hexadecimal device number for an IBM 8232
;     LAN channel station (LCS) device, and IBM 3172 Interconnect
;     Controller, an IBM 2216 Multiaccess Connector Model 400,
;     an IBM FDDI, Ethernet, or Token Ring OSA, or an IBM ATM OSA-2
;     in LAN emulation mode
;   LINK: Defines a network interface link associated with an LCS
;     device; may be for Ethernet Network, Token-Ring Network or
;     PC Network, or FDDI.
;
;   Example: LCS1 is a 3172 model 1 with a Token Ring and Ethernet
;    adapter
;
;DEVICE  LCS1  LCS  BA0  AUTORESTART
;LINK    TR1   IBMTR  0  LCS1
;LINK    ETH1  ETHERNET  1  LCS1
;
;   Example: LCS2 is a 3172 model 2 with a FDDI adapter
;
;DEVICE  LCS2   LCS  BE0  AUTORESTART
;LINK    FDDI1  FDDI  0  LCS2
;
; DEVICE and LINK for MPCIPA QDIO Devices:
;
;   Example: MPCIPA1 is either an IBM OSA-Express Gigabit Ethernet
;    or QDIO Fast Ethernet adapter
;
;DEVICE  MPCIPA1     MPCIPA  NONROUTER AUTORESTART
;LINK    MPCIPALINK1 IPAQENET  MPCIPA1
;
;   Example: MPCIPA2 is either an IBM OSA-Express Gigabit Ethernet
;    or QDIO Fast Ethernet adapter, configured as the PRIMARY router
;
;DEVICE  MPCIPA2     MPCIPA  PRIROUTER AUTORESTART
;LINK    MPCIPALINK2 IPAQENET  MPCIPA2
;
; DEVICE and LINK for HiperSockets CHPID FE
```

```
;
;DEVICE  IUTIQDFE      MPCIPA    AUTORESTART
;LINK    HIPERSOCKFE4  IPAQIDIO  IUTIQDFE
;
;   DEVICE and LINK for MPCPTP devices:
;
;DEVICE  MPCPTP1    MPCPTP   AUTORESTART
;LINK    MPCPTPLINK MPCPTP   MPCPTP1
;
;    DEVICE and LINK for CLAW devices:
;
;DEVICE  RS6K    CLAW  6B2  HOST  PSCA  NONE  26 26 AUTORESTART
;LINK    IPLINK1 IP  0  RS6K
;
;   DEVICE and LINK for SNA LU0 links:
;
;DEVICE  SNALU0  SNAIUCV   SNALINK  LU000000  SNALINK  AUTORESTART
;LINK    SNA1    SAMEHOST  1  SNALU0
;
;    DEVICE and LINK for SNA LU 6.2 links:
;
;DEVICE  SNALU621 SNALU62 SNAPROC AUTORESTART
;LINK    SNA2     SAMEHOST  1  SNALU621
;
;    DEVICE and LINK for X.25 NPSI connections:
;
;DEVICE  X25DEV   X25NPSI  TCPIPX25  AUTORESTART
;LINK    X25LINK  SAMEHOST  1  X25DEV
;
;    DEVICE and LINK for 3745/46 Channel DLC Devices:
;
;DEVICE  CDLC1    CDLC  C00 AUTORESTART
;LINK    CDLCLINK CDLC  1  CDLC1
;
;    DEVICE and LINK for MPC OSA Fast Ethernet Devices:
;
;DEVICE  MENET1   MPCOSA   AUTORESTART
;LINK    ENETLINK OSAENET 0 MENET1
;
;    DEVICE and LINK for MPC OSA FDDI Devices:
;
;DEVICE  MFDDI1   MPCOSA   AUTORESTART
;LINK    FDDILINK OSAFDDI 0 MFDDI1
;
; ----------------------------------------------------------------------
; Virtual device definitions
; ----------------------------------------------------------------------
;
; DEVICE and LINK for Virtual Devices (VIPA):
;
;   DEVICE  VDEV1   VIRTUAL  0
;   LINK    VLINK1  VIRTUAL  0  VDEV1
;
; Dynamic Virtual Devices can be defined on this system.  This system
; can serve as backup for Dynamic Virtual Devices on other systems.
; A predefined range will allow Dynamic Virtual Devices to be defined
; by IOCTL or Bind requests.
;
; VIPADYNAMIC
;   Define two dynamic VIPAs on this stack:
;   VIPADEFINE 255.255.255.192 201.2.10.11 201.2.10.12
;
;   Define two IPv6 dynamic VIPAs on this stack:
;   VIPADEFINE DVIPA1 1::1
;   VIPADEFINE DVIPA2 2::2
;
;   Define Sysplex Distributed statements for these dynamic VIPAs:
```

```
;   VIPADISTribute SYSPLEXPORTS 201.2.10.11 PORT 4011 DESTIP ALL
;   VIPADISTribute TIMEDAFF 200 201.2.10.12 PORT 4012 DESTIP ALL
;   VIPADISTribute DISTMETHOD ROUNDROBIN DVIPA1       DESTIP ALL
;   VIPADISTribute SYSPLEXPORTS DVIPA2      PORT 4013 DESTIP ALL
;
;   Define this stack as backup for these dynamic VIPAs on
;   other TCP/IP stacks:
;   VIPABACKUP 100             201.2.10.13 201.2.10.14
;   VIPABACKUP  80             201.2.10.21 201.2.10.22
;   VIPABACKUP  60             201.2.10.31 201.2.10.33
;   VIPABACKUP  40             201.2.10.32 201.2.10.34
;
;   Define two dynamic VIPA ranges on this stack:
;   VIPARANGE DEFINE 255.255.255.192 201.2.10.192
;   VIPARANGE DAVERANGE1 3::1/100
; ENDVIPADYNAMIC
;
; ---------------------------------------------------------------------
; ATM hardware definitions
; ---------------------------------------------------------------------
;
; ATMLIS: Describes characteristics of an ATM logical IP subnet (LIS).
;
; DEVICE and LINK for ATM devices: (See below)
;
; ATMPVC: Describes a permanent virtual circuit (PVC) to be used by an
;   ATM link.
;
; ATMARPSV: Designates the ATMARP server that will resolve ATMARP
;   requests for a logical IP subnet (LIS).
;
; ATMLIS   LIS1   9.67.100.0 255.255.255.0
; DEVICE   OSA1   ATM  PORTNAME  PORT1
; LINK     LINK1  ATM  OSA1  LIS LIS1
; ATMPVC   PVC1   LINK1
; ATMARPSV ARPSV1 LIS1 PVC PVC1
;
;
; ---------------------------------------------------------------------
; IPv6 INTERFACE statements
; ---------------------------------------------------------------------
;    Virtual interface definitions
;      IPADDR keyword is required for Virtual interfaces
;      Multiple IP addresses can be defined to one interface
;      The prefixes of the IPv6 VIPA addresses should be
;        different than the prefixes used for addresses
;        configured or autoconfigured for real interfaces.
; ---------------------------------------------------------------------
;  INTERFACE VIPAV6 DEFINE
;    VIRTUAL6
;    IPADDR 50C9:C2D4:0:A:9:67:115:66 ; (Global Address)
; ---------------------------------------------------------------------
; To use autoconfiguration, the IPADDR cannot be specified.
; To manally define address(es), use the IPADDR keyword.
; To assign a VIPA address for an interface, use SOURCEVIPAINT
; To have IPv4 and IPv6 share a physical device, define IPv4
;   using DEVICE/LINK/HOME and IPv6 using INTERFACE
; ---------------------------------------------------------------------
;  INTERFACE OSAQDIO26                ; OSA QDIO (Fast Ethernet)
;    DEFINE IPAQENET6
;    PORTNAME OSAQDIO2
;    SOURCEVIPAINT VIPAV6
;    IPADDR 50C9:C2D4:0:1:9:67:115:66 ; (Global Address)
; ---------------------------------------------------------------------
; MPCPTP6 interfaces require manual IPv6 address configuration. If
; you don't code an IPADDR on the MPCPTP6 interface, you'll get only
; the link-local address.  Here, the specified interface ID (INTFID)
```

```
; will be appended to the global prefix, to form the global
; address.
; ---------------------------------------------------------------------
;  INTERFACE MPC1IPV6          ; MPCPTP6 (IPv6 over MPC point to point)
;    DEFINE MPCPTP6
;    TRLENAME TRLE1A
;    INTFID 0012:3456:789A:BCDE
;    IPADDR 2001:0DB8:0:0/64   ; (Global Prefix)
;
; ---------------------------------------------------------------------
; HiperSockets IPv6
; ---------------------------------------------------------------------
;  INTERFACE HIPERSOCKFE6     ; IPAQIDIO6 (HiperSockets IPv6)
;    DEFINE IPAQIDIO6
;    CHPID  FE
;    INTFID 0000:2000:2000:2000
;    IPADDR 50C9:C2D4:0:1:9:67:118:80 ; (Global Address)
;
; ---------------------------------------------------------------------
; To define other Ipv6 Loopback addresses:
; ---------------------------------------------------------------------
;   INTERFACE LOOPBACK6 ADDADDR ::0014:0
; ---------------------------------------------------------------------
; Other device statements
; ---------------------------------------------------------------------
;
;
; TRANSLATE: Indicates a relationship between an internet address and
;   the network address on a specified link.  Only applicable for IPv4
;   devices.
;
; TRANSLATE
;   9.67.43.110   FDDI    FF0000006702   FDDI1
;   9.37.84.49    HCH     FF0000005555   HCHE00
;
;
; =====================================================================
; HOME addresses
; =====================================================================
;
; HOME: Provides the list of home IP addresses and associated link names
;       for IPv4
;
;   - The LOOPBACK statement of 14.0.0.0 should only be used if the
;     installation has applications that require this old loopback
;     address.  The current stack uses 127.0.0.1 as the loopback
;     address.
;
; HOME
;  14.0.0.0       LOOPBACK
;  130.50.75.1    TR1
;  193.5.2.1      ETH1
;  9.67.43.110    FDDI1
;  193.7.2.1      SNA1
;  9.67.113.80    CTCD00
;  9.37.84.49     HCHE00
;  9.67.113.81    MPCIPALINK1
;  9.67.113.82    MPCPTPLINK
;  9.67.113.83    MPCIPALINK2
;  9.67.114.02    IPLINK1
;  9.67.43.03     SNA2
;  9.67.115.85    X25LINK
;  9.67.116.86    VLINK1
;  9.67.117.87    CDLCLINK
;  9.67.100.80    LINK1
;  9.37.112.13    ENETLINK
;  9.37.112.14    FDDILINK
```

```
;  9.67.118.80    HIPERSOCKFE4
;
;
; PRIMARYINTERFACE: Specifies which link is designated as the default
;   local host for use by the GETHOSTID() function.  Only applicable
;   for IPv4 devices.
;
;   - If PRIMARYINTERFACE is not specified, then the first link in
;     the HOME statement is the primary interface, as usual.
;
; PRIMARYINTERFACE TR1
;
; ====================================================================
; Routing configuration
; ====================================================================
; --------------------------------------------------------------------
; Static routing
; --------------------------------------------------------------------
;
; BEGINRoutes: Defines static routes to the IP route table for IPv4
;                  and IPv6
;
;  - Use of BEGINROUTES with OMPROUTE routing daemon is not recommended.
;
BEGINRoutes
;
; Direct Routes - Routes that are directly connected to my interfaces.
;
;      Destination Subnet Mask   First Hop    Link Name Packet Size
;
;ROUTE 130.50.75.0 255.255.255.0 =            TR1       MTU 2000
;ROUTE 193.5.2.0/24               =            ETH1      MTU 1500
;ROUTE 9.67.43.0   255.255.255.0 =            FDDI1     MTU 4000
;ROUTE 193.7.2.2   HOST          =            SNA1      MTU 2000
;
;      Destination Subnet Mask   First Hop  Interface   Packet Size
;
;  ROUTE FE80::1:2:3:4/128         =        OSAQDIO26     MTU 2000
;  ROUTE 2001::0DB8:1/128          =        OSAQDIO26     MTU 2000
;
;
; Indirect Routes - Routes that are reachable through routers on my
;                   network.
;
;      Destination Subnet Mask   First Hop    Link Name Packet Size
;
;ROUTE 193.12.2.0  255.255.255.0 130.50.75.10 TR1       MTU 2000
;ROUTE 10.5.6.4    HOST          193.5.2.10   ETH1      MTU 1500
;
;      Destination Subnet Mask   First Hop  Interface   Packet Size
;
;  ROUTE FEC8::/64           FE80::1:2:3:4    OSAQDIO26   MTU 2000
;
; Default Route - All packets to an unknown destination are routed
;                 through this route.
;
;      Destination               First Hop    Link Name Packet Size
;
;ROUTE DEFAULT                   9.67.43.99   FDDI1     MTU DEFAULTSIZE
;
;      Destination Subnet Mask   First Hop  Interface   Packet Size
;
; ROUTE DEFAULT6             FE80::1:2:3:4  OSAQDIO26   MTU DEFAULTSIZE
ENDRoutes
;
; --------------------------------------------------------------------
; Dynamic routing
```

```
; ----------------------------------------------------------------------
;
; BSDROUTINGPARMS: Defines the characteristics of each link defined at
;   the host.
;
; If not supplied, defaults will be supplied from:
; (1) Static routing definitions in BEGINROUTES
; (2) OMPROUTE configuration (if OMPROUTE is running)
; (3) Stack's interface layer based on hardware capabilites and
;     characteristics of devices and links.
;
;   - OMPROUTE does not require BSDROUTINGPARMS.  However, it will
;     override the parameters with its coded or defaulted values
;     from its configuration.
;
;   - NCPROUTE requires BSDROUTINGPARMS to route Transport PDUs prior
;     to OMPROUTE activation.  If OMPROUTE is also used, the parameters
;     must match the corresponding ones in OMPROUTE configuration for
;     the channel-attached links.
;
; BSDROUTINGPARMS TRUE
;  Link name      MTU    Cost metric  Subnet Mask    Dest address
;   TR1          2000         0       255.255.255.0     0
;   ETH1         1500         0       255.255.255.0     0
;   FDDI1        4000         0       255.255.255.0     0
;   VLINK1    DEFAULTSIZE     0       255.255.255.0     0
;   CTCD00       65527        0       255.255.255.0  9.67.113.90
; ENDBSDROUTINGPARMS
```

Figure 34. Example of physical characteristics in PROFILE.TCPIP

The following section explains several of the statements shown in Figure 34 that are used to set up physical characteristics in PROFILE.TCPIP. For more information about any of these statements, or information on statements not described here, see *z/OS Communications Server: IP Configuration Reference*. For information specific to IPv6 support, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

DEVICE and LINK

Use DEVICE and LINK statements to define each IPv4 network interface to the TCPIP address space. Refer to the *z/OS Communications Server: IP Configuration Reference* for more details about the various network interfaces supported by TCP/IP.

ATM Use the ATM DEVICE and LINK statements to define connectivity to an ATM network. These statements allow for connectivity in either ATM native mode over an ATM virtual circuit (VC) or in ATM LAN Emulation mode.

For ATM native mode, the VC can be either a permanent virtual circuit (PVC) or a switched virtual circuit (SVC). To define a PVC, use the ATMPVC statement. To define SVCs, use the ATMLIS statement to define the ATM logical IP subnet (LIS). Also, for SVCs, use the ATMARPSV statement to define the ATMARP server that will resolve ATMARP requests within the LIS. For ATM LAN emulation mode, the ATM DEVICE and LINK definitions allow you to retrieve SNMP network management data for the device. In this mode, you need to define the device as an LCS.

CDLC The DEVICE CDLC describes the interface between the TCP/IP address space and the 3745/46 devices used.

CLAW Use CLAW DEVICE for RISC System/6000® and SP2.

CTC Use the CTC DEVICE and LINK statements to define connectivity to another z/OS using channel-to-channel.

HYPERchannel A220 DEVICE and LINK
Use the HCH DEVICE and LINK statements to define connectivity via the HYPERchannel A220 adapter.

LAN Channel Station (LCS) DEVICE and LINK
Use the LCS DEVICE and LINK statements to define connectivity to a token-ring, FDDI, or Ethernet LAN. LCS devices can have more than one adapter. Therefore, you can have more than one LINK statement for an LCS DEVICE statement.

In configurations where multiple LCS and/or MPCIPA links onto the same LAN are defined, if the interface targeted by the ARP Request is inactive, one of the other active interfaces on the LAN will automatically take over responsibility for answering ARPs on behalf of the inactive interface. In this way, fault tolerance is achievable on the LAN without requiring a dynamic routing protocol.

TCP/IP supports ARP for VIPAs. In a flat network (one in which traffic flows directly between two endpoints without an intermediate router) using static routing with multiple interfaces onto the same LAN, you can achieve fault tolerance by defining a VIPA in the same subnet as the physical interfaces on the LAN. If a static route specifies a VIPA as the next hop IP address, the host or router will send an ARP for the VIPA. TCP/IP will reply to the ARP with the MAC address of one of the active physical interfaces on that LAN.

MPCIPA
Use the MPCIPA DEVICE and LINK statements to define either of the following:

- LAN connectivity with OSA-Express using the Queued Direct I/O (QDIO) interface
- HiperSockets connectivity

For OSA-Express QDIO, the MPCIPA device name must be the PORT name of the TRLE definition of the QDIO interface as described in *z/OS Communications Server: SNA Resource Definition Reference*. Device specifications for the type of IP routing supported are also specified on the MPCIPA DEVICE statement. These are also described in *z/OS Communications Server: SNA Resource Definition Reference*.

In configurations where multiple LCS and/or MPCIPA links onto the same LAN are defined, if the interface targeted by the ARP Request is inactive, one of the other active interfaces on the LAN will automatically take over responsibility for answering ARPs on behalf of the inactive interface. In this way, fault tolerance is achievable on the LAN without requiring a dynamic routing protocol.

TCP/IP supports ARP for VIPAs. In a flat network (one in which traffic flows directly between two endpoints without an intermediate router) using static routing with multiple interfaces

onto the same LAN, you can achieve fault tolerance by defining a VIPA in the same subnet as the physical interfaces on the LAN. If a static route specifies a VIPA as the next hop IP address, the host or router will send an ARP for the VIPA. TCP/IP will reply to the ARP with the MAC address of one of the active physical interfaces on that LAN.

Use the IPCONFIG DYNAMICXCF statement to cause TCP/IP to automatically define and activate HiperSockets connectivity between each pair of TCP/IP stacks that is in the same CPC in the same z/OS sysplex.

MPCOSA
The MPCOSA DEVICE statements define the MPC OSA Ethernet and FDDI devices.

MPCPTP

MPCPTP can be used to define any of the following:

- A connection to another host over a series of CTCs (in this case, the device name must be the name of a VTAM TRLE).
- An XCF connection to another TCP/IP in the same z/OS sysplex. For an XCF connection, the device name must be the CP name or SSCP name of the target VTAM on the other side of the XCF connection, and the VTAM ISTLSXCF major node must be active to start the device.
- An IUTSAMEH connection (with no need for any I/O devices) to another TCP/IP on the same z/OS system or to VTAM for Enterprise Extender. For an IUTSAMEH connection, the device name must be the reserved name IUTSAMEH. VTAM automatically activates the IUTSAMEH TRLE.

Use the IPCONFIG DYNAMICXCF statement to cause TCP/IP to automatically define and activate XCF connectivity between each pair of TCP/IP stacks in the same sysplex and IUTSAMEH connectivity between multiple TCP/IP stacks on the same z/OS.

SNAIUCV and SNALU62
Use SNAIUCV DEVICE to specify the interface to use for SNA LU0 traffic to the SNALINK started procedures. For example, use this to define the interface between the TCP/IP address space and the SNALINK address space that is using a 3745 running NCPRoute. Similarly, the DEVICE SNALU62 statement defines the interface between the TCP/IP address space and the address space using SNA LU6.2. Refer to *z/OS Communications Server: IP Configuration Reference* for information about how to define multiple LU6.2 connections within the same TCP/IP address space.

X.25 The DEVICE X25DEV defines the interface between the TCP/IP address space and the address space of the X.25 NPSI server.

VIPA and VIPADYNAMIC
Virtual IP Addresses (VIPA) are used to define virtual devices to the TCP/IP address space. There are two types of VIPAs:

- Static
- Dynamic

The static virtual device requires DEVICE and LINK statements to define a device that is always started, can never be stopped, can be known within the network, yet requires no physical adapters. It is very useful to define VIPAs so that if a physical adapter loses its connection to the network, application traffic using the failed physical adapter can be rerouted over another interface to the network. To the network, the VIPA address appears to be one hop away from the TCP/IP address spaces. The network sends and receives datagrams to and from the physical interfaces to get to the VIPA address. For more information about VIPA, see Chapter 7, "Virtual IP Addressing," on page 297.

INTERFACE

Use INTERFACE statements to define each IPv6 network interface to the TCPIP address space. Refer to the *z/OS Communications Server: IP Configuration Reference* for more details about the various network interfaces supported by TCP/IP.

IPAQENET6

Use the IPAQENET6 INTERFACE statement to define IPv6 LAN connectivity through OSA-Express using QDIO.

IPAQIDIO6

Use the IPAQIDIO6 INTERFACE statement to define HiperSockets IPv6 connectivity.

Use the IPCONFIG6 DYNAMICXCF statement to cause TCP/IP to automatically define and activate HiperSockets IPv6 connectivity between each pair of TCP/IP stacks that is in the same CPC in the same z/OS sysplex.

MPCPTP6

Use the MPCPTP6 INTERFACE statement to define any of the following IPv6 connections:

- A connection to another host using ESCON channel-to-channel adapters.
- An XCF connection to another TCP/IP in the same z/OS sysplex.
- An IUTSAMEH connection (with no need for any I/O devices) to another TCP/IP on the same z/OS system or to VTAM for Enterprise Extender.

Use the IPCONFIG6 DYNAMICXCF statement to cause TCP/IP to automatically define and activate XCF connectivity between each pair of TCP/IP stacks in the same sysplex and IUTSAMEH connectivity between multiple TCP/IP stacks on the same z/OS.

VIRTUAL6

Use the VIRTUAL6 INTERFACE statement to define IPv6 static VIPAs.

TRANSLATE

Use TRANSLATE to indicate which LINK has specified network addresses for use as a static ARP table. The first TRANSLATE statement in a configuration data set replaces the entire ARP cache. Subsequent TRANSLATE statements add to the table. If you are using OSPF routing (OMPROUTE), see Chapter 6, "Routing," on page 217 for more information about requirements for the TRANSLATE statement.

HOME

HOME lists the IP addresses and their associated LINK adapter. The first HOME statement within a configuration data set replaces the existing HOME list. If subsequent HOME statements are found within a configuration data set, add entries to the list.

Note: The order of the HOME list is important if IPCONFIG SOURCEVIPA is specified, except for TCP datagram requests with TCPSTACKSOURCEVIPA specified. The source address used will be the preceding VIPA address instead of the physical adapter used to send the datagram. If no VIPA precedes the physical adapter in the HOME list, the physical adapter IP address is used as the source address. Refer to *z/OS Communications Server: IP Configuration Reference* for precautions when either the VIPA address or a physical adapter used as a source for the VIPA has an IP address that is the network address.

PRIMARYINTERFACE

Use PRIMARYINTERFACE to specify which link should be designated as the default local host for use by the GETHOSTID() function. If PRIMARYINTERFACE is not used, the first IP address in the HOME list becomes the default local host address.

BEGINROUTES

Use the BEGINROUTES statement to add static routes to the IP route table.

After an IPv4 interface has a DEVICE, LINK, and HOME statement, it can be started with the START device statement or the VARY TCPIP,,START command.

After an IPv6 interface has an INTERFACE statement, it can be started with the START interface statement or the VARY TCPIP,,START command.

Devices that support ARP offload

Certain devices provide an ARP offload function that offloads all ARP processing to the adapter. The function provided by the adapter impacts the ability of TCP/IP to display ARP cache information or ARP counter statistics for these devices.

Note: ARP processing is relevant only for IPv4 LAN interfaces.

The following devices provide an ARP offload function and provide ARP cache data or ARP counters to TCP/IP.

- MPCIPA (OSA-Express Gigabit Ethernet) with a minimum required microcode level of [MCL] 401
- MPCIPA (OSA-Express Fast Ethernet)
- MPCIPA (OSA-Express Token Ring)

Note: If multiple TCP/IP instances are sharing the device, the ARP data will represent all TCP/IP instances using the device. This information is provided to TCP/IP every 30 seconds from the device.

The following devices provide an ARP offload function and do not provide any ARP cache data or ARP counters to TCP/IP:

- MPCOSA (OSA-2 Fast Ethernet, FDDI)
- MPCIPA (OSA-Express Gigabit Ethernet) with a microcode level earlier than [MCL] 401

Note: For IPv6 LAN interfaces, TCP/IP performs all the neighbor discovery processing, maintains the neighbor cache, and provides the ability to display neighbor cache information.

Interface-layer fault-tolerance for local area networks (interface-takeover function)

The TCP/IP stack in the z/OS Communications Server provides transparent fault-tolerance for failed (or stopped) IPv4 devices or IPv6 interfaces, when the stack is configured with redundant connectivity onto a LAN. This support is provided by the z/OS Communications Server interface-takeover function, and applies to IPv4 MPCIPA and LCS device types and to the IPv6 IPAQENET6 interface type.

At device or interface startup time, TCP/IP dynamically learns of redundant connectivity onto the LAN, and uses this information to select suitable backups in the case of a future failure of the device or interface. This support makes use of ARP flows (for IPv4 devices) or neighbor discovery flows (for IPv6 interfaces), so upon failure (or stop) of a device or interface, TCP/IP immediately notifies stations on the LAN that the original IPv4 or IPv6 address is now reachable through the backup's link-layer (MAC) address. Users targeting the original IP address will see no outage due to the failure, and will be unaware that any failure occurred.

Since this support is built upon ARP or neighbor discovery flows, no dynamic routing protocol in the IP layer is required to achieve this fault tolerance. To enable this support, you only need to configure redundancy onto the LAN:

- You need redundant LAN adapters.
- For IPv4, you must configure and activate multiple LINKs onto the LAN.
- For IPv6, you need to configure and start multiple INTERFACEs onto the LAN.

Restriction: An IPv4 device cannot back up an IPv6 interface, and an IPv6 interface cannot back up an IPv4 device.

The interface-layer fault-tolerance feature can be used in conjunction with VIPA addresses, where applications can target the VIPA address, and any failure of the real LAN hardware is handled by the interface-takeover function. This differs from traditional VIPA usage, where dynamic routing protocols are required to route around real hardware failures.

IPv6 considerations: Stateless autoconfiguration and duplicate address detection

IPv6 provides the capability of autoconfiguring addresses for an interface by using information provided by IPv6 routers. Descriptions of this function can be found in RFC 2461 and RFC 2462. The term *autoconfigured IP address* is used below to mean an IP address that is created as a result of information received from a router advertisement. z/OS TCP/IP allows autoconfiguration if no IP addresses are defined on the profile INTERFACE statement using the IPADDR keyword. If the INTERFACE statement contains IPADDR definitions, this indicates that the installation is defining its own IP addresses and autoconfiguration is not desired. Subsequent descriptions use the term *manually configured addresses* to describe the addresses that are defined using the IPADDR keyword.

TCP creates an autoconfigured IP address for an interface if all three of the following conditions are met:

- The interface is active.

- A valid router advertisement containing prefix information with the autonomous flag on is received over the interface.
- No manually configured home addresses are defined for the interface at the time the router advertisement is received.

The IP address that is created is formed by appending the interface ID generated by the stack to the prefix supplied by the router advertisement. Autoconfigured IP addresses can be identified in the netstat home report by the 'Autoconfigured' flag. For more information on the interface ID generated by the stack, see *z/OS Communications Server: IPv6 Network and Application Design Guide*.

An autoconfigured IP address exists until one of the following occurs:

- The valid lifetime specified by the most recent router advertisement expires. When the valid lifetime expires, the autoconfigured address is removed. Existing connections using this address are terminated when subsequent activity occurs on the connection. The router advertisement that contains the valid lifetime for the autoconfigured address can also specify a preferred lifetime. The preferred lifetime indicates that the IP address can be freely used. When the preferred lifetime expires, the autoconfigured address is considered deprecated. The deprecated state indicates that another IP address should be used if available and provides a transition period before the valid lifetime expires. A deprecated IP address can be identified in the netstat home report by the 'deprecated' flag.
- The installation activates a profile that contains a manually configured IP address on the same interface as the autoconfigured IP address (that is, the INTERFACE statement contains the ADDADDR keyword). If this occurs, any autoconfigured IP addresses on that interface are deleted and existing connections using this address are terminated when subsequent activity occurs on the connection. The manually configured addresses are added and duplicate address detection for the newly added IP addresses initiated, if applicable.

Duplicate address detection is the process described in RFC 2462 which verifies that IPv6 home addresses are unique on the local link before assigning them to an interface. Duplicate address detection is performed on all IPv6 IPAQENET6 home addresses, whether they are manually configured or autogenerated, unless the INTERFACE statement specifies DUPADDRDET 0. Duplicate address detection is not done for LOOPBACK6 or VIRTUAL6 addresses. The duplicate address detection process sends a multicast neighbor solicitation and waits a period of time to see if another neighbor indicates that the address is in use. By default, only one neighbor solicitation is sent and the length of time waited is approximately one second. If no neighbor responds in that interval, the address is considered unique and the interface will start using it. The number of neighbor solicitations sent by duplicate address detection can be modified by the DUPADDRDET keyword on the INTERFACE statement. The duration of the wait interval (awaiting a response from a neighbor already using the address) can be modified by information obtained from routers on the attached network.

Duplicate address detection occurs when the interface is started. Unless the INTERFACE statement indicated duplicate address detection is to be bypassed, IPv6 manually configured addresses are unavailable until the interface is started and duplicate address detection completes without finding another node on the local link with the same address. Prior to activation of the interface, manually configured addresses are shown in the netstat home report as unavailable with a reason of 'DUPLICATE ADDRESS DETECTION PENDING'. While the duplicated address detection is actively in progress for an address, the netstat home report shows the address as unavailable with a reason of 'DUPLICATE ADDRESS

DETECTION IN PROGRESS'. If another neighbor indicates the address is in use during the duplicate address detection process, message EZZ9780I is issued and the address is not made available to the interface. If the address that was found to be in use is a manually configured address, the address shows up in the netstat home report as unavailable with a reason of 'DUPLICATE ADDRESS DETECTED'.

A link-local address is required to activate a QDIO IPv6 interface and will be generated automatically by the stack. The link-local address is generated using the link-local prefix and the interface ID. If the link-local address generated from the interface ID is determined to be a duplicate, the interface is not activated if:

- Autoconfigured addresses are allowed.
- A manually configured home address specifying only the prefix was specified on the INTERFACE statement. In such a case, the interface ID needs to be used to form the complete link-local address, and since the interface ID has been found to be in use on the network, the formed link-local address cannot be used.

If duplicate address detection fails on the link-local address and only fully configured manual addresses were specified on the INTERFACE statement, up to two attempts are made to create a unique link local address using a randomly generated value instead of the interface ID. If duplicate address detection succeeds using the randomly generated link-local address, message EZZ9784I is issued indicating the generated address and the interface is activated.

Setting up reserved port number definitions in PROFILE.TCPIP

Figure 35 on page 205 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. This sample can be copied from SEZAINST(SAMPPROF). Figure 35 on page 205 includes the portion of the sample that shows how to set up reserved port number definitions. Descriptions for the statements follow Figure 35 on page 205. For more information about any of these statements, refer to *z/OS Communications Server: IP Configuration Reference*. For information specific to IPv6 support, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

```
; ====================================================================
; Application configuration
; ====================================================================
;
;
; AUTOLOG: Supplies TCPIP with the procedure names to start and the
; time value to wait at TCP start up for any of those procedures
; to terminate if they are active.
;
; AUTOLOG 5
;  FTPD JOBNAME FTPD1       ; FTP Server
;  LPSERVE                  ; LPD Server
;  NAMED                    ; Domain Name Server
;  NCPROUT                  ; NCPROUTE Server
|;  OMPROUTE                 ; OMPROUTE Server
;  OSNMPD                   ; SNMP Agent Server
;  PORTMAP                  ; Portmap Server (SUN 3.9)
;  PORTMAP JOBNAME PORTMAP1 ; USS Portmap Server (SUN 4.0)
;  RXSERVE                  ; Remote Execution Server
;  SMTP                     ; SMTP Server
|;  TCPIPX25                 ; X25 Server
; ENDAUTOLOG
;
;
; PORT: Reserves a port for specified job names
;
;   - A port that is not reserved in this list can be used by any user.
;     If you have TCP/IP hosts in your network that reserve ports
;     in the range 1-1023 for privileged applications, you should
```

```
;     reserve them here to prevent users from using them.
;     The RESTRICTLOWPORTS option on TCPCONFIG and UDPCONFIG will also
;     prevent unauthorized applications from accessing unreserved
;     ports in the 1-1023 range.
;
;   - A PORT statement with the optional keyword SAF followed by a
;     1-8 character name can be used to reserve a PORT and control
;     access to the PORT with a security product such as RACF.
;     For port access control, the full resource name for the security
;     product authorization check is constructed as follows:
;     EZB.PORTACCESS.sysname.tcpname.safname
;     where:
;       EZB.PORTACCESS is a constant
;       sysname is the MVS system name (substitute your sysname)
;       tcpname is the TCPIP jobname (substitute your jobname)
;       safname is the 1-8 character name following the SAF keyword
;
;     When PORT access control is used, the TCP/IP application
;     requiring access to the reserved PORT must be running under a
;     USERID that is authorized to the resource. The resources
;     are defined in the SERVAUTH class.
;
;     For an example of how the SAF keyword can be used to enhance
;     security, see the definition below for the FTP data PORT 20
;     with the SAF keyword. This definition reserves TCP PORT 20 for
;     any jobname (the *) but requires that the FTP user be permitted
;     by the security product to the resource:
;     EZB.PORTACCESS.sysname.tcpname.FTPDATA in the SERVAUTH class.
;
;   - The BIND keyword is used to force a generic server (one that
;     binds to INADDR_ANY) to bind to the specific IP address that
;     is specified following the BIND keyword. This capability could
;     be used, for example, to allow z/OS UNIX telnet and telnet
;     3270 servers to both bind to TCP port 23.
;     The IP address that follows bind must be in IPv4 (dotted
;     decimal) or IPv6 (hexadecimal notation) format and may be
;     any valid address for the host including VIPA and dynamic
;     VIPA addresses.
;
;   The special jobname of OMVS indicates that the PORT is reserved
;   for any application with the exception of those that use the Pascal
;   API.
;
;   The special jobname of * indicates that the PORT is reserved
;   for any application, including Pascal API socket applications.
;   Jobname may be specified as a prefix of zero to seven characters
;   ending in *.
;
;   The special jobname of RESERVED indicates that the PORT is
;   blocked. It will not be available to any application.
;
;   The special jobname of INTCLIEN indicates that the PORT is
;   reserved for internal stack use.
;
;
PORT
     7 UDP MISCSERV            ; Miscellaneous Server - echo
     7 TCP MISCSERV            ; Miscellaneous Server - echo
     9 UDP MISCSERV            ; Miscellaneous Server - discard
     9 TCP MISCSERV            ; Miscellaneous Server - discard
    19 UDP MISCSERV            ; Miscellaneous Server - chargen
    19 TCP MISCSERV            ; Miscellaneous Server - chargen
    20 TCP *  NOAUTOLOG        ; FTP Server
;   20 TCP *  NOAUTOLOG SAF FTPDATA ; FTP Server
    21 TCP FTPD1               ; FTP Server
    23 TCP INTCLIEN            ; Telnet 3270 Server
;   23 TCP INETD1 BIND 9.67.113.3 ; z/OS UNIX Telnet server
    25 TCP SMTP                ; SMTP Server
    53 TCP NAMED               ; Domain Name Server
    53 UDP NAMED               ; Domain Name Server
   111 TCP PORTMAP             ; Portmap Server (SUN 3.9)
   111 UDP PORTMAP             ; Portmap Server (SUN 3.9)
;  111 TCP PORTMAP1            ; Unix Portmap Server (SUN 4.0)
;  111 UDP PORTMAP1            ; Unix Portmap Server (SUN 4.0)
   123 UDP SNTPD               ; Simple Network Time Protocol Server
   135 UDP LLBD                ; NCS Location Broker
```

```
   161 UDP OSNMPD             ; SNMP Agent
   389 TCP LDAPSRV            ; LDAP Server
   443 TCP HTTPS              ; http protocol over TLS/SSL
   443 UDP HTTPS              ; http protocol over TLS/SSL
;  500 UDP IKED               ; CS IKE daemon
   512 TCP RXSERVE            ; Remote Execution Server
   514 TCP RXSERVE            ; Remote Execution Server
;  512 TCP * SAF OREXECD      ; z/OS UNIX Remote Execution Server
;  514 TCP * SAF ORSHELLD     ; z/OS UNIX Remote Shell Server
;  515 TCP LPSERVE            ; LPD Server
;  515 TCP AOPLPD             ; Infoprint LPD Server
   520 UDP OMPROUTE           ; OMPROUTE Server (IPv4 RIP)
   521 UDP OMPROUTE           ; OMPROUTE Server (IPv6 RIP)
   580 UDP NCPROUT            ; NCPROUTE Server
   750 TCP MVSKERB            ; Kerberos
   750 UDP MVSKERB            ; Kerberos
   751 TCP ADM@SRV            ; Kerberos Admin Server
   751 UDP ADM@SRV            ; Kerberos Admin Server
  3000 TCP CICSTCP            ; CICS Socket
  3389 TCP MSYSLDAP           ; LDAP Server for Msys
; 4500 UDP IKED               ; CS IKE daemon
;
;
; PORTRANGE: Reserves a range of ports for specified jobnames.
;
;   In a common INET (CINET) environment, the port range indicated by
;   the INADDRANYPORT and INADDRANYCOUNT in your BPXPRMxx parmlib member
;   should be reserved for OMVS.
;
;   The special jobname of OMVS indicates that the PORTRANGE is reserved
;   for ANY z/OS UNIX socket application.
;
;   The special jobname of * indicates that the PORTRANGE is reserved
;   for any socket application, including Pascal API socket
;   applications.
;
;   The special jobname of RESERVED indicates that the PORTRANGE is
;   blocked. It will not be available to any application.
;
;   The SAF keyword is used to restrict access to the PORTRANGE to
;   authorized users. See the use of SAF on the PORT statement above.
;
;
;   PORTRANGE 4000 1000 TCP OMVS
;   PORTRANGE 4000 1000 UDP OMVS
;   PORTRANGE 2000 3000 TCP RESERVED
;   PORTRANGE 5000 6000 TCP * SAF RANGE1
;
; SACONFIG: Configures the SNMP TCP/IP subagent
;   The SACONFIG statement specified in this sample prevents the
;   activation of the TCP/IP subagent.  If you want the
;   TCP/IP subagent started during stack initialization, either
;   remove the sample statement, or respecify the statement as
;   follows, supplying a COMMUNITY and AGENT parameter value:
;
;     SACONFIG ENABLED COMMUNITY communityname AGENT portnum
;
;   If you remove the sample statement, the TCP/IP subagent will be
;   started with a COMMUNITY value of 'public', and an
;   AGENT port value of 161.
;
SACONFIG DISABLED
;
```

Figure 35. Example of reserved port number definitions

The following explains the statements shown in Figure 35.

AUTOLOG

Use AUTOLOG to list the procedure names that should start when the TCPIP address space starts. It is also used to supply a timeout value for detecting hung procedures at TCP/IP initialization time. The timeout value

is the time TCP/IP should allow for a procedure to come down when, at startup, it is still active and TCP/IP is attempting to AUTOLOG the procedure again. A hung procedure is active to MVS, but is not listening on the socket that is reserved for it via the PORT statement. When AUTOLOG detects a hung task, TCP/IP checks every 10 seconds (until the timeout value has expired) to see if the procedure has come down. If the procedure comes down during one of these 10 second intervals, it is restarted. If the procedure is still active when the time interval specified by the timeout value expires, then TCP/IP cancels and restarts the procedure.

The AUTOLOG statement shown in Figure 35 on page 205 has a timeout value of five minutes.

In the first AUTOLOG statement the FTP Server shows FTPD JOBNAME FTPD1. This means when the TCPIP address space starts, the FTPD procedure will be started via the MVS START FTPD command. Because FTPD forks a child process that actually listens on PORT 21 (see the PORT statement in this section), the autolog task verifies that FTPD1 is listening on port 21.

Similarly, when the TCPIP address space starts, the autolog task starts the remaining 10 tasks.

Unless the tasks in the AUTOLOG list are in the PORT reservation list, the autolog task does not check for hung tasks every five minutes. Also at startup time, those procedures that are not on the PORT list are first canceled and then started. This occurs because the procedure might have been running from a previous TCP/IP address space and would need to be started and stopped to start listening when the new stack starts.

Notes:

1. If you run multiple TCP/IP address spaces, ensure that the second address space AUTOLOG list does not cancel the procedures of the first. In those cases, an installation might require different procedure names for the servers for each address space. For more information about multiple stacks, see "Port management overview" on page 73.
2. You can use the AUTOLOG statement to automatically start generic servers in a single stack environment, but you should be careful using the AUTOLOG statement to start generic servers in a multiple stack environment. Instead, you could use an operations automation software package (IBM and other vendors provide these) to start generic servers automatically. For a list of generic servers provided by TCP/IP, see "Generic servers in a CINET environment" on page 76.

For those procedures that require parameters to be used on the MVS START command, there is a PARMSTRING option. For more information, refer to *z/OS Communications Server: IP Configuration Reference*.

PORT Use PORT to reserve ports for different jobs. This prevents a rogue application from taking port 21, which is needed by FTP. For each port entry, the port number, protocol, and procedure name are specified. The first port entry shows port 7 UDP reserved for the miscellaneous echo server for procedure MISCSERV. Similarly, port 7 of TCP is also reserved for the same server. In this example, six ports are reserved for the miscellaneous server.

INTCLIEN is an INTernal CLIENt to the TCPIP address space (the TN3270 Telnet server), and it runs continuously. See Chapter 10, "Accessing remote hosts using Telnet," on page 441 for more information about INTCLIEN.

NOAUTOLOG can be specified, as in the port 20 TCP * in Figure 35 on page 205. In this way, the port is reserved for an OMVS forked task so that the FTP server can fork tasks to port 20 as each FTP user logs in.

Use the DELAYACKS and NODELAYACKS options to allow an installation to delay their acknowledgments so they can be combined with data to be sent to foreign hosts. Unless a performance reason is needed, DELAYACKS should be used to delay the transmission of acknowledgments.

Use the SHAREPORT parameter or the SHAREPORTWLM parameter when reserving a port to be shared across multiple TCP listeners. This is not valid for UDP. To understand how these PORT statement parameters are used, refer to *z/OS Communications Server: IP Configuration Reference*. Use the Netstat ALL/-A report to determine whether port sharing is being used for a TCP listener. If port sharing is being used, this report indicates which type is being used.

Typically, reserving a port for a specific job name is sufficient. If the port must instead be reserved for a specific user ID or a set of user IDs, use the SAF keyword to specify the name of a SAF resource to be associated with the port. The user ID associated with the application that attempts to bind to the port must be permitted to the SAF resource.

The BIND keyword is used to force a generic server (one that binds to INADDR_ANY) to bind to the specific IP address that is specified following the BIND keyword. This capability could be used, for example, to allow the z/OS UNIX Telnet and TN3270 Telnet servers to both bind to TCP port 23 on different IP addresses. The IP address that follows bind can be any valid address for the host, including VIPA and dynamic VIPA addresses. The address supplied can be either an IPv4 address (in dotted-decimal format) or an IPv6 address (in colon-hexadecimal format). IPv4-mapped IPv6 addresses and IPv4-compatible IPv6 addresses are not supported. For multiple servers to bind to the same port with this function, the IP address for each server must be unique.

RESERVED indicates that the port is not available for use by any user.

PORTRANGE
PORTRANGE is a statement used to reserve a range of ports for specified job names.

SACONFIG
SACONFIG is the statement used to configure the information about the SNMP TCP/IP subagent. The AGENT keyword on this statement is used to specify the port number to be used when the TCP/IP subagent connects to the SNMP agent. Omission of this statement causes TCPIP to assume the default value of SACONFIG ENABLED COMMUNITY public AGENT 161. For more information on the SACONFIG profile statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Setting up SAF Server Access Authorization (SERVAUTH) (optional)

The TCP/IP address space uses the SERVAUTH System Authorization Facility (SAF) class to protect TCP/IP resources from unauthorized access. The use of SERVAUTH may be optional and is available in degrees so that installations can pick and choose the access needed. Installations may be able to choose to use one, all, or none of the protections provided by SERVAUTH. The customizing described in this section is completely optional when using the IBM security product RACF. Non-IBM security products may require customizing. A template of the commands

and all other SAF commands appears in SEZAINST(EZARACF). Refer to Chapter 3, “Security,” on page 117 for more detailed information.

Configuring the local host table (optional)

Why configure a local host table?

You can set up the local host table to support local host name resolution. If you use the local host table for this purpose, your socket applications will only be able to resolve names and IP addresses that appear in your local host table.

If you need to resolve host names outside your local area, you can configure the resolver to use a domain name server (see the NSINTERADDR statement). If you use a domain name server, you do not need to set up any host definitions in your resolver configuration, but you may still do so.

If you have configured your resolver to use a name server, it will always try to do so, unless your TCP/IP C/C++ API applications were written with a RESOLVE_VIA_LOOKUP symbol in the source code. You can also configure the resolver to only use a local host table by specifying LOOKUP LOCAL in the TCPIP.DATA configuration file. For both cases, all name resolution calls will always use a local host table. This is probably not a technique you will see for standard socket applications, but it may be a technique you could find useful for when you develop your own socket applications or for testing changes before they are placed in your name server.

It might be a good idea to have a local host table available for the resolver to use if the name server is not reachable. If the name server does not respond to name resolution requests, the resolver tries to use the local host table. If the name server is reachable but returns a negative reply for a name resolution request, the resolver tries to resolve the name using the local host table, if such a file is present.

Assume you try to resolve the host name *friendly* and your DOMAINORIGIN is *my.house.com*, the resolver sends a query to the name server for *friendly.my.house.com*. If the name server returns a negative reply (the name is not registered), the resolver looks into the local host table for an entry of *friendly.my.house.com* and, if not found, for an entry of *friendly*.

Due to the flexibility of the Domain Name System, it is recommended you use a domain name server. If you set up a small TCP/IP network, the simplicity of the local host table approach might be preferable.

The following types of local host table can be used:

- HOSTS.LOCAL

  HOSTS.LOCAL is only used for IPv4 requests.
- /etc/hosts

  /etc/hosts is only used for IPv4 requests.
- ETC.IPNODES and /etc/ipnodes

  ETC.IPNODES and /etc/ipnodes can be used for IPv6 requests, and for IPv4 requests when COMMONSEARCH is coded in the resolver setup statements.

Creating HOSTS.LOCAL site host table

The site host table is generated from the *hlq*.HOSTS.LOCAL data set. This data set contains descriptions of local host entries in the HOSTS format. HOSTS.LOCAL

can only contain IPv4 addresses. A sample HOSTS.LOCAL data set is created during installation. The following sections describe how to update the sample *hlq*.HOSTS.LOCAL data set and use it to generate the two data sets, *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO, which function as your site table.

HOST entries

One line of the *hlq*.HOSTS.LOCAL data set is used for each distinct host and ends with four colons (::::). The maximum length of the line is 512 characters. Each host can have multiple IP addresses and multiple names. The line for each host has three essential fields, separated by colons. These fields are:

- The keyword *HOST*
- A list, separated by commas, of IP addresses for that host. A maximum of 6 IP addresses can be specified.
- A list, separated by commas, of fully qualified names for that host. A maximum of 20 host names can be specified. Only the first six host names will be used in the *hlq*.HOSTS.ADDRINFO data set. All twenty host names will be used in the *hlq*.HOSTS.SITEINFO data set.

For example, if you have two local hosts, LOCAL1 (IP addresses 192.6.77.4 and 192.8.4.1) and LOCAL2 (with an alias LOCALB and IP address 192.6.77.2), append the following lines to the *hlq*.HOSTS.LOCAL data set:

```
HOST : 192.6.77.4, 192.8.4.1 : LOCAL1 ::::
HOST : 192.6.77.2 : LOCAL2, LOCALB ::::
```

Notes:

1. The maximum length for a host allowed in the HOST tables is 24 characters.
2. If the HOST entry has more than the maximum of 6 IP addresses for one host name or more than 20 host names for a single IP address, the MAKESITE command will complete successfully with no error message issued.

NET and GATEWAY entries

The NET and GATEWAY statements are not used by TCP/IP for z/OS applications. However, some socket calls require the NET entries. If your programs do not need the NET and GATEWAY statements, delete them before invoking MAKESITE.

Sample HOSTS.LOCAL data set (HOSTS): Following is the sample HOSTS.LOCAL data set provided in SEZAINST(HOSTS):

```
; HOSTS.LOCAL
; -----------
; COPYRIGHT = NONE.
;
; The format of this file is documented in RFC 952, "DoD Internet
; Host Table Specification".
;
; The format for entries is:
;
; NET : ADDR : NETNAME :
; GATEWAY : ADDR, ALT-ADDR : HOSTNAME : CPUTYPE : OPSYS : PROTOCOLS :
; HOST : ADDR, ALT-ADDR : HOSTNAME, NICKNAME : CPUTYPE : OPSYS : PROTOCOLS :
;
; Where:
;   ADDR, ALT-ADDR = IP address in decimal, e.g., 26.0.0.73
;   HOSTNAME, NICKNAME = the fully qualified host name and any nicknames
;   CPUTYPE = machine type (PDP-11/70, VAX-11/780, IBM-3090, C/30, etc.)
;   OPSYS = operating system (UNIX, TOPS20, TENEX, VM/SP, etc.)
```

```
;   PROTOCOLS = transport/service (TCP/TELNET,TCP/FTP, etc.)
;   : (colon) = field delimiter
;   :: (2 colons) = null field
; *** CPUTYPE, OPSYS, and PROTOCOLS are optional fields.
;
;   MAKESITE does not allow continuation lines, as described in
;   note 2 of the section "GRAMMATICAL HOST TABLE SPECIFICATION"
;   in RFC 952.  Entries should be specified on a single line of
;   up to a maximum of 512 characters per line.
;
;
; Note:  The NET and GATEWAY statements are not used by the TCP/IP for
;        MVS applications.  However, some socket calls require the NET
;        entries.  For better performance, if your programs do not need
;        the NET and GATEWAY statements, delete them before running
;        the MAKESITE program.
;
;
HOST : 9.67.43.100 : NAMESERVER ::::
HOST : 9.67.43.126 : RALEIGH ::::
HOST : 129.34.128.245, 129.34.128.246 : YORKTOWN, WATSON ::::
;
NET  : 9.67.43.0 : RALEIGH.IBM.COM :
;
GATEWAY : 129.34.0.0 : YORKTOWN-GATEWAY ::::
;
```

Using MAKESITE

Because many servers and commands allocate *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO, it is important not to overwrite or delete these data sets while TCP/IP is running. To avoid disrupting any active users, use an *HLQ*=parm that is different than your active *hlq*. This allows you to swap names (by renaming the old HOSTS data sets and then renaming the new HOSTS data sets) without starting and stopping TCP/IP.

Use MAKESITE as a TSO command or in a batch job to generate new *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO data sets. The parameters are the same for either a TSO command or a batch job invocation of MAKESITE. Refer to *z/OS Communications Server: IP System Administrator's Commands* for more information.

After you make changes to your *hlq*.HOSTS.LOCAL data set, you must generate and install new *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO data sets.

Guidelines: Use ETC.IPNODES (in the style of etc/ipnodes) as the preferred alternative to the generated local hosts tables from MAKESITE for the following reasons:

- No imposed 24 character restriction on host names.
- No imposed restriction on the first eight characters of the host names having to be unique. However, there are certain applications that require the first eight characters to be unique, such as Network Job Entry (NJE).
- Closely resembles that of other TCP/IP platforms, and eliminates the MAKESITE requirement of file post-processing.
- Allows configuration of both IPv4 and IPv6 addresses.
- Only one file is managed for the system, and that all the APIs can utilize the same single file. The COMMONSEARCH statement in the resolver setup file can be used to reduce IPv6 and IPv4 searching to a single search order, as well as to reduce the z/OS UNIX and native MVS environments to a single search order.

For more information on the use of IPNODES by the resolver to locate IPv4 and IPv6 addresses and sitenames, refer to *z/OS Communications Server: IPv6 Network and Application Design Guide*.

For the search orders used in locating the local host tables, see "Configuration files for TCP/IP applications" on page 38. For a description of the use of IPNODES by the resolver to locate IPv4 and IPv6 addresses and sitenames when the resolver setup statement COMMONSEARCH is specified, see "IPv6/common search order" on page 48 and "IPv6/common search order" on page 52.

Creating /etc/hosts

The /etc/hosts HFS file can be defined as follows:

- The maximum line length is 256 characters. If a line is greater than 256 characters, it is truncated to 256 characters and processed. If trace resolver is active, a warning message is issued.
- The line starts with an IP address, followed by a blank, followed by host names. Host names are separated by one or more blanks.
- Only IPv4 addresses are supported.
- Each IP address can have up to 35 host names. A host name can have up to 35 IP addresses.
- The values for the host name must conform to the following:
  - Maximum of 128 characters.
  - Must contain one or more tokens separated by a period.
  - Each token must be at least 1 character and less than 64 characters.
  - First character in each token must be a letter (A-Z or a-z) or number (0-9).
- A comment is indicated by the # or ; character.

For the search orders used in locating /etc/hosts, see "Configuration files for TCP/IP applications" on page 38.

Creating ETC.IPNODES and /etc/ipnodes

The ETC.IPNODES and /etc/ipnodes file can be defined as follows:

- HFS files can reside in any directory. The maximum line length supported is 256 characters. If a line is greater than 256 characters, it is truncated to 256 characters and processed. If trace resolver is active, a warning message is issued.
- MVS data sets must be partitioned organization (PO) or sequential (PS), RECFM=F or RECFM=FB, a logical record length (LRECL) between 56 and 256, and have any valid blocksize (BLKSIZE) for fixed block.
- It can contain IPv4 and IPv6 addresses, but not IPv4 mapped addresses. Each IP address can have up to 35 host names. A host name can have up to 35 IP addresses.
- The values for the host name must conform to the following:
  - Maximum of 128 characters.
  - Must contain one or more tokens separated by a period.
  - Each token must be at least 1 character and less than 64 characters.
  - First character in each token must be a letter (A-Z or a-z) or number (0-9).
- A comment is indicated by the # or ; character.

The sample IPNODES file provided by z/OS Communications Server follows. It can be found as member EZBREIPN (alias IPNODES) in SEZAINST.

```
;
;  IBM z/OS Communications Server
;  SMP/E distribution name: EZBREIPN
;
;  5694-A01 (C) Copyright IBM Corp. 2002.
;  Licensed Materials - Property of IBM
;
;  Function: Sample ETC.IPNODES file
;
;  The file contains the Internet Protocol (IP) host names
;  and addresses for the local host and other hosts in the
;  Internet network.
;  This file is used to resolve a name into an address (that is, to
;  translate a host name into its Internet address) or resolve
;  an address into a name.
;
;  Comments begin with a # or ; character and continue until the
;  end of the line.
;
;  The following statement defines the Internet Protocol (IP) name
;  and address of the local host and specifies the names and
;  addresses of remote hosts. The maximum line length support is
;  256 characters
;
;  Entries in the hosts file have the following format:
;
;  Address HostName
;
;  Address HostName1 HostName2 HostName3 ..... HostName35
;
;
;  Address: is an IP address, it can be IPV4 or IPV6 address.
;           Note: IPv4-mapped IPv6 address is not allowed.
;
;  HostName: the length of the hostname is up to 128 characters,
;            and each IP address can have up to 35 hostnames.
;
;
   9.67.43.100 NAMESERVER
   9.67.43.126 RALEIGH
   9.67.43.222 HOSTNAME1.RALEIGH.IBM.COM
   129.34.128.245 YORKTOWN WATSON
   1::2 TESTIPV6ADDRESS1
   1:2:3:4:5:6:7:8 TESTIPV6ADDRESS2
;
```

For the search orders used in locating ETC.IPNODES and /etc/ipnodes, see "Configuration files for TCP/IP applications" on page 38.

Verifying your configuration

At this point, your configuration files have been updated.

To verify a configuration, start the TCP/IP address space and ensure that you see the following message:

```
EZB6473I TCP/IP STACK FUNCTIONS INITIALIZATION COMPLETE
```

If the message is not displayed, the messages issued by the TCP/IP address space should describe why TCP/IP did not start.

Verifying TCPIP.DATA statement values in the native MVS environment

To display which TCPIP.DATA statement values are being used and where they are being obtained from, use trace resolver output. You can obtain trace resolver output at your TSO screen by issuing the following TSO commands:

```
alloc f(systcpt) dsn(*)
READY
netstat up
READY
free f(systcpt)
READY
```

Tip: When directing trace resolver output to a TSO terminal, define the screen size to be only 80 columns wide. Otherwise, trace output is difficult to read.

For further information on the trace resolver facility and interpreting its output, refer to *z/OS Communications Server: IP Diagnosis Guide*.

Verifying TCPIP.DATA statement values in the z/OS UNIX environment

To display which TCPIP.DATA statement values are being used and where they are being obtained from, use trace resolver output. You can obtain trace resolver output by issuing the following z/OS UNIX shell commands:

```
export RESOLVER_TRACE=stdout
onetstat -u
set -A RESOLVER_TRACE
```

For further information on the trace resolver facility and interpreting its output, refer to *z/OS Communications Server: IP Diagnosis Guide*.

Verifying PROFILE.TCPIP with TSO NETSTAT or z/OS UNIX onetstat

Many configuration values specified within the PROFILE.TCPIP file can be verified with the TSO NETSTAT or z/OS UNIX onetstat commands. To verify the physical network and hardware definitions, use the NETSTAT DEVLINKS or onetstat -d commands. To see operating characteristics, use NETSTAT CONFIG or onetstat -f. Refer to *z/OS Communications Server: IP System Administrator's Commands* for information about the syntax and output of the commands.

Verifying interfaces with PING and TRACERTE

PING and TRACERTE can be used in the TSO environment to verify adapters or interfaces attached to the z/OS host. In the z/OS UNIX environment, oping and otracert can be used with identical results. For information about the syntax and output of the commands, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Given that your PROFILE.TCPIP file contains the interfaces of your installation and that the TCPIP.DATA file contains the correct TCPIPJOBNAME, the TCPIP address space is configured and you can go on to configuring routes, servers, and so on.

Verifying local name resolution with TESTSITE

Use the TESTSITE command to verify that the *hlq*.HOSTS.ADDRINFO and *hlq*.HOSTS.SITEINFO data sets can correctly resolve the name of a host, gateway,

or net. For more information on the TESTSITE command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Verifying PROFILE.TCPIP and TCPIP.DATA using HOMETEST

Use the HOMETEST command to verify the HOSTNAME, DOMAINORIGIN, SEARCH, and NSINTERADDR TCPIP.DATA statements. HOMETEST will use the resolver to obtain the IP addresses assigned to the HOSTNAME and compare them to the HOME list specified in PROFILE.TCPIP. A warning message will be issued if any HOSTNAME IP addresses are missing from the HOME list.

Activate TRACE RESOLVER if you would like detailed information on how the HOSTNAME is resolved to IP addresses. The information will also include what TCPIP.DATA data set names were used. This can be done by issuing the following TSO command before running HOMETEST. The detailed information will be displayed on your TSO screen.

```
allocate dd(systcpt) da(*)
```

Issue the following TSO command after HOMETEST to turn off TRACE RESOLVER output.

```
free dd(systcpt)
```

If you do not have TRACE RESOLVER turned on before running HOMETEST, the following is displayed:

```
hometest

Running IBM MVS TCP/IP CS V1R6 TCP/IP Configuration Tester

FTP.DATA file not found. Using hardcoded default values.

TCP Host Name is: MVS026

Using Name Server to Resolve MVS026
The following IP addresses correspond to TCP Host Name: MVS026
9.67.113.58

The following IP addresses are the HOME IP addresses defined in PROFILE.TCPIP:
9.67.113.58
127.0.0.1

All IP addresses for MVS026 are in the HOME list!

Hometest was successful - all Tests Passed!
```

Verifying your X Window System installation (Optional)

Note: You cannot verify your X Window System until after routing and DNS setup.

Support is provided for two versions of the X Window System and the corresponding Motif. The current support, provided as part of the base IP support in z/OS Communications Server, is for X Window System Version 11 Release 6 and Motif Version 1.2. Support for X Window System Version 11 Release 4 and Motif Version 1.1 is available as feature HIP614X.

Verifying the X Window X11R4 System installation

X Window X11R4 System is installed with the other target libraries. The macro or headers go into the target library data set SEZACMAC. To verify the installation of the X Window System:

1. Specify your workstation IP address by adding a record (such as the following) to your XWINDOWS.DISPLAY data set.

   ```
   royal.csc.ibm.com:0.0
   ```

   In this example, `royal.csc.ibm.com:0.0` is the name of the host running the X Window System server.

   Note: No leading blanks are allowed in this record.
2. On the workstation running the X Window System server, issue an XHOST command specifying the name of your MVS system.
3. Run the program with the XSAMP1 command.

Verifying the X Window X11R6 System installation

To verify the installation of the X Window X11R6 System:

1. Ensure that a host (the workstation) with an X Window System server that supports X11R6 is properly configured and reachable by the MVS system. From the workstation, use Telnet to access the MVS system, and open a z/OS UNIX shell on the MVS system.
2. From the z/OS UNIX shell, export the DISPLAY environment variable using either the network name or the qualified IP address of the workstation as shown in the following example:

   ```
   export DISPLAY=royal.csc.ibm.com:0.0
   ```

   In this example, *royal.csc.ibm.com* is the name of the workstation running the X Window System server. The display is indicated by *:0.0*, and is specified this way in almost all cases.
3. Authorize the MVS system to access the workstation by executing the XHOST command, and specify either the name of the MVS system or a plus sign (+) as shown in the following example.

   ```
   xhost +
   ```

   Note: The + option turns off security for this workstation and allows any X client to display here.
4. The sample X clients are shipped in the directory */usr/lpp/tcpip/X11R6/Xamples/demos*. Change into this directory. There are four sample program directories, xsamp1, xsamp2, xsamp3, and pexsamp. Change to the xsamp1 directory. Verify that there are files named Makefile and xsamp1.c, and then execute the following command:

   ```
   make
   ```
5. Execute the program using the following command:

   ```
   xsamp1
   ```
6. The z/OS UNIX shell should block as another window is opened. Verify the workstation is displaying a new window. The xsamp1 client displays a blank window for 60 seconds and then exits, taking its window with it. The z/OS UNIX shell should no longer be blocked.

Chapter 6. Routing

The objective of this chapter is to guide you through the steps required to configure static or dynamic routing and explain how to verify the configuration. The contents of this chapter are based on the assumption that you understand your entire network configuration. It also assumes that you have read and completed all of the verification tasks outlined in previous chapters in this document.

After reading this chapter, you should be able to do the following:

- Configure static or dynamic routing
- PING a remote host by IP address
- Use TRACERTE to determine the path that will be taken to reach a particular destination
- Use NETSTAT to display your routing table
- Use DISPLAY commands to display dynamic routing information

Recommendation: The definition or modification of an installation's routing configuration should not be performed without a complete understanding of the entire network design.

Routing terminology

The following list describes some of the more common IP routing-related terms and concepts. If you need more detailed information, refer to *Routing in the Internet* by Christian Huitema.

General terms

Autonomous system (AS)
A group of routers exchanging routing information through a common routing protocol. A single AS can represent a large number of IP networks.

Dynamic routes
IP layer routing table entries that are dynamically managed and can automatically change in response to network topology changes. For IPv4, these routes are managed by a routing daemon. For IPv6, these routes can be managed by a routing daemon, and they can also be learned by listening to router advertisement messages received from routers as part of the router discovery protocol.

Exterior Gateway Protocol (EGP)
A routing protocol spoken by routers belonging to different autonomous systems when those routers are configured to share routing information between autonomous systems. This chapter does not discuss exterior gateway routing.

Interior Gateway Protocol (IGP)
A routing protocol spoken by routers belonging to the same autonomous system. Each AS has a single IGP. A separate AS within a network can be running a different IGP.

Replaceable static routes
IPv4 static routes that can be replaced by OMPROUTE, or IPv6 static

routes that can be replaced by OMPROUTE or by routes learned by listening to router advertisement messages received from routers as part of the router discovery protocol.

Router
A device or host that interprets protocols at the IP layer and forwards datagrams on a path towards their correct destination.

Routing
The process used in an IP network to deliver a datagram to the correct destination.

Routing daemon
A server process that manages the IP route table.

Static routes
IP layer routing table entries that are manually configured (using the BEGINROUTES or GATEWAY configuration statements) and do not change automatically in response to network topology changes, except when:

- The change is due to an ICMP redirect (if not disabled).
- A dynamic routing protocol learns a route to a destination configured as a replaceable static route.

Interior Gateway Protocols (IGP)

An interior gateway protocol is a dynamic route update protocol used between routers that run on TCP/IP hosts within a single autonomous system. The routers use this protocol to exchange information about IP routes.

Some of the more common interior gateway protocols are:

Routing Information Protocol (RIP)
RIP uses a distance vector algorithm to calculate the best path to a destination based on the number of hops in the path. RIP has several limitations. Some of the limitations which exist in RIP Version 1 are resolved by RIP Version 2.

RIP Version 2
RIP Version 2 extends RIP Version 1. Among the improvements are support for multicasting and variable subnetting. Variable subnetting allows the division of networks into variable size subnets. For example, one route can represent addresses from 9.1.1.0 through 9.1.1.255 (the 9.1.1.0/255.255.255.0 subnet) while another can represent addresses from 9.2.0.0 through 9.2.255.255 (the 9.2.0.0/255.255.0.0 subnet).

IPv6 RIP
IPv6 RIP uses the same distance vector algorithm used by RIP to calculate the best path to a destination. It is intended to allow routers to exchange information for computing routes through an IPv6-based network.

Open Shortest Path First (OSPF)
OSPF uses a link state or shortest path first algorithm. OSPF's most significant advantage compared to RIP is the reduced time needed to converge after a network change. In general, OSPF is more complicated to configure than RIP and might not be suitable for small networks.

IPv6 OSPF
IPv6 OSPF also uses a link state or shortest path first algorithm to calculate

the best path to a destination. IPv6 OSPF has the same advantages and more complicated configuration compared to IPv6 RIP, like OSPF compared to RIP.

Table 14. Interior Gateway Protocol characteristics

Feature	RIP-1	RIP-2	IPv6 RIP	OSPF	IPv6 OSPF
Algorithm	Distance vector	Distance vector	Distance vector	Shortest path first	Shortest path first
Network load (1)	High	High	High	Low	Low
CPU processing requirement (1)	Low	Low	Low	High	High
IP network design restrictions	Many	Some	Some	Virtually none	Virtually none
Convergence time	Up to 180 seconds	Up to 180 seconds	Up to 180 seconds	Low	Low
Multicast supported (2)	No	Yes	Yes	Yes	Yes
Multiple equal-cost routes	No (3)	No(3)	No(3)	Yes	Yes

Notes:

1. Depends on network size and stability.
2. Multicast saves CPU cycles on hosts that are not interested in certain periodic updates, such as OSPF link state advertisements or RIP-2 routing table updates. Multicast frames are filtered out either in the device driver or directly on the interface card if this host has not joined the specific multicast group.
3. RIP in OMPROUTE allows multiple equal-cost routes only for directly connected destinations over redundant interfaces. See "Using IPv4 static routing with OMPROUTE" on page 223, and "Using IPv6 static routing with OMPROUTE" on page 226.

Static versus dynamic routing

Whether static or dynamic routing is used, the IP layer routing mechanism is the same. The IP layer routes a packet by searching its routing table for the most specific route known. Route selection occurs in the following order:

1. If a route exists to the destination address (a host route), it is chosen.
2. At this point, the route chosen depends upon the version of IP being used:
   - For IPv4:
     a. If subnet, network, or supernetwork routes exist to the destination, the route with the most specific network mask (the mask with the most bits on) is chosen.
     b. If the destination is a multicast destination and a multicast default route exists, that route is chosen.
   - For IPv6, if prefix routes exist to the destination, the route with the most specific prefix is chosen.
3. Default routes are chosen when no other route exists to a destination.

Multiple equal-cost routes are allowed for both static and dynamic routing, as depicted in Table 14 on page 219.

The sample network

Figure 36 and Figure 37 on page 221 show network diagrams that depict a sample network. This sample network is used in the following sections as the configuration of static and dynamic routing is described:

- "IPv4 static routing" on page 221
- "IPv6 static routing" on page 224
- "IPv4 dynamic routing using OMPROUTE" on page 230
- "IPv6 dynamic routing using OMPROUTE" on page 233

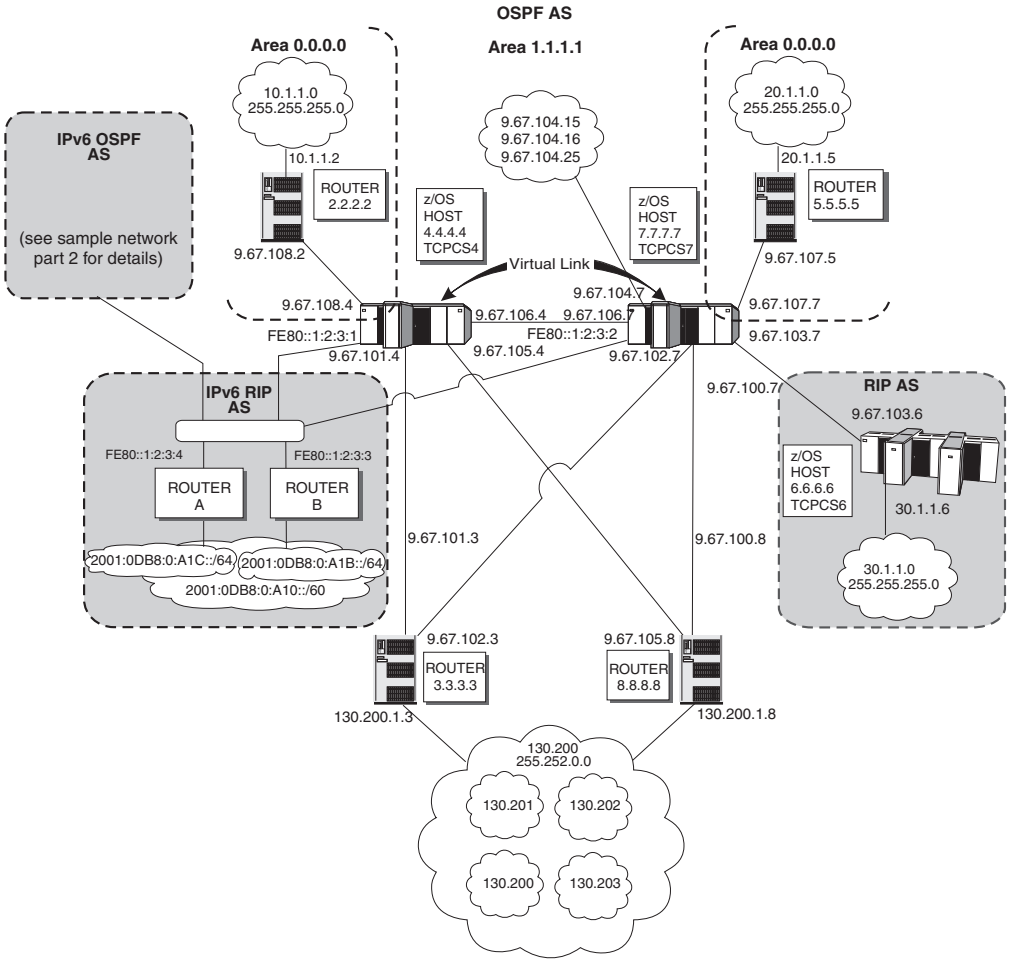


Figure 36. Sample network part 1

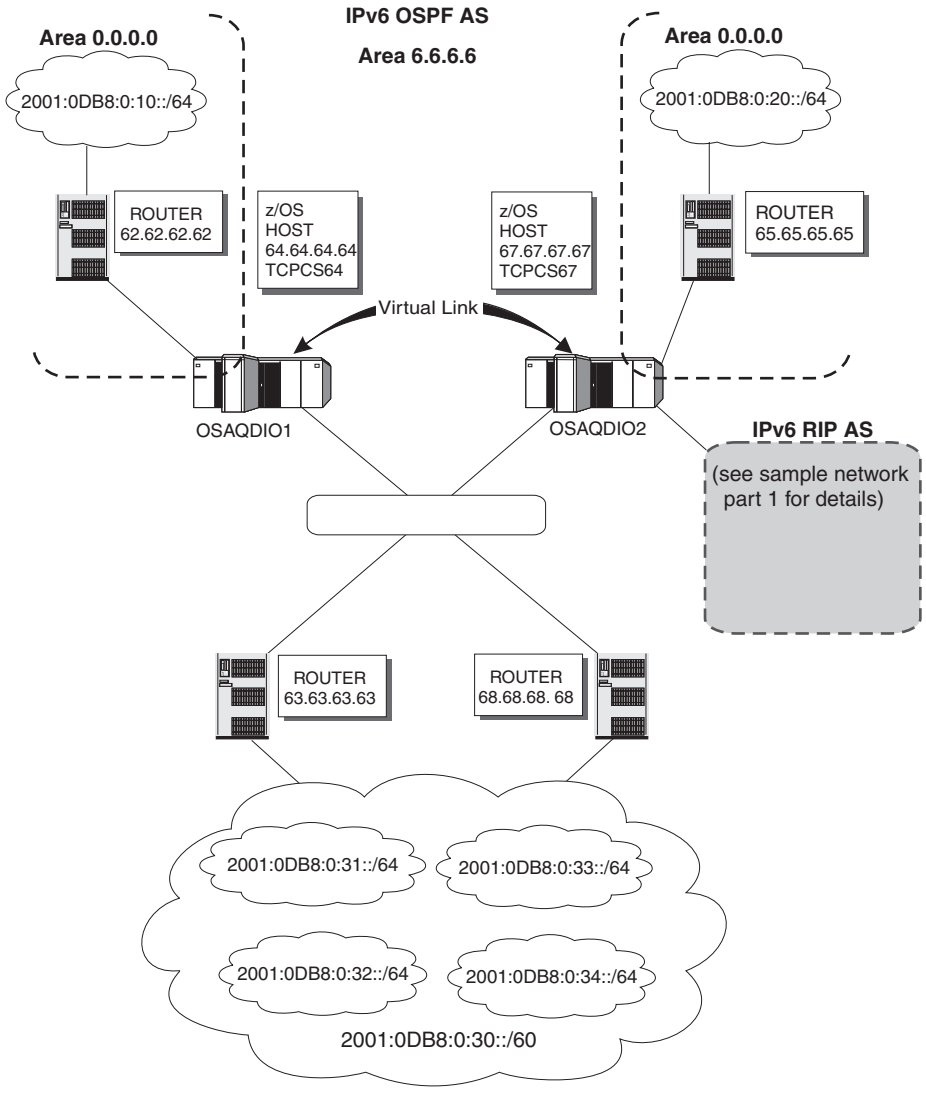


Figure 37. Sample network part 2, IPv6 OSPF AS

In this sample network, TCPCS4 and TCPCS7 are both performing as OSPF area border routers between OSPF areas 0.0.0.0 and 1.1.1.1. TCPCS64 and TCPCS67 are both performing as OSPF area border routers between IPv6 OSPF areas 0.0.0.0 and 6.6.6.6. TCPCS7 is performing as an AS boundary router between the OSPF AS and the RIP AS. TCPCS67 is performing as an AS boundary router between the IPv6 OSPF AS and the IPv6 RIP AS. TCPCS4 and TCPCS7 have interfaces to both the IPv4 network and the IPv6 network.

IPv4 static routing

Static routing requires that routes are configured manually for each router or destination; this is a significant reason system administrators avoid this technique if given a choice. Static routing has the disadvantage that network reachability is not dependent on the state of the network itself. If a destination is down or unreachable over that statically configured route, the static routes remain in the routing table, and traffic continues to be sent toward that destination without success.

To minimize network administrator tasks, configuration of static routes is to be avoided, especially in a large network. However, certain circumstances make static routing more appropriate. For example, static routes can be used:

- To define a default route or a route that is not being advertised within a network using a dynamic routing protocol
- To replace exterior gateway protocols when trying to avoid:
  - The cost of routing protocol traffic between ASs
  - Complex routing policies
- In conjunction with a routing daemon to provide backup routes when the daemon cannot find a dynamic route to the destination. In this case, the static route must be configured as replaceable.

If static routing is used, only the PROFILE.TCPIP data set has to be updated with either the BEGINROUTES or GATEWAY statements. The BEGINROUTES statement is recommended to define static routes due to its ease of use and additional functionality. Additionally, if static routes are to be replaceable by OMPROUTE, the BEGINROUTES configuration statement must be used. GATEWAY does not support definition of replaceable static routes, and a static route defined on a GATEWAY statement will not be replaceable by a routing daemon.

The only ways to modify static routes are:

- Replace the routing table using the VARY TCPIP,,OBEYFILE command
- Use incoming ICMP Redirect packets
- Use ICMP Must Fragment packets
- If a static route is defined on a BEGINROUTES statement as being replaceable, it can be replaced if a dynamic route is discovered by OMPROUTE. This is the only way that a static route can be replaced by a dynamic route.

For more information on the VARY TCPIP,,OBEYFILE command, refer to *z/OS Communications Server: IP System Administrator's Commands*. For more information on the IPCONFIG statement, and the IGNOREREDIRECTS and PATHMTUDISC parameters for the IPCONFIG statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Note that the first BEGINROUTES or GATEWAY statement in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, replaces all static routes in the TCP/IP stack routing table (including those destination addresses specified in the BSDROUTINGPARMS section of the PROFILE.TCPIP). Subsequent statements within the same data set append to the routing table. Also, both BEGINROUTES and GATEWAY statements cannot be used within the same data set.

Every interface must have an IP address to transmit or receive packets. Along with the IP address, each interface must have a subnet mask associated with it for routing purposes. The combination of the address and mask will yield the subnet that the interface belongs to and also determines the broadcast address for the interfaces. For interfaces not involved in dynamic routing, the subnet mask can be specified in one of the following ways:

- On the BSDROUTINGPARMS statement in PROFILE.TCPIP (required if running NCPROUTE).
- On the INTERFACE statement in the OMPROUTE configuration file, if running OMPROUTE. Also, if running OMPROUTE, OMPROUTE will always override the stack's BSDROUTINGPARMS and set the subnet mask for every interface it is aware of, even if you did not code the interface to OMPROUTE. Therefore, it

is highly recommended that you either specify every interface to OMPROUTE with the correct subnet mask, or configure OMPROUTE to ignore undefined interfaces.

If you do not specify the subnet mask by one of the methods above, the subnet mask is determined by the stack using the stack routing table. Specifying the subnet mask rather than letting it default is highly recommended.

For point-to-point links, such as CTC and CLAW, a destination address is used to document the other end of a point-to-point connection. For interfaces not involved in dynamic routing, this can be specified in one of the following ways:

- On the BSDROUTINGPARMS statement in PROFILE.TCPIP.
- On the INTERFACE statement in the OMPROUTE configuration file, if running OMPROUTE. Also, if running OMPROUTE, OMPROUTE will always override the stack's BSDROUTINGPARMS and set the destination address for every point-to-point link it is aware of, even if you did not code the interface to OMPROUTE. Therefore, it is highly recommended that you either specify every interface to OMPROUTE with the correct destination address, or configure OMPROUTE to ignore undefined interfaces.

If you do not specify the destination address by one of the methods above, the stack determines the address by using the stack routing table. Specifying the destination address rather then letting it default is highly recommended.

Replaceable static routes: Because replaceable static routes are intended to be last-resort routes, TCP/IP attempts to use them only if no dynamic routes to the destination are available.

If a non-replaceable static route fails validation, even if the reason for the failure is transient like gateway unreachable, the definition for the non-replaceable static route is discarded. However, if a replaceable static route fails validation for a transient reason, the definition of the route is retained. When there are no dynamic routes to the destination, TCP/IP periodically retries to add the replaceable static route to the routing table. Because of these periodic retries multiple EZZ4333I messages may be seen. Retries will be performed no more often than every 30 seconds, and only as long as there are no active routes to the destination in the routing table, and only if at least one new route has been added to the routing table since the last retry. Retries are terminated as soon as a valid route to the destination is installed into the routing table, whether it is dynamic, static, or replaceable static.

Using IPv4 static routing with OMPROUTE

It is recommended that non-replaceable static routes not be used with OMPROUTE because this will prevent those routes from being dynamically updated in response to network topology changes. An exception is when routes need to be defined to destinations which, for some reason, will not be learned dynamically through the routing protocol. If static routes are required, use the BEGINROUTES or GATEWAY statement in PROFILE.TCPIP to define them.

TCP/IP will treat static routes defined as replaceable on BEGINROUTES as last-resort routes. These routes can be replaced by dynamic routes. Additionally, if a static route is replaced with a dynamic route, TCP/IP will always retain knowledge of the static route and reinstall it if the destination becomes unreachable using dynamic routes. It is not necessary for TCP/IP to relearn

replaceable static routes that have been replaced. For this reason, replaceable static routes can be used with OMPROUTE as backup routes, that is, a route to use if nothing is found dynamically.

Another situation where static routes might be required is when multiple, equal-cost routes to a destination are needed and the RIP routing protocol is being used. This is due to the fact that, with the exception of directly attached resources, the RIP protocol will not create multiple, equal-cost routes to a destination. In other words, if multiple adjacent routers are advertising through RIP that they can reach the same destination, RIP will add a route to the TCP/IP route table through only one of those adjacent routers. If it is required that more than one of these routes exist, they would need to be statically configured using the BEGINROUTES or GATEWAY statement in PROFILE.TCPIP.

If an installation has multiple interfaces to a directly attached network and it wants to use one interface for input packets and one for output packets (traffic splitting), the installation must use static routes. To do this, a static route could be defined for one and only one interface, forcing all output packets to use that interface. The other routers on the directly attached network would have to be defined with a similar static route, but for the other interface. Although this is the easiest way to implement traffic splitting, if one of the interfaces fails, a host might become unreachable even though the other physical connection may still exist.

Tip: A more robust way of accomplishing traffic splitting is to use dynamic routes and make one route preferred over the other using the configured interface costs. See step 6 on page 264 for more information.

The BSDROUTINGPARMS statement in PROFILE.TCPIP is not used when the OMPROUTE routing daemon is used. Instead, the interface characteristics, including subnet mask, are defined in the OMPROUTE configuration file.

Rule: If you are using NCPROUTE with OMPROUTE, the BSDROUTINGPARMS statement is required to route Transport PDUs prior to OMPROUTE activation. Because the BSDROUTINGPARMS parameters are overridden by the interface parameters defined in the OMPROUTE configuration, ensure that the interface parameters for the SNALINK or IP/CDLC channel connections are identical in the BSDROUTINGPARMS statement and the OMPROUTE configuration file.

IPv6 static routing

Static routing requires that routes are configured manually for each router or destination; this is a significant reason system administrators avoid this technique if given a choice. Static routing has the disadvantage that network reachability is not dependent on the state of the network itself. If a destination is down, or unreachable through a statically configured route, the static routes remain in the routing table and traffic continues to be sent toward that destination without success.

To minimize network administrator tasks, configuration of static routes is to be avoided, especially in a large network. However, certain circumstances make static routing more appropriate. For example, static routes can be used:

- To define a route that will not be learned dynamically from OMPROUTE or from router advertisements received from routers as part of the router discovery protocol

- In conjunction with dynamic routes to provide backup routes when a dynamic route to the destination cannot be found. In this case, the static route must be configured as replaceable.

If static routing is used, only the PROFILE.TCPIP data set has to be updated with BEGINROUTES statements. The only ways to modify static routes are:

- Replace the routing table using the VARY TCPIP,,OBEYFILE command
- Use incoming ICMPv6 redirect packets
- If a static route is defined on a BEGINROUTES statement as being replaceable, it can be replaced by a dynamic route

Note that the first BEGINROUTES statement in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, replaces all static routes in the TCP/IP stack routing table. Subsequent statements within the same data set append to the routing table.

Rule: If you use static routes and want to honor ICMPv6 redirect messages (that is, you do not code IPCONFIG6 IGNOREREDIRECTS), then you must code the first hop address using the link-local address of the router. This is required since all redirect messages are sent using the router's link-local address, and if the source address of the redirect message does not match the address of the first hop in the routing table, the redirect message will be ignored.

For more information on the VARY TCPIP,,OBEYFILE command, refer to *z/OS Communications Server: IP System Administrator's Commands*. For more information on the IPCONFIG6 statement, and the IGNOREREDIRECTS parameter on the IPCONFIG6 statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Replaceable static routes: Since replaceable static routes are intended to be last-resort routes, TCP/IP only attempts to use them if no dynamic routes to a destination are available. If a non-replaceable static route fails validation, even if the reason for the failure is transient (for example, gateway unreachable), the definition for the non-replaceable static route is discarded. However, if a replaceable static route fails validation for a transient reason, the definition of the route is retained. When there are no dynamic routes to the destination, TCP/IP periodically retries to add the replaceable static route to the routing table. Because of these periodic retries, multiple EZZ4348I messages might be seen. Retries are performed at the most every 30 seconds, as long as there are no active routes to the destination in the routing table and at least one new route has been added to the routing table since the last retry. Retries are terminated as soon as a valid route to the destination is installed into the routing table, whether it is a dynamic, static, or replaceable static route.

Using IPv6 static routing with router advertisements

The use of non-replaceable static routes with IPv6 router discovery, when those routes are to destinations that will be learned through received router advertisements, is not recommended. Defining these non-replaceable static routes prevents them from being dynamically updated in response to network topology changes. Examples of routes that are not learned through router advertisements are routes for which the destination address is a specific host address and non-default indirect routes.

TCP/IP treats replaceable static routes as last-resort routes. These routes can be replaced by dynamic (router discovery) routes. In addition, if a replaceable static

route is replaced with a dynamic route, TCP/IP always retains knowledge of the replaceable static route and can reinstall it if the destination becomes unreachable using dynamic routes. It is not necessary for TCP/IP to relearn replaceable static routes that have been replaced. For this reason, replaceable static routes can be used with IPv6 router discovery as backup routes, for use if nothing is learned dynamically.

Using IPv6 static routing with OMPROUTE

Non-replaceable static routes are not recommended for use with OMPROUTE because this prevents those routes from being dynamically updated in response to network topology changes. An exception is when routes need to be defined to destinations that, for some reason, will not be learned dynamically through the routing protocol. If static routes are required, use the BEGINROUTES statement in PROFILE.TCPIP to define them.

TCP/IP treats static routes defined as replaceable on BEGINROUTES as last-resort routes. These routes can be replaced by dynamic routes. Additionally, if a replaceable static route is replaced with a dynamic route, TCP/IP always retains knowledge of the replaceable static route and reinstalls it if the destination becomes unreachable using dynamic routes. It is not necessary for TCP/IP to relearn replaceable static routes that have been replaced. For this reason, replaceable static routes can be used with OMPROUTE as backup routes, that is, a route to use if nothing is found dynamically.

Another situation where static routes might be required is when multiple, equal-cost routes to a destination are needed and the IPv6 RIP routing protocol is being used. This is due to the fact that, with the exception of directly attached resources, the IPv6 RIP protocol does not create multiple, equal-cost routes to a destination. In other words, if multiple adjacent routers are advertising through IPv6 RIP that they can reach the same destination, IPv6 RIP will add a route to the TCP/IP route table through only one of those adjacent routers. If it is required that more than one of these routes exist, they would need to be statically configured using the BEGINROUTES statement in PROFILE.TCPIP. An example of this would be, in Figure 36 on page 220, if you wanted TCPCS4 to have two routes to the 2001:0DB8:0:A10::/60 prefix (through router A and router B).

If an installation has multiple interfaces to a directly attached network and wants to use one interface for input packets and one for output packets (traffic splitting), the installation could use static routes. To do this, a static route could be defined for one and only one interface, forcing all output packets to use that interface. The other routers on the directly attached network would have to be defined with a similar static route, but for the other interface. Although this is the easiest way to implement traffic splitting, if one of the interfaces fails, a host might become unreachable even though the other physical connection might still exist.

Tip: A more robust way of accomplishing traffic splitting is to use dynamic routes and make one route preferred over the other through the configured interface costs. For more information, see step 6 on page 264.

Static routing configuration examples

The following sections illustrate static routing configuration examples.

z/OS TCPCS4

Static route statements for z/OS TCPCS4

```
BEGINRoutes ;first BEGINRoutes in the profile
;
;Network/mask                  FirstHop      LinkName         PacketSize
Route 9.67.106.0/24            =             CTC4TO7         MTU 1500       ;route1
Route 9.67.105.0/24            =             CTC4TO8         MTU 1500       ;route2
Route 9.67.101.0/24            =             CTC4TO3         MTU 1500       ;route3
Route 9.67.108.0/24            =             CTC4TO2         MTU 1500       ;route4
Route 9.67.107.0/24            9.67.106.7    CTC4TO7         MTU 1500       ;route5
Route 7.7.7.7/32               9.67.106.7    CTC4TO7         MTU 1500       ;route6
Route 9.67.103.0/24            9.67.101.3    CTC4TO3         MTU 1500       ;route7
Route 9.67.103.0/24            9.67.106.7    CTC4TO7         MTU 1500       ;route8
Route 30.1.1.0/24              9.67.106.7    CTC4TO7         MTU 1500       ;route9
Route 10.1.1.0/24              9.67.108.2    CTC4TO2         MTU 1500       ;route10
Route 130.200.0.0/14           9.67.101.3    CTC4TO3         MTU 1500       ;route11
Route 130.200.0.0/14           9.67.105.8    CTC4TO8         MTU 1500       ;route12
Route 130.203.0.0/16           9.67.105.8    CTC4TO8         MTU 1500       ;route13
Route DEFAULT                  9.67.106.7    CTC4TO7         MTU 1500       ;route14
;
;Destination/PrefixLen         FirstHop      Interface        PacketSize
Route FE80::1:2:3:3/128        =             OSAQDIO46       MTU 5000 REPL  ;route15
Route FE80::1:2:3:4/128        =             OSAQDIO46       MTU 5000 REPL  ;route16
Route 2001:0DB8:0:A1B::/64     FE80::1:2:3:3 OSAQDIO46       MTU 5000 REPL  ;route17
Route 2001:0DB8:0:A1C::/64     FE80::1:2:3:4 OSAQDIO46       MTU 5000 REPL  ;route18
Route DEFAULT6                 FE80::1:2:3:4 OSAQDIO46       MTU 5000 REPL  ;route19
EndRoutes
;
```

Notes:

1. In the BEGINROUTES block, the netmask can be specified by a /xx. This number, denoted by xx, represents the number of significant bits in the netmask. For example:

   ```
   /16 = 16 significant bits = 11111111 11111111 00000000 00000000 = 255.255.0.0
   ```

   For IPv6, you must specify the prefix length of the route using the /xxx notation.
2. For direct routes, use an equals symbol (=) for the first hop.

BSDROUTINGPARMS statements for z/OS TCPCS4

```
BSDRoutingParms TRUE ; Shown only for completeness
; Linkname    MTU    Metric    Subnet Mask     Dest Address
  CTC4TO8     1500        0    255.255.255.0   0
  CTC4TO7     1500        0    255.255.255.0   0
  CTC4TO3     1500        0    255.255.255.0   0
  CTC4TO2     1500        0    255.255.255.0   0
  VIPA1A      1500        0    255.255.255.252 0
  EndBSDRoutingParms
;
```

z/OS TCPCS7

Static route statements for z/OS TCPCS7

```
BEGINRoutes
;
;Network/mask              FirstHop      LinkName   PacketSize
Route 9.67.106.0/24        =             CTC7TO4   MTU 1500       ;route1
Route 9.67.100.0/24        =             CTC7TO8   MTU 1500       ;route2
Route 9.67.102.0/24        =             CTC7TO3   MTU 1500       ;route3
Route 9.67.103.0/24        =             CTC7TO6   MTU 1500       ;route4
Route 9.67.107.0/24        =             CTC7TO5   MTU 1500       ;route5
Route 4.4.4.4/32           9.67.106.4    CTC7TO4   MTU 1500       ;route6
Route 10.1.1.0/24          9.67.106.4    CTC7TO4   MTU 1500       ;route7
Route 20.1.1.0/24          9.67.107.5    CTC7TO5   MTU 1500       ;route8
Route 30.1.1.0/24          9.67.103.6    CTC7TO6   MTU 1500       ;route9
Route 130.200.0.0/14       9.67.100.8    CTC7TO8   MTU 1500       ;route10
```

```
Route 130.200.0.0/14       9.67.102.8    CTC7TO3   MTU 1500       ;route11
Route 130.203.0.0/16       9.67.102.3    CTC7TO3   MTU 1500       ;route12
Route DEFAULT              9.67.107.5    CTC7TO5   MTU 1500       ;route13
;
;Destination/PrefixLen     FirstHop      Interface  PacketSize
Route FE80::1:2:3:3/128    =             OSAQDIO76 MTU 5000 REPL  ;route14
Route FE80::1:2:3:4/128    =             OSAQDIO76 MTU 5000 REPL  ;route15
Route 2001:0DB8:0:A1B::/64 FE80::1:2:3:3 OSAQDIO76 MTU 5000 REPL  ;route16
Route 2001:0DB8:0:A1C::/64 FE80::1:2:3:4 OSAQDIO76 MTU 5000 REPL  ;route17
Route DEFAULT6             FE80::1:2:3:4 OSAQDIO76 MTU 5000 REPL  ;route18
EndRoutes
```

BSDROUTINGPARMS statements for z/OS TCPCS7

```
BSDRoutingParms TRUE
; Linkname    MTU    Metric    Subnet Mask     Dest Address
  CTC7TO8     1500        0    255.255.255.0   0
  CTC7TO3     1500        0    255.255.255.0   0
  CTC7TO6     1500        0    255.255.255.0   0
  CTC7TO4     1500        0    255.255.255.0   0
  CTC7TO5     1500        0    255.255.255.0   0
  VIPA1A      1500        0    255.255.255.252 0
  EndBSDRoutingParms
;
```

The sample configuration has an IPv4 supernet route for 130.200.0.0. An IPv4 supernet route means that the netmask for the route is smaller than the class netmask. In this case, 130.200.0.0 is a class B address. The default netmask for class B is 255.255.0.0. The netmask used for this sample is 255.252.0.0, which is less than 255.255.0.0, hence making this a supernet route. In routing, the stack prefers a route that has the most bits in common. Therefore, the stack chooses a route in the following order:

1. If a route exists to the destination address (a host route), it is chosen.
2. At this point, the route chosen depends upon the version of IP being used:
   - For IPv4:
     a. If subnet, network, or supernetwork routes exist to the destination, the route with the most specific network mask (the mask with the most bits on) is chosen.
     b. If the destination is a multicast destination and a multicast default route exists, that route is chosen.
   - For IPv6, if prefix routes exist to the destination, the route with the most specific prefix is chosen.
3. Default routes are chosen when no other route exists to a destination.

For example, for TCPCS4 (and when trying to reach 130.200.0.0), route12 in the list is used, which is the supernet route 130.200.0.0 with mask 255.252.0.0. If applying the mask of that route, 255.252.0.0, to the destination IP address, 130.200.0.0, the result is 130.200.0.0, which is the IP address of this route. Now, when trying to reach destination 130.203.5.2, the stack would use route13 in the list, which is a network route for 130.203.00 with mask 255.255.0.0. If applying the mask of that route, 255.255.0.0, to the destination IP address, 130.203.5.2, the result is 130.203.00, which is the IP address of this route.

For TCPCS4, route7 and route8 are examples of equal cost multipath routes to get to 9.67.103.0 subnet. This means that TCPCS4 has two different routes to get to this destination. If IPCONFIG MULTIPATH is not enabled, then only route7 will be used as long as it is active. This is because the stack chooses the first route and

ignores route8. If route7 becomes inactive, then the stack will switch and use route8. If MULTIPATH is enabled, then the stack will use both routes according to the MULTIPATH specification.

In the preceding example, all of the IPv4 links have a subnet mask of 255.255.255.0 because this is what is specified for the links in the BSDROUTINGPARMS. Therefore, to determine the broadcast addresses for link CTC4TO3, AND the IP Address, 9.67.101.4, and the subnet mask, 255.255.255.0, to yield the subnet for this link, 9.67.101.0. Then, OR the subnet, 9.67.101.0, with the complement of the subnet mask, 0.0.0.255. This determines that the broadcast address for this link is 9.67.101.255.

For TCPCS4, route15 and route16 would be selected to reach host FE80::1:2:3:3 and host FE80::1:2:3:4 respectively. Route17 and route18 would be selected to reach any IPv6 address that had the first 64 bits of 2001:0DB8:0:A1B and 2001:0DB8:0:A1C respectively. Route19 would be selected for any other IPv6 destination.

Rules:

1. All IPv4 IP addresses must follow Classless Inter-Domain Routing (CIDR) convention that requires the actual mask to be one or more on-bits followed by zero or more off-bits. On-bits cannot be followed by off-bits followed by on-bits. Therefore, a mask of 255.255.254.0 is valid (an actual mask of FFFFE00), but a mask of 255.255.253.0 is not valid (an actual mask of FFFFD00) because 253 is 11111101.
2. VIPA links or VIPA interfaces are not allowed on BEGINROUTES statements.
3. You must have a Direct route to a specific IP Address before using that IP Address as the first-hop for indirect routes. A direct route is a route to a destination that is directly connected to the stack by an interface. An indirect route is a route to a destination that is not directly connected, and therefore a router is used to reach that destination. In the preceeding example, for TCPCS4, the subnet route for 9.67.101.0 is directly connected to TCPCS4 by link CTC4T03, and the host route for FE80::1:2:3:3 is directly connected to TCPCS4 by interface OSAQDIO46. However, the subnet route for 9.67.103.0 is indirectly connected and the router used to reach that destination is 9.67.106.7 and/or 9.67.101.3, depending on the MULTIPATH definition.
4. DEFAULT and DEFAULT6 routes are always indirect routes and therefore must always have a first hop address specified.

IPv4 dynamic routing

This section describes the following for IPv4:

- IPv4 routing daemons
- IPv4 dynamic routing using OMPROUTE

IPv4 routing daemons

Daemon is a UNIX term for a background server process. Daemons are used for dynamic routing. For z/OS Communications Server IP, there is a multiprotocol routing daemon available:

OMPROUTE

For IPv4, OMPROUTE supports the RIP Version 1, RIP Version 2, and OSPF routing protocols. You can send RIP Version 1 or RIP Version 2, but not both at the same time on a single interface. However, you can configure a RIP interface to receive both versions.

Guideline: If OROUTED was used in a prior release and the RIP protocol is still the preferred dynamic routing method in your host configuration, use OMPROUTE to provide RIP support. For more information on migrating from OROUTED to OMPROUTE, refer to *z/OS Migration*.

IPv4 dynamic routing using OMPROUTE

For IPv4, OMPROUTE implements the OSPF protocol described in RFC 1583 (*OSPF Version 2*), the OSPF subagent protocol described in RFC 1850 (*OSPF Version 2 Management Information Base*), and the RIP protocols described in RFC 1058 (*Routing Information Protocol*) and in RFC 1723 (*RIP Version 2 - Carrying Additional Information*). It provides an alternative to the static TCP/IP gateway definitions. The MVS host running with OMPROUTE becomes an active OSPF or RIP router in a TCP/IP network. Either or both of these routing protocols can be used to dynamically maintain the host routing table. For example, OMPROUTE can detect when a route is created, is temporarily unavailable, or if a more efficient route exists. If both OSPF and RIP protocols are used simultaneously, OSPF routes will be preferred over RIP routes to the same destination.

Supported protocols

Open Shortest Path First (OSPF): OSPF is classified as an Interior Gateway Protocol (IGP). This means that it distributes routing information between routers belonging to a single autonomous system (AS), a group of routers all using a common routing protocol. The OSPF protocol is based on link-state or shortest path first (SPF) technology. It has been designed expressly for the TCP/IP Internet environment, including explicit support for IP subnetting and the tagging of externally derived routing information.

OSPF performs the following tasks:

Multiple routes
Provides support for up to 16 equal-cost routes.

Authentication
Provides for the authentication of routing updates.

IP multicast
Uses IP multicast when sending or receiving the updates.

Allows network grouping
Allows sets of networks to be grouped together. Such a grouping is called an area. The topology of an area is hidden from the rest of the autonomous system. This method of hiding information enables a significant reduction in routing traffic. Also, routing within the area is determined only by the area's own topology, lending the area protection from bad routing data. An area is a generalization of an IP subnetted network.

IP subnet configuration
Enables the flexible configuration of IP subnets. Each route distributed by OSPF has a destination and mask. Two different subnets of the same IP network number may have different sizes (that is, different masks). This is commonly referred to as variable length subnetting. A packet is routed to the best (longest or most specific) match. Host routes are considered to be subnets whose masks are all ones (0xFFFFFFFF).

Authenticate OSPF protocol exchanges
Can be configured such that all OSPF protocol exchanges are authenticated. This means that only trusted routers can participate in the

autonomous system's routing. A single authentication scheme is configured for each physical link. This enables some links to use authentication while others do not.

OSPF is a dynamic routing protocol. It quickly detects topological changes in the AS (such as router interface failures) and calculates new loop-free routes after a period of convergence. This period of convergence is short and involves a minimum of routing traffic as compared to the RIP protocol.

In a link-state routing protocol, each router maintains a database describing the autonomous system's topology. Each individual piece of this database is a particular router's local state (for example, the router's usable interfaces and reachable neighbors). The router distributes its local state throughout the autonomous system by flooding.

To generate routes, all routers run the exact same algorithm, in parallel. From the topological database, each router constructs a tree of shortest paths with itself as root. This shortest-path tree gives the route to each destination in the autonomous system. Externally derived routing information (for example, routes learned from the RIP protocol) appears on the tree as leaves. When several equal-cost routes to a destination exist, the routes (up to 16) are added to the TCP/IP stack's route table. The TCP/IP stack uses these equal-cost routes according to the IPCONFIG MULTIPATH statement.

Externally derived routing data (for example, routes learned from the RIP protocol) is passed transparently throughout the autonomous system. This externally derived data is kept separate from the OSPF protocol's link state data. Each external route can also be tagged by the advertising router, but not by OMPROUTE, enabling the passing of additional information between routers on the boundaries of the autonomous system. OMPROUTE does pass tags created by others. For information on configuring OSPF, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249.

Routing Information Protocol (RIP): RIP is an Interior Gateway Protocol (IGP) designed to manage a relatively small network. RIP is based on the Bellman-Ford or the distance-vector algorithm. RIP has many limitations and is not suitable for every TCP/IP environment. Before using the RIP function in OMPROUTE, read RFCs 1058 and 1723 to decide if RIP can be used to manage the routing tables of your network. For more information about RFCs 1058 and 1723, see Appendix F, "Related protocol specifications (RFCs)," on page 1223.

RIP uses the number of hops, or hop count, to determine the best possible route to a host or network. The term hop count is also referred to as the metric. In RIP, a hop count of 16 means infinity, or that the destination cannot be reached. This limits the longest path in the network that can be managed by RIP to 15 gateways.

A RIP router broadcasts routing information to its directly connected networks every 30 seconds. It receives updates from neighboring RIP routers every 30 seconds and uses the information contained in these updates to maintain the routing table. If an update has not been received from a neighboring RIP router in 180 seconds, a RIP router assumes that the neighboring RIP router is down, sets all routes through that router to a metric of 16 (infinity), and stops using those routes when routing IP packets. If an update has still not been received from the neighboring RIP router after another 120 seconds, the RIP router deletes from the routing table all of the routes through that neighboring RIP router.

RIP Version 2 is an extension of RIP Version 1 and provides the following features:

Route Tags
The route tags are used to separate *internal* RIP routes (routes for networks within the RIP routing domain) from *external* RIP routes, which may have been imported from an EGP (external gateway protocol) or another IGP. OMPROUTE does not generate route tags, but preserves them in received routes and readvertises them when necessary.

Variable subnetting support
Variable length subnet masks are included in routing information so that dynamically added routes to destinations outside subnetworks or networks can be reached.

Immediate next hop for shorter paths
Next hop IP addresses, whenever applicable, are included in the routing information to eliminate packets being routed through extra hops in the network.

Multicasting to reduce load on hosts
IP multicast address 224.0.0.9, reserved for RIP Version 2 packets, is used to reduce unnecessary load on hosts which are not listening for RIP Version 2 messages. This support is dependent on interfaces that are multicast-capable.

Authentication for routing update security
Authentication keys can be configured for inclusion in outgoing RIP Version 2 packets. Incoming RIP Version 2 packets are checked against the configured keys.

Configuration switches for RIP Version 1 and RIP Version 2 packets
Configuration parameters allow for controlling which version of RIP packets are to be sent or received over each interface.

Supernetting support
The supernetting feature is part of Classless InterDomain Routing (CIDR). Supernetting provides a way to combine multiple network routes into fewer supernet routes, thus reducing the number of routes in the routing table and in advertisements.

For configuration information for RIP, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249.

IPv6 dynamic routing

This section describes the following for IPv6:

- IPv6 dynamic routing using router discovery
- IPv6 dynamic routing using OMPROUTE

IPv6 dynamic routing using router discovery

Enabling IPv6 router discovery in z/OS Communications Server requires no additional z/OS Communications Server configuration. All that is needed is at least one IPv6 interface that is defined and started, and at least one adjacent router through that interface that is configured for IPv6 router discovery. If these things exist, then z/OS Communications Server begins receiving router advertisements from the adjacent routers. Depending on the configuration in the adjacent routers, the following types of routes may be learned from the received router advertisements:

- Default route for which the originator of the router advertisement is the next hop
- Direct routes (no next hop) to prefixes that reside on the link shared by z/OS Communications Server and the originator of the router advertisement.

Multiple default routes and multiple direct prefix routes to a single prefix may be learned through router advertisements. If an adjacent router resides on a link onto which z/OS Communications Server TCPIP has multiple IPv6 interfaces, there will be multiple routes to each route learned through the adjacent router's router advertisement (one route through each interface onto the link). Also, if default routes are learned from the router advertisements originated by multiple adjacent routers, there will be multiple default routes (one with each of these adjacent routers as next hop). When this condition of multiple routes exists, TCP/IP will use those routes according to the setting of the MULTIPATH parameter on the IPCONFIG6 statement.

If there are non-replaceable static routes to the destinations in the router advertisements, the dynamic routes will not be added to the stack routing table.

IPv6 dynamic routing using OMPROUTE

For IPv6, OMPROUTE implements the IPv6 RIP protocol described in RFC 2080 (*RIPng for IPv6*) and the IPv6 OSPF protocol described in RFC 2740 (*OSPF for IPv6*). It provides an alternative to the static TCP/IP gateway definitions. The MVS host running with OMPROUTE becomes an active OSPF or RIP router in a TCP/IP network. Either or both of these routing protocols can be used to dynamically maintain the host IPv6 routing table. For example, OMPROUTE can detect when a route is created, is temporarily unavailable, or if a more efficient route exists. If both IPv6 OSPF and IPv6 RIP protocols are used simultaneously, IPv6 OSPF routes will be preferred over IPv6 RIP routes to the same destination.

IPv6 OSPF protocol

IPv6 OSPF is classified as an Interior Gateway Protocol (IGP). This means that it distributes routing information between routers belonging to a single autonomous system (AS), a group of routers all using a common routing protocol. The IPv6 OSPF protocol is based on link-state or shortest path first (SPF) technology.

IPv6 OSPF is a dynamic routing protocol. It quickly detects topological changes in the AS (such as router interface failures) and calculates new loop-free routes after a period of convergence. This period of convergence is short and involves a minimum of routing traffic as compared to the IPv6 RIP protocol. However, it does generally require more CPU resources on participating routers.

In IPv6 OSPF, sets of networks can be placed together in groups called areas. The topology of an area is hidden from the rest of the AS. This method of hiding information enables a significant reduction in routing traffic. Also, routing within the area is determined only by the area's own topology, lending the area protection from bad routing data.

Each IPv6 OSPF router maintains a database describing the autonomous system's topology. Each individual piece of this database is a particular router's local state (for example, the router's usable interfaces and reachable neighbors). The router distributes its local state information by flooding. The routing information in an area is summarized and advertised into adjacent areas, allowing for the generation

of interarea routes. This summarization and advertising of routing information between areas is the responsibility of the routers that reside on the border between areas (called area border routers).

To generate routes, all routers run the exact same algorithm, in parallel. From the topological database, each router constructs a tree of shortest paths with itself as root. This shortest-path tree gives the route to each destination in the autonomous system. Externally derived routing information (for example, routes learned from the IPv6 RIP protocol) appears on the tree as leaves. When several equal-cost routes to a destination exist, the routes (up to 16) are added to the TCP/IP stack's route table. The TCP/IP stack uses these equal-cost routes according to the IPCONFIG6 MULTIPATH statement.

Externally derived routing data is passed transparently throughout the autonomous system. This externally derived data is kept separate from the IPv6 OSPF protocol's link state data. Each external route can also be tagged by the advertising router, but not by OMPROUTE, enabling the passing of additional information between routers on the boundaries of the autonomous system. OMPROUTE does pass tags created by others. For information on configuring IPv6 OSPF, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249.

IPv6 RIP protocol

IPv6 RIP is an Interior Gateway Protocol (IGP) designed to manage a relatively small network. IPv6 RIP is based on the Bellman-Ford or the distance-vector algorithm. IPv6 RIP has limitations and is not suited for every TCP/IP environment. Before using the IPv6 RIP function in OMPROUTE, read RFC 2080 to decide if RIP can be used to manage the IPv6 routing tables of your network. For more information about RFC 2080, see Appendix F, "Related protocol specifications (RFCs)," on page 1223.

IPv6 RIP uses the number of hops, or hop count, to determine the best possible route to a host or network. The term hop count is also referred to as the metric. In IPv6 RIP, a hop count of 16 means infinity, or that the destination cannot be reached. This limits the longest path in the network that can be managed by IPv6 RIP to 15 gateways.

A IPv6 RIP router broadcasts routing information to its directly connected networks every 30 seconds. It receives updates from neighboring IPv6 RIP routers every 30 seconds and uses the information contained in these updates to maintain the IPv6 routing table. If an IPv6 RIP update has not been received from a neighboring router in 180 seconds, an IPv6 RIP router assumes that the neighboring router is down, sets all IPv6 RIP routes through that router to a metric of 16 (infinity), and stops using those routes when routing IP packets. If an update has still not been received from the neighboring router after another 120 seconds, the IPv6 RIP router deletes from the IPv6 routing table all of the IPv6 RIP routes through that neighboring router.

Next hop IP addresses, whenever applicable, are included in IPv6 RIP updates to eliminate packets being routed through extra hops in the network. IPv6 multicast address FF02::9, reserved for IPv6 RIP packets, is used to reduce unnecessary load on hosts that are not listening for IPv6 RIP messages.

For configuration information for IPv6 RIP, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249.

OMPROUTE configuration

Run-time environment

OMPROUTE is a z/OS UNIX application, and it requires the Hierarchical File System (HFS) to operate. It can be started from an MVS started procedure, from the z/OS shell, or from AUTOLOG (see step 2 on page 240 for restrictions on using AUTOLOG to start OMPROUTE). OMPROUTE must be started by a RACF-authorized user ID, and it must reside in an APF authorized library.

OMPROUTE uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and syslogd are used for major events such as initialization, termination, and error conditions. CTRACE is used for tracing the receipt and transmission of OSPF/RIP packets and communications between OMPROUTE and the TCP/IP stack. In addition, OMPROUTE can be configured to use CTRACE for detailed tracing and debugging. STDOUT is used for detailed tracing and debugging.

When syslogd is active and /dev/console is defined in the syslog.conf file, OMPROUTE logging messages are directed to syslogd and to the MVS console. If syslogd is active and /dev/console is not defined, OMPROUTE logging messages are directed to syslogd only.

When syslogd is inactive, OMPROUTE logging messages are directed to the OMPROUTE JES joblog and to the MVS console.

If OMPROUTE tracing or debugging is enabled, the output can be directed to STDOUT, the JES joblog, or OMPROUTE CTRACE. Additionally, the logging messages (that is, interface initialization messages) can be found in syslogd, provided it is active at the time of tracing or debugging. Syslogd directs urgent OMPROUTE messages to the MVS console.

Tip: If you enable OMPROUTE tracing without syslogd active, large amounts of data can be written to the console.

OMPROUTE uses a standard message catalog. The message catalog must be in the HFS. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

Configuration of OMPROUTE is accomplished using an OMPROUTE configuration file. For details on the statements in the OMPROUTE configuration file, refer to *z/OS Communications Server: IP Configuration Reference*.

Display of OMPROUTE information is performed using the DISPLAY or MODIFY command. Modification of OMPROUTE information is performed using the MODIFY command. For details on OMPROUTE's DISPLAY and MODIFY commands, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Language Environment run-time considerations

When starting OMPROUTE from a started or cataloged procedure, it is usually recommended that OMPROUTE be started directly from the SEZALOAD data set using PGM=OMPROUTE. However, there is a situation where it might be desirable to start OMPROUTE using BPXBATCH.

When OMPROUTE is started using PGM=OMPROUTE, the STDENV DD card, if used, is passed directly to the OMPROUTE program. The Language Environment does not get access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS= environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM= parameter and the options must be specified before any OMPROUTE options. However, the PARM= statement allows a maximum of 100 characters. If the desired Language Environment run-time options plus OMPROUTE parameters exceeds 100 characters, consideration should be given to using BPXBATCH to start OMPROUTE. When PGM=BPXBATCH is used, the Language Environment environment variable _CEE_RUNOPTS can be included on the STDENV DD card to specify run-time options in excess of 100 characters long.

OMPROUTE tuning considerations

It might be desirable to tune OMPROUTE's storage usage when running in complex network environments or when running with traces enabled for long periods of time. To determine if tuning is necessary, use the Language Environment run-time options RPTSTG and RPTOPTS. For information on these options and the reports they generate, refer to z/OS Language Environment documentation.

Multiple TCP/IP stacks

A one-to-one relationship exists between an instance of OMPROUTE and a stack. OSPF, RIP, IPv6 OSPF, and IPv6 RIP support on multiple stacks requires multiple instances of OMPROUTE.

TCP/IP stack routing table management

OMPROUTE's job is limited to the management of the TCP/IP stack routing table. OMPROUTE is not involved in the actual routing decisions made by the TCP/IP stack when routing a packet to its destination.

If IPv4 interfaces are defined in the OMPROUTE configuration file as OSPF or RIP interfaces, all IPv4 dynamic routes are deleted from the stack's IPv4 routing table upon initialization of OMPROUTE. If IPv6 interfaces are defined in the OMPROUTE configuration file as OSPF or RIP interfaces, all IPv6 dynamic routes, excluding those dynamic routes learned through IPv6 router discovery, are deleted from the stack's IPv6 routing table upon initialization of OMPROUTE. OMPROUTE then repopulates the stack routing tables that it cleared, using information learned through the routing protocols.

IPv4 Internet Control Message Protocol (ICMP) redirects are ignored when OMPROUTE is active and there are IPv4 interfaces configured to OMPROUTE as RIP or OSPF interfaces. IPv6 ICMP redirects are ignored when OMPROUTE is active and there are IPv6 interfaces configured to OMPROUTE as RIP or OSPF interfaces.

OMPROUTE does not make use of the BSDROUTINGPARMS statement. Instead, the IPv4 Maximum Transmission Unit (MTU), subnet mask, and destination address parameters are configured using the OSPF_INTERFACE, RIP_INTERFACE, and INTERFACE statements in the OMPROUTE configuration file. The MTU for IPv6 interfaces is learned from the TCP/IP stack and therefore is not a parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, and IPV6_INTERFACE statements.

Result: The NETSTAT DEVlinks report displays these parameters under `BSD Routing Parms` even though the BSDROUTINGPARMS statement is not defined.

Using RIP, IPv6 RIP, OSPF, and IPv6 OSPF with OMPROUTE

When OMPROUTE is initialized, it uses the OMPROUTE configuration file to determine which routing protocols will be enabled. If at least one OSPF interface is configured, the OSPF protocol is enabled. If at least one RIP interface is configured, RIP is enabled. If at least one IPv6 RIP interface is configured, IPv6 RIP is enabled. If at least one IPv6 OSPF interface is configured, IPv6 OSPF is enabled. If OMPROUTE is started with no interfaces defined for a particular protocol, that protocol is disabled until one of the following occurs:

- OMPROUTE is stopped and restarted with a configuration file containing at least one interface of the specific type.
- OMPROUTE is dynamically reconfigured using the MODIFY command with a configuration file containing at least one interface of the specific type.

When OMPROUTE is configured for both the OSPF and RIP protocols (either IPv4 or IPv6), routes that are learned through the OSPF protocol take precedence over routes learned through the RIP protocol.

The OSPF and RIP protocols are communicated over IPv4 interfaces that are defined with the OSPF_INTERFACE and RIP_INTERFACE configuration statements, respectively. An IPv4 interface involved in the communication of neither the RIP nor the OSPF protocol should be configured to OMPROUTE with the INTERFACE configuration statement, unless Ignore_Undefined_Interfaces=YES is coded on the Global_Options configuration statement. OMPROUTE supports a maximum of 255 real, physical, IPv4 interfaces (that is, interfaces on which data can actually be sent and received). There is no theoretical limit on how many VIPAs can be configured, though there are practical limits imposed by network design. For special VIPA considerations, see step 4 on page 253.

The IPv6 OSPF and IPv6 RIP protocols are communicated over IPv6 interfaces that are defined with the IPV6_OSPF_INTERFACE and IPV6_RIP_INTERFACE configuration statements, respectively. An IPv6 interface not involved in the communication of the IPv6 OSPF or IPv6 RIP protocol can be configured to OMPROUTE with the IPV6_INTERFACE configuration statement. If default values are acceptable and you do not need to define additional prefixes to an IPv6 interface not involved in the communication of the IPv6 OSPF or IPv6 RIP protocol, it is not necessary to configure the interface to OMPROUTE at all. OMPROUTE will learn about the interface and its MTU value from the stack and use default values for other parameters. This is different from IPv4, where all interfaces should be configured to prevent OMPROUTE from using default values for MTU size and subnet mask, unless Global_Options is coded with Ignore_Undefined_Interfaces=YES.

OMPROUTE allows for the generation of multiple, equal-cost routes to a destination. For OSPF and IPv6 OSPF, up to 16 multiple equal-cost routes are allowed. For RIP and IPv6 RIP, multiple equal-cost routes are supported only to directly connected destinations over redundant interfaces.

Special considerations

Token-ring multicast

If OMPROUTE will be communicating through the OSPF or RIP Version 2 protocol over a token ring media, and there will be routers attached to that token ring that

are not listening (at the DLC layer) for the token ring multicast MAC address 0xC000.0004.0000, the following TRANSLATE statement is required in the PROFILE.TCPIP:

```
TRANSLATE 224.0.0.0 IBMTR FFFFFFFFFFFF linkname
```

Without this statement, OSPF and RIP Version 2 multicast packets are discarded at the DLC layer by those routers that are not listening for the token ring multicast MAC address.

Virtual IP addresses (VIPA)

OMPROUTE is enhanced with virtual IP addressing (VIPA) to handle network interface failures by switching to alternate paths. The VIPA routes are included in the OSPF, RIP, IPv6 OSPF, and IPv6 RIP advertisements to adjacent routers. Adjacent routers learn about VIPA routes from the advertisements and can use them to reach the destinations at the MVS host.

Service policy

If service policy is going to be used to restrict access to neighbors on point-to-multipoint interfaces (for example MPCPTP interfaces including XCF and IUTSAMEH connections) for temporary intervals, those neighbors must be explicitly defined on the OSPF_INTERFACE or RIP_INTERFACE statement. Otherwise, OMPROUTE might not be able to communicate with those neighbors when the access restriction expires.

Multiple equal-cost routes

When IPCONFIG MULTIPATH or IPCONFIG6 MULTIPATH is specified in PROFILE.TCPIP and multiple routes exist in the TCP/IP route table for a destination, outbound traffic for that destination will be spread across all of the routes. This traffic spreading will be done on either a packet-basis or connection-basis, depending on the parameter specified on IPCONFIG MULTIPATH or IPCONFIG6 MULTIPATH. When OMPROUTE is being used to provide dynamic routing for a TCP/IP stack, multiple routes to the same destination can be dynamically added to the TCP/IP stack's route table, based upon the routing information learned from other routers. These multiple routes will be added when the route calculation for each has resulted in the same route cost value. No more than 16 equal-cost routes will be added for each destination. For RIP and IPv6 RIP, multiple equal-cost routes will be added only to directly-connected destinations over redundant interfaces. The RIP and IPv6 RIP protocols will generate no more than one indirect route to a destination.

Table 15. Multipath route limitations

Multipath route type	BEGINROUTES (Static)	OMPROUTE (OSPF and IPv6 OSPF)	OMPROUTE (RIP and IPv6 RIP)
Direct host	Yes (no limit)	Yes (up to 16)	No
Indirect host	Yes (no limit)	Yes (up to 16)	No
Direct network	Yes (no limit)	Yes (up to 16)	Yes (up to 16 for redundant interfaces)
Indirect network	Yes (no limit)	Yes (up to 16)	No
Default (indirect)	Yes (no limit)	Yes (up to 16)	No

Note: If more than 16 equal-cost multipath routes are desired for OSPF or if multiple equal-cost indirect routes are desired for RIP, use the GATEWAY or BEGINROUTES statement. If more than 16 equal-cost multipath routes are

desired for IPv6 OSPF or if multiple equal-cost indirect routes are desired for IPv6 RIP, use the BEGINROUTES statement.

Sysplex autonomics

If you enable sysplex autonomics, it is very important that you ensure that the WLM policy for the OMPROUTE address space receives sufficient resources in relationship to other work being managed on the system. Under high load conditions it is possible that OMPROUTE, if not properly classified, can trigger an autonomic response from the TCP/IP stack it has an affinity with, resulting in the TCP/IP address space removing itself from the sysplex group. It is recommended that the TCP/IP and OMPROUTE address spaces be placed in the SYSSTC service classification. Classification in another service class will leave the system vulnerable to a Sysplex Distributor outage. For further information on sysplex autonomics, see "Sysplex problem detection and recovery" on page 376.

Steps for configuring OMPROUTE

The steps to configure OMPROUTE are:

1. Create the OMPROUTE configuration file.
2. Reserve the ports, and ensure loopback availability.
3. Update the resolver configuration file.
4. Update the OMPROUTE cataloged procedure.
5. Specify the RIP UDP port numbers in the SERVICES file or data set (if using the RIP or IPv6 RIP protocol).
6. RACF authorize user IDs for starting OMPROUTE.
7. Start syslogd.
8. Update the OMPROUTE environment variables (optional).
9. Create static routes (optional).
10. Configure OSPF authentication (optional, if using the IPv4 OSPF protocol).

Following are the details for the steps to configure OMPROUTE.

1. Create the OMPROUTE configuration file.

   The OMPROUTE configuration file provides information about the host's routing capabilities and TCP/IP interfaces. For more detail about the contents of this file, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249. The following is the search order used by OMPROUTE to locate the configuration data set or file:

   a. If the environment variable, OMPROUTE_FILE, has been defined, OMPROUTE uses the value as the name of an MVS data set or HFS file to access the configuration data. The syntax for an MVS data set name is `//mvs.dataset.name`. The syntax for an HFS file name is `/dir/subdir/file.name`.
   b. `/etc/omproute.conf`
   c. *hlq*.ETC.OMPROUTE.CONF

   A sample configuration file is provided in SEZAINST(EZAORCFG). The configuration files for TCPCS4, TCPCS6, TCPCS7, and TCPCS64 in the sample network are shown in "Sample OMPROUTE configuration files" on page 290. For a description of the syntax rules for the OMPROUTE configuration file, as well as details on each of the configuration statements, refer to *z/OS Communications Server: IP Configuration Reference*.

2. Reserve the ports, and ensure loopback availability.

   RIP protocol

   If the RIP protocol of OMPROUTE is going to be used, UDP port 520 should be reserved for OMPROUTE. If the IPv6 RIP protocol of OMPROUTE is going to be used, UDP port 521 should be reserved for OMPROUTE. This is done by adding the name of the member containing the OMPROUTE cataloged procedure to the PORT statement in PROFILE.TCPIP:

   ```
   PORT
       520 UDP OMPROUTE
       521 UDP OMPROUTE
   ```

   If you want to be able to start OMPROUTE from the z/OS shell, use the special name OMVS as follows:

   ```
   PORT
       520 UDP OMVS
       521 UDP OMVS
   ```

   Autolog considerations for OMPROUTE when using the OSPF protocol

   If a procedure in the AUTOLOG list also has a PORT statement reserving a TCP or UDP port but does not have a listening connection on that port, TCP/IP periodically attempts to cancel that procedure and start it again.

   Therefore, if OMPROUTE is being started with AUTOLOG and only the OSPF or IPv6 OSPF protocol is being used (no RIP or IPv6 RIP protocol, so no listening connection on the RIP or IPv6 RIP UDP port), it is important to do one of the following:

   – Ensure that the RIP UDP port (520) and the IPv6 RIP UDP port (521) are not reserved by the PORT statement in the PROFILE.TCPIP.
   – Add the NOAUTOLOG parameter to the PORT statement in the PROFILE.TCPIP. For example,

     ```
     PORT
      520 UDP OMPROUTE NOAUTOLOG
      521 UDP OMPROUTE NOAUTOLOG
     ```

   Tip: When using only the OSPF or IPv6 OSPF protocol, the auto-start feature of AUTOLOG can be used as described above. However, the monitoring and auto-restart features of AUTOLOG are unavailable due to AUTOLOG's dependence on listening to a TCP or UDP connection, which does not exist with OSPF and IPv6 OSPF.

   If you fail to take one of the above actions, OMPROUTE will be periodically canceled and restarted by TCP/IP.

   IP filtering considerations

   If running IP filtering through firewall technologies, you must allow TCP traffic to and from loopback. This is required because OMPROUTE opens a TCP socket to the loopback address to communicate with the TCP/IP stack. If this communication is blocked by filters or firewalls, OMPROUTE will terminate, as it cannot do any work if it cannot communicate with the TCP/IP stack.

---

3. Update the resolver configuration file.

   The resolver configuration file contains keywords (DATASETPREFIX and TCPIPjobname) used by OMPROUTE. The value assigned to DATASETPREFIX will determine the high-level qualifier (*hlq*). The *hlq* is used in the search order for the OMPROUTE configuration file. If no DATASETPREFIX keyword is found, a default of TCPIP is used. The value

assigned to TCPIPjobname will be used as the name of the TCP/IP stack with which OMPROUTE establishes a connection.

For a description of the search order used by the resolver to locate the resolver configuration file, see "Resolver configuration files" on page 40.

---

4. Update the OMPROUTE cataloged procedure.

   If OMPROUTE is to be started by a procedure, create the cataloged procedure by copying the sample in SEZAINST(OMPROUTE) to your system or recognized PROCLIB. Specify OMPROUTE parameters and change the data set names to suit your local configuration.

```
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: EZBORPRC
//*
//*        5694-A01 (C) Copyright IBM Corp. 1998, 2003
//*        Licensed Materials - Property of IBM
//*        This product contains "Restricted Materials of IBM"
//*        All rights reserved.
//*        US Government Users Restricted Rights -
//*        Use, duplication or disclosure restricted by
//*        GSA ADP Schedule Contract with IBM Corp.
//*        See IBM Copyright Instructions.
//*
//OMPROUTE PROC
//OMPROUTE EXEC PGM=OMPROUTE,REGION=10M,TIME=NOLIMIT,
// PARM=('POSIX(ON)',
//       'ENVAR("_CEE_ENVFILE=DD:STDENV")/')
//*
//* Example of start parameters to OMPROUTE:
//*
//* PARM=('POSIX(ON)',
//*       'ENVAR("_CEE_ENVFILE=DD:STDENV")/-t1 -6t1')
//*
//*        Provide environment variables to run with the
//*        desired stack and configuration.  As an example,
//*        the file specified by STDENV could have these
//*        five lines in it:
//*
//*           RESOLVER_CONFIG=//'SYS1.TCPPARMS(TCPDATA2)'
//*           OMPROUTE_FILE=/u/usernnn/config.tcpcs2
//*           OMPROUTE_DEBUG_FILE=/tmp/logs/omproute.debug
//*           OMPROUTE_IPV6_DEBUG_FILE=/tmp/logs/omprout6.debug
//*           OMPROUTE_DEBUG_FILE_CONTROL=1000,5
//*
//*        For information on the above environment variables,
//*        refer to the IP CONFIGURATION GUIDE.
//*
//STDENV   DD PATH='/u/usernnn/envcs2',
//         PATHOPTS=(ORDONLY)
//*
//*        The stdout stream may be redirected to a HFS file as
//*        shown below.
//*        The PATHOPTS OTRUNC option will clear the stdout file
//*        every time OMPROUTE is started.  If you want to retain
//*        previous stdout information, change it to OAPPEND.
//*
//SYSPRINT DD SYSOUT=*
//*SYSPRINT DD PATH='/tmp/omproute.stdout',
//*            PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//*            PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
//*
//*        The stderr stream may be redirected to a HFS file as
//*        shown below.
//*        The PATHOPTS OTRUNC option will clear the stderr file
//*        every time OMPROUTE is started.  If you want to retain
```

```
//*        previous stderr information, change it to OAPPEND.
//*
//SYSOUT   DD SYSOUT=*
//*SYSOUT   DD PATH='/tmp/omproute.stderr',
//*            PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//*            PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
//*
//CEEDUMP  DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
```

---

5. Specify the RIP UDP port numbers in the SERVICES file or data set (if using the RIP or IPv6 RIP protocol).

   The services file contains the relationship between services and port numbers as described in *z/OS Communications Server: IP Configuration Reference*. The portion of the services file relevant to OMPROUTE is:

   ```
   route           520/udp         router routed
   route           521/udp         ipv6rip
   ```

   The file must exist for the RIP and IPv6 RIP protocols of OMPROUTE to operate.

   For a description of the search order used to locate the services file, see "Configuration files for TCP/IP applications" on page 38.

---

6. RACF authorize user IDs for starting OMPROUTE.

   To reduce risk of an unauthorized user starting OMPROUTE and affecting the contents of the routing table, users who start OMPROUTE must be RACF-authorized to the entity MVS.ROUTEMGR.OMPROUTE and require a UID of zero. To RACF-authorize, the following commands must be entered from a RACF user ID, substituting the authorized user ID on the ID (userid) parameter. The commands in the following example are taken from SEZAINST(EZARACF).

   ```
   RDEFINE OPERCMDS (MVS.ROUTEMGR.OMPROUTE) UACC(NONE)
   PERMIT MVS.ROUTEMGR.OMPROUTE ACCESS(CONTROL) CLASS(OPERCMDS) ID(userid)
   SETROPTS RACLIST(OPERCMDS) REFRESH
   ```

   Rule: OMPROUTE requires UID=0 for correct installation, configuration, and operation.

---

7. Start syslogd.

   To write only the urgent OMPROUTE messages to the z/OS console, syslogd should be running while OMPROUTE is running. Syslogd sends the non-urgent messages to the HFS message log.

---

8. Update the OMPROUTE environment variables (optional).

   The following environment variables are used by OMPROUTE and can be tailored to a particular installation:

   RESOLVER_CONFIG
   The RESOLVER_CONFIG variable is used by OMPROUTE to locate the resolver configuration file. For more information on OMPROUTE's use of the resolver configuration file, see step 3 on page 240. For more information about the RESOLVER_CONFIG environment variable, see "Setting environment variables for configuration files" on page 43.

OMPROUTE_FILE

The OMPROUTE_FILE variable is used by OMPROUTE in the search order for the OMPROUTE configuration file. For details on the search order used for locating this configuration file, see step 1 on page 239.

OMPROUTE_OPTIONS

The OMPROUTE_OPTIONS variable is used by OMPROUTE to set various controls for OMPROUTE processing. Currently only the hello_hi option is supported. The syntax of this variable is:

```
OMPROUTE_OPTIONS=hello_hi
```

Specifying OMPROUTE_OPTIONS=hello_hi changes the way OMPROUTE processes the IPv4 OSPF hello packets. These packets are then given a higher priority than other updates and processed by the first available OMPROUTE task ahead of other received IPv4 OSPF packets. Prior to specifying this parameter, customers must be aware of the impact to their network of processing hello packets out of the received order sequence.

Tips:

a. Specifying OMPROUTE_OPTIONS=hello_hi only helps to keep adjacencies up when OMPROUTE is running and getting flooded with protocol packets. It does not provide any help for the case when adjacencies are not staying up because OMPROUTE is not getting enough cycles (that is, swapped out or running in too low a priority).

b. IPv6 OSPF always gives hello packets higher priority than other IPv6 OSPF traffic, so this option is not necessary for IPv6 OSPF.

OMPROUTE_DEBUG_FILE

The OMPROUTE_DEBUG_FILE variable is used by OMPROUTE to override the debug output destination for IPv4 dynamic routing protocols and for processing common to both IPv4 and IPv6 routing protocols.

Tip: When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is active, debug output is written to CTRACE and not to the destination specified by this variable.

For more information on using this environment variable, see "OMPROUTE parameters" on page 246.

OMPROUTE_DEBUG_FILE_CONTROL

The OMPROUTE_DEBUG_FILE_CONTROL variable is used by OMPROUTE to control the size and quantity of trace files created when the OMPROUTE_DEBUG_FILE or OMPROUTE_IPV6_DEBUG_FILE variable is specified. The syntax of this variable is:

```
OMPROUTE_DEBUG_FILE_CONTROL=<size of file>,<num of files>
```

The default values for <size of file> and <num of files> are 200 (KB) and 5 respectively. In general, these values are sufficient for most installations.

If OMPROUTE_DEBUG_FILE and OMPROUTE_IPV6_DEBUG_FILE are both specified with different output destinations, the values specified on the OMPROUTE_DEBUG_FILE_CONTROL variable will

apply to both the IPv4 debug files and the IPv6 debug files. The total number of files created in this environment would be <num of files> multiplied by 2.

OMPROUTE_IPV6_DEBUG_FILE
The OMPROUTE_IPV6_DEBUG_FILE variable is used by OMPROUTE to override the debug output destination for IPv6 dynamic routing protocols.

Tip: When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is active, debug output is written to CTRACE and not to the destination specified by this variable.

For more information on using this environment variable, see "OMPROUTE parameters" on page 246.

OMPROUTE_CTRACE_MEMBER
The OMPROUTE_CTRACE_MEMBER variable is used by OMPROUTE to specify the name of the parmlib member containing CTRACE default settings. If not specified, the default value is CTIORA00. Use this environment variable to set different CTRACE options and buffer sizes for multiple OMPROUTE instances.

---

9. Create static routes (optional).

   OMPROUTE does not use the environment variable GATEWAYS_FILE to initialize static routes. To create IPv4 static routes, use the BEGINROUTES or GATEWAY statement in PROFILE.TCPIP. To create IPv6 static routes, use the BEGINROUTES statement in PROFILE.TCPIP. For information on the syntax of the GATEWAY and BEGINROUTES statements, refer to *z/OS Communications Server: IP Configuration Reference*.

   During initialization, OMPROUTE learns of static routes by reading the internal routing table set up by TCP/IP. If static routes are changed during execution by VARY TCPIP,,OBEYFILE command processing, OMPROUTE is dynamically notified of the changes by TCP/IP. OMPROUTE will advertise active static routes to other routers if allowed by configuration (for example, the IMPORT_STATIC_ROUTES parameter of the AS_BOUNDARY_ROUTING or IPV6_AS_BOUNDARY_ROUTING configuration statements).

   Static routes can be defined as replaceable or nonreplaceable, with nonreplaceable being the default.

   A nonreplaceable static route cannot be replaced or modified by OMPROUTE, even if a better dynamic route can be learned and even if the static route is not actually available (but a static route that is not available will not be advertised by OMPROUTE). Because of this, the use of nonreplaceable static routes with OMPROUTE is not recommended unless it is to provide routing over an interface over which no routing protocol is being communicated.

   A replaceable static route will be replaced by OMPROUTE if it dynamically learns a route to the destination. Any dynamically learned route will be considered more desirable than a replaceable static route. A replaceable static route should be considered as a last resort route, to be used by TCP/IP when no dynamic route to a destination can be found.

   For detailed information, see "Using IPv4 static routing with OMPROUTE" on page 223 and "Using IPv6 static routing with OMPROUTE" on page 226.

---

10. Configure OSPF authentication (optional, if using the IPv4 OSPF protocol).

OMPROUTE supports defining the IPv4 OSPF authentication type by area or by interface. All interfaces attached to an area default to the type of authentication defined for that area on the AREA configuration statement, unless overridden on the OSPF_INTERFACE configuration statement. The values of authentication keys must be defined on OSPF_INTERFACE statements in any case. All routers which could become neighbors of each other must use the same authentication type and key, or OSPF communication between the routers will not be possible.

Virtual links behave similarly to interfaces for authentication purposes. A virtual link will default to use the same type of authentication that is specified for the backbone area unless overridden on the VIRTUAL_LINK configuration statement. When the authentication type is not NONE, the value of the authentication key must be specified on the VIRTUAL_LINK configuration statement. There is no requirement for a virtual link to have the same authentication key value as its underlying real interface.

OSPF authentication does not protect the contents of an OSPF packet. These packets are not encrypted. However, it does provide verification that the packet is genuine.

Tip: Unlike IPv4 OSPF, IPv6 OSPF does not provide its own, protocol-layer authentication. It relies instead on authentication provided by IPv6 IPSec.

There are two methods of IPv4 OSPF authentication, password and MD5 cryptographic.

Password authentication is very basic: an 8-byte password is appended to all OSPF packets and sent in the clear with the rest of the packet. If the sent password matches that defined by the packet receiver, the packet is accepted.

MD5 authentication is more sophisticated. The combination of the OSPF packet and the MD5 key is summarized into a 16-byte message digest, which is appended to the packet and sent. The keys are never sent, only the message digests. The receiver then attempts to re-create the message digest from the combination of its defined key and the OSPF packet. If the digest is successfully re-created, the packet is accepted; otherwise it is rejected. MD5 authentication also contains a monotonic increasing counter to protect against replay attacks.

If MD5 cryptographic authentication is being used, a 16-byte MD5 key must be defined on the OSPF_INTERFACE configuration statement. This key is defined as a hexadecimal string and may be obtained in several ways. One method for obtaining MD5 keys is provided in the pwtokey utility, which converts a password into an MD5 key. This UNIX System Services utility implements the algorithm defined in RFC 2574, *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*. Since OSPF does not support localization of keys, it is only necessary to provide a password to this utility to generate a single, 16-byte key. If multiple sites have this utility, MD5 keys can easily be generated from passwords, which are easier to remember and communicate than 16-byte hexadecimal strings.

Starting and controlling OMPROUTE

After the necessary RACF authorization has been defined (see step 6 on page 242), OMPROUTE can be started from an MVS procedure, from the z/OS shell, or from AUTOLOG.

- You can start OMPROUTE from the MVS operators console by starting the OMPROUTE start procedure. A sample start procedure is provided with the product in SEZAINST(OMPROUTE).

- You can start OMPROUTE from the z/OS shell by starting OMVS and then issuing the OMPROUTE command and, optionally, any parameters. For information on parameters, see "OMPROUTE parameters."
- You can use the AUTOLOG statement to start OMPROUTE automatically during TCP/IP initialization. Insert the name of the OMPROUTE start procedure in the AUTOLOG statement of the PROFILE.TCPIP data set.

```
AUTOLOG
  OMPROUTE
ENDAUTOLOG
```

  Note: For special considerations when using AUTOLOG to start OMPROUTE with the OSPF or IPv6 OSPF protocol, see step 2 on page 240.

In a Common INET environment, OMPROUTE will attempt to connect to a stack whose name is determined by the TCPIPjobname keyword from the resolver configuration data set or file. In configurations with multiple stacks, a copy of OMPROUTE must be started for each stack that requires OMPROUTE services. To associate OMPROUTE with a particular stack, use the environment variable RESOLVER_CONFIG to point to the data set or file that defines the unique TCPIPjobname.

When running from an MVS procedure, the environment variables can be set by using the STDENV DD statement in the OMPROUTE procedure. For information concerning the environment variables used by OMPROUTE, see step 8 on page 242.

OMPROUTE parameters

OMPROUTE accepts five command line parameters, which govern tracing and debug information. OMPROUTE's trace and debug information is written to stdout with three exceptions:

- When the routing application was started with no tracing, and then a MODIFY command is issued to enable tracing. In this case, the output destination defaults to the file omproute_debug in the current temporary directory (the default is /tmp).
- When the debug output destination has been overridden with the use of an environment variable (OMPROUTE_DEBUG_FILE or OMPROUTE_IPV6_DEBUG_FILE).

  Rule: OMPROUTE_DEBUG_FILE and OMPROUTE_IPV6_DEBUG_FILE can specify only a z/OS UNIX file. To have debug information go into an MVS data set, instead of coding these environment variables, use the SYSPRINT DD card to redirect stdout to the desired MVS data set.
- When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is enabled. In this case, the output is sent to the CTRACE facility.

If OMPROUTE is to be started from an MVS procedure, add your parameters to PARM=() in the OMPROUTE cataloged procedure. For example:

```
//* PARM=('POSIX(ON)',
//*       'ENVAR("_CEE_ENVFILE=DD:STDENV")/-t1 -6t1')
//*
```

If OMPROUTE is to be started from a z/OS shell command line, enter the parameters on the command line.

For either method of starting OMPROUTE, parameters can be specified in mixed case.

Note: Use of the *-tn*, *-dn*, *-6tn*, *-6dn*, and *-sn* parameters affects OMPROUTE performance and might require increasing the Dead_Router_Interval on OSPF interfaces to keep neighbor adjacencies from collapsing.

The -tn and -6tn command line parameters

The *-tn* option specifies the external tracing level for OMPROUTE initialization and IPv4 routing protocols, where *n* is a supported trace level. The *-6tn* option specifies the external tracing level for IPv6 routing protocols, where *n* is a supported trace level. The *-tn* and *-6tn* traces are intended for customers, testers, service, or developers, and provide information on the operation of the routing application. These options can be used for many purposes, such as debugging a configuration, education on the operation of the routing application, verification of test cases, and so on. The following levels are supported:

1 Informational messages

2 Formatted packet trace

These option levels are cumulative—level 2 includes level 1. For example, -t2 provides formatted packet trace and informational messages.

The -dn, -6dn, and -sn command line parameters

These options specify the internal debugging levels. They are intended for service and provide internal debugging information needed for debugging problems. Use of these parameters can significantly impact performance and are not recommended unless needed to debug a problem. For more information about the use of these parameters, refer to *z/OS Communications Server: IP Diagnosis Guide*.

Controlling OMPROUTE

You can control OMPROUTE from the operator's console using the MODIFY command. The syntax of the MODIFY command can be found in *z/OS Communications Server: IP System Administrator's Commands*. MODIFY commands are available to perform the following functions:

Stopping OMPROUTE

OMPROUTE can be stopped in several ways:

- From MVS, issue STOP <procname> or MODIFY <procname>,KILL.

  If OMPROUTE was started from a cataloged procedure, procname is the member name of that procedure. If OMPROUTE was started from the z/OS shell, procname is useridX, where X is the sequence number set by the system. To determine the sequence number, from the SDSF LOG window on TSO, issue /D OMVS,U=userid. This will show the programs running under this user ID. The procname can also be set using the environment variable _BPX_JOBNAME and then starting OMPROUTE in the shell background.
- From a z/OS shell superuser ID, issue the kill command to the process ID (PID) associated with OMPROUTE. To determine the PID, use one of the following methods:

- From the MVS console, issue D OMVS,U=userid, or issue /D OMVS,U=userid at the SDSF LOG window on TSO (where userid is the user ID that started omproute from the shell).
- Issue the ps -ef command from the z/OS shell.
- Record the PID when you start OMPROUTE.

For information on the environment variable _BPX_JOBNAME, refer to *z/OS UNIX System Services Planning*. For information on the D OMVS,U=userid command, refer to *z/OS MVS System Commands*.

Tip: When OMPROUTE is taken down, it should be kept down for at least 3 times the largest dead router interval of any IPv4 OSPF interfaces using MD5 authentication. The same applies to routers adjacent to interfaces using MD5 authentication. Do not stop and start OMPROUTE without waiting for this required time interval.

Rereading the configuration file

The MODIFY <procname>,RECONFIG command is used to reread the OMPROUTE configuration file. This command will accept and apply only the following new statements:

- OSPF_INTERFACE
- RIP_INTERFACE
- INTERFACE
- IPV6_OSPF_INTERFACE
- IPV6_OSPF (ROUTERID parameter only)
- IPV6_RIP_INTERFACE
- IPV6_INTERFACE

These new configuration statements must be reread from the configuration file through this command prior to any new interfaces referred to by new OMPROUTE configuration statements being configured to the TCP/IP stack.

Restriction: The IPV6_OSPF statement (ROUTERID parameter only) is recognized by this command only if there is no existing IPv6 OSPF router ID value set in OMPROUTE. This command cannot be used to change the value of an existing IPv6 OSPF router ID.

Enabling or disabling the OMPROUTE subagent

Use the MODIFY <procname>,ROUTESA=ENABLE command or the MODIFY <procname>,ROUTESA=DISABLE command to enable or disable the OMPROUTE subagent.

Rule: To change any other value on the ROUTESA_CONFIG statement, the OMPROUTE application must be recycled.

The OMPROUTE subagent implements RFC 1850, *OSPF Version 2 Management Information Base*, for the OSPF protocol. The ROUTESA_CONFIG statement is used in the OMPROUTE configuration file to configure the OMPROUTE subagent. For details on ROUTESA_CONFIG, refer to *z/OS Communications Server: IP Configuration Reference*.

Changing the cost of OSPF links

The cost of an OSPF interface can be dynamically changed using the MODIFY <procname>,OSPF,WEIGHT,NAME=<if_name>,COST=<cost> command. The cost

of an IPv6 OSPF interface can be dynamically changed using the MODIFY <procname>,IPV6OSPF,WEIGHT,NAME=<if_name>,COST=<cost> command. This new cost is flooded quickly throughout the OSPF routing domain, and modifies the routing immediately.

The cost of the interface reverts to its configured value whenever OMPROUTE is restarted. To make the cost change permanent, you must change the appropriate OSPF_INTERFACE or IPV6_OSPF_INTERFACE statement in the configuration file.

Controlling OMPROUTE tracing and debugging

The following commands are used to start, stop, or change the level of OMPROUTE tracing and debugging:

- MODIFY <procname>,TRACE=n : for OMPROUTE tracing for initialization and IPv4 routing protocols; n can be 0–2.
- MODIFY <procname>,DEBUG=n : for OMPROUTE debugging for initialization and IPv4 routing protocols; n can be 0–4.
- MODIFY <procname>,SADEBUG=n : for OMPROUTE subagent debugging; n can be 0 or 1.
- MODIFY <procname>,TRACE6=n : for OMPROUTE tracing for IPv6 routing protocols; n can be 0–2.
- MODIFY <procname>,DEBUG6=n : for OMPROUTE debugging for IPv6 routing protocols; n can be 0–4.

Tip: Use of OMPROUTE tracing and debugging affects OMPROUTE performance and might require increasing the Dead_Router_Interval on OSPF interfaces to keep neighbor adjacencies from collapsing.

Steps for configuring OSPF and RIP (IPv4 and IPv6)

The steps for configuring OSPF and RIP are:

1. Set the OSPF router ID, if the OSPF protocol is used.
2. Define OSPF areas, if the OSPF protocol is used.
3. Limit information exchange between OSPF areas, if the OSPF protocol is used.
4. Define IPv4 interfaces, if the IPv4 OSPF or IPv4 RIP protocol is used.
5. Define IPv6 interfaces, if the IPv6 OSPF or IPv6 RIP protocol is used.
6. Define interface costs (OSPF_INTERFACE, RIP_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE).
7. Configure virtual links, if the OSPF protocol is used.
8. Manage high-cost links, if the OSPF protocol is used.
9. Define RIP filters, if the RIP protocol is used.
10. Define route precedence in a multiprotocol environment, if the OSPF protocol is used.

Following are the detailed steps to configure OSPF and RIP (IPv4 and IPv6).

1. Set the OSPF router ID, if the OSPF protocol is used.

   Every router in an OSPF autonomous system must be assigned a unique router ID.

   IPv4 OSPF

   The ROUTERID parameter of the OSPF configuration statement should be coded within the OMPROUTE configuration file to assign the router ID. The value must be the IP address of one of the OSPF_INTERFACEs

defined in the OMPROUTE configuration file. If the ROUTERID parameter of the OSPF configuration statement is not coded, OMPROUTE chooses the IP address from one of the OSPF_INTERFACE statements as the router ID. With the advent of Dynamic VIPAs (DVIPAs) that can move between z/OS hosts within a sysplex, it is highly recommended that the ROUTERID be a physical interface or a static VIPA, not a Dynamic VIPA.

In the example network shown in Figure 36 on page 220, the router ID is set to the static VIPA address that represents each OMPROUTE router. TCPCS4 has ROUTERID=4.4.4.4, and TCPCS7 has ROUTERID=7.7.7.7.

IPv6 OSPF

The ROUTERID parameter of the IPV6_OSPF configuration statement is coded within the OMPROUTE configuration file to assign the IPv6 OSPF router ID. This router ID is any IPv4-style dotted-decimal value except for 0.0.0.0, with care taken to assure its uniqueness across routers within the IPv6 autonomous system. If the ROUTERID parameter of the IPV6_OSPF configuration statement is not coded and the IPv4 OSPF protocol is also being used, OMPROUTE will use the IPv4 OSPF router ID as the IPv6 OSPF router ID. If the IPv4 OSPF protocol is not being used, the ROUTERID parameter of the IPV6_OSPF statement must be specified.

In the example network shown in Figure 37 on page 221, the ROUTERID parameter of the IPV6_OSPF statement on TCPCS64 will be explicitly configured using ROUTERID=64.64.64.64.

---

2. Define OSPF areas, if the OSPF protocol is used.

   The sample network shown in Figure 36 on page 220 and Figure 37 on page 221 depicts an IPv4 network and an IPv6 network, both divided using two methods. The first division is between IP subnetworks within the OSPF autonomous system (AS) and IP subnetworks external to the OSPF AS (those within the RIP AS). The subnetworks included within each OSPF AS are further subdivided into regions called areas. OSPF areas are collections of contiguous IP subnetworks. The function of areas is to reduce the OSPF overhead required to compute routes to destinations in different areas. Overhead is reduced because less information is exchanged and stored by routers and because fewer CPU cycles are required for a less complex route table calculation.

   Every OSPF AS that will have multiple areas must have at least a backbone area. The backbone is always identified by area number 0.0.0.0. For networks with multiple areas, the backbone provides a core that connects the areas. Unlike other areas, the backbone's subnets can be physically separate. In this case, logical connectivity of the backbone is maintained by configuring virtual links between backbone routers across intervening non-backbone areas. See step 7 on page 265 for more information on this subject.

   Routers that attach to more than one area function as area border routers. All area border routers are part of the backbone, so they must either attach directly to a backbone IP subnet or be connected to another backbone router over a virtual link.

   The information and algorithms used by OSPF to calculate routes vary according to whether the destination is within the same area, in a different area within the OSPF AS, or external to the OSPF AS. Every router maintains a database of all links within its area. A shortest path first algorithm is used to calculate the best routes to destinations within the area from this database. Routes between areas are calculated from summary advertisements

originated by area border routers for destinations located in other areas of the OSPF AS. External routes (for example, routes to destinations that lie within a RIP AS) are calculated from AS external advertisements originated by AS boundary routers and flooded throughout the OSPF AS.

IPv4 OSPF

Use the AREA configuration statement to define the areas to which a router attaches. If you do not use the AREA statement, the default is that all OSPF interfaces attach to the backbone area. In the sample network, TCPCS4 and TCPCS7 are both area border routers belonging to both the backbone area (0.0.0.0) and area 1.1.1.1.

```
AREA
  Area_Number=0.0.0.0;

AREA
  Area_Number=1.1.1.1;
```

IPv6 OSPF

Use the IPV6_AREA configuration statement to define the areas to which a router attaches. If you do not use the IPV6_AREA statement, the default is that all IPv6 OSPF interfaces attach to the backbone area. In the sample network, TCPCS64 and TCPCS67 are area border routers belonging to both the backbone area (0.0.0.0) and area 6.6.6.6.

```
IPV6_AREA
   Area_Number=0.0.0.0;
IPV6_AREA
   Area_Number=6.6.6.6;
```

---

3. Limit information exchange between OSPF areas, if the OSPF protocol is used.

   When area border routers are configured, the OSPF route information that crosses the area boundary can be controlled using configuration statements. For recommendations regarding the usefulness of multiple areas in the z/OS CS environment, see "Network design considerations with z/OS Communications Server" on page 270.

   One option is to define an area as a stub area. AS external advertisements are never flooded into stub areas. In addition, the origination into the stub area of summary advertisements for interarea routes can be suppressed, creating what is commonly known as a totally stubby area.

   Destinations external to the stub area are still reachable due to the area border routers advertising default routes into stub areas. Traffic within the stub area for unknown destinations is forwarded to the area border router (using the default route). It is acceptable for routers within the area to use a default route for traffic destined outside the AS. The border router uses its more complete routing information to forward the traffic on an appropriate path toward its destination.

   The following requirements must be met for an area to be defined as a stub area:

   - No virtual links are configured through the area to maintain backbone connectivity.
   - No routers within the area are AS boundary routers (OSPF routers that advertise routes from external sources as AS external advertisements).

     Tip: Static routes and RIP routes are AS external.

   Another option is to use IP address ranges to limit the number of summary advertisements originated out of an area. In IPv4, a range is defined by an IP

address and an address mask. Destinations are considered to fall within the range if the destination address and the range IP address match after the range mask has been applied to both addresses. In IPv6, a range is defined by an IP address prefix and a prefix length, and destinations are considered to fall within the range if a destination address and a range IP address match for the length of the range's prefix.

When a range is configured for an area at an area border router, the border router suppresses summary advertisements for destinations within that area that fall within the range. The suppressed advertisements would have been originated into the other areas to which the border router attaches. Instead, the area border router may originate a single summary advertisement for the range or no advertisement at all, depending on the option chosen with the configuration statement.

Rules:

a. If the range is not advertised, there will be no interarea routes for any destination that falls within the range.

b. Ranges cannot be used for areas through which virtual links are configured to maintain backbone connectivity.

IPv4 OSPF

The following statement configures an OSPF area as a stub area. The Import_Summaries=No parameter will result in the suppression of summary advertisements for interarea routes into the stub area, creating a totally stubby area:

```
AREA
  Area_Number=2.2.2.2
  Stub_area=Yes
  Import_Summaries=No;
```

In the sample network shown in Figure 36 on page 220, the following RANGE statement could be configured on TCPCS7 to prevent TCPCS7 from advertising destinations in the 9.67.101.0 subnet into the backbone area (Area 0.0.0.0):

```
RANGE
   IP_Address=9.67.101.0
   Subnet_Mask=255.255.255.0
   Area_Number=1.1.1.1
   Advertise=No;
```

IPv6 OSPF

The following statement configures an IPv6 OSPF area as a stub area. The Import_Summaries=No parameter will result in the suppression of summary advertisements for interarea routes into the stub area, creating a totally stubby area:

```
IPV6_AREA
   Area_Number=1.1.1.1
   Stub_area=Yes
   Import_Prefixes=No;
```

In the sample network, the following IPV6_RANGE statement could be configured on TCPCS67 to cause TCPCS67 to advertise all destinations in the 2001:0DB8:0:31::/64 prefix into the backbone area (Area 0.0.0.0) as a single route to the 2001:0DB8:0:31::/64 prefix:

```
IPV6_RANGE
   Prefix=2001:0DB8:0:31::/64
   Area_Number=6.6.6.6
   Advertise=Yes;
```

4. Define IPv4 interfaces, if the IPv4 OSPF or IPv4 RIP protocol is used.

   Each interface in use by the stack should be defined to OMPROUTE using an OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE statement. This section describes the differences between interface types that you should consider when configuring interfaces to OMPROUTE. In general, use the following guidelines:

   - An interface over which the OSPF protocol is communicated with other routers must be configured with the OSPF_INTERFACE statement.
   - An interface over which the RIP protocol is communicated with other routers must be configured with the RIP_INTERFACE statement.
   - Either all other interfaces should be configured with the INTERFACE statement, or OMPROUTE should be configured to ignore undefined interfaces using the IGNORE_UNDEFINED_INTERFACES parameter of the GLOBAL_OPTIONS statement in the OMPROUTE configuration file. It is important that one or the other be done.

     If OMPROUTE is not configured to ignore undefined interfaces, it configures stack interfaces that are not defined to OMPROUTE with default values. These values might not be desirable. For example, the class mask will be used as the subnet mask and 576 will be used as the MTU value. Furthermore, OMPROUTE overrides stack values with its default values. To prevent this from happening, either configure every interface, even those that are not using RIP or OSPF, or configure OMPROUTE to ignore undefined interfaces.

   A VIPA interface is an exception to these guidelines and is discussed in more detail later in this step.

   z/OS Communications Server enforces RFC rules against using either a subnetwork's broadcast or network address as a host address. (An address that has all ones in the host portion is a subnet broadcast address. An address that has all zeros in the host portion is the subnet's network address.) Therefore, the subnet_mask on an OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE statement should have enough zero bits such that no home address in that subnet has all zeros or all ones in the host portion of the address. For example if a subnet has two home addresses 10.1.1.1 and 10.1.1.2, the subnet mask must have zeros in at least two bits; for example, 255.255.255.252. However, if a subnet has four home addresses 10.1.1.1, 10.1.1.2, 10.1.1.3, and 10.1.1.4, the subnet mask must have zeros in at least three bits; for example, 255.255.255.248. In this case, there could be up to 6 home addresses in that subnet (10.1.1.1 through 10.1.1.6). In general, if a subnet mask has n zero bits, then there can be up to ((2**n)-2) home addresses in that subnet. This limit applies even if the home addresses are configured on different TCP/IP stacks.

   Rules:

   a. OMPROUTE supports a maximum of 255 real, physical, IPv4 interfaces (that is, interfaces on which data can actually be sent and received). There is no theoretical limit on how many VIPAs can be configured, though there are practical limits imposed by network design.
   b. RIP version 1 uses broadcast and RIP version 2 uses multicast. RIP version 1 will not run on a medium that supports multicast but not broadcast, which is the default QDIO configuration. To configure OSA-Express operating in QDIO mode (for example, gigabit Ethernet) to send and receive broadcast packets, use the IPBCAST parameter on the LINK statement in your TCP/IP profile. The OSA-Express microcode level must support broadcast to use this parameter. Some levels of

OSA-Express microcode support multicasting but not broadcasting. In this case, RIP version 2, which relies on multicast, is recommended over RIP version 1, which relies on broadcast.

Configuring multi-access parallel interfaces

Whenever configuring multi-access parallel interfaces (primary and backup redundant interfaces having IP addresses in the same network) for OMPROUTE (OSPF), use the Parallel_OSPF parameter on the OSPF_INTERFACE statement to designate whether each OSPF interface is primary or backup. If the IP_address parameter on an OSPF_INTERFACE statement is using wildcards (*), also include the Name parameter for the interface to distinguish the primary from the backups. For additional information on the OSPF_INTERFACE statement and its parameters, refer to *z/OS Communications Server: IP Configuration Reference*.

Point-to-point (For example CTC and CLAW)

For point-to-point interfaces, the destination IP address must be known to OMPROUTE. If OSPF or RIP is run on the interface, the destination IP address is learned when the router at the other end comes up and shares information with OMPROUTE. Additionally, defining a destination address on an interface that is running OSPF or RIP allows OMPROUTE to learn and advertise a route to that destination address before a neighboring router has become fully adjacent. This can be beneficial if the neighboring router takes a long time to come up, or is otherwise not expected to be promptly and reliably available.

Tip: For a point-to-point interface running OSPF, OMPROUTE does not advertise a host route to its own home address on the point-to-point interface. By default, OMPROUTE advertises a host route to the link destination, and relies on the router at the other end to advertise a host route to OMPROUTE's home address. This behavior is described by the OSPF architecture, RFC 1583 (OSPF version 2), section 12.4.1. This means that OMPROUTE's home address on the point-to-point link will not be advertised as reachable unless there is a neighboring router available to advertise it. OMPROUTE implements an extension described in RFC 2328 to overcome this limitation. If a neighboring router will not be reliably available over the point-to-point link, you might want to code the parameter SUBNET=YES on the OSPF_INTERFACE statement for the point-to-point interface. This causes OMPROUTE to implement option 2 described in RFC 2328, section 12.4.1.1, and advertise a route to the point-to-point link's subnet address. This makes both endpoints reachable regardless of the status of a neighboring router.

If the interface is simply an INTERFACE (not running OSPF or RIP), specify the DESTINATION_ADDR parameter to allow for the creation of a host route to the address at the remote end of the interface.

Sample OSPF_INTERFACE

```
OSPF_INTERFACE
 IP_Address=9.67.106.7
 Name=CTC7TO4
 Subnet_mask=255.255.255.0
 Attaches_to_Area=1.1.1.1
 Destination_Addr=9.67.106.4;
```

Sample RIP_INTERFACE

```
RIP_INTERFACE
 IP_Address=9.67.103.7
 Name= CTC7TO6
 Subnet_mask=255.255.255.0
 Destination_Addr=9.67.103.6
 RIPV2=Yes;
```

Sample INTERFACE

```
INTERFACE
 IP_Address=9.67.111.1
 Name=CTCX
 Subnet_mask=255.255.255.0
 Destination_addr=9.67.111.2;
```

Note: If another router is directly attached through a CLAW device, and the OSPF protocol is being communicated with that router, the other router must also be configured to view the CLAW device as a point-to-point interface. Failure to do this results in a failure to add any routes through that router.

Point-to-multipoint

For point-to-multipoint capable interfaces (for example MPCPTP interfaces including XCF and IUTSAMEH connections), OMPROUTE must know the IP addresses of the other routers (neighbors) with which it needs to communicate the OSPF or RIP packets. However, due to underlying signaling that takes place when a host connects to these network types, the stack is able to learn the required addresses. In turn, OMPROUTE learns those IP addresses from the stack. As a result, it is not necessary to configure the IP addresses of the other routers on the interface statements.

Sample OSPF_INTERFACE

```
OSPF_INTERFACE
 IP_Address=9.27.13.81
 Name=XCFD00
   Attaches_to_Area=1.1.1.1
 Subnet_mask=255.255.255.0;
```

Sample RIP_INTERFACE

```
RIP_INTERFACE
 IP_Address=9.27.23.81
 Name=MPCA01
 Subnet_mask=255.255.255.0
 RIPV2=Yes;
```

Sample INTERFACE

```
INTERFACE
 IP_Address=9.27.33.81
 Name=XCFB00
 Subnet_mask=255.255.255.0;
```

Non-broadcast network interfaces (For example, Hyperchannel and ATM)

If the OSPF or RIP protocol communicates with one or more routers over a non-broadcast network interface, OMPROUTE must know the IP addresses of the other routers (neighbors) with which it needs to communicate. For non-broadcast network interfaces, there is no underlying signaling that allows the stack to learn the required IP addresses. As a result, the neighbor addresses must be configured to OMPROUTE with the parameters configured as follows:

– DR_NEIGHBOR and/or the NO_DR_NEIGHBOR parameters on the OSPF_INTERFACE statement

– NEIGHBOR parameter on the RIP_INTERFACE statement
– NON_BROADCAST=YES and ROUTER_PRIORITY parameters on the OSPF_INTERFACE statement

In the OSPF case, DR_NEIGHBOR defines which routers within the non-broadcast network can become the designated router. NO_DR_NEIGHBOR defines which routers cannot become the designated router. ROUTER_PRIORITY defines the priority of this router on the non-broadcast network so that the designated router can be elected for the network. Note that multiple DR_NEIGHBOR and NO_DR_NEIGHBOR parameters can be coded on one statement.

Sample OSPF_INTERFACE

```
OSPF_INTERFACE
 IP_Address=9.37.84.49
 Name=HCHE00
 Subnet_mask=255.255.255.0
   Attaches_to_Area=1.1.1.1
 Non_Broadcast=Yes
 DR_Neighbor=9.37.84.53
 No_DR_Neighbor=9.37.84.63
 Cost0=3
 Router_Priority=2;
```

Sample RIP_INTERFACE

```
RIP_INTERFACE
 IP_Address=9.37.104.79
 Name=ATME00
 Subnet_mask=255.255.255.0
 RIPV2=Yes
 Neighbor=9.37.104.85
 Neighbor=9.37.104.53;
```

Sample INTERFACE

```
INTERFACE
 IP_Address=9.77.13.49
 Name=ATMB00
 Subnet_mask=255.255.255.0;
```

Broadcast network interfaces (For example, Token Ring, Ethernet, and FDDI)

When the OSPF or RIP protocol is communicated over a broadcast medium such as Token Ring, Ethernet, or FDDI, these networks allow for broadcasting and multicasting. Therefore, it is not necessary for OMPROUTE to know the IP addresses of the other routers on the network for OSPF or RIP packets to be communicated with those routers. OMPROUTE sends packets to the other routers on the network by using appropriate broadcast or multicast addresses. The IP addresses of the other routers are learned as OSPF/RIP packets are received from them. The OSPF_INTERFACE must include the ROUTER_PRIORITY parameter to assist in electing a designated router for the network.

Sample OSPF_INTERFACE

```
OSPF_INTERFACE
 IP_Address=9.59.101.5
 Name=TR1
   Subnet_mask=255.255.255.0
 Attaches_to_Area=1.1.1.1
 Cost0=2
 Router_Priority=1;
```

Sample RIP_INTERFACE

```
RIP_INTERFACE
 IP_Address=9.29.107.3
 Name=TR2
 Subnet_mask=255.255.255.0
 RIPV2=Yes;
```

Sample INTERFACE

```
INTERFACE
 IP_Address=9.77.14.49
 Name=ETHB00
 Subnet_mask=255.255.255.0;
```

If OMPROUTE will be communicating with the OSPF or RIP Version 2 protocol over a token ring media where an attached router does not listen for multicast MAC address 0xC000.0004.0000, see "Token-ring multicast" on page 237.

For interfaces into broadcast media which contain routers that do not support multicast, it is possible to configure the interfaces as non-broadcast network interfaces. This would cause OMPROUTE to unicast to the neighbor addresses rather than using a multicast address. However, it would also be necessary to configure all the routers on the network to unicast. Otherwise, their multicast packets would never be received.

Note that it is possible to define neighbors using DR_NEIGHBOR and/or NO_DR_NEIGHBOR parameters for OSPF_INTERFACEs and using NEIGHBOR parameters for RIP_INTERFACEs that are broadcast capable, but it is not required or recommended. If you define neighbors on these interfaces, you must define all of them, as OMPROUTE will not communicate RIP or OSPF to undefined neighbors if any are defined on an interface.

VIPA interfaces (Static VIPA and Dynamic VIPA)

If only the RIP protocol is used by OMPROUTE, VIPA interfaces should be defined with the INTERFACE statement. If only OSPF or if both OSPF and RIP are used by OMPROUTE, VIPA interfaces should be defined with the OSPF_INTERFACE statement.

Sample OSPF_INTERFACE

```
OSPF example:
 OSPF_INTERFACE
 IP_Address=4.4.4.4
 Name=VIPA1
   Subnet_mask=255.255.255.252;
```

Sample INTERFACE

```
non-OSPF example:
 INTERFACE
 IP_Address=6.6.6.6
 Name=VIPA1
 Subnet_mask=255.255.255.252;
```

Rule: The most specific subnet mask you can specify is 255.255.255.252.

If the name in an OSPF_INTERFACE or INTERFACE statement refers to a link of type VIRTUAL, then OMPROUTE generates and advertises the following routes whenever applicable:

- A network route to the network specified in that statement
- A subnet route to the subnet specified in that statement
- A host route to the IP_address specified in that statement

Following are the conditions for advertising these routes on a physical network interface to a network:

– Network route - if VIPA is not in the same network as the physical network interface and is allowed by filters or RANGE.
– Subnet route - VIPA subnet routes are advertised in OMPROUTE in all conditions, except for RIP when filters prevent it.
– Host route - as allowed by filters or RANGE. Advertisement of the host route for a VIPA defined on an OSPF_INTERFACE statement can be controlled by the SUBNET parameter on the OSPF_INTERFACE statement that defines that VIPA. If SUBNET=YES, then the host route is not advertised. If SUBNET=NO (the default), the host route is advertised. Take care in using this parameter. VIPA host routes should not be suppressed for dynamic VIPAs or for VIPAs whose subnet might exist on multiple hosts. It is up to the user to ensure these restrictions are enforced, as they are not and cannot be enforced by OMPROUTE.

On the RIP_INTERFACE statement for a physical network interface, the VIPA routes are allowed to be advertised by the following filter parameters:

– Send_Net_Routes
– Send_Subnet_Routes
– Send_Host_Routes, and Send_Only

In addition, the global FILTER and Send_Only statements for RIP can be used to specify which routes are advertised or not.

For OSPF, the RANGE statement can be used to advertise or not to advertise the VIPA routes external to an area in terms of address range based on a subnet mask.

Note: For RIP, the Send_Only = (VIRTUAL) filter in conjunction with the Send_Net_Routes, Send_Subnet_Routes, and Send_Host_Routes filters, or the FILTER statement with VIPA routes, indicates whether or not VIPA routes can be advertised over a RIP interface. Unlike RIP, there are no routing filters for OSPF. For OSPF, the RANGE statement can be used to control which address range of routes can be advertised or not advertised external to an area; however, it is not granular enough for use as a routing filter. In area-border router configurations, if there are multiple VIPA addresses that are uniquely subnetted, the RANGE statement can be used to specify which VIPA subnet address range of routes can be advertised or not advertised external to an area.

For Dynamic VIPA (DVIPA), link names are assigned programmatically by the stack when the DVIPA is created. Therefore, the name field set on the INTERFACE or OSPF_INTERFACE statement is ignored by OMPROUTE for DVIPAs.

Because a stack could have a large number of DVIPAs defined, as well as DVIPA ranges, additional wildcard capabilities exist on the OSPF_INTERFACE and INTERFACE statements for use only with DVIPAs.

Ranges of DVIPA interfaces can be defined using the Subnet_Mask parameter on the OSPF_INTERFACE or INTERFACE statement. The range defined in this way will be all the IP addresses that fall within the subnet defined by the mask and the IP address. The IP address parameter must be the subnet number of the range being defined, not a host address within that range. Further information on the Subnet_Mask parameter is available earlier in this step.

In the following example, DVIPA interfaces in the range of 10.138.65.81 through 10.138.65.94 are defined:

Sample OSPF_INTERFACE

```
OSPF example:
 OSPF_INTERFACE
 IP_Address=10.138.65.80
 Name=DVIPAs
   Subnet_mask=255.255.255.240;
```

Sample INTERFACE

```
non-OSPF example:
 INTERFACE
 IP_Address=10.138.65.80
 Name=DVIPAs
 Subnet_mask=255.255.255.240;
```

In the following example, only an interface with home address 10.138.65.98 is being defined, because the subnet number (obtained by performing a binary AND operation of the IP_Address parameter and the Subnet_mask parameter) for this definition is 10.138.65.96. Since the IP_Address parameter does not equal that subnet number, this definition will not be treated as a DVIPA wildcard.

```
OSPF_INTERFACE
IP_Address=10.138.65.98
Name=DVIPA
Subnet_mask = 255.255.255.240;
```

You must consider an additional issue when VIPAs are being moved between TCP/IP stacks and dynamic routing is provided for those stacks by OMPROUTE. This movement of VIPAs can be done manually or automatically using Dynamic VIPAs. For the VIPAs to be correctly processed and advertised by the routing protocols, they (like all other interfaces) must be configured to OMPROUTE at the time that they become active on the TCP/IP stack. This configuration of VIPAs to OMPROUTE can be accomplished by:

- Explicitly configuring each VIPA with its own OSPF_INTERFACE or INTERFACE statement
- Configuring a range of DVIPAs with a single OSPF_INTERFACE or INTERFACE statement, using the method described above
- Configuring a group of VIPAs with a single OSPF_INTERFACE or INTERFACE statement, using the wildcarding feature available on the interface statements

The recommended approach for configuring OMPROUTE for VIPAs that might move is to preconfigure the OMPROUTE on each TCP/IP stack with all VIPAs that could potentially exist on that stack at some time. Preconfiguring in this way prepares each OMPROUTE for the possible addition of the VIPAs to its stack. During times when the VIPAs do not exist on a particular OMPROUTE's stack, the configuration information will not be used. However, during periods when the VIPAs do exist on that OMPROUTE's stack, the configuration information will be available for use by OMPROUTE. This method is recommended because of its ability to respond to movement of the VIPAs between TCP/IP stacks without modification of the OMPROUTE configuration with each move.

If the pre-configuration of VIPAs described in this section has not been done, it is still possible to define a VIPA to OMPROUTE such that it is properly processed and advertised when it becomes active on the corresponding TCP/IP stack. To do this, add the appropriate OSPF_INTERFACE or INTERFACE statement to the OMPROUTE

configuration file and then cause OMPROUTE to reread the configuration file by issuing the MODIFY <procname>,RECONFIG command.

Rule: You must modify the OMPROUTE configuration file and issue the RECONFIG command prior to the movement of the VIPA to the corresponding TCP/IP stack.

Method of assigning interface definitions to stack interfaces (wildcard and explicit):

Wildcard interface definitions can be a convenient way of making your interface definitions easier. However, to avoid unintended results, be sure to understand how they are parsed, and how different types of interface definitions interact with each other. Following is the outline of the algorithm OMPROUTE uses to find the matching definitions in the OMPROUTE configuration file for an IPv4 stack interface.

a. Search for a RIP_Interface definition for the interface as follows:
   1) Search for an explicit matching RIP_Interface definition (IP address matches exactly if a dynamic VIPA, otherwise name and IP address match exactly). If found, use that definition and go to step b.
   2) If the interface is a dynamic VIPA, search for a dynamic VIPA wildcard RIP_Interface definition that fits. This is a definition where the IP address is the subnet number and the subnet mask is the subnet mask of the subnet that the interface's home address falls into, ignoring the name parameter. If found, use that definition and go to step b.
   3) Search for a one-octet wildcard RIP_Interface definition (n.n.n.*), where the first three octets match the first three octets of the interface's home address and the name matches the interface's link name. If found, use that definition and go to step b.
   4) Search for a two-octet wildcard RIP_Interface definition (n.n.*.*), where the first two octets match the first two octets of the interface's home address and the name matches the interface's link name. If found, use that definition and go to step b.
   5) Search for a three-octet wildcard RIP_Interface definition (n.*.*.*), where the first octet matches the first octet of the interface's home address and the name matches the interface's link name. If found, use that definition and go to step b.
   6) Search for a one-octet wildcard RIP_Interface definition (n.n.n.*), where the first three octets match the first three octets of the interface's home address, ignoring the name parameter if coded. If found, use that definition and go to step b.
   7) Search for a two-octet wildcard RIP_Interface definition (n.n.*.*), where the first two octets match the first two octets of the interface's home address, ignoring the name parameter if coded. If found, use that definition and go to step b.
   8) Search for a three-octet wildcard RIP_Interface definition (n.*.*.*), where the first octet matches the first octet of the interface's home address, ignoring the name parameter if coded. If found, use that definition and go to step b.
   9) If there is a four-octet wildcard RIP_Interface definition (*.*.*.* or ALL), use that definition and go to step b.

      Restriction: Only one four-octet wildcard of each type (OSPF_INTERFACE or RIP_INTERFACE) is allowed.

b. Search for an OSPF_Interface definition for the interface. Note that this step is done regardless of the outcome of step a. The steps for searching OSPF_Interface definitions are the same as the steps for searching RIP_Interface definitions described above, except that OSPF_Interface definitions are searched.

c. If either a RIP_Interface or an OSPF_Interface definition, or both, are found, the algorithm is complete. In this case, Interface definitions are not searched. If neither a RIP_Interface nor an OSPF_Interface definition was found, go to step d.

d. Search for an Interface definition for the interface. The steps for searching Interface definitions are the same as the steps for searching RIP_Interface statements described above, except that Interface definitions are searched.

e. If no definitions are found, check the value of Global_Options Ignore_Undefined_Interfaces. If this option is turned on, the interface is ignored. If it is not turned on, the interface is treated as if it were defined with an Interface statement, with an MTU value of 576 and a subnet mask of the class mask.

The algorithm is complete. Key conclusions of this algorithm are as follows:

- A wildcard interface definition with a matching name is preferred over a wildcard interface definition of the same type without a matching name, regardless of which definition is a more specific wildcard.
- If the interface is not a dynamic VIPA, the subnet mask coded on the definition statement has no influence on which definition, wildcard or otherwise, is selected for an interface. The subnet mask is a characteristic of the definition, not a selection criterion.
- If a RIP_Interface or an OSPF_Interface definition, or both, are found, Interface definitions are not considered. This means that any matching RIP_Interface or OSPF_Interface definition supersedes all Interface definitions, even if the Interface definitions are explicit or are more specific wildcard matches. For example, a three-octet wildcard OSPF_Interface definition supersedes an explicit Interface definition.
- An interface can be both a RIP_Interface and an OSPF_Interface. OMPROUTE supports running both protocols over the same interface. However, an interface cannot be both an interface that runs no routing protocol (that is, defined with the Interface statement) and one that runs RIP, OSPF, or both.

---

5. Define IPv6 interfaces, if the IPv6 OSPF or IPv6 RIP protocol is used.

   Each IPv6 interface in use by the stack can be defined to OMPROUTE using an IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, or IPV6_INTERFACE statement. Defining IPv6 interfaces to OMPROUTE is much simpler than defining IPv4 interfaces, because you do not specify IP addresses or MTU sizes to OMPROUTE. Instead, you simply define interfaces by their names and OMPROUTE learns IP addresses and MTU sizes from the TCP/IP stack. Also, because multicast support is a basic requirement of IPv6, there are no non-broadcast multiaccess considerations or other considerations that might require definitions of neighbors or destination addresses.

   Use the following guidelines when configuring IPv6 interfaces to OMPROUTE:

   - An interface over which the IPv6 OSPF protocol is communicated with other routers must be configured with the IPV6_OSPF_INTERFACE statement.

- An interface over which the IPv6 RIP protocol is communicated with other routers must be configured with the IPV6_RIP_INTERFACE statement.
- All other interfaces can be configured with the IPV6_INTERFACE statement, if OMPROUTE default values are not acceptable or you wish to define additional prefixes on the interface.
- The interface name must be coded on the IPV6_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE statements. The value of the NAME parameter must match the interface name coded on the INTERFACE statement in the TCP/IP profile. Wildcard names ending with an asterisk (*) can be coded. For example, OSAQDIO* would match stack interfaces named OSAQDIO1, OSAQDIO2, OSAQDIOABC, and so on.

  If the interface is dynamically generated by the TCP/IP stack, its name parameter must match what is generated by the TCP/IP stack. Interfaces that are dynamically generated by the TCP/IP stack are named as follows:
  - An IPv6 dynamic XCF interface that is generated to attach to other TCP/IP stacks within the same z/OS host is always named EZ6SAMEMVS.
  - An IPv6 dynamic XCF interface that is generated to attach to TCP/IP stacks in another z/OS image is always named EZ6XCF*xx*, where *xx* is the SYSCLONE value of the other z/OS host.

  If the routing parameters for your dynamic XCF interfaces will all be the same, you can use wildcard definitions to avoid having to know and build definitions for every possible dynamic XCF interface on your system. For example, a wildcard definition for name EZ6* would match all dynamic XCF interfaces that could be generated on a TCP/IP stack. A wildcard definition for EZ6XCF* would match all dynamic XCF interfaces that could be generated to attach to other z/OS images.

  The following definition would define all IPv6 dynamic XCF interfaces to OMPROUTE, with the hello and dead router intervals changed from the default values:

  ```
  IPV6_OSPF_INTERFACE
   NAME=EZ6*
   HELLO_INTERVAL = 30
   DEAD_ROUTER_INTERVAL = 120;
  ```

  If the default values of 10 and 40 for the hello and dead router intervals are acceptable, the above definition can be simplified even more:

  ```
  IPV6_OSPF_INTERFACE
   NAME=EZ6*;
  ```
- To define one or more prefixes on an interface, use the PREFIX parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, and IPV6_INTERFACE statements. You should only need to do this for prefixes that you need to define to an interface, which will not be learned using IPv6 router discovery. Also note that prefixes defined to OMPROUTE in this manner are not used by TCP/IP to autoconfigure home addresses on the interface.

  The following sample shows an IPv6 OSPF interface with prefixes defined:

  ```
  IPV6_OSPF_INTERFACE
   NAME=OSAQDIO4L6
   PREFIX=2001:0DB8:1::/48
   PREFIX=2001:0DB8:2::/48;
  ```

  The prefixes defined in this manner on an IPv6 OSPF interface are advertised as reachable, and are also included in the link LSA generated by OMPROUTE, so all IPv6 OSPF routers on the link will know they are

local prefixes. If OMPROUTE is also running IPv6 RIP, they are also advertised into the IPv6 RIP autonomous system as reachable, if IPv6 RIP filters permit it.

The following sample shows an IPv6 RIP interface with prefixes defined:

```
IPV6_RIP_INTERFACE
 NAME=OSAQDIO3L6
 PREFIX=2001:0DB8:3::/48
 PREFIX=2001:0DB8:4::/48;
```

The prefixes defined in this manner on an IPv6 RIP interface are advertised into the IPv6 RIP autonomous system as reachable if IPv6 RIP filters permit it. They are also advertised into the IPv6 OSPF autonomous system as reachable, if OMPROUTE is running IPv6 OSPF and is configured as an IPv6 AS boundary router importing RIP routes.

The following sample shows an IPv6 generic interface with prefixes defined:

```
IPV6_INTERFACE
 NAME=OSAQDIO2L6
 PREFIX=2001:0DB8:5::/48
 PREFIX=2001:0DB8:6::/48;
```

The prefixes defined in this manner on an IPv6 generic interface are advertised into the RIP autonomous system as reachable, if OMPROUTE is running IPv6 RIP and IPv6 RIP filters permit it. They are also advertised into the IPv6 OSPF autonomous system as reachable, if OMPROUTE is running IPv6 OSPF and is configured as an IPv6 AS boundary router importing direct routes.

Method of assigning interface definitions to stack interfaces (wildcard and explicit):

For IPv6 interfaces, interface-name wildcards can be used to simplify definitions. However, be sure to understand how they are parsed, and how different types of interface definitions interact with each other, to avoid unintended results. Following is the outline of the algorithm OMPROUTE uses to find the matching definitions in the OMPROUTE configuration file for an IPv6 stack interface.

a. Search for an IPv6_RIP_Interface definition for the interface as follows:
   1) Search for an explicit matching IPv6_RIP_Interface statement for the interface. This is one where the name parameter exactly matches the interface's name. If one is found, use that definition and go to step b.
   2) Search for the best IPv6_RIP_Interface wildcard match for the name. The IPv6_RIP_Interface wildcard definitions are searched, starting with the most specific (longest wildcard name string) and checking each in order of declining specificity until a match is found. As soon as a match is found, use that definition and go to step b.

b. Search for an IPv6_OSPF_Interface definition for the interface. Note that this step is done regardless of the outcome of step a. The steps for searching IPv6_OSPF_Interface definitions are the same as the steps for searching IPv6_RIP_Interface definitions described above, except that IPv6_OSPF_Interface definitions are searched.

c. If either an IPv6_RIP_Interface or an IPv6_OSPF_Interface definition, or both, are found, the algorithm is complete. In this case, IPv6_Interface definitions are not searched. If neither an IPv6_RIP_Interface nor an IPv6_OSPF_Interface definition was found, go to step d.

d. Search for an IPv6_Interface definition for the interface. The steps for searching IPv6_Interface definitions are the same as the steps for

searching IPv6_RIP_Interface statements described above, except that IPv6_Interface definitions are searched.

e. If no definitions are found, check the value of Global_Options Ignore_Undefined_Interfaces. If this option is turned on, the interface is ignored. If it is not turned on, the interface is treated as if it were defined with an IPv6_Interface statement. Default values will be used for all parameters.

The algorithm is complete. Key conclusions of this algorithm are as follows:

- If an IPv6_RIP_Interface definition, an IPv6_OSPF_Interface definition, or both, are found, IPv6_Interface definitions are not considered. This means that any matching IPv6_RIP_Interface or IPv6_OSPF_Interface definition supersedes all IPv6_Interface definitions, even if the IPv6_Interface definitions are explicit or more specific wildcard matches. For example, an IPv6_OSPF_Interface definition with a name parameter of V* supersedes any IPv6_Interface, explicit or wildcard, with a name parameter that begins with V. In this case, the IPv6_Interface definition is redundant and will never be used. If OMPROUTE detects this case, it issues message EZZ8068I and deletes the redundant IPv6_Interface definition.

  Note: If an IPv6_Interface definition has already been selected for an interface that is installed in the stack, and then an IPv6_OSPF_Interface or IPv6_RIP_Interface definition that would make that IPv6_Interface definition redundant is added using RECONFIG, OMPROUTE issues message EZZ8069I and retains the IPv6_Interface definition.

- An interface can be both an IPv6_RIP_Interface and an IPv6_OSPF_Interface. OMPROUTE supports running both protocols over the same interface. However, an interface cannot be both an interface that runs no routing protocol (that is, defined with the IPv6_Interface statement) and one that runs IPv6 RIP, IPv6 OSPF, or both.

---

6. Define interface costs (OSPF_INTERFACE, RIP_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE).

   Both the OSPF and RIP protocols have a cost value associated with interfaces. With both protocols, the cost of a route to reach a destination is the sum of the costs of each link that will be traversed on the way to the destination. In the sample network shown in Figure 36 on page 220, the cost of a route to get from TCPCS7 to router 3.3.3.3 through TCPCS4 is the cost of the link from TCPCS7 to TCPCS4 plus the cost of the link from TCPCS4 to router 3.3.3.3.

   The method for configuring cost values differs between the OSPF and RIP protocols. The cost values of OSPF links should be configured to ensure that preferred routes to destinations will have a lower cost than less preferable routes. The less preferable routes, with the higher cost, will not be used except upon failure of the preferred routes.

   For the purpose of the following examples, the sample network shown in Figure 36 on page 220 is used and the convention stack (interface) is used to refer to the cost configured for a particular interface on a stack. For instance TCPCS7(9.67.106.7) refers to the cost configured for interface 9.67.106.7 on TCPCS7. While these examples use the IPv4 portion of the sample network, the same methods for computing route costs would also be used by the IPv6 portion.

   There are three possible routes from TCPCS7 to router 3.3.3.3. They are:

- Direct (TCPCS7 —> 3.3.3.3),
- Through TCPCS4 (TCPCS7 —> TCPCS4 —> 3.3.3.3)
- Through router 8.8.8.8 and TCPCS4 (TCPCS7 —> 8.8.8.8 —> TCPCS4 —> TCPCS3)

If the preferred route from TCPCS7 to router 3.3.3.3 is through TCPCS4, then interface costs must be configured such that the following are true:

```
TCPCS7(9.67.106.7) + TCPCS4(9.67.101.4) < TCPCS7(9.67.102.7)
TCPCS7(9.67.106.7) + TCPCS4(9.67.101.4) < TCPCS7(9.67.100.7) +
8.8.8.8(9.67.105.8) + TCPCS4(9.67.101.4)
```

The reasons for preferring one route over another are numerous. One approach for assigning OSPF link costs would be to set the costs to values inversely proportional to the bandwidth of the physical media. This would result in higher bandwidth routes having lower costs, thus becoming the preferred routes.

The cost values of RIP links are generally set to a value of 1. This results in the cost of a route to a destination being the number of hops to reach the destination. In the sample network, this would result in the three possible RIP routes from TCPCS7 to router 3.3.3.3 having the following costs:

- Direct (TCPCS7 -> 3.3.3.3), cost = 1
- Through TCPCS4 (TCPCS7 -> TCPCS4 -> 3.3.3.3), cost = 2
- Through router 8.8.8.8 and TCPCS4 (TCPCS7 -> 8.8.8.8 -> TCPCS4 -> TCPCS3), cost = 3

If it were desired that the route through TCPCS4 be the preferred route, this could be accomplished by increasing the cost of getting directly from TCPCS7 to router 3.3.3.3. This could be done by increasing either the out metric for 9.67.102.3 on router 3.3.3.3 or the in metric for 9.67.102.7 on TCPCS7. Take care when increasing in metric and out metric values to be sure that the cost to reach any destination does not exceed the RIP maximum of 15.

IPv4 OSPF and RIP

The cost value of an OSPF interface is set using the COST0 parameter of the OSPF_INTERFACE statement. The in metric and out metric of a RIP interface are set using the IN_METRIC and OUT_METRIC parameters of the RIP_INTERFACE statement.

IPv6 OSPF and RIP

The cost value of an IPv6 OSPF interface is set using the COST parameter of the IPV6_OSPF_INTERFACE statement. The in metric and out metric of an IPv6 RIP interface are set using the IN_METRIC and OUT_METRIC parameters of the IPV6_RIP_INTERFACE statement.

---

7. Configure virtual links, if the OSPF protocol is used.

The OSPF protocol is dependent upon complete connectivity of the backbone area. To maintain backbone connectivity, each backbone router must be interconnected. If the configuration of an OSPF autonomous system is such that the backbone area will become separated into two or more disconnected sections, connectivity must be restored for the protocol to work correctly. This can be done using a virtual link. An OSPF virtual link should not be confused with a VIPA link. Virtual links can be configured between any two backbone routers that have an interface to a common non-backbone area.

IPv4 OSPF

The VIRTUAL_LINK statements specify the router ID of the link endpoint and must be configured at both endpoints. In the sample network shown in Figure 36 on page 220, a virtual link is configured between TCPCS4 and TCPCS7 to restore backbone connectivity through area 1.1.1.1.

```
TCPCS4:
VIRTUAL_LINK
   Virtual_Endpoint_RouterID=7.7.7.7
   Links_Transit_Area=1.1.1.1;
TCPCS7:
VIRTUAL_LINK
   Virtual_Endpoint_RouterID=4.4.4.4
   Links_Transit_Area=1.1.1.1;
```

IPv6 OSPF

The IPV6_VIRTUAL_LINK statements specify the router ID of the link endpoint and must be configured at both endpoints. In the sample network, a virtual link is configured between TCPCS64 and TCPCS67 to restore backbone connectivity through area 6.6.6.6.

```
TCPCS64:
IPV6_VIRTUAL_LINK
   Virtual_Endpoint_RouterID=67.67.67.67
   Links_Transit_Area=6.6.6.6;
TCPCS67:
IPV6_VIRTUAL_LINK
   Virtual_Endpoint_RouterID=64.64.64.64
   Links_Transit_Area=6.6.6.6;
```

---

8. Manage high-cost links, if the OSPF protocol is used.

   The periodic nature of OSPF routing traffic requires a link's underlying data-link connection to be constantly open. This can result in unwanted usage charges on network segments whose costs are very high. There are two configuration steps that can be taken to inhibit the periodic nature of the protocol.

   The first step that can be taken is to define the link as a demand circuit. When this is done, link state advertisements (LSAs) sent over the interface will not be periodically refreshed. Only LSAs with real changes will be readvertised. In addition, aging of these LSAs will be disabled such that they will not age out of the link state database.

   Another step that can be taken is to define hello suppression for the link. Hello suppression is only meaningful if the link is a demand circuit and is either point-to-point or point-to-multipoint. Hello suppression will inhibit the periodic transmission of OSPF hello packets.

   IPv4 OSPF

   To define OSPF interfaces as demand circuits, the Demand_Circuit=YES parameter must first be specified on the global OSPF configuration statement. Then, the OSPF_INTERFACE statement for each interface to be configured as a demand circuit must be specified with the Demand_Circuit=YES parameter. Use the Hello_Suppression parameter of the OSPF_INTERFACE statement to configure hello suppression. For more information on configuring the Hello_Suppression parameter on the OSPF_INTERFACE statement, refer to *z/OS Communications Server: IP Configuration Reference*. If hello suppression is implemented, the PP_Poll_Interval parameter of the OSPF_INTERFACE statement can be used to specify the interval at which OMPROUTE should attempt to contact a neighbor to reestablish a neighbor relationship when the relationship has failed, but the interface is still available.

IPv6 OSPF

To define IPv6 OSPF interfaces as demand circuits, the Demand_Circuit=YES parameter must first be specified on the global IPV6_OSPF configuration statement. Then, the IPV6_OSPF_INTERFACE statement for each interface to be configured as a demand circuit must be specified with the Demand_Circuit=YES parameter. Use the Hello_Suppression parameter of the IPV6_OSPF_INTERFACE statement to configure hello suppression. For more information on configuring the Hello_Suppression parameter on the IPV6_OSPF_INTERFACE statement, refer to *z/OS Communications Server: IP Configuration Reference*. If hello suppression is implemented, the PP_Poll_Interval parameter of the IPV6_OSPF_INTERFACE statement can be used to specify the interval at which OMPROUTE should attempt to contact a neighbor to reestablish a neighbor relationship when the relationship has failed, but the interface is still available.

---

9. Define RIP filters, if the RIP protocol is used.

   RIP Filters can be configured to OMPROUTE such that certain RIP routing information will not be broadcast out to other routers and/or accepted from other routers. The filters can be applied to individual RIP interfaces or to all RIP interfaces. When defining a filter, a filter type (sending or receiving) is specified along with values identifying the route information to be filtered. By using filters, an installation can limit the amount of RIP routing information broadcast into the network and/or the amount of RIP routing information maintained by OMPROUTE. In addition, filters can be used to hide destination addresses from portions of the network.

   IPv4 RIP

   To configure a filter for an individual RIP interface, use the FILTER parameter of the RIP_INTERFACE statement. To configure a filter that applies to all IPv4 RIP interfaces, use the global FILTER statement. In the sample network shown in Figure 36 on page 220, if you wanted to hide the 10.1.1.0 subnet from TCPCS6 (as well as all routers and hosts on the remote side of TCPCS6), you could define the following filter on TCPCS7:

   ```
   Filter=(nosend,10.1.1.0,255.255.255.0);
   ```

   IPv6 RIP

   To configure a filter for an individual IPv6 RIP interface, use the FILTER parameter of the IPV6_RIP_INTERFACE statement. To configure a filter that applies to all IPv6 RIP interfaces, use the global IPV6_RIP_FILTER statement. In the sample network shown in Figure 36 on page 220, for example, if you wanted to hide the 2001:0DB8:0:A1B/64 prefix from TCPCS4, you could define the following filter on TCPCS4:

   ```
   IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1B/64);
   ```

---

10. Define route precedence in a multiprotocol environment, if the OSPF protocol is used.

    Note that this discussion of route precedence is quite complicated. If only the OSPF or IPv6 OSPF routing protocol, or both, are used in your network, route precedence is less of a concern. If, in addition, none of your OSPF or IPv6 OSPF routers are configured as AS boundary routers, the route precedence concern is entirely eliminated. For environments with multiple protocols or AS boundary routers, the following information is provided. Note that in this discussion, RIP is meant to apply to both RIP and IPv6 RIP,

OSPF is meant to apply to both OSPF and IPv6 OSPF, and the OSPF configuration statement is meant to apply to both the OSPF statement and the IPV6_OSPF statement.

OMPROUTE applies an order of precedence in choosing between two routes to the same destination that were learned through different routing protocols or using information provided by an OSPF AS boundary router. To describe this order of precedence applied by OMPROUTE, a few terms must first be defined.

RIP route
A route learned through the RIP protocol. A RIP route is generated using information provided in a RIP packet from a neighboring router. For example, in the sample network shown in Figure 36 on page 220, the route from TCPCS7 to destination subnet 30.1.1.0 is a RIP route.

OSPF internal route
A route learned through the OSPF protocol where the entire path traversed to reach the destination lies within the OSPF autonomous system. For example, in the sample network shown in Figure 36 on page 220, the route from TCPCS7 to destination 9.67.108.2 on Router 2.2.2.2 is an OSPF internal route.

OSPF external route
A route learned through the OSPF protocol where part of the path traversed to reach the destination does not lie within the OSPF autonomous system. The path will leave the autonomous system if it uses information brought into the OSPF autonomous system by an AS boundary router. This information brought into the OSPF AS may be information imported from a different autonomous system (for example, RIP) or information about destinations statically configured on or directly connected to the AS boundary router. For example, in the sample network shown in Figure 36 on page 220, the route from TCPCS4 to destination 9.67.103.6 on TCPCS6 is an OSPF external route. TCPCS7, configured as an AS boundary router, has imported information about that destination into the OSPF AS from the RIP AS.

OSPF external routes fall into two categories based upon the setting of the multiprotocol comparison value. If the comparison value is set to Type1 on the AS boundary router that imports the external information into the OSPF AS, then OSPF external routes generated using this information will be OSPF type 1 external routes. If the comparison value is set to Type2 on the AS boundary router, then the generated routes will be OSPF type 2 external routes. For example, in the sample network shown in Figure 36 on page 220, if the comparison value on TCPCS7 (an AS boundary router) is set to Type1, the route from TCPCS4 to destination 9.67.103.6 on TCPCS6 is an OSPF type 1 external route. If the comparison value on TCPCS7 is set to Type2, the route is an OSPF type 2 external route.

Multiprotocol comparison

You can configure this comparison value to allow for the specification of how route costs from different autonomous systems should be treated when they coexist. In OMPROUTE, you can configure this value using the COMPARISON parameter on the OSPF or IPV6_OSPF configuration statements. When COMPARISON=Type1 is configured, the route cost values

used within different autonomous systems (for example, the OSPF AS and the RIP AS) are considered comparable. With COMPARISON=Type2 configured, the route cost values used with the different autonomous systems are considered non-comparable.

The comparison value can be used in several different ways, depending on the function being performed by a router:

- As an AS boundary router, OMPROUTE uses the comparison value to determine the type of external routes (type 1 or type 2) generated by routers in the OSPF AS using routing information that the AS boundary router imports into the OSPF AS.
- As an AS boundary router, OMPROUTE also uses the comparison value in determining how route cost values will be assigned when importing routes from the OSPF AS into the RIP AS.
  - When COMPARISON=Type1 is configured (indicating that cost values are comparable), an OSPF route imported into the RIP AS will be advertised with the actual cost of the OSPF route. For example, in the sample network, if TCPCS7 is configured with COMPARISON=Type1 and the OSPF route from TCPCS7 to destination 9.67.108.2 on TCPCS2 has a cost of 7, then TCPCS7 will advertise into the RIP AS a RIP route to that destination with a cost of 7.

    Notes:

    a. An exception to this rule (defining how OSPF routes are advertised into the RIP AS when COMPARISON=Type1) occurs when the OSPF route to be imported is an OSPF type 2 external route. When this is the case, the route is not advertised into the RIP AS at all.

    b. It is important to remember the requirement that all destinations in the RIP AS must be reachable with a cost no greater than 15. Using COMPARISON=Type1 requires that the cost values of OSPF routes be low. Any destinations in the OSPF AS that can only be reached from the RIP AS with a cost greater than 15 will become unreachable.
  - When COMPARISON=Type2 is configured (indicating that cost values are non-comparable), an OSPF route imported into the RIP AS is advertised with a cost of 1. If a router in the RIP AS has two possible routes to a destination, one internal to the RIP AS and another that was imported from OSPF, this approach results in the route imported from OSPF being favored. For example, in the sample network shown in Figure 36 on page 220, if TCPCS7 is configured with COMPARISON=Type2 and TCPCS7 can somehow reach a destination in the 30.1.1.0 subnet without passing through TCPCS6 (using links not shown in the sample), then TCPCS7 advertises into the RIP AS a RIP route to the destination with a cost of 1. As a result, TCPCS6 determines that the destination can be reached through TCPCS7 with a cost of 2. If the cost of the route for TCPCS6 to reach the destination internal to the RIP AS is greater than 2, then the route through TCPCS7 is chosen.

    Note: An exception to this rule (defining how OSPF routes are advertised into the RIP AS when COMPARISON=Type2) occurs when the OSPF route to be imported is an OSPF type 2 external route. When this is the case, the route is advertised into the RIP AS with the actual cost of the OSPF type 2 external route.

- As any router that has routing information from different autonomous systems, OMPROUTE uses the comparison value while choosing between the routes generated using the information from the different autonomous systems. How the comparison value is used in this case is shown in Table 16.

Given these definitions, the order of precedence used in choosing between multiple routes to the same destination, which were learned through the different protocols or by using information provided by an OSPF AS boundary router, can be shown in Table 16. In Table 16, *Source comparison* refers to the setting of the comparison value (using the COMPARISON parameter on the OSPF configuration statement) on the router that is using the order of precedence to choose between the multiple routes. *Route 1* and *Route 2* are the two possible routes being chosen between.

Table 16. Route precedence

Source comparison	Route 1 type	Route 2 type	Route chosen
Type 1	OSPF internal	RIP	OSPF internal
Type 1	OSPF internal	OSPF type 1 external	OSPF internal
Type 1	OSPF internal	OSPF type 2 external	OSPF internal
Type 1	RIP	OSPF type 1 external	Lowest cost route
Type 1	RIP	OSPF type 2 external	RIP route
Type 1	OSPF type 1 external	OSPF type 2 external	OSPF type 1 external
Type 2	OSPF internal	RIP	OSPF internal
Type 2	OSPF internal	OSPF type 1 external	OSPF internal
Type 2	OSPF internal	OSPF type 2 external	OSPF internal
Type 2	RIP	OSPF type 1 external	OSPF type 1 external
Type 2	RIP	OSPF type 2 external	Lowest cost route
Type 2	OSPF type 1 external	OSPF type 2 external	OSPF type 1 external

Network design considerations with z/OS Communications Server

OMPROUTE may be run on z/OS Communications Server for a variety of reasons. If the z/OS Communications Server host is being used as an application or server host and the routing daemon is being run primarily to provide access to network resources, or to provide network resources access to the z/OS Communications Server host, then care must be taken to ensure that the z/OS Communications Server host is not overly burdened with routing work. Unlike routers or other network boxes whose sole purpose is routing, an application host z/OS Communications Server will be doing many things other than routing, and it is not desirable for a large percentage of machine resources (memory and CPU) to be used for routing tasks, as can happen in very complex or unstable networks. In this case the z/OS Communications Server should not be configured as a backbone router, either intentionally or inadvertently. Careful network design can minimize the routing burdens on the z/OS Communications Server application host without compromising the accessibility of z/OS Communications Server resources to the network and vice versa. If care is not taken to minimize the routing work required by the z/OS Communications Server host, OMPROUTE may consume excessive cycles or memory processing huge numbers of routing updates from the network. Or the burden of routing updates may become so large that the z/OS

Communications Server cannot keep up because of other workloads on the machine. Since OSPF is heavily timer-driven, this could cause loss of adjacencies and routing problems.

The primary way to reduce the routing burdens on the z/OS Communications Server host is by use of OSPF areas. See step 2 on page 250 for more information. A z/OS Communications Server application host or sysplex can be placed into a non-backbone area with dedicated routers acting as area-border routers. The area-border routers would advertise the z/OS Communications Server's resources to other attached areas (for example the backbone) and would summarize the network outside the local area to the z/OS Communications Server hosts. If possible, this can be further refined to reduce routing protocol traffic by use of interarea route summarization, accomplished in OMPROUTE area-border routers by the RANGE and IPV6_RANGE statements, and in Cisco routers with the area range command. For more information, refer to *z/OS Communications Server: IP Configuration Reference* and step 3 on page 251.

An even further, and ideal, optimization would be to make the area containing the z/OS Communications Server application host or sysplex a stub area. A stub area can be configured such that route summaries (IPv4) or prefixes (IPv6) from other areas are not flooded into the stub area by the area border routers. When this is done, only routes to destinations within the stub area are shared among the hosts. Default routes are used to represent all destinations outside the stub area. The stub area's resources are still advertised to the network at large by the area-border routers. You can use this optimization, sometimes referred to as a totally stubby area, if the following apply to your network:

- It is acceptable to use default routes to reach destinations outside the stub area. This means that either there is only one area-border router connecting the stub area to the rest of the network, or if there are multiple such connections they are redundant, so that it does not matter which one is used to get outside the stub area.
- You have no non-OSPF destinations to advertise to the network at large. Stub areas do not permit importation of OSPF external routes. This means for example that you do not have a RIP network attached to the stub area, or if you do, you do not want its destinations reachable from the stub area. Other types of routes that cannot be imported into stub areas include direct routes (for example, for networks attached to interfaces that are not running the OSPF protocol) and static routes. RIP or static routes can be used by the z/OS Communications Server that learns them, they just cannot be advertised.

  Tip: If you define your VIPAs as OSPF interfaces in your OMPROUTE configuration file, routes to their addresses will be considered OSPF routes, and therefore importable into the stub area and able to be advertised by the area-border routers to the network at large.

It is highly recommended to put z/OS Communications Server application hosts or sysplexes into stub areas if at all possible.

A further optimization is to prevent z/OS Communications Server from becoming the designated router on multiaccess media, when pure routers that can perform this function are present. On a multiaccess medium, the designated router and the backup designated router will carry the majority of the routing protocol load for all hosts on the medium. While z/OS Communications Server is capable of performing this role, it does impose additional routing overhead on the system. It would be preferable to allow pure routers to perform this role, if they are available. This is accomplished by ensuring that the pure routers' interfaces onto the medium have higher ROUTER_PRIORITY values than the z/OS

Communications Server interfaces on the same medium. However, if the only hosts on a medium are z/OS Communications Server, such as in HiperSockets communications, then one or two of them will have to be designated router or backup designated router.

Verification of OMPROUTE IPv4 configuration and state

The following sections show sample output from each of the commands that can be used to display OMPROUTE IPv4 information. The syntax of these DISPLAY commands, as well as detailed information about the data displayed, can be found in *z/OS Communications Server: IP System Administrator's Commands*.

Note: All commands that include the LIST subparameter indicate that the information being displayed is configured information only and does not necessarily mean that the information is currently being used by OMPROUTE. To display information in current use, use related commands to display current, run-time statistics, and parameters. There are cases when the configured information will not match the in-use information due to some undefined or unresolved information in the OMPROUTE configuration. For example, undefined interfaces or parameters in the OMPROUTE configuration or an incorrect sequence of dynamic reconfiguration with the MODIFY OMPROUTE,RECONFIG command can result in no update of the in-use information at all. Information defined on wildcard interfaces is not displayed in the LIST commands; it is only displayed in the corresponding non-LIST commands when wildcard information is resolved to actual physical interfaces.

Displaying all OSPF configuration information

To display all of the OSPF configuration information, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,ALL
EZZ7831I GLOBAL CONFIGURATION 735
    TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
    STACK AFFINITY:          TCPCS7
    OSPF PROTOCOL:           ENABLED
    EXTERNAL COMPARISON:     TYPE 2
    AS BOUNDARY CAPABILITY:  ENABLED
    IMPORT EXTERNAL ROUTES:  RIP SUB
    ORIG. DEFAULT ROUTE:     NO
    DEFAULT ROUTE COST:      (1, TYPE 2)
    DEFAULT FORWARD. ADDR.:  0.0.0.0
    LEARN HIGHER COST DFLT:  NO
    DEMAND CIRCUITS:         ENABLED

EZZ7832I AREA CONFIGURATION
AREA ID          AUTYPE          STUB? DEFAULT-COST IMPORT-SUMMARIES?
0.0.0.0          0=NONE           NO          N/A           N/A
1.1.1.1          0=NONE           NO          N/A           N/A

--AREA RANGES--
AREA ID          ADDRESS          MASK             ADVERTISE?
1.1.1.1          9.67.101.0       255.255.255.0    NO

EZZ7833I INTERFACE CONFIGURATION
IP ADDRESS      AREA             COST  RTRNS  TRNSDLY PRI  HELLO   DEAD    DB_EX
7.7.7.7         1.1.1.1             1    N/A    N/A   N/A    N/A    N/A     N/A
9.67.104.7      1.1.1.1             1      5      1     1     10     40      40
9.67.100.7      1.1.1.1             1      5      1     1     10     40      40
9.67.102.7      1.1.1.1             1      5      1     1     10     40      40
9.67.106.7      1.1.1.1             1      5      1     1     10     40      40
```

```
9.67.107.7      0.0.0.0             1      5      1     1     10     40      40

EZZ7836I VIRTUAL LINK CONFIGURATION
VIRTUAL ENDPOINT     TRANSIT AREA      RTRNS  TRNSDLY HELLO  DEAD  DB_EX
4.4.4.4              1.1.1.1             10      5     30   180    180

EZZ7835I NBMA CONFIGURATION
                INTERFACE ADDR      POLL INTERVAL
                          9.67.104.7          180
EZZ7834I NEIGHBOR CONFIGURATION
                NEIGHBOR ADDR     INTERFACE ADDRESS   DR ELIGIBLE?
                9.67.104.15       9.67.104.7          YES
                9.67.104.25       9.67.104.7          NO
                9.67.104.16       9.67.104.7          NO
```

Displaying information about configured OSPF areas

To display information about configured OSPF Areas, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,AREAS
EZZ7832I AREA CONFIGURATION 737
AREA ID          AUTYPE          STUB? DEFAULT-COST IMPORT-SUMMARIES?
0.0.0.0          0=NONE           NO          N/A           N/A
1.1.1.1          0=NONE           NO          N/A           N/A

--AREA RANGES--
AREA ID          ADDRESS          MASK             ADVERTISE?
1.1.1.1          9.67.101.0       255.255.255.0    NO
```

Displaying configuration information about configured OSPF interfaces

To display configuration information about configured OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,IFS
EZZ7833I INTERFACE CONFIGURATION 739
IP ADDRESS       AREA             COST  RTRNS  TRNSDLY PRI  HELLO   DEAD   DB_EX
7.7.7.7          1.1.1.1             1    N/A    N/A   N/A    N/A    N/A    N/A
9.67.104.7       1.1.1.1             1      5      1     1     10     40     40
9.67.100.7       1.1.1.1             1      5      1     1     10     40     40
9.67.102.7       1.1.1.1             1      5      1     1     10     40     40
9.67.106.7       1.1.1.1             1      5      1     1     10     40     40
9.67.107.7       0.0.0.0             1      5      1     1     10     40     40
```

Note: Wildcard interface definitions are not displayed. However, when an actual interface is resolved to a wildcard definition, its information is displayed.

Displaying information about configured Non-broadcast Multiple Access OSPF interfaces

To display information about configured Non-broadcast Multiple Access OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,NBMA
EZZ7835I NBMA CONFIGURATION 745
                INTERFACE ADDR      POLL INTERVAL
                9.67.104.7          180
```

Displaying information about configured OSPF virtual links

To display information about configured OSPF virtual links, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,VLINKS
EZZ7836I VIRTUAL LINK CONFIGURATION 747
VIRTUAL ENDPOINT     TRANSIT AREA      RTRNS  TRNSDLY HELLO  DEAD  DB_EX
4.4.4.4              1.1.1.1             10      5     30   180    180
```

Displaying information about configured OSPF neighbors

To display information about configured OSPF neighbors enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,NBRS
EZZ7834I NEIGHBOR CONFIGURATION 749
               NEIGHBOR ADDR     INTERFACE ADDRESS   DR ELIGIBLE?
               9.67.104.15       9.67.104.7          YES
               9.67.104.25       9.67.104.7          NO
               9.67.104.16       9.67.104.7          NO
```

Displaying the contents of a single OSPF link state advertisement

To display the contents of a single OSPF link state advertisement, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LSA,LSTYPE=1,LSID=7.7.7.7,ORIG=7.7.7.7,AREAID=1.1.1.1
EZZ7880I LSA DETAILS 751
        LS AGE:          521
        LS OPTIONS:      E,DC
        LS TYPE:         1
        LS DESTINATION (ID): 7.7.7.7
        LS ORIGINATOR:   7.7.7.7
        LS SEQUENCE NO:  0X80000013
        LS CHECKSUM:     0XA9A
        LS LENGTH:       120
        ROUTER TYPE:  ABR,ASBR,V
        # ROUTER IFCS:   8
                LINK ID:          7.7.7.4
                LINK DATA:        255.255.255.252
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0  METRIC:  1
                LINK ID:          8.8.8.8
                LINK DATA:        9.67.100.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          3.3.3.3
                LINK DATA:        9.67.102.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          4.4.4.4
                LINK DATA:        9.67.106.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          7.7.7.7
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
                LINK ID:          9.67.100.8
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
                LINK ID:          9.67.102.3
                LINK DATA:        255.255.255.255
```

```
INTERFACE TYPE:   3
        NO. OF METRICS: 0
        TOS 0 METRIC:   1
LINK ID:          9.67.106.4
LINK DATA:        255.255.255.255
INTERFACE TYPE:   3
        NO. OF METRICS: 0
        TOS 0 METRIC:   1
```

Displaying statistics and parameters for OSPF areas

To display statistics and parameters for all OSPF areas attached to the router, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,AREASUM
EZZ7848I AREA SUMMARY 757
AREA ID        AUTHENTICATION   #IFCS  #NETS  #RTRS  #BRDRS DEMAND
0.0.0.0           NONE             2      0      4      2  ON
1.1.1.1           NONE             5      0      4      2  ON
```

Displaying the list of AS external advertisements

To display a list of AS external advertisements that are in the OSPF link state database, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,EXTERNAL
EZZ7853I AREA LINK STATE DATABASE 759
TYPE LS DESTINATION     LS ORIGINATOR     SEQNO     AGE   XSUM
  5 @6.6.6.6            7.7.7.7         0X80000007  825  0X1B5C
  5 @9.67.103.6         7.7.7.7         0X80000007  831  0XE1F3
  5 @10.1.1.0           2.2.2.2         0X80000003 1690  0X2775
  5 @10.1.1.1           2.2.2.2         0X80000003 1690  0X1D7E
  5 @20.1.1.0           5.5.5.5         0X80000003 1616  0X4A3C
  5 @20.1.1.1           5.5.5.5         0X80000003 1616  0X4045
  5 @30.0.0.0           7.7.7.7         0X80000006  831  0XB0C0
  5 @30.1.1.0           7.7.7.7         0X80000006  831  0X99D5
  5 @30.1.1.4           7.7.7.7         0X80000001  825  0X7BF4
  5 @30.1.1.8           7.7.7.7         0X80000001  825  0X5319
  5 @130.200.0.0        3.3.3.3         0X80000003 1695  0X98C0
  5 @130.200.0.0        8.8.8.8         0X80000003 1630  0X243
  5 @130.200.1.1        3.3.3.3         0X80000003 1695  0X83D3
  5 @130.200.1.18       8.8.8.8         0X80000003 1630  0X42EF
  5 @130.201.0.0        3.3.3.3         0X80000003 1695  0X8CCB
  5 @130.201.0.0        8.8.8.8         0X80000003 1630  0XF54E
  5 @130.202.0.0        3.3.3.3         0X80000003 1694  0X80D6
  5 @130.202.0.0        8.8.8.8         0X80000003 1629  0XE959
               # ADVERTISEMENTS:       18
               CHECKSUM TOTAL:         0X83472
```

Displaying a list of non-AS external advertisements

To display a list of non-AS external advertisements that are in the OSPF link state database for a particular OSPF area, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,DATABASE,AREAID=1.1.1.1
EZZ7853I AREA LINK STATE DATABASE 761
TYPE LS DESTINATION     LS ORIGINATOR     SEQNO     AGE   XSUM
  1 @3.3.3.3            3.3.3.3         0X8000000F  879  0X8B11
  1 @4.4.4.4            4.4.4.4         0X8000001A  713  0XA020
  1 @7.7.7.7            7.7.7.7         0X80000013  711  0XA9A
  1 @8.8.8.8            8.8.8.8         0X8000000D  861  0XBD81
  3 @2.2.2.2            4.4.4.4         0X80000003 1676  0XC45C
  3 @5.5.5.4            7.7.7.7         0X80000003  880  0XE327
  3 @5.5.5.5            7.7.7.7         0X80000003  880  0XDF29
  3 @7.7.7.4            7.7.7.7         0X80000001  710  0X956E
  3 @9.67.107.5         7.7.7.7         0X80000006  881  0X4A14
  3 @9.67.107.7         7.7.7.7         0X80000003  880  0X4618
```

```
3 @9.67.108.2         4.4.4.4         0X80000003 1667  0XBDB1
3 @9.67.108.4         4.4.4.4         0X80000003 1658  0XB3B8
4 @2.2.2.2            4.4.4.4         0X80000003 1658  0XAC74
4 @5.5.5.5            7.7.7.7         0X80000003  880  0XC741
              # ADVERTISEMENTS:       14
              CHECKSUM TOTAL:         0X884B0
```

Displaying current, run-time statistics and parameters for OSPF interfaces

To display current, run-time statistics and parameters for OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,INTERFACE
EZZ7849I INTERFACES 763
IFC ADDRESS     PHYS         ASSOC. AREA     TYPE   STATE  #NBRS  #ADJS
7.7.7.7         VIPA1A       1.1.1.1         VIPA    N/A    N/A    N/A
9.67.104.7      NBMA7        1.1.1.1         MULTI     1      3      0
9.67.100.7      CTC7TO8      1.1.1.1         P-P      16      1      1
9.67.102.7      CTC7TO3      1.1.1.1         P-P      16      1      1
9.67.106.7      CTC7TO4      1.1.1.1         P-P      16      1      1
9.67.107.7      CTC7TO5      0.0.0.0         P-P      16      1      1
UNNUMBERED      VL/0         0.0.0.0         VLINK    16      1      1
```

Displaying current, run-time statistics and parameters for a specific OSPF interface

To display current, run-time statistics and parameters for a specific OSPF interface, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,IF,NAME=CTC7TO4
EZZ7850I INTERFACE DETAILS 769
               INTERFACE ADDRESS:      9.67.106.7
               ATTACHED AREA:          1.1.1.1
               PHYSICAL INTERFACE:     CTC7TO4
               INTERFACE MASK:         255.255.255.0
               INTERFACE TYPE:         P-P
               STATE:                  16
               DESIGNATED ROUTER:      N/A
               BACKUP DR:              N/A

 DR PRIORITY:     N/A  HELLO INTERVAL:   10  RXMT INTERVAL:     5
 DEAD INTERVAL:    40  TX DELAY:          1  POLL INTERVAL:     0
 DEMAND CIRCUIT:  OFF  HELLO SUPPRESS:  OFF  SUPPRESS REQ:    OFF
 MAX PKT SIZE:   1024  TOS 0 COST:        1  DB_EX INTERVAL:   40
 AUTH TYPE:  PASSWORD

 # NEIGHBORS:       1  # ADJACENCIES:     1  # FULL ADJS.:      1
 # MCAST FLOODS:   15  # MCAST ACKS:      4

 NETWORK CAPABILITIES:
  POINT-TO-POINT
  DEMAND-CIRCUITS
```

Displaying current, run-time statistics and parameters for OSPF neighbors

To display current, run-time statistics and parameters for OSPF neighbors, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,NBR
EZZ7851I NEIGHBOR SUMMARY 771
NEIGHBOR ADDR   NEIGHBOR ID     STATE  LSRXL DBSUM LSREQ HSUP IFC
9.67.104.16     0.0.0.0             1     0     0     0  OFF NBMA7
9.67.104.25     0.0.0.0             1     0     0     0  OFF NBMA7
9.67.104.15     0.0.0.0             1     0     0     0  OFF NBMA7
```

```
9.67.100.8      8.8.8.8           128     0     0     0  OFF CTC7TO8
9.67.102.3      3.3.3.3           128     0     0     0  OFF CTC7TO3
9.67.106.4      4.4.4.4           128     0     0     0  OFF CTC7TO4
9.67.107.5      5.5.5.5           128     0     0     0  OFF CTC7TO5
VL/0            4.4.4.4           128     0     0     0  OFF *
```

Displaying current run-time statistics and parameters for a specific OSPF neighbor

To display current run-time statistics and parameters for a specific OSPF neighbor, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,NBR,IPADDR=9.67.106.4
EZZ7852I NEIGHBOR DETAILS 779
               NEIGHBOR IP ADDRESS:    9.67.106.4
               OSPF ROUTER ID:         4.4.4.4
               NEIGHBOR STATE:         128
               PHYSICAL INTERFACE:     CTC7TO4
               DR CHOICE:              0.0.0.0
               BACKUP CHOICE:          0.0.0.0
               DR PRIORITY:            1
               NBR OPTIONS:            E
  DB SUMM QLEN:      0  LS RXMT QLEN:      0  LS REQ QLEN:       0
  LAST HELLO:        4  NO HELLO:        OFF
  # LS RXMITS:       1  # DIRECT ACKS:     0  # DUP LS RCVD:     6
  # OLD LS RCVD:     0  # DUP ACKS RCVD:   1  # NBR LOSSES:      0
  # ADJ. RESETS:     0
```

Displaying routes to other routers that have been calculated by OSPF

To display routes to other routers that have been calculated by OSPF, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,ROUTERS
EZZ7855I OSPF ROUTERS 781
DTYPE RTYPE DESTINATION       AREA           COST       NEXT HOP(S)
ASBR  SPF   2.2.2.2          0.0.0.0          2        9.67.106.4
  BR  SPF   4.4.4.4          0.0.0.0          1        9.67.106.4
ASBR  SPF   5.5.5.5          0.0.0.0          1        9.67.107.5
ASBR  SPF   3.3.3.3          1.1.1.1          1        9.67.102.3
  BR  SPF   4.4.4.4          1.1.1.1          1        9.67.106.4
ASBR  SPF   8.8.8.8          1.1.1.1          1        9.67.100.8
```

Displaying the number of LSAs currently in the link state database

To display the number of LSAs currently in the link state database, categorized by type, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,DBSIZE
EZZ7854I LINK STATE DATABASE SIZE 783
               # ROUTER-LSAS:            8
               # NETWORK-LSAS:           0
               # SUMMARY-LSAS:           37
               # SUMMARY ROUTER-LSAS:    7
               # AS EXTERNAL-LSAS:       18
               # INTRA-AREA ROUTES:      24
               # INTER-AREA ROUTES:      1
               # TYPE 1 EXTERNAL ROUTES: 0
               # TYPE 2 EXTERNAL ROUTES: 0
```

Displaying statistics generated by the OSPF routing protocol

To display statistics generated by the OSPF routing protocol, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,STATS
EZZ7856I OSPF STATISTICS 785
                OSPF ROUTER ID:         7.7.7.7
                EXTERNAL COMPARISON:    TYPE 2
                AS BOUNDARY CAPABILITY: YES
                IMPORT EXTERNAL ROUTES: RIP SUB
                ORIG. DEFAULT ROUTE:    YES
                DEFAULT ROUTE COST:     (1, TYPE 2)
                DEFAULT FORWARD. ADDR.: 0.0.0.0
                LEARN HIGHER COST DFLT: NO
ATTACHED AREAS:                  2  OSPF PACKETS RCVD:              821
OSPF PACKETS RCVD W/ERRS:        0  TRANSIT NODES ALLOCATED:         55
TRANSIT NODES FREED:            47  LS ADV. ALLOCATED:              263
LS ADV. FREED:                 201  QUEUE HEADERS ALLOC:             96
QUEUE HEADERS AVAIL:            96  MAXIMUM LSA SIZE:               976
# DIJKSTRA RUNS:                 9  INCREMENTAL SUMM. UPDATES:        4
INCREMENTAL VL UPDATES:          0  MULTICAST PKTS SENT:            746
UNICAST PKTS SENT:             107  LS ADV. AGED OUT:                 0
LS ADV. FLUSHED:                22  PTRS TO INVALID LS ADV:           0
INCREMENTAL EXT. UPDATES:       49
```

Displaying all of the RIP configuration information

To display all of the RIP configuration information, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,ALL
EZZ7843I RIP CONFIGURATION 800
TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
STACK AFFINITY: TCPCS7
RIP: ENABLED
RIP DEFAULT ORIGINATION: ALWAYS, COST = 1
PER-INTERFACE ADDRESS FLAGS:
CTC7TO6          9.67.103.7      RIP-2 MULTICAST.
                                 SEND NET AND SUBNET ROUTES
                                 RECEIVE NO DYNAMIC HOST ROUTES
                                 RIP INTERFACE INPUT METRIC: 1
                                 RIP INTERFACE OUTPUT METRIC: 0

EZZ7844I RIP ROUTE ACCEPTANCE
ACCEPT RIP UPDATES ALWAYS FOR:
  30.1.1.8          30.1.1.4
IGNORE RIP UPDATES FROM:
   NONE
```

Displaying information about configured RIP interfaces

To display information about configured RIP interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,IFS
EZZ7843I RIP CONFIGURATION 806
TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
STACK AFFINITY: TCPCS7
RIP: ENABLED
RIP DEFAULT ORIGINATION: ALWAYS, COST = 1
PER-INTERFACE ADDRESS FLAGS:
CTC7TO6          9.67.103.7      RIP-2 MULTICAST.
                                 SEND NET AND SUBNET ROUTES
                                 RECEIVE NO DYNAMIC HOST ROUTES
                                 RIP INTERFACE INPUT METRIC: 1
                                 RIP INTERFACE OUTPUT METRIC: 0
                                 RIP RECEIVE CONTROL: ANY
```

Displaying the routes to be unconditionally accepted

To display the routes to be unconditionally accepted, as configured with the Accept_RIP_Route statement, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,ACCEPTED
EZZ7844I RIP ROUTE ACCEPTANCE 808
ACCEPT RIP UPDATES ALWAYS FOR:
  30.1.1.8          30.1.1.4
```

Displaying current run-time information about RIP interfaces

To display current, run-time information about RIP interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,IF
EZZ7859I RIP INTERFACES 810
IFC ADDRESS      IFC NAME         SUBNET MASK     MTU   DESTINATION
9.67.103.7       CTC7TO6          255.255.255.0   1024  0.0.0.0
```

Displaying current run-time information about a specific RIP interface

To display current, run-time information about a specific RIP interface, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,IF,NAME=CTC7TO6
EZZ7860I RIP INTERFACE DETAILS 812
INTERFACE ADDRESS:      9.67.103.7
INTERFACE NAME:         CTC7TO6
SUBNET MASK:            255.255.255.0
MTU                     1024
DESTINATION ADDRESS:    0.0.0.0

RIP VERSION:            2     SEND POIS. REV. ROUTES: YES
IN METRIC:              1     OUT METRIC:             0
RECEIVE NET ROUTES:     YES   RECEIVE SUBNET ROUTES:  YES
RECEIVE HOST ROUTES:    NO    SEND DEFAULT ROUTES:    NO
SEND NET ROUTES:        YES   SEND SUBNET ROUTES:     YES
SEND STATIC ROUTES:     NO    SEND HOST ROUTES:       NO

SEND ONLY: ALL

RIP RECEIVE CONTROL:    ANY
```

Displaying the global RIP filters

To display the global RIP filters, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,FILTERS
EZZ8016I GLOBAL RIP FILTERS 684
SEND ONLY: ALL

IGNORE RIP UPDATES FROM:
                     9.67.103.10              9.67.103.9

FILTERS: NOSEND          10.1.1.0           255.255.255.0
```

Displaying the routes in the OMPROUTE routing table

To display all of the routes in the OMPROUTE routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE
EZZ7847I ROUTING TABLE 796
TYPE   DEST NET         MASK      COST    AGE     NEXT HOP(S)

SBNT   2.0.0.0          FF000000  1       1368    NONE
 SPF   2.2.2.0          FFFFFFFC  3       1380    9.67.106.4
 SPF   2.2.2.2          FFFFFFFF  3       1380    9.67.106.4
SBNT   3.0.0.0          FF000000  1       1549    NONE
 SPF   3.3.3.0          FFFFFFFC  2       1561    9.67.102.3
 SPF   3.3.3.3          FFFFFFFF  2       1561    9.67.102.3
```

```
SBNT   4.0.0.0          FF000000  1       1549    NONE
 SPF   4.4.4.4          FFFFFFFC  2       1561    9.67.106.4
 SPF   4.4.4.4          FFFFFFFF  2       1561    9.67.106.4
SBNT   5.0.0.0          FF000000  1       1549    NONE
 SPF   5.5.5.4          FFFFFFFC  2       1567    9.67.107.5
 SPF   5.5.5.5          FFFFFFFF  2       1567    9.67.107.5
SBNT   6.0.0.0          FF000000  1       1549    NONE
 RIP   6.6.6.4          FFFFFFFC  2       30      9.67.103.6
SBNT   7.0.0.0          FF000000  1       1368    NONE
SPIA*  7.7.7.4          FFFFFFFC  3       1380    9.67.106.4
 DIR*  7.7.7.7          FFFFFFFF  1       1574    VIPA1A
SBNT   8.0.0.0          FF000000  1       1549    NONE
 SPF   8.8.8.8          FFFFFFFC  2       1545    9.67.100.8
 SPF   8.8.8.8          FFFFFFFF  2       1545    9.67.100.8
SBNT   9.0.0.0          FF000000  1       1368    NONE
 DIR*  9.67.100.0       FFFFFF00  1       1576    9.67.100.7
 SPF   9.67.100.7       FFFFFFFF  2       1545    CTC7TO8
 SPF   9.67.100.8       FFFFFFFF  1       1572    9.67.100.8
 SPF   9.67.101.3       FFFFFFFF  2       1561    9.67.106.4
 SPF   9.67.101.4       FFFFFFFF  2       1561    9.67.102.3
 DIR*  9.67.102.0       FFFFFF00  1       1575    9.67.102.7
 SPF   9.67.102.3       FFFFFFFF  1       1566    9.67.102.3
 SPF   9.67.102.7       FFFFFFFF  2       1561    CTC7TO3
 DIR*  9.67.103.0       FFFFFF00  1       1575    9.67.103.7
 RIP   9.67.103.6       FFFFFFFF  1       30      9.67.103.6
 SPF   9.67.105.4       FFFFFFFF  2       1545    9.67.100.8
 SPF   9.67.105.8       FFFFFFFF  2       1561    9.67.106.4
 DIR*  9.67.106.0       FFFFFF00  1       1576    9.67.106.7
 SPF   9.67.106.4       FFFFFFFF  1       1566    9.67.106.4
 SPF   9.67.106.7       FFFFFFFF  2       1561    CTC7TO4
 DIR*  9.67.107.0       FFFFFF00  1       1577    9.67.107.7
 SPF   9.67.107.5       FFFFFFFF  1       1574    9.67.107.5
 SPF   9.67.107.7       FFFFFFFF  2       1566    CTC7TO5
 SPF   9.67.108.2       FFFFFFFF  2       1380    9.67.106.4
 SPF   9.67.108.4       FFFFFFFF  3       1380    9.67.106.4
SBNT   10.0.0.0         FF000000  1       1368    NONE
SPE2   10.1.1.0         FFFFFF00  0       1379    9.67.106.4
SPE2   10.1.1.1         FFFFFFFF  0       1379    9.67.106.4
SBNT   20.0.0.0         FF000000  1       1549    NONE
SPE2   20.1.1.0         FFFFFF00  0       1379    9.67.107.5
SPE2   20.1.1.1         FFFFFFFF  0       1379    9.67.107.5
 RIP   30.0.0.0         FF000000  2       30      9.67.103.6
 RIP   30.1.1.0         FFFFFF00  2       30      9.67.103.6
 RIP % 30.1.1.4         FFFFFFFF  2       30      9.67.103.6
 RIP % 30.1.1.8         FFFFFFFF  2       30      9.67.103.6
SPE2   130.200.0.0      FFFF0000  0       1379    9.67.100.8      (2)
SPE2   130.200.1.1      FFFFFFFF  0       1379    9.67.102.3
SPE2   130.200.1.18     FFFFFFFF  0       1379    9.67.100.8
SPE2   130.201.0.0      FFFF0000  0       1379    9.67.100.8      (2)
SPE2   130.202.0.0      FFFF0000  0       1379    9.67.100.8      (2)
                     0 NETS DELETED, 4 NETS INACTIVE
```

Displaying the routes to a specific destination

To display information about the routes to a specific destination, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE,DEST=130.201.0.0
EZZ7874I ROUTE EXPANSION 798
DESTINATION:    130.201.0.0
MASK:           255.255.0.0
ROUTE TYPE:     SPE2
DISTANCE:       0
```

```
AGE:            1485
NEXT HOP(S):    9.67.100.8        (CTC7TO8)
                9.67.102.3        (CTC7TO3)
```

Displaying all of the generic configuration information

To display all of the IPv4 configuration information that is not related to any routing protocol, enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,LIST,ALL
EZZ8053I IPV4 GENERIC CONFIGURATION
TRACE: 2, DEBUG: 3, SADEBUG LEVEL: 0
IPV4 TRACE DESTINATION: /TMP/AMPROUT3.DBG
STACK AFFINITY: TCPCS3

EZZ8056I IPV4 GEN INT CONFIGURATION
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU DESTADDR
NSQDIO3L         9.67.120.3      255.255.255.0     576  N/A
CTC3TO4          9.67.101.3      255.255.255.0   10000  9.67.101.4
```

Displaying information about configured generic interfaces

To display information about configured generic interfaces (that is, interfaces defined to OMPROUTE with the INTERFACE statement, or not defined to OMPROUTE but learned from the stack), enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,LIST,IFS
EZZ8056I IPV4 GEN INT CONFIGURATION
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU DESTADDR
NSQDIO3L         9.67.120.3      255.255.255.0     576  N/A
CTC3TO4          9.67.101.3      255.255.255.0   10000  9.67.101.4
```

Displaying current run-time information about generic interfaces

To display current run-time information about generic interfaces, enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,IFS
EZZ8060I IPV4 GENERIC INTERFACES
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU  CFG  IGN
NSQDIO3L         9.67.120.3      255.255.255.0     576  YES   NO
CTC3TO1          130.200.1.3     N/A               N/A   NO  YES
VIPA03           3.3.3.103       N/A               N/A   NO  YES
CTC3TO4          9.67.101.3      255.255.255.0   10000  YES   NO
```

Verification of OMPROUTE IPv6 configuration and state

The following sections show sample output from each of the commands that can be used to display OMPROUTE IPv6 information. The syntax of these DISPLAY commands, as well as detailed information about the data displayed, can be found in *z/OS Communications Server: IP System Administrator's Commands*.

Displaying all IPv6 OSPF information

To display a comprehensive list of IPv6 OSPF information, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,ALL
EZZ7970I IPV6 OSPF INFORMATION 322
TRACE6: 0, DEBUG6: 0
STACK AFFINITY            TCPCS67
IPV6 OSPF PROTOCOL:       ENABLED
IPV6 OSPF ROUTER ID:      67.67.67.67
DFLT IPV6 OSPF INST ID:   0
```

```
EXTERNAL COMPARISON:      TYPE 2
AS BOUNDARY CAPABILITY:   ENABLED
IMPORT EXTERNAL ROUTES:   RIP
ORIG. DEFAULT ROUTE:      NO
DEMAND CIRCUITS:          ENABLED

EZZ7973I IPV6 OSPF AREAS
AREA ID         STUB DFLT-COST IMPORT-PREF DEMAND IFCS NETS RTRS ABRS
6.6.6.6          NO        N/A     N/A       OFF     2    1    4    2
0.0.0.0          NO        N/A     N/A       OFF     2    0    4    2

--AREA RANGES--
AREA ID         ADVERTISE  PREFIX
6.6.6.6             NO     2001:DB8:0:101::/64

EZZ7958I IPV6 OSPF INTERFACES
NAME             AREA            TYPE   STATE COST HELLO DEAD NBRS ADJS
VIPA1A6          6.6.6.6         VIPA    N/A     1  N/A   N/A  N/A  N/A
MPCPTP7TO5       0.0.0.0         P-2-MP   16     1   10    40    1    1
NSQDIO1L6        6.6.6.6         BRDCST   32     1   10    40    3    2
VL/0             0.0.0.0         VLINK    16     1   30   180    1    1

EZZ7972I IPV6 OSPF VIRTUAL LINKS
ENDPOINT        TRANSIT AREA   STATE COST HELLO DEAD NBRS ADJS
64.64.64.64     6.6.6.6          16     1   30   180    1    1

EZZ8129I IPV6 OSPF NEIGHBORS
ROUTER ID       STATE LSRXL DBSUM LSREQ HSUP RTR-PRI IFC
65.65.65.65       128     0     0     0  OFF       1 MPCPTP7TO5
64.64.64.64       128     0     0     0  OFF       1 NSQDIO1L6
63.63.63.63       128     0     0     0  OFF       1 NSQDIO1L6
68.68.68.68       128     0     0     0  OFF       1 NSQDIO1L6
64.64.64.64       128     0     0     0  OFF       1 *
```

Displaying IPv6 OSPF area statistics and parameters

To display the statistics and parameters for all IPv6 OSPF areas attached to the router, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,AREASUM
EZZ7973I IPV6 OSPF AREAS 536
AREA ID         STUB DFLT-COST IMPORT-PREF DEMAND IFCS NETS RTRS ABRS
6.6.6.6          NO        N/A     N/A       OFF     2    1    4    2
0.0.0.0          NO        N/A     N/A       OFF     2    0    4    2

--AREA RANGES--
AREA ID         ADVERTISE  PREFIX
6.6.6.6             NO     2001:DB8:0:101::/64
```

Displaying IPv6 OSPF interface statistics and parameters

To display current, run-time statistics and parameters related to IPv6 OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,IFS
EZZ7958I IPV6 OSPF INTERFACES 575
NAME             AREA            TYPE   STATE COST HELLO DEAD NBRS ADJS
VIPA1A6          6.6.6.6         VIPA    N/A     1  N/A   N/A  N/A  N/A
MPCPTP7TO5       0.0.0.0         P-2-MP   16     1   10    40    1    1
NSQDIO1L6        6.6.6.6         BRDCST   32     1   10    40    3    2
VL/0             0.0.0.0         VLINK    16     1   30   180    1    1
```

Displaying statistics and parameters for a specific IPv6 OSPF interface

To display current, run-time statistics and parameters for a specific IPv6 OSPF interface, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,IF,NAME=NSQDIO1L6
EZZ7959I IPV6 OSPF INTERFACE DETAIL 677
INTERFACE NAME:     NSQDIO1L6
INTERFACE ID:       20
INSTANCE ID:        0
INTERFACE ADDRESS:  FE80::7
                    2001:DB8:0:120::7
INTERFACE PREFIX:   STAT 2001:DB8:0:120::/64
ATTACHED AREA:      6.6.6.6
INTERFACE TYPE:     BRDCST
STATE:              32
DESIGNATED ROUTER:  68.68.68.68
BACKUP DR:          64.64.64.64

DR PRIORITY:       1  HELLO INTERVAL:   10  RXMT INTERVAL:     5
DEAD INTERVAL:    40  TX DELAY:          1  POLL INTERVAL:   N/A
DEMAND CIRCUIT:  OFF  HELLO SUPPRESS:  N/A  SUPPRESS REQ:    N/A
MTU:            9000  COST:             1  DB_EX INTERVAL:   40

# NEIGHBORS:       3  # ADJACENCIES:     2  # FULL ADJS.:      2
# MCAST FLOODS:    7  # MCAST ACKS:      9

NETWORK CAPABILITIES:
 BROADCAST
 DEMAND-CIRCUITS
 MULTICAST
```

Displaying IPv6 OSPF virtual link statistics and parameters

To display current, run-time statistics and parameters related to IPv6 OSPF virtual links, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,VLINK
EZZ7972I IPV6 OSPF VIRTUAL LINKS 703
ENDPOINT        TRANSIT AREA   STATE COST HELLO DEAD NBRS ADJS
64.64.64.64     6.6.6.6          16     1   30   180    1    1
```

Displaying statistics and parameters for a specific IPv6 OSPF virtual link

To display current, run-time statistics and parameters for a specific IPv6 OSPF virtual link, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,VLINK,ENDPT=64.64.64.64
EZZ7971I IPV6 VIRTUAL LINK DETAILS 713
VIRTUAL LINK ENDPOINT:     64.64.64.64
PHYSICAL INTERFACE NAME:   NSQDIO1L6
VL TRANSIT AREA:           6.6.6.6
STATE:                     16

HELLO INTERVAL:     30  DEAD INTERVAL:     180  DB_EX INTERVAL:    180
RXMT INTERVAL:      10  TX DELAY:            5  COST:                1
DEMAND CIRCUIT:     ON  HELLO SUPPRESS:    OFF  SUPPRESS REQ:       ON

# NEIGHBORS:         1  # ADJACENCIES:       1  # FULL ADJS.:        1
```

Displaying IPv6 OSPF neighbor statistics and parameters

To display the statistics and parameters related to IPv6 OSPF neighbors, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,NBRS
EZZ8129I IPV6 OSPF NEIGHBORS 715
ROUTER ID       STATE LSRXL DBSUM LSREQ HSUP RTR-PRI IFC
65.65.65.65       128     0     0     0  OFF       1 MPCPTP7TO5
63.63.63.63         8     0     0     0  OFF       1 NSQDIO1L6
```

```
64.64.64.64       128     0     0     0  OFF       1 NSQDIO1L6
68.68.68.68       128     0     0     0  OFF       1 NSQDIO1L6
64.64.64.64       128     0     0     0  OFF       1 *
```

Tip: On multiaccess media (LANs), attached routers become adjacent only with the designated router and the backup designated router. Routers that are not performing a designated router role do not become adjacent to each other on LAN networks. Therefore, in the example above, the state of 8 for 63.63.63.63 does not necessarily represent a problem or an incomplete adjacency. You can conclude that 64.64.64.64 and 68.68.68.68 are the designated routers, and 63.63.63.63 is simply another router on the network. State 8 is the final state in this case.

Displaying statistics and parameters for a specific IPv6 OSPF neighbor

To display current, run-time statistics and parameters for a specific IPv6 OSPF neighbor, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,NBR,ID=64.64.64.64,IFNAME=NSQDIO1L6
EZZ8130I IPV6 OSPF NEIGHBOR DETAILS 737
NEIGHBOR IP ADDRESS:    FE80::4
OSPF ROUTER ID:         64.64.64.64
NEIGHBOR STATE:         128
PHYSICAL INTERFACE:     NSQDIO1L6
DR CHOICE:              68.68.68.68
BACKUP CHOICE:          64.64.64.64
DR PRIORITY:            1
NBR OPTIONS:            (0X00)

DB SUMM QLEN:      0  LS RXMT QLEN:      0  LS REQ QLEN:       0
LAST HELLO:        5  NO HELLO:        OFF
# LS RXMITS:       1  # DIRECT ACKS:     5  # DUP LS RCVD:     4
# OLD LS RCVD:     0  # DUP ACKS RCVD:   3  # ADJ. RESETS:     1
```

Displaying IPv6 OSPF link state database statistics

To display the number of LSAs currently in the link state database, categorized by type, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,DBSIZE
EZZ8128I IPV6 OSPF LS DATABASE SIZE 841
# ROUTER-LSAS:              8
# NETWORK-LSAS:             1
# INTER-AREA PREFIX LSAS:   50
# INTER-AREA ROUTER LSAS:   6
# AS EXTERNAL-LSAS:         6
# LINK LSAS:                6
# INTRA-AREA PREFIX LSAS:   21
# UNKNOWN LSAS:             0
# INTRA-AREA ROUTES:        24
# INTER-AREA ROUTES:        0
# TYPE 1 EXTERNAL ROUTES:   0
# TYPE 2 EXTERNAL ROUTES:   0
```

Displaying IPv6 OSPF link state advertisement

To display the contents of a single link state advertisement contained in the IPv6 OSPF database, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,LSA,LSTYPE=2001,LSID=0,ORIG=64.64.64.64,
AREAID=6.6.6.6
EZZ7880I LSA DETAILS 834
        LS AGE:          61
        LS TYPE:         0X2001 (ROUTER LSA)
        LS ID:           0
        LS ORIGINATOR:   64.64.64.64
```

```
        LS SEQUENCE NO:  0X8000000F
        LS CHECKSUM:     0X3886
        LS LENGTH:       40
        ROUTER TYPE:     (0X01) ABR
        LS OPTIONS:      (0X000033) V6,E,R,DC
INTERFACES:
 TYPE  METRIC  INTERFACE ID   NBR INTERFACE ID   NBR ROUTER ID
    2       1            16                 14     68.68.68.68
```

Displaying IPv6 OSPF external advertisements

To display the AS external advertisements belonging to the IPv6 OSPF routing domain, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,EXTERNAL
EZZ8127I IPV6 OSPF AS EXTERNAL LSDB 555
             AS EXTERNAL LSAS (LS TYPE=4005)
LS ORIGINATOR   LS ID      SEQNO       AGE PREFIX
67.67.67.67     5          0X80000001  565 6:6:6:6:6:6:6:6/128
67.67.67.67     6          0X80000001  561 2001:DB8:0:A1C::6/128
67.67.67.67     7          0X80000001  558 2001:DB8:0:103::6/128
67.67.67.67     8          0X80000001  222 2001:DB8:0:A10::/60
67.67.67.67     9          0X80000001  222 2001:DB8:0:A1B::/64
67.67.67.67     10         0X80000001  222 2001:DB8:0:A1C::/64
   # ADVERTISEMENTS:    6   CHECKSUM TOTAL: 0X000271C6
```

Displaying IPv6 OSPF area link state database

To display the contents of a particular IPv6 OSPF area link state database, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,DATABASE,AREAID=6.6.6.6
EZZ8126I IPV6 OSPF AREA LS DATABASE 829
             ROUTER LSAS (LS TYPE=2001)
LS ORIGINATOR   LS ID      SEQNO       AGE LINKS  RTR-TYPE
63.63.63.63     0          0X80000001  376 1
64.64.64.64     0          0X80000002  321 1      ABR,V
67.67.67.67     0          0X80000004  320 1      ABR,ASBR,V
68.68.68.68     0          0X80000002  595 1
   # ADVERTISEMENTS:    4   CHECKSUM TOTAL: 0X0001D024

             NETWORK LSAS (LS TYPE=2002)
LS ORIGINATOR   LS ID      SEQNO       AGE ROUTERS
68.68.68.68     14         0X80000004  375 4
   # ADVERTISEMENTS:    1   CHECKSUM TOTAL: 0X0000F5CC

             INTER-AREA PREFIX LSAS (LS TYPE=2003)
LS ORIGINATOR   LS ID      SEQNO       AGE PREFIX
64.64.64.64     4          0X80000002  395 2001:DB8:0:108::4/128
64.64.64.64     8          0X80000001  395 2001:DB8:0:108::2/128
64.64.64.64     9          0X80000001  395 2001:DB8:0:10::2/128
64.64.64.64     10         0X80000001  395 2001:DB8:0:10::/64
64.64.64.64     11         0X80000001  395 2:2:2:2:2:2:2:2/128
64.64.64.64     22         0X80000001  375 2001:DB8:0:120::4/128
64.64.64.64     26         0X80000001  321 2001:DB8:0:107::7/128
64.64.64.64     27         0X80000001  321 2001:DB8:0:120::7/128
64.64.64.64     28         0X80000001  321 2001:DB8:0:107::5/128
64.64.64.64     29         0X80000001  321 2001:DB8:0:20::5/128
64.64.64.64     30         0X80000001  321 2001:DB8:0:20::/64
67.67.67.67     15         0X80000002  358 2001:DB8:0:107::7/128
67.67.67.67     16         0X80000001  358 2:2:2:2:2:2:2:2/128
67.67.67.67     19         0X80000001  358 2001:DB8:0:107::5/128
67.67.67.67     20         0X80000001  358 2001:DB8:0:20::5/128
67.67.67.67     21         0X80000001  358 2001:DB8:0:20::/64
67.67.67.67     25         0X80000001  356 2001:DB8:0:120::7/128
67.67.67.67     26         0X80000001  317 2001:DB8:0:108::4/128
67.67.67.67     27         0X80000001  317 2001:DB8:0:108::2/128
67.67.67.67     28         0X80000001  317 2001:DB8:0:10::2/128
```

```
67.67.67.67     29         0X80000001  317 2001:DB8:0:10::/64
67.67.67.67     30         0X80000001  317 2001:DB8:0:120::4/128
   # ADVERTISEMENTS:   22   CHECKSUM TOTAL: 0X000E7320

              LINK LSAS (LS TYPE=0008)
LS ORIGINATOR   LS ID      SEQNO       AGE INTERFACE
63.63.63.63     34         0X80000001  387 NSQDIO1L6
64.64.64.64     16         0X80000001  402 NSQDIO1L6
67.67.67.67     20         0X80000002  640 NSQDIO1L6
68.68.68.68     14         0X80000002  638 NSQDIO1L6
   # ADVERTISEMENTS:    4   CHECKSUM TOTAL: 0X000295E4

              INTRA-AREA PREFIX LSAS (LS TYPE=2009)
LS ORIGINATOR   LS ID      SEQNO       AGE REF-LSTYPE REF-LSID
63.63.63.63     34         0X80000001  387 0X2001     0
63.63.63.63     36         0X80000001  387 0X2001     0
63.63.63.63     38         0X80000001  387 0X2001     0
64.64.64.64     16         0X80000001  402 0X2001     0
64.64.64.64     20         0X80000001  402 0X2001     0
67.67.67.67     20         0X80000002  639 0X2001     0
67.67.67.67     26         0X80000002  639 0X2001     0
68.68.68.68     14         0X80000003  595 0X2001     0
68.68.68.68     16         0X80000001 1738 0X2001     0
68.68.68.68     18         0X80000002  638 0X2001     0
68.68.68.68     65550      0X80000004  375 0X2002     14
   # ADVERTISEMENTS:   11   CHECKSUM TOTAL: 0X00068473
```

Displaying IPv6 OSPF router routes

To display all routes to other routers that have been calculated by IPv6 OSPF and are now present in the routing table, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,ROUTERS
EZZ8125I IPV6 OSPF ROUTERS 820
DEST: 68.68.68.68
  NEXT HOP: FE80::8
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 64.64.64.64
  NEXT HOP: FE80::4
  DTYPE:   BR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 65.65.65.65
  NEXT HOP: FE80::5:7
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 0.0.0.0
DEST: 63.63.63.63
  NEXT HOP: FE80::3
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 62.62.62.62
  NEXT HOP: FE80::4
  DTYPE:  RTR   RTYPE: SPF    COST: 2        AREA: 0.0.0.0
DEST: 64.64.64.64
  NEXT HOP: FE80::4
  DTYPE:   BR   RTYPE: SPF    COST: 1        AREA: 0.0.0.0
```

Displaying IPv6 OSPF routing protocol statistics

To display statistics generated by the IPv6 OSPF routing protocol, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,STATS
EZZ8124I IPV6 OSPF STATISTICS 839
ATTACHED AREAS:                   2  # DIJKSTRA RUNS:                12
OSPF PACKETS RCVD:              619  OSPF PACKETS RCVD W/ERRS:        0
TRANSIT NODES ALLOCATED:         26  TRANSIT NODES FREED:            17
LS ADV. ALLOCATED:              275  LS ADV. FREED:                 175
QUEUE HEADERS ALLOC:             64  QUEUE HEADERS AVAIL:            64
INCREMENTAL SUMM. UPDATES:        5  INCREMENTAL VL UPDATES:          0
```

```
INCREMENTAL EXT. UPDATES:        27  PTRS TO INVALID LS ADV:          0
MULTICAST PKTS SENT:            421  UNICAST PKTS SENT:              40
LS ADV. AGED OUT:                 0  LS ADV. FLUSHED:                41
```

Displaying all of the IPv6 RIP information

To display all of the IPv6 RIP information, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,ALL
EZZ8030I IPV6 RIP CONFIGURATION
TRACE6: 2, DEBUG6: 3
STACK AFFINITY: TCPCS4
IPV6 RIP: ENABLED
IPV6 RIP DEFAULT ORIGINATION: DISABLED

EZZ8027I IPV6 RIP INTERFACES
                                    ---------SEND-----------  --RCV--
NAME               MTU STATE IN OUT PRF HST STA DEF RADV PSN  PRF HST
OSAQDIO46         9000    UP  1   0 YES  NO  NO  NO  YES YES  YES  NO


EZZ8031I IPV6 RIP ROUTE ACCEPTANCE
ACCEPT IPV6 RIP UPDATES ALWAYS FOR:
  2001:0DB8:0:A1C::2
  2001:0DB8:0:A1C::1

EZZ8029I GLOBAL IPV6 RIP FILTERS

SEND ONLY: ALL

IGNORE IPV6 RIP UPDATES FROM:
  FE80::1:2:3:8

FILTERS: NORECEIVE       2001:0DB8:0:A1C::/64
```

Displaying information about IPv6 RIP interfaces

To display information about IPv6 RIP interfaces, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,IFS
EZZ8027I IPV6 RIP INTERFACES
                                    ---------SEND-----------  --RCV--
NAME               MTU STATE IN OUT PRF HST STA DEF RADV PSN  PRF HST
OSAQDIO46         9000    UP  1   0 YES  NO  NO  NO  YES YES  YES  NO
```

Displaying information about a specific IPv6 RIP interface

To display information about a specific IPv6 RIP interface, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,IF,NAME=OSAQDIO46
EZZ8028I IPV6 RIP INTERFACE DETAILS
INTERFACE NAME:     OSAQDIO46
INTERFACE ADDRESS:  FE80::1:2:3:1
INTERFACE PREFIX:   RADV 2001:0DB8:1::/48
MTU:                    9000     STATE:                  UP
IN METRIC:              1        OUT METRIC:             0
SEND PREFIX ROUTES:     YES      SEND HOST ROUTES:       NO
SEND STATIC ROUTES:     NO       SEND DEFAULT ROUTES:    NO
SEND RTR. ADV. ROUTES:  YES      SEND POIS. REV. ROUTES: YES
RECEIVE PREFIX ROUTES:  YES      RECEIVE HOST ROUTES:    NO

SEND ONLY:  ALL

FILTERS: NONE
```

Displaying the routes to be unconditionally accepted by IPv6 RIP

To display the routes to be unconditionally accepted by IPv6 RIP, as configured with the IPv6_Accept_RIP_Route statement, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,ACCEPTED
EZZ8031I IPV6 RIP ROUTE ACCEPTANCE
ACCEPT IPV6 RIP UPDATES ALWAYS FOR:
  2001:0DB8:0:A1C::2
  2001:0DB8:0:A1C::1
```

Displaying the global IPv6 RIP filters

To display the global IPv6 RIP filters, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,FILTERS
EZZ8029I GLOBAL IPV6 RIP FILTERS

SEND ONLY: ALL

IGNORE IPV6 RIP UPDATES FROM:
  FE80::1:2:3:8

FILTERS: NORECEIVE       2001:0DB8:0:A1C::/64
```

Displaying the routes in the OMPROUTE IPv6 routing table

To display all of the routes in the OMPROUTE IPv6 routing table, enter the following command:

```
D TCPIP,TCPCS67,OMP,RT6TABLE
EZZ7979I IPV6 ROUTING TABLE 641
DESTINATION: 2:2:2:2:2:2:2:2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 3:3:3:3:3:3:3:3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 4:4:4:4:4:4:4:4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 2170
DESTINATION: 5:5:5:5:5:5:5:5/128
  NEXT HOP: FE80::5:7
  TYPE: SPF            COST:  1         AGE: 2197
DESTINATION: 6:6:6:6:6:6:6:6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 7:7:7:7:7:7:7:7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 59
DESTINATION: 8:8:8:8:8:8:8:8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:10::/64
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  3         AGE: 32
DESTINATION: 2001:DB8:0:10::2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:30::/60
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:31::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:32::/64
  NEXT HOP: FE80::3 (2)
```

```
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:33::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:33::3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 2001:DB8:0:34::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:38::8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:103::6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:103::7/128
  NEXT HOP: ::
  TYPE: DIR *          COST:  1         AGE: 2209
DESTINATION: 2001:DB8:0:107::5/128
  NEXT HOP: FE80::5:7
  TYPE: SPF            COST:  1         AGE: 2198
DESTINATION: 2001:DB8:0:107::7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 2198
DESTINATION: 2001:DB8:0:108::2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:108::4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 32
DESTINATION: 2001:DB8:0:120::/64
  NEXT HOP: ::
  TYPE: SPF *          COST:  1         AGE: 2172
DESTINATION: 2001:DB8:0:120::3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 2001:DB8:0:120::4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 2170
DESTINATION: 2001:DB8:0:120::7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 2172
DESTINATION: 2001:DB8:0:120::8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:A10::/60
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1B::/64
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1C::/64
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1C::6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
                       0 NETS DELETED, 5 NETS INACTIVE
```

Displaying the routes to a specific IPv6 destination

To display information about the routes to a specific IPv6 destination, enter the following command:

```
D TCPIP,TCPCS4,OMP,RT6TABLE,DEST=2001:0DB8:0:A10::
EZZ7980I IPV6 ROUTE EXPANSION
DESTINATION: 2001:0DB8:0:A10::/60
ROUTE TYPE:  RIP
COST:        2
AGE:         352
NEXT HOP(S): FE80::1:2:3:3                          (OSAQDIO46)
             FE80::1:2:3:4                          (OSAQDIO46)
```

Displaying all of the IPv6 generic information

To display all of the IPv6 generic information, enter the following command:

```
D TCPIP,TCPCS4,GENERIC6,ALL
EZZ8053I IPV6 GENERIC CONFIGURATION
TRACE6: 2, DEBUG6: 3
IPV6 TRACE DESTINATION: /TMP/OMPROUT6.DBG
STACK AFFINITY: TCPCS4

EZZ8060I IPV6 GENERIC INTERFACES
NAME               MTU STATE CONFIGURED
VIPA16           65535    UP     YES
```

Displaying information about IPv6 generic interfaces

To display information about IPv6 generic interfaces (that is, interfaces defined in the OMPROUTE configuration file using IPv6_INTERFACE statements, or IPv6 interfaces not defined to OMPROUTE but learned from the stack), enter the following command:

```
D TCPIP,TCPCS4,OMP,GENERIC6,IFS
EZZ8060I IPV6 GENERIC INTERFACES
NAME               MTU STATE CONFIGURED
VIPA16           65535    UP     YES
```

Displaying information about a specific IPv6 generic interface

To display information about a specific IPv6 generic interface, enter the following command:

```
D TCPIP,TCPCS4,OMP,GENERIC6,IF,NAME=VIPA16
EZZ8065I IPV6 GEN INTERFACE DETAILS
INTERFACE NAME:     VIPA16
INTERFACE ADDRESS:  2001:0DB8:6:6:6:6:6:6
INTERFACE PREFIX:   STAT 2001:0DB8:6::/48
MTU:               65535
STATE:              UP
CONFIGURED:         YES
```

Sample OMPROUTE configuration files

The following is an example of an OSPF and IPv6 RIP environment (from TCPCS4 in the Figure 36 on page 220).

```
;*********************************
; OSPF Configuration Statements   *
;*********************************
OSPF
     RouterID=4.4.4.4;
Area
     Area_Number = 0.0.0.0;
Area
     Area_Number = 1.1.1.1;
OSPF_Interface
     IP_Address=9.67.108.4
     Name = CTC4TO2
     Subnet_Mask=255.255.255.0
```

```
     Attaches_To_Area=0.0.0.0
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.106.4
     Name = CTC4TO7
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.105.4
     Name = CTC4TO8
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.101.4
     Name = CTC4TO3
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=4.4.4.4
     Name = VIPA1A
     Subnet_Mask=255.255.255.252
     Attaches_To_Area=1.1.1.1
     Cost0 = 1;
Virtual_Link
     Virtual_Endpoint_RouterID=7.7.7.7
     Links_Transit_Area=1.1.1.1;
;************************************
; IPv6 RIP Configuration Statements *
;************************************
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1C::1;
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1C::2;
IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1C::/64);
IPv6_RIP_Interface
     Name = OSAQDIO46;
IPv6_Default_Route
     Name=OSAQDIO46
     Next_Hop=FE80::1:2:3:4;
```

The following is an example of mixed OSPF, IPv4 RIP, and IPv6 RIP environments (from TCPCS7 in Figure 36 on page 220).

```
;*********************************
;  OSPF Configuration Statements  *
;*********************************
OSPF
     RouterID=7.7.7.7;
Area
     Area_Number = 0.0.0.0;
Area
     Area_Number = 1.1.1.1;
AS_Boundary_Routing
     Import_Subnet_Routes=YES
     Import_RIP_Routes=YES;
OSPF_Interface
     IP_Address=9.67.107.7
     Name = CTC7TO5
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=0.0.0.0
```

```
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.106.7
     Name = CTC7TO4
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.102.7
     Name = CTC7TO3
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.100.7
     Name = CTC7TO8
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.104.7
     Name = NBMA7
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     Non_Broadcast=YES
     NB_Poll_Interval=180
     MTU = 1024
     Cost0 = 1
     DR_Neighbor=9.67.104.15
     No_DR_Neighbor=9.67.104.16
     No_DR_Neighbor=9.67.104.25;
OSPF_Interface
     IP_Address=7.7.7.7
     Name = VIPA1A
     Subnet_Mask=255.255.255.252
     Attaches_To_Area=1.1.1.1
     Cost0 = 1;
Range
     IP_Address=9.67.101.0
     Subnet_Mask=255.255.255.0
     Area_Number=1.1.1.1
     Advertise=NO;
Virtual_Link
     Virtual_Endpoint_RouterID=4.4.4.4
     Links_Transit_Area=1.1.1.1;
;*********************************
;  RIP Configuration Statements  *
;*********************************
Originate_RIP_Default
     Condition=Always;
Accept_RIP_Route
     IP_Address=30.1.1.4;
Accept_RIP_Route
     IP_Address=30.1.1.8;
Filter=(nosend,10.1.1.0,255.255.255.0);
RIP_Interface
     IP_Address=9.67.103.7
     Name = CTC7TO6
     Subnet_Mask=255.255.255.0
     Receive_Dynamic_Hosts=NO
     MTU = 1024
     RipV2=YES;
;***********************************
```

```
; IPv6 RIP Configuration Statements *
;************************************
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1B::1;
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1B::2;
IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1B::/64);
IPv6_RIP_Interface
     Name = OSAQDIO76;
```

The following is an example of a pure RIP environment (from TCPCS6 in Figure 36 on page 220).

```
;*********************************
;  RIP Configuration Statements  *
;*********************************
RIP_Interface
     IP_Address=9.67.103.6
     Name = CTC6TO7
     Subnet_Mask=255.255.255.0
     MTU = 1024
     Send_Static_Routes=YES
     Send_Host_Routes=YES
     RipV2=YES;
Interface
     IP_Address=6.6.6.6
     Name = VIPA1A
     Subnet_Mask=255.255.255.252;
```

The following is an example of a pure IPv6 OSPF environment (from TCPCS64 in Figure 37 on page 221).

```
;************************************
; IPv6 OSPF Configuration Statements *
;************************************
IPv6_OSPF
     RouterID = 64.64.64.64;
IPv6_Area
     Area_Number = 0.0.0.0;
IPv6_Area
     Area_Number = 6.6.6.6;
IPv6_OSPF_Interface
     Name = NSQDIO4L6
     Prefix = 2001:0DB8:0:120::/64
     Attaches_to_Area = 6.6.6.6;
IPv6_OSPF_Interface
     Name = VIPA1A6
     Attaches_to_Area = 6.6.6.6
     Cost = 1;
IPv6_OSPF_Interface
     Name = MPCPTP4TO2
     Attaches_to_Area = 0.0.0.0
     Cost = 1;
IPv6_Virtual_Link
     Virtual_Endpoint_RouterID = 67.67.67.67
     Links_Transit_Area = 6.6.6.6;
```

Verification of routing (Static and dynamic)

- If static routes are used, an indirect route must not be defined before the route to its first hop is defined. The following example shows an incorrect configuration.

  ```
  BEGINRoutes ;first BEGINRoutes in the profile
  ;Network/mask              FirstHop       LinkName  PacketSize
  Route 9.67.104.0/24        9.67.105.8     CTC4TO8   MTU 1500
  ```

```
Route 9.67.105.0/24        =              CTC4TO8   MTU 1500
Route 2001:0DB8:0:A1B::/64 FE80::1:2:3:3  OSAQDIO46 MTU 5000
Route FE80::1:2:3:3/128    =              OSAQDIO46 MTU 5000
ENDRoutes
```

When configured incorrectly, the following error messages are displayed:

```
EZZ0657I ROUTE LIST ENTRY ON LINE 28 FOR DESTINATION 9.67.104.0 IS
UNREACHABLE THROUGH INTERFACE 9.67.105.8 ON CTC4TO8
EZZ0657I ROUTE LIST ENTRY ON LINE 30 FOR DESTINATION 2001:0DB8:0:A1B:: IS
UNREACHABLE THROUGH INTERFACE FE80::1:2:3:3 ON OSAQDIO46
```

- If OMPROUTE is used for the OSPF protocol only and AUTOLOG is not configured correctly (see step 2 on page 240), OMPROUTE will be periodically restarted and the following messages are displayed:

```
$HASP100 OMPROUTE ON STCINRDR
$HASP373 OMPROUTE STARTED
IEF403I OMPROUT1 - STARTED
      OMPROUT1    OMPROUTE    BPXBATCH    0000
EZZ7800I OMPROUTE STARTING
EZZ7872I OMPROUTE FOUND ANOTHER ROUTING APPLICATION ALREADY ACTIVE
EZZ8074I OMPROUTE PROCESSING ERROR
EZZ7805I OMPROUTE EXITING ABNORMALLY - RC(11)
OMPROUT1    *OMVSEX     BPXPRECP    0011
IEF404I OMPROUT1 - ENDED
$HASP395 OMPROUT1 ENDED
```

- If a configuration statement in the OMPROUTE configuration file has a missing semicolon, the syntax checker might issue the following message:

```
EZZ7830I SYNTAX ERROR AT LINE 22 OF OMPROUTE CONFIGURATION FILE
PROCESSING END OF FILE
```

Verifying connections with NETSTAT, PING, and TRACERTE

The interfaces were verified with the instructions in Chapter 2, "Configuration overview," on page 11. The first thing to verify is that the devices and interfaces are started. In the case of point-to-point links like the CTCs in TCPCS4, the following message is written to the z/OS console when the device starts:

```
EZZ4313I INITIALIZATION COMPLETE FOR DEVICE CTCE02
```

In the case of IPv6 interfaces like OSAQDIO46 in TCPCS4, the following message is written to the z/OS console when the interface starts:

```
EZZ4340I INITIALIZATION COMPLETE FOR INTERFACE OSAQDIO46
```

The same information can be determined from NETSTAT DEV. Following is a portion of the output of NETSTAT DEV with the CTCE02 device shown as ready. The NETSTAT DEV can be issued on TCPCS4 and TCPCS7 to verify that the devices on both systems are ready.

```
DevName: CTCE02            DevType: CTC       DevNum: 0E00
  DevStatus: Ready
  LnkName: CTC4TO7           LnkType: CTC         LnkStatus: Ready
    NetNum: 0    QueSize: n/a
    ActMtu: 32760
  BSD Routing Parameters:
    MTU Size: 01500             Metric: 01
    DestAddr: 0.0.0.0           SubnetMask: 255.255.255.0
  Multicast Specific:
    Multicast Capability: Yes
  GROUP             REFCNT
  -----             ------
  224.0.0.5         0000000001
  224.0.0.1         0000000001
  Link Statistics:
```

```
    BytesIn                          = 488
    Inbound Packets                  = 0
    Inbound Packets In Error         = 0
    Inbound Packets Discarded        = 0
    Inbound Packets With No Protocol  = 0
    BytesOut                         = 1092
    Outbound Packets                 = 0
    Outbound Packets In Error        = 0
    Outbound Packets Discarded       = 0
```

Following is a portion of the output of NETSTAT DEV with an IPv6 interface (OSAQDIO46) shown as ready.

```
DevName: OSAQDIO2          DevType: MPCIPA
  DevStatus: Ready
  IntfName: OSAQDIO46        IntfType: IPAQENET6 IntfStatus: Ready
    NetNum: n/a  QueSize: 0   Speed: 0000001000
   MacAddress: 0002559A3F65
    DupAddrDet: 1
    CfgRouter: Non                   ActRouter: Non
    CfgMtu: None                     ActMtu: 8992
  Multicast Specific:
    Multicast Capability: Yes
    RefCnt      Group
    ------      -----
    0000000001  ff02::1:ff03:1
    0000000002  ff02::1
  Interface Statistics:
    BytesIn                          = 592
    Inbound Packets                  = 0
    Inbound Packets In Error         = 0
    Inbound Packets Discarded        = 0
    Inbound Packets With No Protocol  = 0
    BytesOut                         = 1008
    Outbound Packets                 = 0
    Outbound Packets In Error        = 0
    Outbound Packets Discarded       = 0
```

If the devices do not have a LnkStatus or IntfStatus of Ready, this must be resolved before continuing. There are several things that might cause the LnkStatus or IntfStatus to not be ready. For example, the device might not be defined to z/OS correctly, the device might not be defined in PROFILE.TCPIP correctly, and so on.

You can PING each others hosts within the network to verify indirect routes exist.

```
ping 9.67.107.7
 CS V1R6: Pinging host 9.67.107.7
 Ping #1 response took 0.048 seconds.
 READY
ping 2001:0db8:0:a1b:2:559a:3f65:3
 CS V1R6: Pinging host 2001:0db8:0:a1b:2:559a:3f65:3
 Ping #1 response took 0.051 seconds.
 READY
```

Use TRACERTE to verify that the correct route is being taken for each indirectly attached host:

```
tracerte 9.67.107.5
 CS V1R6: Traceroute to 9.67.107.5 (9.67.107.5)
 1 9.67.106.7 (9.67.106.7)  40 ms  7 ms  6 ms
 2 9.67.107.5 (9.67.107.5)  9 ms  8 ms  9 ms
 READY
```

Following is an IPv6 example for indirectly attached hosts:

```
tracerte 2001:0db8:0:a1c:2:36a4:b39a:7
 CS V1R6: Traceroute to 2001:0db8:0:a1c:2:36a4:b39a:7
 at IPv6 address: 2001:0db8:0:a1c:2:36a4:b39a:7
 1 fe80::1:2:3:4
 (fe80::1:2:3:4) 13 ms 25 ms 40 ms
 2 2001:0db8:0:a1c:2:36a4:b39a:7
 (2001:0db8:0:a1c:2:36a4:b39a:7) 29 ms 263 ms 196 ms
```

Chapter 7. Virtual IP Addressing

This chapter contains information about the following topics:

- Terminology
- Introduction to VIPA
- Moving VIPA (Upon outage of TCP/IP)
- Static VIPAs, dynamic VIPAs (DVIPAs), and distributed dynamic VIPAs
- Using static VIPAs
- Using dynamic VIPAs (DVIPAs)
- Choosing which form of dynamic VIPA to use
- Configuring distributed DVIPAs — Sysplex Distributor
- Resolution of DVIPA conflicts
- IPv6 considerations
- Other considerations
- Example of configuring dynamic and distributed VIPAs
- Verifying the DVIPAs in a sysplex
- Verifying Sysplex Distributor workload
- DVIPAs and routing protocols

Note: This chapter applies to both IPv4 and IPv6, unless otherwise noted.

Terminology

Virtual IP Address (VIPA)
A VIPA is a generic term that refers to an internet address on a z/OS host that is not associated with a physical adapter. There are two types of VIPAs:

- A *Static VIPA* cannot be changed except through a VARY TCPIP,,OBEYFILE operator command.
- A *dynamic VIPA (DVIPA)* can move to other TCP/IP stack members in a sysplex or it can be activated by an application program or by a supplied utility. Dynamic VIPAs are used to implement Sysplex Distributor as described in "Considerations for VIPA" on page 83.

Distributed DVIPA
A distributed DVIPA, which is a special type of DVIPA, can distribute connections within a sysplex.

Dynamic routing
VIPAs are designed to interoperate with a dynamic routing daemon. Therefore, it is highly recommended that a routing daemon be used on a z/OS host that uses VIPAs.

Introduction to VIPA

Traditionally, an IP address is associated with each end of a physical link (or each point of access to a shared-medium LAN), and the IP addresses are unique across the entire visible network, which can be the Internet or a closed intranet. The majority of IP hosts have a single point of attachment to the network, but some

hosts (particularly large server hosts) have more than one link into the network. A TCP/IP host with multiple points of attachment also has multiple IP addresses, one for each link.

Within the IP routing network, failure of any intermediate link or adapter disrupts end user service only if there is not an alternate path through the routing network. Routers can route IP traffic around failures of intermediate links in such a way that the failures are not visible to the end applications or IP hosts. However, because an IP packet is routed based on ultimate destination IP address, if the adapter or link associated with the destination IP address fails, there is no way for the IP routing network to provide an alternate path to the stack and application. Endpoint (source or destination) IP adapters and links thus constitute single points of failure. While this might be acceptable for a client host, where only a single user will be cut off from service, a server IP link might serve hundreds or thousands of clients, all of whose services would be disrupted by a failure of the server link.

The virtual IP address (VIPA) removes the adapter as a single point of failure by providing an IP address that is associated with a stack without associating it with a specific physical network attachment. Because the virtual device exists only in software, it is always active and never experiences a physical failure. A VIPA has no single physical network attachment associated with it. Also, the TCP/IP stack does not maintain interface counters for VIPA interfaces (VIRTUAL links).

To the routing network, a VIPA appears to be a host address indirectly attached to the z/OS. When a packet with a VIPA destination reaches the stack, the IP layer recognizes the address and passes it to the protocol layer in the stack.

The failure of the physical interface can be extended to the failure of the TCP/IP address space, the entire z/OS, or for planned outages. A VIPA just needs to move to a backup stack, and the routes to the VIPA need to be updated. Then clients can transparently connect to the backup stack. This process is known as VIPA takeover.

VIPA takeover improved with the introduction of dynamic virtual IP address (DVIPA) and distributed dynamic virtual IP address (distributed DVIPA). The DVIPA function improves VIPA takeover by allowing a system programmer to plan for system outages and provide for backup systems to take over without operator intervention or external automation. The distributed DVIPA function allows the connections for a single DVIPA to be serviced by applications on several stacks listed in the configuration statement (the distribution list). This adds the benefit of limiting the scope of an application or stack failure, while also providing enhanced work load balancing.

In general, z/OS configured with VIPA provides the following advantages:

- Automatic and transparent recovery from device and adapter failure.

  When a device (for example, a channel-attached router) or adapter (for example, an OSA-Express adapter) fails, if there is another device or link that provides the alternate paths to the destination:

  – IP will detect the failure, find an alternate path for each network, and route outbound traffic to hosts and routers on those networks via alternate paths.
  – Inbound and outbound traffic will not need to reestablish the active TCP connections that were using the failed device.
  – For connection requests originating at a z/OS TCP/IP stack, tolerance of device and adapter failures can be achieved by using source VIPA addressing.

- Recovery from z/OS TCP/IP stack failure (where an alternate z/OS TCP/IP stack has the necessary redundancy).

  Assuming that an alternate stack is installed to serve as a backup, the use of VIPAs enables the backup stack to activate the VIPA address of the failed stack.

  Connections on the failed primary stack will be disrupted but they can be reestablished on the backup using the same IP address as the destination. In addition, the temporarily reassigned VIPA address can be restored to the primary stack after the cause of failure has been removed.
- Limited scope of a stack or application failure.

  If a DVIPA is distributed among several stacks, the failure of only one stack affects only the subset of clients connected to that stack. If the distributing stack experiences the failure, a backup assumes control of the distribution and maintains all existing connections.
- Enhanced workload management through distribution of connection requests.

  With a single DVIPA being serviced by multiple stacks, connection requests and associated workloads can be spread across multiple z/OS images according to Workload Manager (WLM) and Service Level Agreement policies (for example, QOS).
- Allows the non-disruptive movement of an application server to another stack so that workload can be drained from a system in preparation for a planned outage.

Moving a VIPA (for TCP/IP outage)

While a VIPA provides non-disruptive rerouting of IP data during failure of a physical interface, termination of the stack or the associated z/OS (including planned outages) will disrupt connections or UDP sessions to applications on the terminated stack. While failure of the TCP connection or UDP session will be visible to the clients, the duration of the outage is determined by how long the client application is unable to reconnect to an appropriate server application. Because it is common in large enterprises to have multiple instances of an application residing on different z/OS images, if the VIPA address can be moved to another stack that supports the application, the clients can reconnect and the perceived outage will be over.

An IP address associated with a particular physical device is unavailable until the owning stack is restarted; however, a VIPA is not associated with any particular physical interface. If termination of a stack is detected and a suitable application already is active on another stack, the VIPA can be moved. Connections on the terminated stack will be disrupted, but they can be reestablished on the backup stack using the original VIPA.

Movement of a static VIPA to a backup stack can be accomplished by using VARY TCPIP,,OBEYFILE commands on the backup. The data set specified on the command must contain an appropriate set of DEVICE, LINK, HOME, and optionally, BSDROUTINGPARMS statements for IPv4 static VIPAs or INTERFACE statements for IPv6 static VIPAs. If OMPROUTE is used as the routing daemon, an appropriate interface statement is needed in the OMPROUTE configuration file. If the TCP/IP configuration file with the statements defining the VIPA is created in advance, the transfer can be accomplished via automation. This procedure is documented in *z/OS Communications Server: IP Configuration Reference*. Movement of a DVIPA, on the other hand, can be accomplished by configuring a stack to backup a specific DVIPA that is defined on another stack. In this case, failure of the defining stack causes the DVIPA to move without operator intervention or extra

automation. See "Planning for dynamic VIPA takeover" on page 306 for more information. Regardless of the type of VIPA to be moved, it is up to the system programmer or operator to ensure that the VIPA is moved to a backup stack that has the appropriate server applications.

In the absence of a failure, a VIPA is just like any other IP address, and routing for a VIPA is the same as for an IP address associated with a physical link.

Static VIPAs, dynamic VIPAs (DVIPAs), distributed DVIPAs

z/OS TCP/IP stack supports two types of VIPAs: static and dynamic. Dynamic VIPAs (DVIPAs) can be used to distribute connections in a sysplex. This is referred to as a distributed DVIPA.

All three VIPAs can coexist on a given stack, but there are differences in how these VIPAs are configured and used.

Static VIPAs have the following characteristics:

- They can be activated during TCP/IP initialization or VARY TCPIP,,OBEYFILE command processing, and are configured using an appropriate set of DEVICE, LINK, HOME, and optionally, OMPROUTE configuration statements or BSDROUTINGPARMS statements for IPv4 Static VIPAs or INTERFACE statements for IPv6 Static VIPAs.
- Using the SOURCEVIPA configuration option, static VIPAs can be used as the source IP address for outbound datagrams for TCP, RAW, UDP (except routing protocols), and ICMP requests. For IPv6 static VIPAs to be used as source addresses, the SOURCEVIPA configuration option must be enabled and the VIPA interface must appear on the SOURCEVIPAINT keyword on some other INTERFACE statement. This provides tolerance of device and adapter failures for connection requests originating at a z/OS TCP/IP stack.
- They can be specified as the source IP address for outbound TCP connection requests for all applications using this stack with TCPSTACKSOURCEVIPA.
- They can be specified as the source IP address for outbound TCP connection requests for a specific job through the use of the SRCIP profile statement block.
- The number of static VIPAs on a stack is limited only by the range of host IP addresses that are available for that host.
- They can be moved to a backup stack after the original owning stack has failed, by using VARY TCPIP,,OBEYFILE command processing to configure the VIPA on the backup stack and updating the routers.

Dynamic VIPAs have the following characteristics:

- They can be configured to be moved dynamically from a failing stack to a backup stack within the same sysplex without operator intervention or external automation.
- They can be moved manually by deactivating or reactivating them with the VARY TCPIP,,SYSPLEX operator command.
- They can be dynamically activated by an application program.
- They can distribute connections within a sysplex.
- They can be specified on a TCPSTACKSOURCEVIPA statement. This allows a user to specify one VIPA to be used as the source IP address for outbound datagrams for TCP-only requests.
- They can be specified as the source IP address for outbound TCP connection requests for a specific job through the use of the SRCIP profile statement block.

- Unlike static VIPAs, dynamic VIPAs:
  - Are limited to 1024 per stack.
  - Cannot be specified as the VIPA used by Enterprise Extender for connectivity purposes. (See "Configuring static VIPAs for Enterprise Extender" on page 303 for details.)

Distributed DVIPAs have the following characteristics:

- Have all the characteristics of DVIPAs, but cannot be dynamically activated by an application program.
- One stack defines a DVIPA and advertises its existence to the network. Stacks in the target distribution list activate the DVIPA and accept connection requests.
- Connection workload can be spread across several stacks.

See "Configuring distributed DVIPAs — Sysplex Distributor" on page 316 for more detailed descriptions.

Recommendation: OSA-Express devices have a limit on the number of IP addresses (both IPv4 and IPv6 addresses) that can be registered to the device. The limit is dependent on the microcode level of the OSA-Express device. This limit applies across all TCP/IP stacks that share the OSA-Express device. When defining a large number of VIPAs, take care not to exceed this limit. If the limit is exceeded, IP addresses beyond the limit will not be registered with the OSA-Express devices, and incoming packets with those IP addresses will not be routed to the correct stack unless that stack is designated as the primary router.

Using static VIPAs

The following sections describe how to configure static VIPAs, the special case of static VIPAs and Enterprise Extender, and how to implement static VIPA takeover.

Because a VIPA is associated with a z/OS TCP/IP stack and it is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex or *even* to any z/OS TCP/IP stack if the address fits into the network configuration.

Steps for configuring static VIPAs for a z/OS TCP/IP stack

Perform the following steps to configure a static VIPA address in one stack:

1. When configuring static VIPAs for the IPv4 network, add DEVICE, LINK, HOME, and optionally BSDROUTINGPARMS statements for each static VIPA to be defined. When configuring static VIPAs for the IPv6 network, add INTERFACE statements of type VIRTUAL6 for each static VIPA to be defined.

   Notes:

   a. A VIPA link or VIPA interface cannot be coded on a static route in the GATEWAY or BEGINROUTES statements.

   b. If you want to add a static VIPA in an IPv4 network with an address that already exists in the HOME list, you must first delete the existing address using the VARY TCPIP,,OBEYFILE command with the existing address omitted from the HOME list. Then use the VARY TCPIP,,OBEYFILE command with a new and complete HOME list.

2. For IPv4 networks, if tolerance of device and adapter failures is desired for connection requests originating at a z/OS TCP/IP stack, specify the SOURCEVIPA option on the IPCONFIG statement.

   Tips:

   - For the SOURCEVIPA option to work properly, the receiving nodes in the network must be configured to recognize the SOURCEVIPA addresses using the static or dynamic routing protocols. Otherwise, timeouts for the connection or request responses will occur as a result of the VIPA addresses being network unreachable.
   - If TCPSTACKSOURCEVIPA is specified on the IPCONFIG statement, it overrides SOURCEVIPA for outbound IPv4 TCP connections. If the SRCIP profile statement block is defined to establish one or more job-specific source IPv4 addresses, these IP addresses override TCPSTACKSOURCEVIPA or the VIPAs in the HOME list for IPCONFIG SOURCEVIPA, or both, for the specified job names.

   For more information on configuring IPv4 SOURCEVIPA or TCPSTACKSOURCEVIPA addresses on the IPCONFIG statement, or SRCIP addresses in the SRCIP statement block, refer to *z/OS Communications Server: IP Configuration Reference*.

---

3. For IPv6 networks, if tolerance of device and adapter failures is desired for connection requests originating at a z/OS TCP/IP stack, specify the following:

   - The SOURCEVIPA option on the IPCONFIG6 statement.
   - The SOURCEVIPAINT keyword with a VIPA interface name on the INTERFACE statements of the real (physical) interfaces, or on the DYNAMICXCF specification on the IPCONFIG6 statement.

   Tips:

   - For the SOURCEVIPA option to work properly, the receiving nodes in the network must be configured to recognize the SOURCEVIPAINT addresses using the static or dynamic routing protocols. Otherwise, timeouts for the connection or request responses will occur as a result of the VIPA addresses being network unreachable.
   - If TCPSTACKSOURCEVIPA is specified on the IPCONFIG6 statement, it overrides SOURCEVIPA for outbound IPv6 TCP connections. If the SRCIP profile statement block is defined to establish one or more job-specific source IPv6 addresses, these IP addresses override TCPSTACKSOURCEVIPA or SOURCEVIPAINT, or both, for the specified job names.

   For more information on configuring IPv6 SOURCEVIPA or TCPSTACKSOURCEVIPA addresses on the IPCONFIG6 statement, or SRCIP addresses in the SRCIP statement block, refer to *z/OS Communications Server: IP Configuration Reference*.

---

4. For host name resolution of a VIPA address, configure the domain name servers to associate the host name with the VIPA.

---

5. Configure the routing daemon to advertise the presence of the VIPA.

   As described in prior steps, remember that the VIPA to be advertised can be determined by the SOURCEVIPA or TCPSTACKSOURCEVIPA parameters on

the IPCONFIG and IPCONFIG6 statements, or by the SRCIP statement. For more information, refer to *z/OS Communications Server: IP Configuration Reference*.

Figure 38 illustrates a simple configuration showing multiple network attachments using a single static VIPA address. Since any other network interface can be used with static VIPA's, refer to "Setting up physical characteristics in PROFILE.TCPIP" on page 190 for descriptions of other network interfaces. The simple configuration will be used as the TCPCS6 system throughout this chapter.

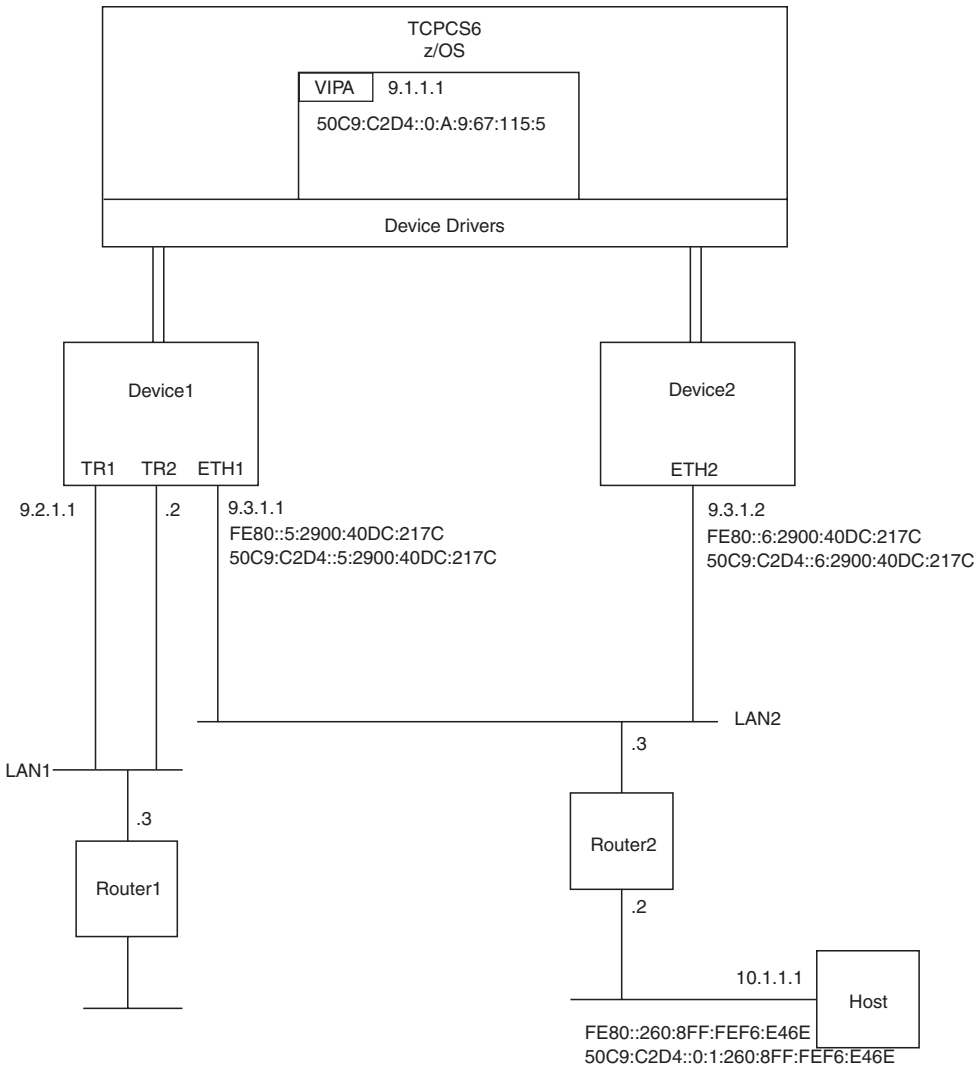


Figure 38. Static VIPA configuration

Configuring static VIPAs for Enterprise Extender

Defining at least one static VIPA is required by VTAM to access the IP network. Since VTAM does not move within a sysplex, a dynamic VIPA cannot be used. Enterprise Extender supports the use of multiple static VIPA addresses, and a VIPA address is chosen as follows:

1. VTAM uses the IPADDR value specified on the Enterprise Extender XCA major node GROUP definition statement, or the VIPA address obtained when

name-to-address resolution is performed on the HOSTNAME value specified on the Enterprise Extender XCA major node GROUP definition statement.

2. If the GROUP definition statement specifies neither IPADDR nor HOSTNAME, VTAM uses the static VIPA address specified on the VTAM IPADDR start option, or the static VIPA address returned by the resolver when name-to-address resolution is performed on the value of the VTAM HOSTNAME start option.
3. If neither the IPADDR nor the HOSTNAME start option is specified, VTAM uses the first IPv4 static VIPA in the HOME list.

If remote APPN nodes use a host name and not a host address to define the destination of an Enterprise Extender connection, the domain name server must return the VIPA address used by VTAM for the host name. Alternatively, if the Enterprise Extender endpoints reside across a firewall or for another reason require network address translation, the domain name server should return a Network Address Translation (NAT) address. This NAT address should in turn map, on the destination side, to the static VIPA address of the intended target VTAM host.

Rule: If you plan on using IPv4 protocols for your Enterprise Extender communication, you must define DEVICE, LINK, and optionally HOME statements for IUTSAMEH, or specify IPCONFIG DYNAMICXCF in the appropriate TCPIP profile data set. If you plan on using IPv6 protocols for your Enterprise Extender communication, you must define an INTERFACE statement for IUTSAMEH or specify IPCONFIG6 DYNAMICXCF in the appropriate TCPIP profile data set.

For more information about Enterprise Extender, refer to the following:

z/OS Communications Server: SNA Network Implementation Guide

`http://www.ibm.com/software/network/library/whitepapers/eextender.html`

`http://www.ibm.com/software/network/library/whitepapers/eemsthtm/eemst.htm`

IBM Redbook, *SNA and TCP/IP Integration* (SG24–5291–00)
Note: This document is also available at http://www.redbooks.ibm.com.

Considerations when using static VIPAs with IPv6

When static VIPAs are configured for use with IPv6, it is recommended that the prefixes of the IPv6 VIPA addresses be different than the prefixes used for addresses assigned to real interfaces. This reduces the likelihood of address collisions between the manually configured VIPA addresses and the autoconfigured addresses of the real interfaces.

To allow other hosts that share links with the z/OS TCP/IP stack to access the IPv6 VIPA addresses, without the need for manual route configuration, a router on each of the links should include the VIPA prefix in its router advertisements. The router advertisements should define the prefix as being on-link and should indicate that the prefix should not be used for autoconfiguration.

Planning for static VIPA takeover and takeback

Because a VIPA is associated with a z/OS TCP/IP stack and is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex or even to any z/OS TCP/IP stack as long as the address fits into the network configuration. Moving a static VIPA can be done manually by an operator or by customer-programmed automation. Movement of the static VIPA allows other hosts that were connected to the primary stack to reestablish connections

with a backup TCP/IP stack using the same VIPA. After the primary TCP/IP stack has been restored, the reassigned VIPA address can be moved back.

Consider the following when backing up and restoring a z/OS TCP/IP stack:

- All connections on the failing host will be disrupted.
- The client can use any ephemeral port number when reestablishing the connection to the backup server.
- Having a different port number for the backup and primary server is not recommended. For example, if the primary FTP used port 21 and the backup FTP used port 1021, when backing up and restoring a z/OS TCP/IP stack, the client would have to know whether to use port 21 or 1021.

Using dynamic VIPAs (DVIPAs)

DVIPA support allows:

- Dynamic movement of a VIPA from a failing TCP/IP stack to a backup stack
- Dynamic allocation of a VIPA by an application program
- Manual movement of a VIPA by deactivating or reactivating it with the VARY TCPIP,,SYSPLEX command

Dynamic VIPAs (DVIPAs) are IP addresses like all other IP addresses associated with a TCP/IP, and they appear as though they had been defined at the end of the HOME list.

Dynamic VIPAs can be either IPv4 or IPv6, and both can be configured within the same VIPADYNAMIC block. A single statement (for example, VIPADEFINE or VIPABACKUP) must contain either IPv4 addresses or IPv6 addresses, but not both. However, statements containing IPv4 addresses can be intermixed with statements containing IPv6 addresses within the same VIPADYNAMIC block in any manner desired.

Configuring dynamic VIPA (DVIPA) support

Unlike static VIPAs, DVIPAs are not configured using DEVICE, LINK, and HOME statements (for IPv4) or using INTERFACE statements (for IPv6). The configuration statements for the DVIPA support are contained within the VIPADYNAMIC block and consist of the following:

- VIPADEFINE and VIPABACKUP statements used to configure DVIPAs to be dynamically moved from a failing TCP/IP to a backup TCP/IP
- VIPARANGE used to specify a range of IP addresses which may be dynamically activated as a VIPA by an application program
- VIPADELETE used to delete existing DVIPAs
- VIPADISTRIBUTE used to configure a DVIPA as a distributed DVIPA and designate the target stacks

The following sections discuss how these statements are used to provide the DVIPA support. For syntax details, see *z/OS Communications Server: IP Configuration Reference*. For further reference, refer to *Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 5: Availability, Scalability, and Performance*, SG24–6517.

When dynamically created DVIPAs are deleted, any applications bound to those DVIPAs (VIPARANGE or MODDVIPA) will receive an asynchronous error, EUNATCH (3448) - the protocol required to support the address family is unavailable.

When dynamic VIPAs (DVIPAs) are used for VIPA takeover together with DNS/WLM in a sysplex, code all of the DVIPAs in the sysplex under each host name in the DNS forward domain data file for the cluster zone. This will circumvent manual intervention in the DNS data files when a DVIPA is taken over or given back and will not cause any undesirable effects in DNS/WLM function.

Planning for dynamic VIPA takeover

Movement by network management automation or operator intervention is not always desirable. Operator intervention takes time and is subject to errors. Automation requires proper detection of the failure and is also prone to error if the failure does not produce the exact console messages anticipated.

The dynamic VIPA takeover function addresses this problem. It is important to understand that dynamic VIPA takeover does not introduce functions that could not be accomplished by operator action or automation. It just removes the dependency on human detection of the error or customer programming for automation. Dynamic VIPA takeover is completely accomplished by the TCP/IP stacks.

DVIPA takeover is possible when a DVIPA is configured as active (using VIPADEFINE) on one stack and as backup (using VIPABACKUP) on another stack within the sysplex. When the stack on which the DVIPA is active terminates or leaves the sysplex group, the backup stack automatically activates the DVIPA and notifies the routing daemon. For information on what causes a stack to leave the sysplex group, see "Sysplex problem detection and recovery" on page 376.

For DVIPA takeover to be useful, the applications that service the DVIPA addresses must be available on the backup stacks. In the absence of the application, the DVIPA will be active, but client connections to the application will still fail. If OMPROUTE is used, it is recommended that GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN be configured. This causes DVIPA takeover to be delayed until OMPROUTE is active and able to advertise DVIPAs on the takeover stack. For more information on how DELAYJOIN works, see "Sysplex problem detection and recovery" on page 376.

A determination of how the workload will be distributed among the backup stacks when the primary stack fails should be made. It is possible to designate a single stack as a backup and move all the workload to it, or the workload can be spread among several stacks. In the first case, you need to configure only one DVIPA with a VIPADEFINE statement on the primary stack, and only one VIPABACKUP statement is required on the backup stack. For the second option, you must configure multiple DVIPAs with a VIPADEFINE statement on the primary stack.

After determining the workload distribution, each of the secondary stacks will require a VIPABACKUP statement for the DVIPA it will be supporting.

The following example shows how to implement a single stack backup for multiple applications.

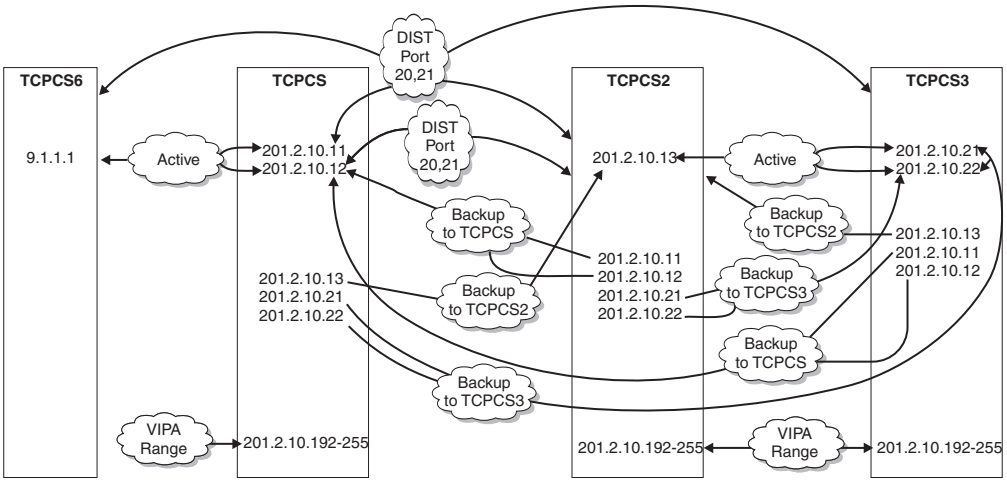


Figure 39. Sample DVIPA addressing in a sysplex environment

Stack TCPCS:

```
Uses VIPADEFINE to define 201.2.10.11
Has a Web server running that binds to INADDR_ANY.
  Web client programs use 201.2.10.11 as their destination address.
Has an FTP server running that binds to INADDR_ANY.
  FTP client programs use 201.2.10.11 as their destination address.
```

Stack TCPCS3:

```
Uses VIPABACKUP to define 201.2.10.11 as backup for stack TCPCS.
Has a Web server running that binds to INADDR_ANY.
Has an FTP server running that binds to INADDR_ANY.
```

In the preceding scenario, when stack TCPCS goes down, stack TCPCS3 receives all new connection requests for both the Web and FTP servers. FTP and Web client programs continue to use 201.2.10.11 as their destination address, but they now connect to stack TCPCS3.

The following example shows how to implement a multiple stack backup for multiple applications.

Stack TCPCS:

```
Uses VIPADEFINE to define 201.2.10.11 and 201.2.10.12
Has a Web server running that binds to INADDR_ANY.
  Web client programs use 201.2.10.11 as their destination address.
Has an FTP server running that binds to INADDR_ANY.
  FTP client programs use 201.2.10.12 as their destination address.
```

Stack TCPCS2:

```
Uses VIPABACKUP to define 201.2.10.11 as backup for stack TCPCS.
Has a Web server running that binds to INADDR_ANY.
```

Stack TCPCS3:

```
Uses VIPABACKUP to define 201.2.10.12 as backup for stack TCPCS.
Has an FTP server running that binds to INADDR_ANY.
```

In the preceding scenario, when stack TCPCS goes down, new connections for the Web server at 201.2.10.11 will connect with stack TCPCS2, and new connections for the FTP server at 201.2.10.12 will connect with stack TCPCS3.

Manually initiating takeover for an individual dynamic VIPA

There might be times when you want a takeover to occur for an individual DVIPA for temporary, operational purposes, such as dealing with temporary shifts in available capacity. An operator command, VARY TCPIP,,SYSPLEX,DEACTIVATE, is provided to deactivate an active VIPADEFINE DVIPA so that it appears to have been deleted from the stack. This enables an already-configured backup stack to takeover the DVIPA. This command also saves the original configuration definition.

When you want to move the DVIPA back to the original stack, you can issue another operator command, VARY TCPIP,,SYSPLEX,REACTIVATE, on the original stack to cause the DVIPA to be redefined with its saved configuration. This causes a VIPADEFINE DVIPA to be activated again on the original stack, and causes the backup stack to relinquish ownership of the DVIPA and return to being a backup stack.

The following example shows how to deactivate a DVIPA:

```
VARY TCPIP,,SYSPLEX,DEACTIVATE,DVIPA=203.1.1.99
```

The following example shows how to reactivate a DVIPA that has been deactivated:

```
VARY TCPIP,,SYSPLEX,REACTIVATE,DVIPA=203.1.1.99
```

Deactivating a DVIPA can be done only for DVIPAs that have been configured on the stack with VIPADEFINE or VIPABACKUP. You cannot deactivate a VIPARANGE DVIPA created by BIND, the SIOCSVIPA or SIOCSVIPA6 ioctl command, or the MODDVIPA utility. When you deactivate a DVIPA, if there are any existing connections to the DVIPA on this stack and there is another stack able to maintain the connections, the DVIPA is kept in QUIESCING status until the last connection terminates, and then the DVIPA is deactivated.

You can also use these operator commands to deactivate and reactivate a backup DVIPA. If you deactivate a DVIPA that is in backup status, it makes that stack ineligible to takeover the DVIPA. Reactivating a VIPABACKUP DVIPA makes the stack eligible again to takeover the DVIPA.

While a DVIPA is deactivated it still appears in the Netstat VIPADCFG/-F report, where it is identified as deactivated, but it does not appear in any other Netstat reports unless the DVIPA is in QUIESCING status or this stack is a target for that DVIPA from some other distributing stack.

Different application uses of IP addresses and DVIPAs

Not all IP-based server applications relate to IP addresses in the same way. Automated movement of DVIPAs, and the planning for dynamic VIPA takeover, must take this difference into account.

Some IPv4 or IPv6 applications, such as TN3270 or FTP, will accept client requests on any IP address by binding to INADDR_ANY or IN6ADDR_ANY. The distinguishing feature of such an application is the function it provides, such as the particular set of SNA applications for TN3270. If the function is replicated across multiple z/OS images in the sysplex, as is often the case for distributed workload, the DVIPA must merely be moved to a stack supporting the application. This scenario is called the multiple application-instance scenario. For the multiple application-instance scenario, the stacks in the sysplex do all the work of activating a DVIPA in the event of a failure.

For other types of applications, each application instance must have a unique IP address. This scenario is called the unique application-instance scenario and uses DVIPAs that are activated with an ioctl or a bind().

To maintain the relationship between an application instance and its DVIPA, the application must indicate to the stack that the DVIPA needs to be activated. This occurs in the following cases:

- When the application instance issues a bind() function call and specifies an IP address that is not active on the stack. The stack will activate the address as a DVIPA, provided it meets certain criteria. When the application instance closes the socket, the DVIPA is deleted.
- Some applications cannot be configured to issue bind() for a specific IP address, but are unique application-instance scenario applications. For such applications, a utility is provided (MODDVIPA), which issues SIOCSVIPA or SIOCSVIPA6 ioctl() to activate or deactivate the DVIPA. This utility can be included in a JCL procedure or OMVS script to activate the DVIPA before initiating the application. As long as the same JCL package or script is used to restart the application instance on another node in the event of a failure, the same DVIPA will be activated on the new stack. For information about the authorization required to execute the MODDVIPA utility, see "Using the MODDVIPA utility" on page 313.
- An alternative for unique application-instance scenario server applications that cannot themselves bind to a unique IP address is to use the BIND parameter on the PORT reservation statement. It is always a good practice to reserve a port for the listening socket of a server application. If the BIND parameter and an IP address are specified on the PORT reservation statement for a server application, and the application binds a socket to that port and INADDR_ANY, z/OS TCP/IP will convert the bind to be specific to the IP address specified on the PORT reservation statement. From that point on, it will appear as though the application itself had issued the bind() specifying the designated IP address, including having the IP address deleted when the server application closes the socket.

Configuring dynamic VIPAs

To allow continued and unchanged operation of static VIPAs in z/OS TCP/IP, DVIPAs are defined with configuration statements in the PROFILE.TCPIP data set. An overview of the relevant configuration statements is provided in the following sections, and also see "Verifying the DVIPAs in a sysplex" on page 348 for a description of the configuration statements. For an example of the VIPADYNAMIC configuration statement and display commands for dynamic VIPAs, see *z/OS Communications Server: IP Configuration Reference*.

Configuring the multiple application-instance scenario

For the multiple application-instance scenario, each instance is assigned a unique DVIPA. The VIPADEFINE keyword of the VIPADYNAMIC configuration statement is used to create the DVIPA on the stack where the DVIPA is normally expected to be active. When the VIPADEFINE statement is processed in a TCP/IP profile, corresponding DEVICE, LINK, HOME, and BSDROUTINGPARMS statements are generated automatically for IPv4 addresses. For IPv6 addresses, an INTERFACE statement is automatically generated. Routing daemons are automatically informed.

Additional configuration is required on other stacks in the sysplex to indicate which stack should take over the DVIPA in the event of a failure. The VIPADYNAMIC statement has a VIPABACKUP keyword for this purpose. A

VIPABACKUP configures the DVIPA but does not activate it until it is necessary. Because more than one TCP/IP can backup a single DVIPA, a rank parameter on the VIPABACKUP statement determines the order in which several stacks will assume responsibility for a DVIPA.

The stacks in the sysplex exchange information on all VIPADEFINEs and VIPABACKUPs defined in the sysplex, so that all are aware of which stack should take over a particular DVIPA. The list of backup stacks for a specific DVIPA can be different from the list of backup stacks for all other DVIPAs.

In the multiple application-instance scenario, instances of the application in question are activated among sysplex nodes according to some plan, presumably related to balancing workload across available capacity. This activation is done independently of VIPA takeover. Configure the associated DVIPAs as follows:

1. For each instance of a particular application to be supported via DVIPA, add a VIPADEFINE statement to the TCP/IP profile for the TCP/IP associated with the application instance.
2. For each of the dynamic VIPAs in step 1, determine which application instance or instances should take over the workload (considering probable capacity and any other application-related considerations). If more than one TCP/IP is to provide backup for a DVIPA, determine the order in which the selected TCP/IPs should be designated as backup. Add a VIPABACKUP statement to each TCP/IP that is to provide backup for the DVIPA, with appropriate rank values to determine the order. Do this for each of the DVIPAs in step 1.
3. Perform steps 1 and 2 for each other application to be supported by DVIPAs.

   Note: It is possible to share a dynamic VIPA among several different applications, but in doing so, ensure that instances of all such applications will exist together on any TCP/IP to which the DVIPA may be moved in case of a failure.

After these steps are complete, start the affected TCP/IPs (or modify their configuration using the VARY TCPIP,,OBEYFILE command), if applicable, configure DNS for the application names, and start the application instances. From that point on, the TCP/IPs in the sysplex will collaborate to ensure that each dynamic VIPA is kept active somewhere within the sysplex as long as there is at least one functioning TCP/IP which has been designated as backup for the dynamic VIPA.

Configuring the unique application-instance scenario

The unique application-instance scenario ties a DVIPA to a specific instance of an application. To isolate errors in configuring applications, TCP/IP needs a mechanism to identify permissible DVIPAs. This is provided with one or more VIPARANGE statements. The VIPARANGE statement identifies a range of IP addresses which can be activated as DVIPAs by an application instance. Each TCP/IP stack can have up to 256 VIPARANGE definition statements for IPv4 and up to 256 for IPv6. The VIPARANGE statement consists of an IPv4 address and subnet mask, or an interface name and an IPv6 address and prefix length, and defines a subnet for DVIPAs. More than one VIPARANGE statement with different ranges can be defined on a TCP/IP. VIPARANGE does not itself cause any DVIPAs to be activated. Rather, DVIPAs are activated either by an application issuing a bind() for a specific IP address, by use of the SIOCSVIPA or SIOCSVIPA6 ioctl() command issued by an authorized application, or by the MODDVIPA utility.

When an application issues bind() for a specific IP address or an address was selected using the BIND keyword on the PORT statement, the receiving stack checks it against addresses in the HOME list. If the IP address has already been activated on this stack (whether for a physical device, a static VIPA, or a dynamic VIPA), the bind() execution is successful. If the IP address is not active on this TCP/IP, the current VIPARANGEs are checked to see if the IP address falls within one of them. If an appropriate VIPARANGE is found, the IP address is activated as a DVIPA and the operation succeeds. If no appropriate VIPARANGE is found, or if the IP address is active elsewhere in the sysplex other than by a NONDISRUPTIVE bind(), the request fails and bind() returns EADDRNOTAVAIL.

When an authorized application issues the SIOCSVIPA or SIOCSVIPA6 ioctl() command to create a DVIPA, or when the MODDVIPA utility is executed in JCL or an OMVS script to activate a DVIPA on behalf of an application instance, the current VIPARANGES are checked to see whether the IP address falls within one of them. If an appropriate VIPARANGE is found, and the IP address is not currently active on this TCP/IP or elsewhere in the sysplex as an IP address or a VIPADEFINE/VIPABACKUP dynamic VIPA, then the IP address is activated as a DVIPA. However if no appropriate VIPARANGE is found on this TCP/IP, or if the IP address is currently defined on this TCP/IP or configured elsewhere in the sysplex as an IP address or a VIPADEFINE/VIPABACKUP dynamic VIPA, then the request fails with errno and errnojr set to indicate the reason for the failure and the utility ends with a nonzero condition code. See “Dynamic VIPA creation results” on page 336 for more information.

Note: If the requested IP address has been activated as a dynamic VIPA by a bind(), SIOCSVIPA ioctl, or SIOCSVIPA6 ioctl elsewhere in the sysplex, the result depends on how the stacks were configured. See “Dynamic VIPA creation results” on page 336 for more information.

In the unique application-instance scenario, each application instance is assigned a unique IP address as its DVIPA. Before defining individual DVIPAs, one or more blocks of IP addresses must be defined for these DVIPAs, and the individual DVIPAs must be defined from within the blocks. Each block should be represented as a subnet, so that a VIPARANGE statement can be defined for it.

Follow these steps when setting up any unique application instances:

1. For each application instance, assign a DVIPA from one of the blocks of IP addresses for this purpose. Do not assign an IP address which is also assigned to another application instance, or which is defined by VIPADEFINE for the multiple application-instance scenario. Configure the application to use this DVIPA [if it issues bind() for a particular IP address], or add a BIND parameter to the PORT reservation statement for the port of the application’s listening socket to cause the listening socket to be bound to this DVIPA. Alternatively, add the MODDVIPA utility to the JCL or OMVS script and configure the MODDVIPA utility to activate the DVIPA before starting the application. When the application ends, use the MODDVIPA utility to delete the DVIPA.
2. For each application instance, determine on which stack the application instance will normally be executed and to which stacks the application instance could be moved in case of failure of the normal stack or the application itself. For each such stack, add a valid VIPARANGE statement to the profile.

   Note: The dynamic VIPA must be within the VIPARANGE subnet. The broadcast address and the net prefix cannot be used.

3. Perform steps 1 and 2 until all application instances have been allocated a unique DVIPA.

The application restart strategy should ensure that the worst-case failure scenario does not attempt to activate more than 1024 DVIPAs on a single stack. If such an attempt is made, activation of the 1025th DVIPA will fail, with possible resulting loss of connectivity from clients to the server application.

Note: The limit of 1024 DVIPAs on a single TCP/IP applies to all DVIPAs, whether defined by VIPADEFINE/ VIPABACKUP configuration statements, through a VIPADISTRIBUTE statement on another stack, by a bind() call, or by executing the MODDVIPA utility.

Defining a single block makes the definition process easier, but also provides less individual control. Alternatively, since the smallest subnet consists of four IP addresses, defining a unique subnet for each DVIPA in this scenario *wastes* three other IP addresses that could have been used for DVIPAs.

Using the SIOCSVIPA or SIOCSVIPA6 ioctl command

An ioctl command, SIOCSVIPA or SIOCSVIPA6, allows an application to create or delete a dynamic VIPA on the stack where the application is running. The application issuing the SIOCSVIPA or SIOCSVIPA6 ioctl command must be APF authorized and be running under a user ID with superuser authority. If the new profile for the MODDVIPA program has been defined under RACF, any user ID can be allowed to issue the SIOCSVIPA or SIOCSVIPA6 ioctl simply by being permitted to use this profile. For more information, see "Defining a RACF profile for MODDVIPA" on page 314.

To create a new dynamic VIPA, the requested IP address must be within a subnet that has been previously specified by a VIPARANGE configuration statement in the PROFILE.TCPIP data set for this stack. The SIOCSVIPA or SIOCSVIPA6 ioctl command can be used to delete any existing dynamic VIPA on the stack, except for distributed DVIPAs.

The following example shows how to set up the SIOCSVIPA ioctl command for applications that will use IPv4 addresses.

```
#include "ezbzdvpc.h"               /* header that contains
                                      the structure for
                                      SIOCSVIPA ioctl
                                       and needed constants*/
struct dvreq dv;                     /* the structure passed
                                        on the ioctl command*/
dv.dvr_version = DVR_VER1;            /*version               */
dv.dvr_length = sizeof(struct dvreq);  /* structure length     */
dv.dvr_option = DVR_DEFINE;            /* to define a new
                                        dynamic VIPA.  Use
                                        DVR_DELETE to delete
                                        a dynamic VIPA        */
dv.dvr_addr.s_addr = inet_addr(my_ipaddr);  /* where my_ipaddr is
                                               a character string
                                               in standard
                                               dotted-decimal
                                               notation */
```

The ioctl command is then invoked as follows:

```
rc = ioctl(s, SIOCSVIPA, &dv);
```

The following example shows how to set up the SIOCSVIPA6 ioctl command using the input parameter list that supports IPv6 addresses.

```
#include "ezbzdvpc.h"           /* header that contains
                                   the structure for
                                   SIOCSVIPA6 ioctl
                                   and needed constants */
 struct dvreq6 dv6;              /* the structure passed on
                                   the ioctl command */
 dv6.dvr6_version = DVR_VER2;   /* version */
 dv6.dvr6_length = sizeof(struct dvreq6); /* structure length */
 dv6.dvr6_option = DVR_DEFINE;  /* to define a new dynamic
                                   VIPA.  Use DVR_DELETE to
                                   delete a dynamic VIPA. */
inet_pton(AF_INET6, my_ipv6addr, dv6.dvr6_addr6.s6_addr);
                                 /* where my_ipv6addr
                                   is a character string in
                                   standard IPv6 address
                                   notation, representing a
                                   fully qualified IPv6
                                   address */
```

The ioctl command is then invoked as follows:

```
rc = ioctl(s, SIOCSVIPA6, &dv6);
```

The SIOCSVIPA or SIOCSVIPA6 ioctl command sets nonzero errno and errnojr values to indicate error conditions. Refer to *z/OS Communications Server: IP and SNA Codes* for a description of the errnojr values returned.

Using the MODDVIPA utility

You can use the MODDVIPA utility to activate or delete a dynamic VIPA. The utility can be initiated from JCL or an OMVS script. MODDVIPA must be loaded from an APF authorized library and be executed under a user ID with SuperUser authority. The user ID must also have an OMVS segment defined (or defaulted). If the new profile for the MODDVIPA program has been defined under RACF, any user ID can execute the MODDVIPA program simply by being permitted to use this profile. For more information, see "Defining a RACF profile for MODDVIPA" on page 314.

Input parameters: The input parameters for the utility are:

-p <tcpipname>
Specifies the TCP/IP which is to create or delete a DVIPA.

-c <IPaddress> or -d <IPaddress>
Specifies to create (-c) or delete (-d) the address (IP address) specified.

Notes:

1. The input parameters -p, -c, and -d must be entered in lowercase.
2. <*tcpipname*> must be entered in upper case.
3. For IPv4 addresses, <*IPaddress*> is dotted-decimal notation. For IPv6 addresses, <*IPaddress*> is standard colon-hexadecimal notation for specifying IPv6 addresses.
4. To create a DVIPA, the specified IP address must be within a subnet that has been previously specified by a VIPARANGE configuration statement in the PROFILE.TCPIP data set for the specified TCP/IP.

Output: The MODDVIPA utility sets the following exit (completion) codes for create (-c):

0 Success: The DVIPA was activated.

4 Warning: The requested DVIPA was not activated because the specified IP address is already active on this stack.

8 Error: The IP address was not defined as a DVIPA on this TCP/IP.

12 An error was found in the input parameters

The MODDVIPA utility sets the following exit (completion) codes for delete (-d):

0 Success: The dynamic VIPA was deleted.

8 Error: The requested DVIPA was not deleted.

12 An error was found in the input parameters

Notes:

1. When an error is detected, the ernno text and errnojr value are printed to stderr.
2. If the IP address requested for the DVIPA is not within a VIPARANGE configured on this stack, completion code 8 is returned even if the IP address is currently active on this stack

Examples

Within JCL:

```
//TCPDVP EXEC  PGM=MODDVIPA,REGION=0K,TIME=1440,                        X
//             PARM='POSIX(ON) ALL31(ON)/-p TCPCS3 -c 1.2.3.4'
```

From OMVS shell:

```
moddvipa -p TCPCS3 -c 1.2.3.4
```

Defining a RACF profile for MODDVIPA

Note: Wherever SIOCSVIPA is used in this section, SIOCSVIPA6 can be used as well.

You can restrict access to the MODDVIPA (EZBXFDVP) program by defining a RACF profile under the SERVAUTH class and specifying the user IDs that are authorized to execute the SIOCSVIPA ioctl or the MODDVIPA utility program. You can decide on the level of control that is appropriate for your installation.

To restrict access to the SIOCSVIPA ioctl (and thus the MODDVIPA utility), you can define a RACF profile using the following example:

```
RDEFINE SERVAUTH (EZB.MODDVIPA.sysname.tcpname)
        UACC(NONE)

PERMIT EZB.MODDVIPA.sysname.tcpname
       ACCESS(READ) CLASS(SERVAUTH) ID(USER1)
```

In the example above, *sysname* is the name of the MVS system where the ID will execute the MODDVIPA utility or issue the SIOCSVIPA ioctl, and *tcpname* is the job name of the TCP/IP started task. The job name for started tasks, such as TCP/IP, is derived depending on how it is started:

- If the START command is issued with the name of a member in a cataloged procedure library (for example, S TCPIPX), the job name will be the member name (for example, TCPIPX).
- If the member name on the START command is qualified by a started task identifier (for example, S TCPIPX.ABC), the job name will be the started task

identifier (for example, ABC). The started task identifier is not visible to all MVS components, but TCP/IP uses it to build the RACF resource name.
- The JOBNAME parameter can also be used on the START command to identify the job name (for example, S TCPIPX,JOBNAME=XYZ).
- The JOBNAME can also be included on the JOB card.

In this example, user ID *USER1* is being permitted to invoke the MODDVIPA utility (and thus the SIOCSVIPA ioctl).

If this RACF profile is created, the user ID must be permitted to access this profile or else the SIOCSVIPA ioctl (and thus the MODDVIPA utility) will fail with a 'permission denied' error, regardless of SuperUser authority.

Also note that before the RACF profiles take effect, a refresh of these profiles might be required. This can be accomplished by the following RACF command:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

For more information, refer to *z/OS Security Server RACF Security Administrator's Guide*.

Choosing which form of dynamic VIPA support to use

The following sections explain which of the features should be used for the type of application being used.

When should VIPADEFINE and VIPABACKUP be used to define a dynamic VIPA?

- One or more applications bind to INADDR_ANY or IN6ADDR_ANY, and the applications exist on multiple TCP/IPs.
- Dynamic VIPA takeover is desired.
- The DVIPA does not need to be deleted when the application is stopped.

When should VIPARANGE and bind() be used to define a dynamic VIPA?

- The application cannot bind to INADDR_ANY or IN6ADDR_ANY, or dynamic VIPA takeover is not desired.
- The IP address to which the application binds can be controlled by the user. The application's first explicit bind (the listening socket) will remain for the life of the application. Otherwise, the DVIPA will be removed everytime the application's DVIPA owning socket is closed, and re-added everytime there is a new DVIPA owning socket (another explicit bind has been done and the DVIPA does not exist).
- Automatic deletion of the dynamic VIPA when the application is stopped is acceptable.
- A specific dynamic VIPA address must be associated with a specific application.
- The application is not APF authorized, or not run under a user ID with superuser authority.

When should VIPARANGE and the MODDVIPA utility (or ioctl command SIOCSVIPA or SIOCSVIPA6) be used to define a dynamic VIPA?

- The application cannot bind to INADDR_ANY or IN6ADDR_ANY, or dynamic VIPA takeover is not desired.
- The IP address to which the application binds is known but cannot be controlled by the user.

- Automatic deletion of the dynamic VIPA when the application is stopped is not acceptable.
- The MODDVIPA utility (or application issuing the ioctl command) will be run from an APF authorized library and under a user ID with superuser authority.

Configuring distributed DVIPAs — Sysplex Distributor

A distributed DVIPA exists on several stacks, but is advertised outside the sysplex by only one stack. This stack receives all incoming connection requests and routes them to all the stacks in the distribution list for processing. This provides the benefit of distributing the workload of incoming requests and providing additional fail-safe precautions in the event of a server failure.

You can distribute connections destined for a dynamic VIPA (DVIPA) by adding a VIPADISTRIBUTE configuration statement for a previously defined dynamic VIPA. The order of the statements is important. The VIPA is first defined with the VIPADEFINE statement and then included on a VIPADISTRIBUTE statement. Another TCP/IP can act as a backup for the distributed DVIPA by properly coding a VIPABACKUP statement; the backup will perform the routing function in the event of a failure. The options specified on a VIPADISTRIBUTE statement are inherited by a backup stack unless the second stack has its own VIPADISTRIBUTE statement for that DVIPA, in which case it will use its own VIPADISTRIBUTE statement for distributing. You can also code a VIPADISTRIBUTE statement with just the VIPABACKUP statement and not for the VIPADEFINE statement. This would allow workload distribution only during a primary outage.

You can change the distribution of a DVIPA after a backup stack has activated it. However, if the backup stack did not not have its own distribution defined by a VIPADISTRIBUTE statement before it activated the DVIPA, any distribution changes made while the DVIPA is active on the backup stack are temporary. Those changes will be in effect while the DVIPA remains active on the backup stack, but will not be remembered if this stack takes over the DVIPA again in the future.

Following is an example of a properly coded distributed DVIPA:

```
IPCONFIG SYSPLEXROUTING DYNAMICXCF 193.9.200.4 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2000::93:9:200:4
VIPADYNAMIC
  VIPADEFINE  255.255.255.192 9.67.240.02
  VIPADISTRIBUTE DEFINE  9.67.240.02 PORT 20 21 8000 9000 DESTIP
          193.9.200.2
          193.9.200.4
          193.9.200.6
  VIPADEFINE V6DVIPA1 2000::9:67:240:2
  VIPADISTRIBUTE DEFINE V6DVIPA1 PORT 20 21 8000 9000 DESTIP
          2000::93:9:200:2
          2000::93:9:200:4
          2000::93:9:200:6
ENDVIPADYNAMIC
```

Prior to z/OS V1R6 Communications Server, the TCP/IP stack that was configured as a distributor of dynamic VIPAs was required to enable IP forwarding using the IPCONFIG (or IPCONFIG6) DATAGRAMFWD TCP/IP profile statement. For installations that do not wish to configure their TCP/IP stack as a forwarding node, it is no longer a requirement for distributing dynamic VIPAs. However, if your installation is configured such that target TCP/IP stacks only have XCF connectivity, datagram forwarding still needs to be configured on the distributor, as all packets originating from the target will be forwarded by the distributor.

There are several configuration changes that can be made to affect the method the distributing stack will use to forward connections to the target stacks. In each of the following items, *all participating stacks* is used to refer to the distributing stack and all target stacks.

WLM-based forwarding based on target system workload
If the DISTMethod BASEWLM parameter is specified on the respective VIPADISTRIBUTE statement, or if the DISTMethod parameter is not specified, this distribution method is enabled. This is the default distribution method. To enable the distributing stack to forward connections based upon the workload of each of the target stacks, specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks. This registers all participating stacks with WLM and enables the distributing stack to request workload information from WLM.

WLM-based forwarding based on server-specific workload
If the DISTMethod SERVERWLM parameter is specified on the respective VIPADISTRIBUTE statement, the distributing stack selects from the available servers for a DVIPA/port and forwards connections based on a WLM recommendation indicating how well each server is executing on its system. To enable the distributing stack to forward connections based on server-specific workload, specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks.

WLM/QoS-based forwarding
Regardless of whether BASEWLM or SERVERWLM weights are being used, to enable the distributing stack to forward connections based upon a combination of workload information and network performance information (TCP retransmissions and time-outs), specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks, and also define a Sysplex Distributor performance policy on the target stacks. For information on configuring these policies, see "Sysplex Distributor policy example" on page 761.

Round-robin forwarding
If the DISTMethod ROUNDROBIN parameter is specified on the respective VIPADISTRIBUTE statement, the distributing stack uses a round-robin mechanism to select one of the DVIPA/port targets for each connection.

If you have not made the changes needed to enable WLM-based forwarding (SYSPLEXROUTING has not been specified for all participating stacks, or all participating stacks are not configured for WLM GOAL mode), even if BASEWLM or SERVERWLM is specified, the distributing stack will use round-robin forwarding to distribute connections.

Tip: The weights received from WLM are returned based on the value of HOSTNAME. The HOSTNAME value is determined by the search path at initialization time. Verify that all systems have a unique HOSTNAME value to ensure proper distribution.

Whether the distributing stack is performing WLM-based forwarding, WLM/QoS-based forwarding, or round-robin forwarding, Sysplex Distributor routing policies can further affect the distribution of connections. Sysplex Distributor routing policies, configured on the distributing stack, are used to specify a set of target stacks for a given set of traffic. For example, all traffic destined to a given port/DVIPA from a specified subnet can be assigned one group of target stacks, while traffic for the same port/DVIPA from another subnet can be

assigned to a different group of target stacks. For more information on configuring these types of policies, see "Sysplex Distributor policy example" on page 761.

Distribution of connections to target servers can also be affected by the responsiveness of the target stacks or target servers. The Sysplex Distributor stack monitors how well target servers respond to connection setup requests at intervals of approximately 1 minute, calculating a target server responsiveness (TSR) value for each server. For WLM-based forwarding based on target system workload, WLM/QoS-based forwarding, or WLM-based forwarding based on server-specific workload, new connection setup requests are diverted away from target servers that are handling connection setup requests relatively poorly. For round-robin forwarding, if the Sysplex Distributor determines that a target server is not successfully accepting connection setup requests at all, that target server is bypassed. Periodically, the distributor sends a new connection request to a target with a TSR of 0, to check whether the responsiveness of that target has improved.

Each distributing stack and each target stack must have an IPv4 or IPv6 DYNAMICXCF address, or both. This address is used by other distributing stacks as a destination point. When using Sysplex Distributor, do not define an IUTSAMEH link. These links will be created automatically from the DYNAMICXCF statement. Refer to *z/OS Communications Server: IP Configuration Reference* for directions for coding DYNAMICXCF on the IPCONFIG or IPCONFIG6 statements. For more information on additional configuration parameters required, also see the usage notes related to the DYNAMICXCF parameter under the IPCONFIG or IPCONFIG6 statements in *z/OS Communications Server: IP Configuration Reference*.

The VIPADISTRIBUTE statement specifies how new connection requests are routed to a set of candidate target stacks. The VIPADISTRIBUTEd DVIPA can be followed by up to 64 ports. The preceding example shows the well-known ports for FTP and the ports for a custom application.

Up to 32 target TCP/IPs follow the DESTIP keyword and are identified by their respective dynamic XCF IP addresses. The VIPADISTRIBUTE statement can also specify DESTIP ALL, in which case all current and future stacks with activated dynamic XCF can participate in the distribution as candidate target stacks. As an application listens to one of the specified ports on each listed TCP/IP, the routing TCP/IP begins to forward connections to that stack.

For further reference on Sysplex Distributor, refer to *Communications Server for z/OS V1R2 TCP/IP Implementation Guide Volume 5: Availability, Scalability, and Performance*, SG24–6517.

Manually quiescing DVIPA Sysplex Distributor server applications

You can stop a particular server application, or all server applications, on a target stack from receiving new DVIPA Sysplex Distributor workload using the VARY TCPIP,,SYSPLEX command with the QUIESCE parameter and either the PORT= or TARGET parameter.

This command, entered on the target stack where the application exists, prevents the application from receiving any new DVIPA Sysplex Distributor connections but does not affect existing connections. This command can be issued to gracefully quiesce a target application or target system, before the application or system is

brought down for planned maintenance. It can also be used to temporarily divert new workload requests from a particular application or a target stack.

The VARY TCPIP,,SYSPLEX command with the RESUME parameter and either the PORT= or TARGET parameter can then be used to resume using the application as a target for new DVIPA Sysplex Distributor connections.

For more details on the VARY TCPIP,,SYSPLEX command and its parameters, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Route selection for distributing packets

Sysplex Distributor uses the dynamic XCF interfaces (DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statements) to distribute all incoming packets to target stacks. Using dynamic XCF interfaces has several advantages, such as simplified configuration, because you can leverage the existing XCF communication links in your sysplex environment and do not have to define separate communication paths to all target systems. This is especially useful in scenarios where target systems do not have direct network connectivity, but rather rely on the network connectivity of the routing stacks.

Sysplex Distributor also includes some optimization logic that enables it to select alternative network paths for forwarding DVIPA packets. This optimization logic is automatically performed by the Sysplex Distributor in the following configurations:

- When the routing stack and the target stacks reside in the same MVS image, the routing stack uses internal communication links (IUTSAMEH) to forward DVIPA packets to the target, avoiding an external network altogether.
- When the routing stack and a target stack are running in different LPARs but on the same zSeries central processor complex (CPC), and HiperSockets connectivity is available, the HiperSockets connectivity is automatically used for forwarding DVIPA packets instead of the dynamic XCF interfaces.

These optimizations provide the best performance characteristics for forwarding DVIPA packets within the same CPC.

For configurations where the routing and target stacks reside in different CPCs, there are some scenarios where it might be desirable to use interfaces other than the dynamic XCF interfaces for forwarding distributed DVIPA packets:

- When sysplex XCF communication interface links, whether configured through CTCs or through the coupling facility, are constrained or are heavily used by other, non-DVIPA communications.
- When the routing and target stacks have direct network connectivity over high speed, low latency, and wide bandwidth networks, such as access to the same Gigabit Ethernet segments using OSA-Express.

In these configurations, forwarding DVIPA packets over these alternative network connections can actually improve performance (that is, reduce latency and CPU cost) while reducing the utilization of sysplex XCF interfaces.

You can use the VIPAROUTE statement in the VIPADYNAMIC block in the TCP/IP profile to influence the Sysplex Distributor's logic in selecting a route and interface to forward DVIPA packets to a target stack. Using the VIPAROUTE statement, you can indicate which IP address on the target stack is to be used as the destination or target IP address during the route lookup selection. When Sysplex Distributor processes the incoming packet, it determines whether a

matching VIPAROUTE statement has been defined, and if it has, the TCP/IP stack returns the best available route to reach the target IP address. The incoming packet is then encapsulated using a destination of the VIPAROUTE target IP address in the outer header, and forwarded to the target stack. Generic routing encapsulation (GRE) is used to encapsulate IPv4 packets, while an outer IPv6 header is used to encapsulate IPv6 packets. This enables you to select the optimal interface for forwarding DVIPA packets across different CPCs, while retaining the optimized communication paths when the routing and target stacks reside on the same MVS image, the same CPC, or both.

If you do not want to use the dynamic XCF interfaces at all, you must define the dynamic routes so that the cost to reach the IP address on the target stack using the dynamic XCF interfaces is higher than the cost to reach that IP address using other interfaces. For more information, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 249. If this is not done, it is possible that dynamic XCF interfaces will be used if the normal routing tables select those interfaces. By ensuring that the route using the dynamic XCF interfaces has a higher cost, you can exploit the other interfaces for forwarding most DVIPA traffic, yet maintain the dynamic XCF interface as a backup should the preferred network interfaces fail.

There are some situations where specifying a dynamic XCF address as the target IP address on the VIPAROUTE statement is desirable. For example, if a distributing stack uses non-HiperSockets interfaces to reach its target stacks, but the backup stack uses HiperSockets interfaces to reach the same targets, it is preferable to use the HiperSockets interfaces if the backup stack takes over. In this instance, it would be preferable to define VIPAROUTE statements with the dynamic XCF address as the target IP address on the backup stack, which causes the HiperSockets interfaces to be used.

In the following cases, even though VIPAROUTE was specified, the dynamic XCF interface is used for distribution:

- A target IP address is specified that is not owned by the target stack.
- The defined dynamic XCF address is for a pre-V1R7 target stack.

When these conditions are detected, messages are issued at the distributing stack, as well as when the distributing stack first attempts to route a connection request to the target stack.

The dynamic XCF address must still be configured even if VIPAROUTE definitions are used, as Sysplex Distributor continues to use a dynamic XCF address to identify every target TCP/IP stack in the sysplex. In addition, there are several functions that continue to depend on dynamic XCF connectivity for intra-sysplex communications:

- Sysplex-wide security associations (SWSA)
- Multilevel security packets
- QoS performance data collection of Policy Agent

Generic routing encapsulation (GRE)

Generic routing encapsulation (GRE) is a standard protocol described by RFC 1701, and is used for several TCP/IP functions. GRE is implemented for only IPv4 packets. GRE enables a wrapper to be placed around a packet during transmission of the data. A receiving stack that supports GRE removes the GRE wrapper, enabling the original packet to be processed by the receiving stack. GRE is often used to deliver a packet to a stack using an alternate destination IP address.

Fragmentation considerations

For IPv4, when incoming packets destined for a DVIPA address need to be forwarded to a target TCP/IP stack using a route that was determined by a VIPAROUTE statement, the packet is encapsulated using GRE prior to being forwarded. This enables the packet to be forwarded through the network to the target stack while preserving the original packet's destination IP address (that is, the DVIPA address). The GRE encapsulation process increases the size of the forwarded packet by 28 bytes. As a result, if the size of the encapsulated GRE packets are larger than the maximum transmission unit (MTU) of the network interface that will be used for forwarding the packet, the TCP/IP stack might need to perform fragmentation, creating two or more packets that are forwarded to the target stack. The target stack then reassembles the fragmented packets.

While fragmentation and reassembly processing is not unusual in an IP network, it is generally desirable to eliminate the need for this processing, thereby optimizing performance. In the case of fragmentation resulting from GRE encapsulation, the cost of the fragmentation and reassembly processing might become a concern if a large percentage of the incoming DVIPA packets to be forwarded require fragmentation. Fortunately, configuration options do exist that can help eliminate the need for this fragmentation and reassembly processing, including the following:

- Enable path MTU discovery on the client IP hosts (that is, client hosts sending packets to the DVIPA). This enables the client hosts to dynamically discover the smallest MTU along the path from the client to the server. In the case of forwarded DVIPA traffic, the path MTU is adjusted (reduced) by the length of the GRE header, if the addition of this header would have resulted in fragmentation being required. For IPv6, path MTU discovery is automatically enabled for all hosts, and no explicit configuration should be required.
- Ensure that the MTU size of the routes over the network interfaces that are used to forward the DVIPA packets is large enough to account for the largest client packet plus the length of the GRE header. This might be an option if the clients are connected to networks with a smaller MTU size than what is available in the network paths between the z/OS hosts (that is, the TCP/IP stack forwarding the DVIPA packets and the target TCP/IP stack). Consider the following example:
  - The majority of the client hosts are connected to the network using Fast Ethernet, and as result use an MTU of 1492.
  - The route selected using the VIPAROUTE statement specifies a network path over OSA Express Gigabit Ethernet between the two z/OS hosts. Gigabit Ethernet accommodates a much larger MTU size than Fast Ethernet (8992 versus 1492). This larger MTU size with Gigabit Ethernet is often referred to as jumbo frames.

  In this configuration, the z/OS hosts can be configured to take advantage of the larger MTU size when communicating with each other over the Gigabit Ethernet network. As a result, fragmentation is avoided for these forwarded DVIPA packets, as the larger MTU easily accommodates the increased packet size resulting from the GRE encapsulation. However, it is important to ensure that this larger MTU size is used only for communications among hosts where the entire network path supports the larger MTU size. Otherwise, packets sent from the z/OS hosts using the larger MTU size might need to be fragmented as they cross network boundaries that support only lower MTU sizes. As a result, when configuring larger MTU sizes, such as jumbo frames for Gigabit Ethernet, it is also important to consider enabling path MTU discovery on the hosts using the larger MTU size. This enables these hosts (in this example, the z/OS hosts) to use the larger MTU size only where appropriate, without introducing

fragmentation. For more information on specifying MTU sizes on z/OS, see "Maximum Transmission Unit (MTU) considerations" on page 100.

Dynamic port assignment

Sysplex Distributor can also react dynamically to servers binding to the distributed DVIPA and creating a listening socket, adding a port to the list of ports for which connection workload balancing will occur. If the PORT parameter is omitted from the VIPADISTRIBUTE statement for the distributed DVIPA, any server that binds to the distributed DVIPA and a nonzero port will be eligible for workload distribution. If only one server binds to the distributed DVIPA and port and establishes a listening socket, that server will get all of the work. When the second server binds to the distributed DVIPA and the same port and establishes a listening socket, it will immediately become eligible to participate in connection workload balancing. TCP/IP will not enforce a limit on the number of ports that can participate in connection workload balancing per distributed DVIPA, other than the total number of allowable ports.

Sysplex-wide source VIPA

Sysplex Distributor addresses the requirement of providing to clients outside a parallel sysplex a single-IP-address appearance to application instances spread across the sysplex, and also the distribution of the incoming work among the various instances. Many applications are part of a cooperative network of applications, and the sysplex applications that serve as clients to end users might also have to initiate (client-like) outbound connection requests to cooperating applications. The SOURCEVIPA feature allows applications to attain independence of any physical adapter, but SOURCEVIPA is limited to statically defined VIPAs within a stack. Different instances of the same application using Sysplex Distributor, and thus having a single IP address for inbound connection requests, will use different IP addresses for their outbound connection requests.

These problems are resolved by allowing a single sysplex-wide VIPA to be used as the source IP address for TCP applications and to have the sysplex stacks collaborate on assigning ephemeral ports to prevent duplicate connection 4-tuples (combination of source and destination IP addresses and ports). These solutions are provided by SYSPLEXPORTS, in conjunction with either job-specific source IP addressing or sysplex-wide source VIPAs for TCP connections.

For information on job-specific source IP addressing using the SRCIP statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Sysplex-wide source VIPAs for TCP connections

The TCPSTACKSOURCEVIPA keyword on the IPCONFIG or IPCONFIG6 statements allows users to specify a single VIPA, static or dynamic, to be used as a source IP address for TCP applications that initiate outbound connections on that stack. TCPSTACKSOURCEVIPA is only in effect when SOURCEVIPA is enabled and an application issues a connect() without a bind(). TCPSTACKSOURCEVIPA overrides other forms of source IP selection for all TCP applications that issue the connect() without a bind(). Applications that perform an explicit bind() do not use the TCPSTACKSOURCEVIPA address. Applications using the Pascal API TcpOpen() call have an explicit bind() performed for them.

If you specify TCPSTACKSOURCEVIPA and do not specify SOURCEVIPA in a profile, a warning message is issued and TCPSTACKSOURCEVIPA will not be enabled. The address specified does not need to be active on the stack at profile

processing time. For example, a valid TCPSTACKSOURCEVIPA address can be an address that will be a target address on this stack for sysplex distribution.

If the TCPSTACKSOURCEVIPA address is not an active VIPA on the stack at the time a connect() is issued, the connect() call goes through normal source IP selection. A warning message will be issued no more than once every five minutes (to avoid flooding the system console), indicating an attempt to use the address specified in TCPSTACKSOURCEVIPA failed.

TCPSTACKSOURCEVIPA can be coded on all target stacks. The target TCPSTACKSOURCEVIPA statements can specify individual unique addresses, or can be duplicates of those specified on the distributing stack (a target DVIPA). Specifying the same DVIPA address for TCPSTACKSOURCEVIPA on the distributor and all target stacks creates a sysplex-wide source VIPA and raises the concern for coordination of ephemeral ports across the sysplex.

For information on diagnosing sysplex-wide source VIPAs for TCP connections problems, see *z/OS Communications Server: IP Diagnosis Guide*.

SYSPLEXPORTS

Whenever two or more application instances use the same source IP address and initate connections to the same destination IP address and port, sysplex-wide coordination of assignment of ephemeral ports is required so that the 4-tuple for each connection remains unique. As long as the source IP address is on a single stack, this coordination is not a problem because the stack manages assignment of ephemeral ports. However, with Sysplex Distributor applications, multiple application instances might desire to initiate connections using the same distributed DVIPA, potentially to the same destination IP address and port, so uniqueness of the connection 4-tuples cannot be guaranteed unless the stacks collaborate across the sysplex for ephemeral port assignment for distributed DVIPAs. This can be done by adding the optional SYSPLEXPORTS parameter to the VIPADISTRIBUTE statement.

SYSPLEXPORTS must be specified on the first VIPADISTRIBUTE statement processed for a particular DVIPA. It cannot be enabled once a DVIPA has been configured for distribution. Once enabled, it cannot be disabled until all distribution has been deleted for the DVIPA.

When a distributed DVIPA can be active on more than one target stack, SYSPLEXPORTS can be specified to cause the stacks to collaborate in the assignment of ephemeral ports for outbound initiated TCP connections. This ensures that two different connections do not end up with the same connection 4-tuple.

At profile processing time, a stack whose profile contains a SYSPLEXPORTS parameter on a VIPADISTRIBUTE statement will connect to the coupling facility EZBEPORT structure containing sysplex port assignment information. (The structure will be a list structure with an entry for each DVIPA address with a VIPADISTRIBUTE with SYSPLEXPORTS specified anywhere in the sysplex. The first stack to connect to the EZBEPORT structure for a particular DVIPA will create an entry for that DVIPA in the coupling facility.) The stack will create and initialize, in the EZBEPORT structure, a sublist for this DVIPA of assigned ports for this stack. For more information on setting up EZBEPORT, refer to Setting up the sysplex environment for VTAM and TCP/IP functions in *z/OS Communications Server: SNA Network Implementation Guide*.

The stack also maintains a list of allowable ephemeral ports on this stack, which is basically any port number above 1023 that has not been reserved for TCP by a PORT or PORTRANGE statement. Only port number values in this list will be allocated for use by this stack for its SYSPLEXPORTS DVIPAs. Since this list is unique to a particular stack and determined by stack configuration, a port number that is not permissible for one stack because it has been reserved might be allowable for another stack, and could in fact be allocated for use by that stack for a SYSPLEXPORTS DVIPA.

Under the following conditions, the first time an application issues a TCP bind() with port 0 or a connect() request, TCP/IP will receive an unassigned group of ports from the coupling facility structure that are allowable as ephemeral ports on the stack (not otherwise reserved by PORT or PORTRANGE):

- The bind() or connect() request uses a distributed DVIPA as the source address (whether by the application explicitly binding the socket to the designated DVIPA or by the stack assigning the TCPSTACKSOURCEVIPA address).
- The distributed DVIPA is designated as SYSPLEXPORTS.

The stack will assign an ephemeral port from the obtained group as the source port for the TCP connection request, and the coupling facility structure will be updated to show the group of ports as assigned. The stack will assign ports from within this group for each subsequent equivalent bind() request. If all the ports within the group are used, the stack will obtain another group of ports from the coupling facility.

This means that the maximum number of simultaneously active outbound connections using sysplex-wide ephemeral port assignment is approximately 63000 for a particular distributed DVIPA, and is exactly equal to all port numbers between 1024 and 65535 that have not been reserved with a PORT or PORTRANGE configuration statement on all stacks at the same time. This is the same as for ephemeral port assignment within a single stack. That is, a single z/OS TCP stack supports no more than about 63000 simultaneously active, locally initiated TCP connections whose source ports are ephemeral ports assigned by the stack. If a stack is unable to successfully obtain an ephemeral port from the coupling facility for a SYSPLEXPORTS DVIPA, the connection request will be terminated with an error indication.

For information on diagnosing SYSPLEXPORTS problems, see *z/OS Communications Server: IP Diagnosis Guide*.

Timed affinities

Sysplex Distributor normally distributes each connection request, as it arrives to one of the candidate server instances, based on available capacity and policies in effect when the connection request arrives. In particular, each connection is assumed to be independent of all other existing and future connections, with respect to the server instance that will receive the incoming connection request.

Some applications, however, establish an affinity between a client and a particular server instance that needs to span multiple connections. TN3270 printer sessions, for example, are based on a connection request from a TN3270 client, and that printer session connection request needs to be routed to the same TN3270 server that is serving the LU2 session. Similarly, Web-based applications, such as shopping carts, might need to have all connections from a particular client come to the server instance that has the contents of the shopping cart stored as session state.

The TIMEDAFFinity parameter on the VIPADISTRIBUTE statement indicates to Sysplex Distributor that connections to a particular distributed DVIPA need to take into account the client origin. Connections from the same client, as identified by IP address, need to be routed to the same server instance, even when multiple server instances are hosted by a single target stack.

If a nonzero value for TIMEDAFFinity is specified on a VIPADISTRIBUTE statement, the first connection from a particular client is routed as normal to a target stack and listening application. At that time, both the Sysplex Distributor routing stack and the target stack establish an affinity to govern subsequent connection requests from the same client. This affinity maintains a connection count, initially one. As subsequent connection requests for the same distributed DVIPA and port come in from the same client IP address, they are routed to the same server instance and the affinity connection count is incremented. As affinity-based connections, including the first one, are closed, the connection count is decremented.

When the last existing connection is closed and the count gets to zero, a timer of the duration (in seconds) specified by the TIMEDAFFinity parameter is started. If another connection request is received from the same client to the same distributed DVIPA and port, the timer is stopped and the connection request is routed to the designated server instance. If no connection request is received from that client for the designated distributed DVIPA and port before the timer expires, the affinity is removed from the Sysplex Distributor routing and target stacks. The next connection request from that client for the distributed DVIPA and port will be routed according to normal Sysplex Distributor considerations of relative capacity and policies.

Using the VARY TCPIP,,OBEYFILE command with a profile containing a VIPADISTRIBUTE statement that specifies the TIMEDAFFINITY value as 0 prevents new affinities from being established. However, an existing affinity remains until all connections using the affinity end and the timer for the affinity expires. The timer value is not changed for an existing affinity by the new statement; instead, it continues to use the configured TIMEDAFFINITY value that was active when the affinity was established. To remove existing affinities, issue the VARY TCPIP,,OBEYFILE command with a profile containing a VIPADISTRIBUTE DELETE statement, followed by a VIPADISTRIBUTE DEFINE statement with a TIMEDAFFINITY value of 0.

Note that under some circumstances, a client's affinity with a specific target application server instance might be terminated prior to the specified time interval if key resources needed to satisfy new client TCP connection requests are not available. This includes the following scenarios:

- A target application server instance terminates, is quiesced for DVIPA Sysplex Distributor workload balancing, or is no longer listening on its specified port. Any affinities to the target application instance are terminated and new connections for that port are no longer routed to this server.

  One exception to this scenario occurs for configurations where multiple applications reside on the same system and TCP/IP stack, and also listen on the same port (that is, the set of applications comprises a shareport group). In this case, when one of the application instances in the shareport group terminates, is quiesced for DVIPA Sysplex Distributor workload balancing, or is no longer listening on the specified port, the routing stack continues to maintain any existing affinities, but only to this target system and TCP/IP stack, as long as at least one application instance in the shareport group is active, is not quiesced,

and is listening to the specified port. When a new TCP connection is received from a client that previously had an affinity to the server that is no longer active, the request is routed to the same target system and TCP/IP stack. One of the other available servers in the shareport group is selected to process the request and a new affinity is established.

- A target system or TCP/IP stack is no longer available. All affinities associated with applications running on that system and TCP/IP stack are terminated.
- A target TCP/IP stack is no longer reachable across dynamic XCF links (that is, the dynamic XCF link to the target stack is no longer active). Any existing client affinities to applications residing on that target stack are terminated if new connection requests from these clients are received while the dynamic XCF link is not active. Note that this scenario only occurs if a dynamic XCF link becomes inactive and all attempts to automatically restart the link are unsuccessful.
- A target server is found to be unresponsive to accepting new connection requests. All affinities to the target server are terminated and new connections for that port are no longer routed to that server.

Note that Sysplex Distributor's notion of *client* is tied exclusively to source IP address on the connection request. This is not new, and is also true of other functions such as Policy Agent and IPSec. However, it can present problems in situations where many different connections from truly different client instances all appear with the same source IP address. Examples include the following:

- Proxy applications, such as a Web HTTP server proxy, that initiate secondary connections on behalf of a large number of different clients. The secondary connections into Sysplex Distributor all look as though they come from the same client, and any affinity for those connections directs all connection requests to the same server.
- Firewall Network Address Translation (NAT), which keeps connection tables so it can map multiple client-side addresses into a single server-side source IP address.
- Clients on z/OS configured for SOURCEVIPA, or in general, many instances of a client on the same network node with one or very few IP addresses (as far as outbound connection requests are concerned).
- Clients running on the Sysplex Distributor routing stack, which is a special instance of the previous scenario. The source IP address for a client running on the Sysplex Distributor routing node is the distributed DVIPA, the same as the destination IP address. This is not new, and is true for any connection request issued for a server on the same stack, if the connection request socket was not bound to a particular IP address before issuing the connect() call.

In cases like this, distribution might be significantly less than optimal, because Sysplex Distributor has no way to distinguish among connection requests from different real clients all using the same source IP address. In the worst case, where all source IP addresses are the same (for example, where there is a single proxy instance in a firewall), there will be no load balancing at all as long as an affinity exists.

Essentially, timed affinity was created to allow connection workload distribution where it was not possible before, due to the client/server application model requiring a particular client to go to a particular server instance for some period of time. Such applications are not, at present, workload balanced. If a network configuration is such that a reasonable number of source IP addresses do not allow workload balancing, techniques to partition the work at the source must be implemented (configuring the clients to go to unique server addresses, or

configuring the proxy to spread the workload among multiple servers, as the WebSphere HTTP server plug-in does). Where the anticipated number of different source IP addresses, and the connection request arrival rate, is large enough to provide reasonable balancing and reaction to changing sysplex workloads, this provides a reasonable solution while respecting affinities between clients and servers as required for the application. The main reason for configuring the affinity on the basis of each distributed DVIPA and port pair is that affinity requirements differ from application (port number) to application, and some applications do not need affinity at all. You must take into account your specific network configuration, and specifically the arrival rate of connections from different IP addresses, in determining whether timed affinity with Sysplex Distributor is appropriate for a particular application in your network.

Affinity information needs to be handled on the Sysplex Distributor routing stack, backup stacks, and target stacks. If a target stack is not z/OS V1R5 or later, server application instances that are part of a shareport group on that stack will not work properly, and affinities for that target stack will not be communicated to the backup routing stack on failure of a primary routing stack. If a backup routing stack is not z/OS V1R5 or later, affinity information will not be sent to it from surviving target stacks if it takes over from a failed routing stack. Therefore, it is strongly recommended that all TCP/IP stacks participating in distribution for a distributed DVIPA with affinities be at least z/OS V1R5.

The Netstat commands (TSO NETSTAT, z/OS UNIX netstat, and MVS console DISPLAY TCPIP,,NETSTAT) have a VCRT (-V) DETAIL report to show the affinity related information for each connection, as follows:

```
MVS TCP/IP NETSTAT CS V1R7       TCPIP Name: TCPCS           15:10:28
Dynamic VIPA Connection Routing Table:
Dest:      203.1.10.18..21               (1)
  Source:  193.10.1.118..0
  DestXCF: 193.1.1.108
    CfgTimAff: 0200  TimAffCnt: 0000000003  TimAffLft: 0000
Dest:      203.1.10.18..21               (2)
  Source:  193.10.1.118..1026
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21               (2)
  Source:  193.10.1.118..1027
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21               (2)
  Source:  193.10.1.118..1028
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21               (3)
  Source:  193.10.1.119..0
  DestXCF: 193.1.2.108
    CfgTimAff: 0200  TimAffCnt: 0000000001  TimAffLft: 0000
Dest:      203.1.10.18..21               (4)
  Source:  193.10.1.119..1030
  DestXCF: 193.1.2.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21               (5)
  Source:  193.10.1.120..0
  DestXCF: 193.1.1.108
    CfgTimAff: 0200  TimAffCnt: 0000000000  TimAffLft: 0099
Dest:      204.2.10.11..21               (6)
  Source:  193.10.1.199..1031
```

```
  DestXCF: 193.1.6.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      205.2.10.11..21               (6)
  Source:  193.10.1.199..1032
  DestXCF: 193.1.6.108
    PolicyRule:    *NONE*
    PolicyAction:  *NONE*
```

The following notes apply to the preceding example.

Notes:

1. Affinity CRT entry for the three regular CRT entries that follow. If there is an affinity entry, it is shown before the regular CRT entries.
2. Three regular CRT entries, associated with the single preceding affinity entry.
3. Affinity CRT entry for the regular CRT entry that follows.
4. Regular CRT entry associated with the previous affinity entry.
5. An affinity CRT entry that has no connection associated with it. The use count is zero. There are 99 seconds affinity time left before this affinity entry is removed.
6. Regular CRT entry. There is no affinity associated with it.

Sysplex-wide security associations

To enable sysplex-wide security associations (SWSA) on a stack that has IP security enabled, add the subparameter DVIPSEC to the IPSEC keyword in the IPSEC statement block of the TCP/IP profile. To enable SWSA on a stack that has firewall support enabled, add the subparameter DVIPSEC to the FIREWALL parameter of the IPCONFIG statement. To take advantage of the functions described here, you must add this subparameter to your primary (including Sysplex Distributor hosts) and backup hosts. It is not necessary to add DVIPSEC to hosts that serve only as targets for Sysplex Distributor. For more information on configuring SWSA, see *z/OS Communications Server: IP Configuration Reference*.

SWSA also requires the use of a coupling facility structure, EZBDVIPA. For information on the setup and use of the EZBDVIPA coupling facility structure, see *z/OS Communications Server: SNA Network Implementation Guide*.

Dynamic IPSec security associations (SA), negotiated by IKE, can use a DVIPA address as the SA endpoint. Manually configured SAs are not supported by SWSA. For more information on IPSec, see Chapter 17, "IP security," on page 819 or refer to *z/OS Integrated Security Services Firewall Technologies*.

When using SWSA, there are two possible configurations to consider:

- DVIPA takeover
- Sysplex Distributor

To support IPSec in conjunction with DVIPA takeover and Sysplex Distributor, some IKE and IPSec configuration is required. Loss of access to the coupling facility is also discussed in the following subsections.

IPv6 sysplex-wide security associations are not supported.

For information on diagnosing SWSA problems, see *z/OS Communications Server: IP Diagnosis Guide*.

DVIPA takeover

When a DVIPA is moved during DVIPA takeover (planned or unplanned), SWSA automatically reestablishes new IPSec SAs with the same security service characteristics as the SAs that existed on the host that previously owned the DVIPA. The SA reestablishment is transparent to the client that owns the other end of the SA. That is, the SA reestablishment looks like a normal SA refresh. For example, as shown in Figure 40, during DVIPA takeover, DVIPA 192.168.253.4 is taken over by the backup host, and SAs are transparently reestablished between the client and the backup host.

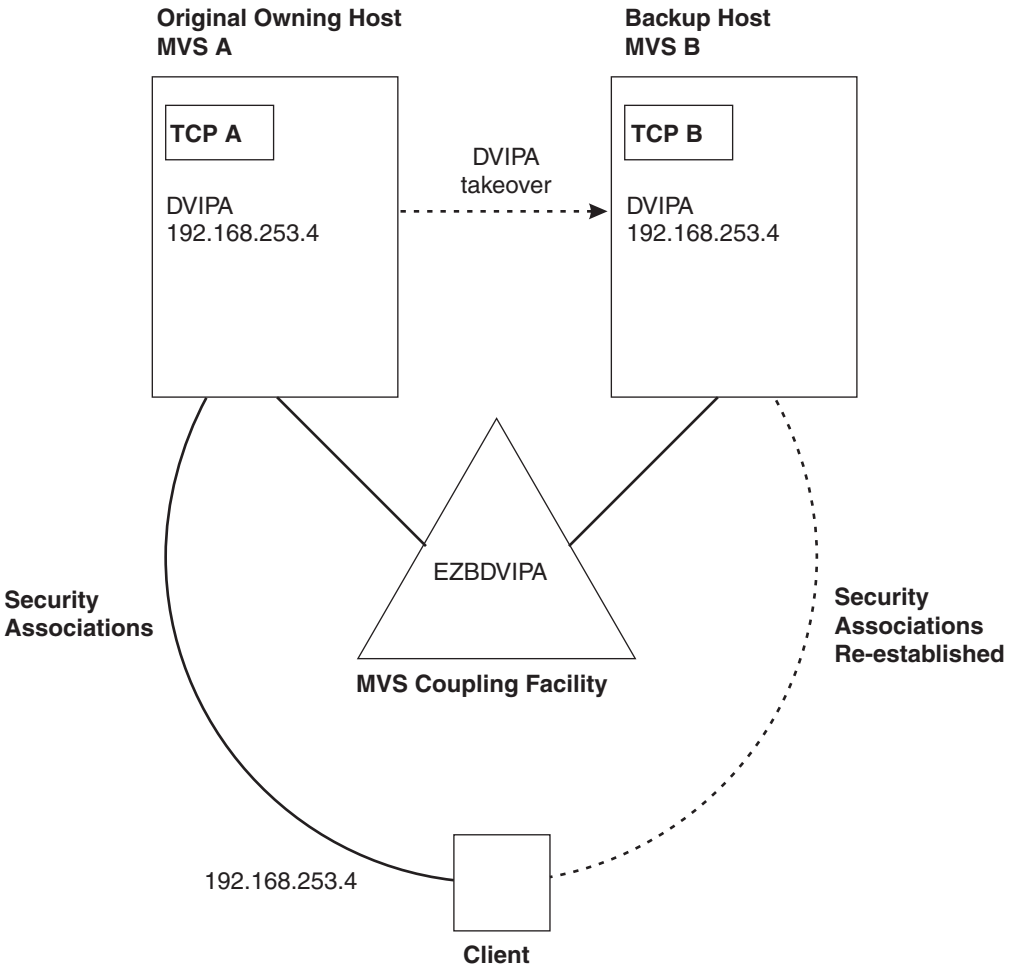


Figure 40. DVIPA takeover with SWSA

The IKE running on behalf of the TCP stack of the DVIPA owner is responsible for all IKE SA negotiations. The TCP stack owning the DVIPA is responsible for keeping the coupling facility updated with information needed to reestablish the SAs in the event of a DVIPA takeover. When a takeover occurs, the IKE on the backup host assumes responsibility for renegotiating new SAs based on the stored information read from the coupling facility during the takeover by the TCP stack of the new DVIPA owner.

Sysplex Distributor

TCP traffic protected by an IPSec SA with a sysplex-distributed DVIPA endpoint can be distributed to target hosts. IPSec cryptography for inbound traffic is performed on the target host whenever possible. If not possible, the distributor performs the cryptography before forwarding the packet to the target stack. IPSec

cryptography for outbound traffic is performed on the target host, and then sent directly into the network without being routed through the distributor. Figure 41 shows the target stack performing the cryptography for the inbound and outbound traffic.

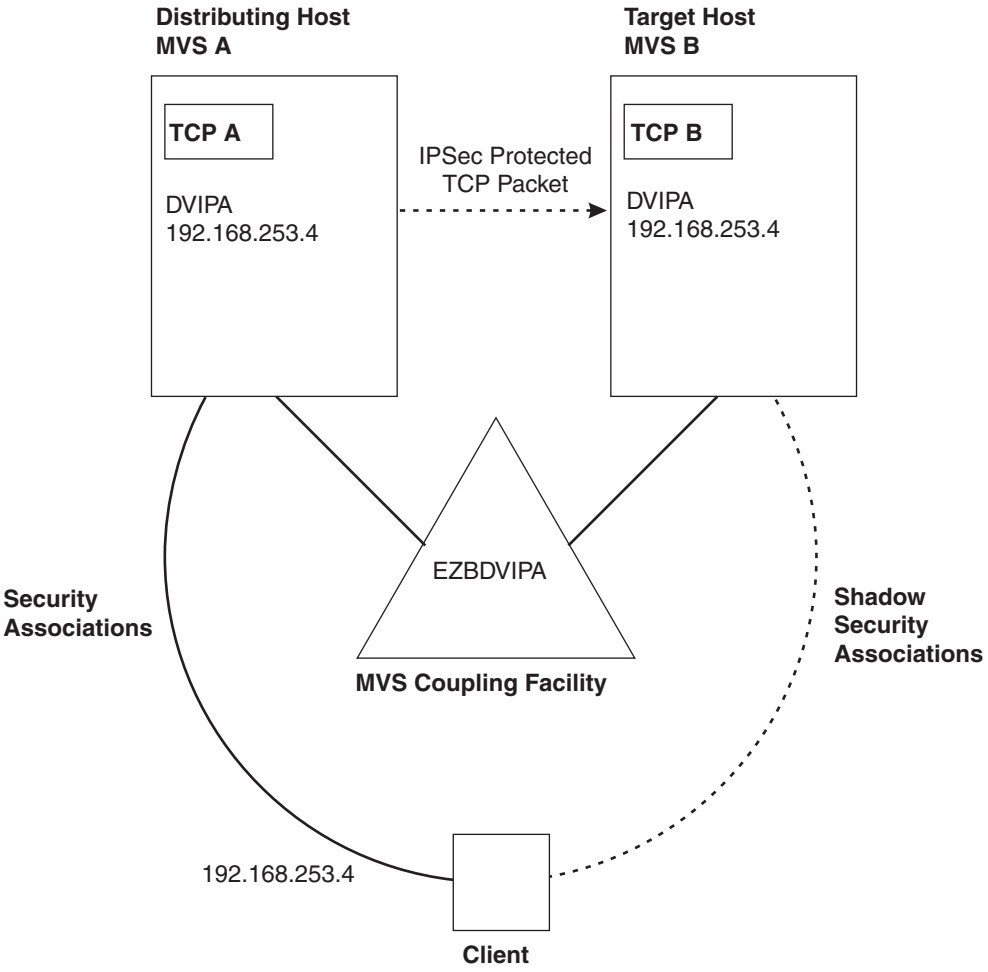


Figure 41. Sysplex Distributor with SWSA

The IKE running on behalf of the distributor TCP stack (the DVIPA owner) is responsible for all IKE SA negotiations. The distributor stack keeps the master copy of the SA associated with the DVIPA. Whenever a new SA is negotiated or refreshed and the SA is installed in the distributor stack, a copy (shadow) of the SA, which contains information necessary to perform IPSec cryptography, is sent within the sysplex to the target hosts. The shadow SAs enable the distribution of cryptography to the target stacks. The coupling facility is used as a central repository for SA replay protection sequence numbers used for outbound operations. The SA lifesizes (bytes sent and received over an SA) are maintained in the master SA.

Using IPSec with DVIPAs and Sysplex Distributor

To support IPSec in conjunction with DVIPA takeover and Sysplex Distributor, some IKE and IPSec configuration on the original or distributing host must be replicated onto all systems that can either serve as a backup host for a VIPA takeover or a target host for Sysplex Distributor. This includes IP Security policy that affects traffic using distributed DVIPA (from an IKE definition perspective).

- From a stack perspective, all anchor rules that are applicable to distributed DVIPA traffic must be identical on all systems. In addition, the ordering of the rules must allow for consistent application of security policy on all systems.
- To be considered a sysplex-wide SA, the SA negotiated that applies to DVIPAs must be at a granularity no coarser than host for the local address. That is, a dynamic SA cannot use a subnet or range that encompasses a DVIPA address. This rule ensures that on a DVIPA Giveback the SA can be moved from host to host without concerns about an SA being applicable to both the backup and primary host simultaneously. If such a dynamic SA is negotiated, the IPSec traffic using it cannot be distributed or recovered through the DVIPA takeover support.

Loss of access to coupling facility

If access is lost to the coupling facility containing the DVIPA structure EZBDVIPA, it is possible the TCP connections using this DVIPA could terminate and new connections needing IPsec will fail to establish. Loss of access could be caused by any of the following:

- A disconnect from the coupling facility structure.
- The structure is rebuilt.
- The structure encounters a critical storage shortage.

Loss of coupling facility access should affect only Sysplex Distributor connections that are being encrypted or authenticated. When access to EZBDVIPA is restored, the sessions can be re-established.

Resolution of dynamic VIPA conflicts

The same dynamic VIPA can exist on more than one stack in the sysplex, playing different roles on the different stacks. The TCP/IP stacks collaborate to prevent conflicting definitions. For example, at any given time only one stack will advertise a given dynamic VIPA to the routers.

Potentially conflicting dynamic VIPA definitions can arise during profile processing or as the result of changes within the sysplex due to a stack or application failure or as the result of movement of workload to a different stack. The following scenarios are examples of dynamic VIPA conflict resolution handled automatically by the TCP/IP stacks. For a summary of dynamic VIPA conflict identification and resolution, see “Dynamic VIPA creation results” on page 336.

Restart of the original VIPADEFINE TCP/IP after an outage

When a dynamic VIPA is defined using VIPADEFINE on one TCP/IP, and other stacks are designated as backup using VIPABACKUP statements for the same dynamic VIPA, the stack with the highest backup rank for that DVIPA will activate it if or when the VIPADEFINE stack fails.

If the failed stack is later restarted with the same VIPADEFINE profile statement, it is likely that connections to that DVIPA will exist on the backup stack that now has the DVIPA activated and advertised to the routers. How and when ownership of the DVIPA is returned to the restarted stack is determined by how the DVIPA was originally configured.

VIPADEFINE MOVEABLE IMMEDIATE

If the DVIPA is an IPv6 DVIPA, or is an IPv4 DVIPA that was originally configured with MOVEABLE IMMEDIATE, the following occurs:

- The DVIPA ownership is immediately transferred to the restarting stack which adds the DVIPA to its HOME list and the routers are dynamically notified. The restarted stack receives all new connections for that DVIPA. The stack also can receive packets for existing connections, and it routes these to the backup stack to preserve those connections.
- At the same time, the backup stack notifies the routers that it no longer is the owner of the DVIPA.
  - If there are no current connections to the DVIPA, it is removed from the HOME list on the backup stack and it reverts to backup status.
  - If there are any existing connections, the DVIPA remains in the HOME list of the backup stack and the DVIPA is put into *Moving* status until the last existing connection is terminated. At that time, the DVIPA is removed from the HOME list and reverts to backup status.

IBM recommends this form of a planned DVIPA takeback occur only during low periods of connection activity. This gives the attached routers time to update their routing tables and avoid connections being reset due to receiving an ICMP_HOST_UNREACH or ICMP6_DST_UNREACH from the router.

Tip: If OMPROUTE is used, it is recommended that GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN be configured. This causes DVIPA takeback to be delayed until OMPROUTE is active and able to advertise DVIPAs on the takeback stack. For more information on using DELAYJOIN, see "Sysplex problem detection and recovery" on page 376.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. MOVEABLE IMMEDIATE is the default for IPv4 DVIPAs, and the only behavior for IPv6 DVIPAs.

VIPADEFINE MOVEABLE WHENIDLE

If an IPv4 DVIPA was originally configured with MOVEABLE WHENIDLE, the following occurs:

- If it appears that there are no active connections to the DVIPA on the backup stack:
  - The DVIPA is removed from the HOME list on the backup stack and reverts to backup status.
  - The restarted stack assumes ownership of the DVIPA by adding it to its HOME list and notifying the routers.
- If there are existing connections to the DVIPA on the backup stack:
  - Ownership of the DVIPA remains with the backup stack. The DVIPA on the restarting stack is placed in backup status at the head of the backup list for the DVIPA.
  - The backup stack periodically checks to see if it has any active connections to the DVIPA.

    When or if it appears that there are no active connections for the DVIPA, the following occurs:
    - The DVIPA is removed from the HOME list on the backup stack and reverts to backup status.
    - The restarted stack assumes ownership of the DVIPA by adding it to its HOME list and notifying the routers.

Notes:

1. WHENIDLE is supported only for IPv4 DVIPAs.
2. A small period of time exists between the check for connections and the movement of the dynamic VIPA to the restarted stack. If connections are made to the old host (the backup stack) in this interval, they will be broken.
3. During the time that TCP/IP is periodically checking for connections, TCP/IP does *not* refuse new connections because this would be the same as an outage. If moving the work back to the restarted stack is more important than maintaining uninterrupted service to all clients, then the system operator can use the VARY TCPIP,,OBEYFILE command to delete the dynamic VIPA on the backup stack with the VIPADELETE profile statement. This causes the restarted stack to immediately activate the DVIPA. Optionally, the data set specified on the command can contain a VIPABACKUP statement following the VIPADELETE statement. This will restore the stack as a backup stack.

Movement of unique application-instance (BIND)

A dynamic VIPA is created when any application binds to a nonexistent, specific IP address falling within a configured VIPARANGE on that stack.

In the case of a stack failure, the same application could be started on another stack and (assuming the new stack also has an appropriate VIPARANGE configured) when the application binds to the same IP address, the dynamic VIPA is created on the second stack. Future client connections to that IP address are routed to the second stack where the application is now running.

However, if the same (or a different) application is started on a second stack and attempts to create the same dynamic VIPA using a bind() while it exists on the first stack, the end result is determined by how the VIPARANGE was configured on the stack where the first bind() occurred.

VIPARANGE (DEFINE) MOVEABLE NONDISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE NONDISRUPTIVE, the following occurs:

- If the RACF profile EZB.BINDDVIPARANGE.*sysname.tcpname* has been defined, the application issuing the bind() must be permitted to the profile.

  Note: If the RACF profile is not defined but the same VIPARANGE is configured on another stack, any application on that stack can cause the DVIPA to move there by issuing a BIND() to that DVIPA.
- The DVIPA ownership is immediately transferred to the second stack which adds the DVIPA to its HOME list and dynamically notifies the routers. This stack will now receive all new connections for the DVIPA.
- At the same time, the first stack notifies the routers that it no longer is the owner of the DVIPA, and puts the DVIPA into *moving* status. The DVIPA remains in *moving* status (and in the first stacks HOME list) until the application closes the socket.
- Existing connections on the first stack are preserved. If the second stack receives packets intended for existing connections, it routes the packets to the first stack.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. NONDISRUPTIVE is the default for V2R10 and later, and is the only option supported for IPv6.
3. Both stacks must be running V2R10 or later to get non-disruptive behavior. If either stack is running V2R8, the result will be as described in "VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE."

VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE DISRUPTIVE (or if either stack is running V2R8), the following occurs:

- The bind() request for the application on the second stack will fail.
- The DVIPA on the first stack is not affected.

Notes:

1. If movement of the application from the first to the second stack is intended, the application must be ended on the first stack before it is started on the second stack.
2. DISRUPTIVE is supported only for IPv4 DVIPAs.

Defining a RACF profile for VIPARANGE

A RACF SERVAUTH class profile can control which applications can bind() to define a DVIPA using VIPARANGE. To restrict applications that can bind() to a DVIPA within a VIPARANGE, define the EZB.BINDDVIPARANGE.*sysname*.*tcpname* profile in RACF, and permit the applications that are allowed to bind() as shown in the following example. If the profile is not defined, then the bind() will be processed.

To restrict access to all the IP addresses in the VIPARANGE definitions, you can define a RACF profile using the following example:

```
RDEFINE SERVAUTH (EZB.BINDDVIPARANGE.sysname.tcpname)
   UACC(NONE)

PERMIT EZB.BINDDVIPARANGE.sysname.tcpname
   ACCESS(READ) CLASS(SERVAUTH) ID(userid)
```

In the example above, *sysname* is the name of the MVS system, *userid* is the user ID that the application is running on, and *tcpname* is the job name of the TCP/IP started task.

The job name for started tasks, such as TCP/IP, is derived depending on how it is started:

- If the START command is issued with the name of a member in a cataloged procedure library (for example, S TCPIPX), the job name will be the member name (for example, TCPIPX).
- If the member name on the START command is qualified by a started task identifier (for example, S TCPIPX.ABC), the job name will be the started task identifier (for example, ABC). The started task identifier is not visible to all MVS components, but TCP/IP uses it to build the RACF resource name.
- The JOBNAME parameter can also be used on the START command to identify the job name (for example, S TCPIPX,JOBNAME=XYZ).
- The JOBNAME can also be included on the JOB card.

If this RACF profile is created, the user ID must be permitted to access this profile, or bind() requests will fail with a permission denied error, regardless of superuser authority.

Note also that before the RACF profiles take effect, a refresh of these profiles might be required. This can be accomplished by the following RACF command:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Movement of a unique APF-authorized application instance (ioctl)

APF-authorized applications running under a user ID with SuperUser authority have the ability to activate a dynamic VIPA with the SIOCSVIPA or SIOCSVIPA6 ioctl() either within the application itself or by invoking the MODDVIPA utility. Because this is a controlled environment, it is assumed configuration errors are minimized or avoided and the usage is correct. Thus, even if the requested DVIPA is currently active on another TCP/IP stack via BIND() or ioctll(), the DVIPA will be immediately activated on this stack. What happens on the other stack is determined by how the VIPARANGE was configured on that stack.

VIPARANGE (DEFINE) MOVEABLE NONDISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE NONDISRUPTIVE, the following occurs:

- The DVIPA ownership is immediately transferred to the second stack which adds the DVIPA to its HOME list and dynamically notifies the routers.
- At the same time, the first stack notifies the routers that it no longer is the owner of the DVIPA, and puts the DVIPA into *moving* status. The DVIPA remains in *moving* status (and in the first stack's HOME list) until the DVIPA is deleted on that stack with the VIPADELETE profile statement or the SIOCSVIPA or SIOCSVIPA6 ioctl DELETE option.
- Existing connections on the first stack are preserved. If the second stack receives packets intended for existing connections, it will route the packets to the first stack.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. NONDISRUPTIVE is the default for V2R10 and later.
3. Both stacks must be running V2R10 or later to get non-disruptive behavior. If either stack is running V2R8, the result will be as described in "VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE."

VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE DISRUPTIVE (or if either stack is running V2R8), the following occurs:

- The ioctl request for the application on the second stack succeeds. The DVIPA is added to the HOME list on the second stack, and the routers are dynamically notified.
- The DVIPA on the first stack is deleted.

Notes:

1. Any existing connections to the DVIPA on the first stack are broken.
2. DISRUPTIVE is IPv4 only.

Same dynamic VIPA for VIPADEFINE and BIND(), SIOCSVIPA or SIOCSVIPA6 ioctl, or MODDVIPA utility

Regardless of careful implementation, it is possible that the same IP address is inadvertently selected for VIPADEFINE and for use with BIND(), SIOCSVIPA or SIOCSVIPA6 ioctl, or the MODDVIPA utility. Because the application scenarios are quite different, this must be an error.

If this duplicate DVIPA address conflict occurs on the same TCP/IP, the second attempt might fail. If an IP address is specified in a VIPADEFINE, and that same IP address has already been activated on the TCP/IP by an application using BIND(), the SIOCSVIPA or SIOCSVIPA6 ioctl, or the MODDVIPA utility, the VIPADEFINE will be rejected during VARY TCPIP,,OBEYFILE command processing. If an IP address is activated with VIPADEFINE, and the application does a BIND(), ioctl(), or the MODDVIPA utility is used, the BIND() will succeed, but the ioctl() will fail with a nonzero errno and the MODDVIPA utility will set a nonzero condition to indicate that the IP address already exists.

The same situation could also occur on two different TCP/IPs in the sysplex. Because the TCP/IPs are exchanging information among themselves, if the two attempts are far enough apart in time, the second attempt will be caught immediately and rejected. However, it is possible that the attempt will be made almost simultaneously on two different TCP/IPs, such that neither TCP/IP is yet aware of the attempt on the other TCP/IP. If both attempt such an activation, and the exchange of information then shows a conflict, the internal sysplex time stamps are used to determine which attempt was really first. The one that appears to be first is allowed to continue, and the dynamic VIPA is deleted from the *later* TCP/IP. While such a simultaneous attempt is somewhat unpredictable in respect to which one will succeed, the dynamic VIPA will remain active on only one TCP/IP, and examination of messages will indicate which TCP/IP successfully created the DVIPA and on which TCP/IP it was rejected.

Dynamic VIPA creation results

Table 17 on page 337 summarizes the results of attempting to create a dynamic VIPA when it (or the same IP address for HOME statement) already exists in the sysplex.

Table 17. Summary of dynamic VIPA creation results

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
bind()	bind()	Second bind() succeeds, but no new VIPA is created.	If both stacks are running V2R10 or later, and the first BIND DVIPA was created with MOVEABLE NONDISRUPTIVE: • On stack 2, bind() succeeds • On stack 1, the BIND VIPA remains in the HOME list (unadvertised) and any existing connections are preserved • New connections to that IP address go to the application on stack 2. Otherwise, second bind fails.
bind()	ioctl()	ioctl() fails with warning condition code, but the application associated with the ioctl is still able to use the dynamic VIPA.	ioctl() succeeds, bind is deleted (even if BIND DVIPA was created as MOVEABLE NONDISRUPTIVE)
bind()	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
bind()	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
bind()	HOME	See note.	See note.
ioctl()	bind()	bind() succeeds, no new VIPA is created.	bind() fails.
ioctl()	ioctl()	Second ioctl() fails with warning condition code, but the application associated with the ioctl is still able to use the dynamic VIPA.	Second ioctl() succeeds. If both stacks are running V2R10 or later, and the ioctl DVIPA on stack 1 was created with MOVEABLE NONDISRUPTIVE, the DVIPA on stack 1 remains in the HOME list (unadvertised) and any existing connections are preserved. Otherwise, the ioctl DVIPA on stack 1 is deleted and any existing connections are broken.
ioctl()	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
ioctl()	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
ioctl()	HOME	See note.	See note.
VIPADEFINE	bind()	bind() succeeds, but no new VIPA is created.	bind() fails.

Table 17. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
VIPADEFINE	ioctl()	ioctl() fails.	ioctl() fails.
VIPADEFINE	VIPADEFINE	If the second VIPADEFINE statement is an exact duplicate of the first, the second VIPADEFINE is ignored with no error message. For IPv4, the second VIPADEFINE fails if it has different options or a different mask than the first VIPADEFINE specified. For IPv6, the second VIPADEFINE fails if the interface name is already defined with a different address or the address is already defined for a different interface name.	For IPv4, the second VIPADEFINE succeeds but activation on stack 2 might be deferred. If both stacks are running V2R10 or later, and the DVIPA was created on stack 1 as MOVEABLE IMMEDIATE: • Second VIPADEFINE is activated immediately • Any connections to the DVIPA on stack 1 are preserved. (DVIPA stays in HOME list unadvertised) Otherwise, the second VIPADEFINE activation is deferred until there are no connections on stack 1, at which point, stack 1 reverts to backup status. For IPv6, if both the interface name and address on the second VIPADEFINE match the first VIPADEFINE, the DVIPA is activated on stack 2. Any connections to the DVIPA on stack 1 are preserved. If either the interface name or address is the same but the other is not, the second VIPADEFINE fails.
VIPADEFINE	VIPABACKUP	VIPABACKUP fails.	Both succeed.
VIPADEFINE	HOME	See note.	See note.
VIPABACKUP	bind()	bind() fails.	If the IP address is already active on the bind() stack, the bind() will succeed. Otherwise, the bind() fails.
VIPABACKUP	ioctl()	ioctl() fails.	ioctl() fails.

Table 17. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
VIPABACKUP is backup status	VIPADEFINE	VIPADEFINE succeeds, replaces the VIPABACKUP. For IPv6, both the interface name and the address on the VIPADEFINE must match the VIPABACKUP. If one matches and the other does not match, the VIPADEFINE fails.	VIPADEFINE succeeds. For IPv6, both the interface name and the address on the VIPADEFINE must match the VIPABACKUP. If one matches and the other does not match, the VIPADEFINE fails.
VIPABACKUP in active status (after takeover)	VIPADEFINE	VIPADEFINE rejected	For IPv4, the VIPABACKUP DVIPA is MOVEABLE IMMEDIATE or WHENIDLE depending how the original VIPADEFINE DVIPA was created. If both stacks are running V2R10 or later, and the VIPABACKUP DVIPA is MOVEABLE IMMEDIATE: • The VIPADEFINE is activated immediately. • Any connections to the DVIPA on stack 1 are preserved (DVIPA stays in HOME list unadvertised). • When there are no more connections, stack 1 reverts to backup status. Otherwise, the VIPADEFINE activation on stack 2 is deferred until there are no active connections on stack 1, at which point stack 1 reverts to backup status. For IPv6, if both the interface name and the address on the VIPADEFINE match the VIPABACKUP, the DVIPA is activated on stack 2. Any connections to stack 1 are preserved. If only one matches, the VIPADEFINE fails.

Table 17. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
VIPABACKUP	VIPABACKUP	If the DVIPA is in backup status on this stack, the second VIPABACKUP succeeds. If the DVIPA is in active status on this stack (for example, after a takeover), a VIPABACKUP with a different rank will be rejected. For IPv6, both the interface name and address must match the first VIPABACKUP. If one matches and the other is different, the second VIPABACKUP fails.	Second VIPABACKUP succeeds. For IPv6, both the interface name and address must match the first VIPABACKUP. If one matches and the other is different, the second VIPABACKUP fails.
VIPABACKUP	HOME	See note.	See note.
HOME	bind()	bind() succeeds, but no new VIPA is created.	bind() fails.
HOME	ioctl()	ioctl() fails.	ioctl() fails.
HOME	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
HOME	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
Note: Defining the same IP address in the HOME list as an existing dynamic VIPA will not be rejected by the TCP/IP stack, but it is likely to cause routing problems.			

IPv6 considerations

This section discusses special considerations for IPv6.

VIPARANGE

MOVeable DISTRUPTive is not supported for IPv6. To prevent an unauthorized application from creating a dynamic VIPA with a bind() call, a RACF profile (called EZB.BINDDVIPARANGE.*sysname*.*tcpname*) can be added. If the RACF check fails, the bind() request fails, and if the RACF check is successful, the bind() request is accepted.

VIPADEFINE and VIPABACKUP

VIPABACKUP must specify an interface name and IPv6 address. The interface name and IPv6 address pair must match any preexisting dynamic VIPA definitions for either that interface name or IPv6 address. If using VIPABACKUP for an initial DVIPA activation, neither the interface name nor the IPv6 address can preexist. Furthermore, for the takeover and backup functions to succeed, if a corresponding VIPADEFINE is to be subsequently activated, the interface name and IPv6 address on the VIPADEFINE statement must match the VIPABACKUP statement.

Unique application-instance scenario and IPv6-enabled applications

A single instance of an IPv6-enabled TCP application can typically handle communications with IPv4 or IPv6 partner applications. IPv6-enabled TCP server applications usually accomplish this by using a single AF_INET6 socket to accept connections from both IPv4 and IPv6 clients. As a result, IPv6-enabled applications have some additional configuration considerations when compared to IPv4-only applications, as they require both an IPv4 DVIPA and an IPv6 DVIPA to be configured. If the application is using the MODDVIPA utility or the SIOCSVIPA IOCTL to define the DVIPA, it needs to be modified as follows:

- If using the MODDVIPA utility, an additional set of invocations for this utility is needed to define and delete the IPv6 DVIPA to be associated with the application.
- If the application is using the SIOCSVIPA IOCTL, it needs to be modified to also issue the SIOCSVIPA6 IOCTL to define and delete the IPv6 DVIPA to be associated with the application.

Applications that are currently relying on defining the DVIPA dynamically using the bind() socket API or the BIND parameter on the PORT reservation statement have some additional considerations, as a single socket cannot be bound to more than one IP address.

Following are some of the options that can help alleviate this issue:

- Instead of relying on the bind to trigger the DVIPA definition, the application or configuration can be modified to allow the application to bind to IN6ADDR_ANY. The MODDVIPA utility can then be used to define and delete both the IPv4 and IPv6 DVIPAs associated with the application. Another similar alternative would be for the application to issue the SIOCSVIPA or SIOCSVIPA6 IOCTL for both of these DVIPAs, but this is probably less desirable since it involves application changes.
- Start two instances of the application, one binding to the IPv4 DVIPA and the other binding to the IPv6 DVIPA. This assumes that multiple instances of the same application can be started within the same MVS image or across different images in the sysplex. This configuration allows IPv4 clients to reach the application instance bound to the IPv4 DVIPA and IPv6 clients to reach the application instance bound to the IPv6 DVIPA.
- In some cases it might be possible to modify the application so that it can create two sockets, one for IPv6 communications and one for IPv4 communications. Each socket could then be bound to the IPv4 or IPv6 DVIPA as appropriate, and a single instance of the application could handle both IPv4 and IPv6 clients. This solution might be less desirable since it also involves application code changes.

Other considerations

The following sections describe other considerations you should understand regarding dynamic VIPA support.

Dynamic VIPAs, OSA-Express QDIO, and Spanning Tree Protocol

Spanning Tree Protocol (STP) can potentially impact environments where both OSA-Express QDIO and dynamic VIPA are deployed.

With dynamic VIPA, a TCP/IP address takeover occurs when either a TCP/IP stack fails or a dynamic VIPA is manually moved by operator intervention. In either case, when a dynamic VIPA moves, the IP address and respective workloads are taken over by another TCP/IP stack through other OSA-Express Ethernet devices running on a different server. When the original TCP/IP stack and respective OSA-Express QDIO devices are returned to operation, both the IP address and respective workload traffic are taken back by the recovered TCP/IP stack.

If the network bridge or switch is not configured properly, packets can get lost in the network or be blocked by the networking equipment. This is a result of a physical looping condition identified by the STP, or expired OSA-Express QDIO timers due to the increased latencies associated with blocked ports or delayed packets. In these cases, dynamic VIPA connectivity can fail.

When configuring STP, use care in the bridge or switch configuration to avoid or minimize potential loop conditions. For example, if the STP is not an integral component of the overall network, disabling the STP on all of the Virtual LANs (VLANs) that connect to OSA-Express QDIO devices will help avoid the problem.

Also, some networking switches provide a mechanism for suppressing the STP's listening and learning states on specific switch ports. For example, Cisco Systems provides an STP configuration feature called PortFast that can place specific switch ports into forwarding state as soon as a link is detected. Without PortFast enabled, a switch port has to transition through the listening and learning stages (30 seconds total) of Spanning Tree reconvergence before the switch port can actually pass valid traffic. For PortFast capable Cisco switches, enabling PortFast on all switch ports that OSA-Express QDIO is connected into allows network administrators to both preserve their original Spanning Tree configuration by not having to change it, while also providing a viable mechanism to avoid potential dynamic VIPA problems.

For more information on configuring STP and immediate port forwarding, refer to the bridge or switch operational manual.

Mixture of types of dynamic VIPAs within subnets

Any particular IP address can be used in only one way as a dynamic VIPA. As described in previous sections, a dynamic VIPA can be defined either via VIPADEFINE or by application action within a valid VIPARANGE, but not both. However, within a subnet defined as a VIPARANGE, some IP addresses can be used for VIPADEFINE, and others may be assigned to unique application instances, without conflict, as long as the limit of a total of 1024 active and backup dynamic VIPAs on a single TCP/IP is not exceeded. TCP/IP will make no attempt to reject a VIPADEFINE dynamic VIPA that also falls within a VIPARANGE. This allows installations with limited availability of IP addresses to assign individual addresses to either application scenario, without having to define separate subnets and use up additional IP addresses in that manner.

MVS failure and sysplex failure management

The TCP/IPs in a sysplex use MVS XCF messaging to exchange information about dynamic VIPAs. When a TCP/IP fails or is ended by operator command, but the underlying MVS remains active, the other TCP/IPs are immediately notified, and takeover of VIPADEFINE dynamic VIPAs is automated and very fast. In addition, each TCP/IP stack monitors for local problem conditions that might cause it to

leave the TCP/IP sysplex group, at which point the other TCP/IPs will be immediately notified. For more information, see "Sysplex problem detection and recovery" on page 376.

However, when an MVS fails, there is normally an operator message on the console requiring a response (WTOR). Until this response is made by an operator or automation, the other MVS systems do not notify the remaining TCP/IPs in the sysplex of the failure of the TCP/IP on the failing MVS. This can delay automated backup of dynamic VIPAs defined with VIPADEFINE. Sysplex Failure Management (SFM) can be used to automate the required response to the console message of the failing MVS. Refer to *z/OS MVS Setting Up a Sysplex* for information on how to set up SFM to avoid the requirement for a manual response and speed backup of dynamic VIPAs defined with VIPADEFINE.

For more information, refer to *z/OS Communications Server: IP Diagnosis Guide*.

Applications and dynamic VIPAs

While most applications support multiple instances in a sysplex, very few applications expect IP addresses to move around under them. TCP applications use TCP connections to form a relationship between particular client and server instances to exchange data over an extended period. They rely on notification of TCP connection termination to initiate recovery and to reestablish a new relationship (possibly from a client to a different server). Conversely, most UDP applications do the equivalent function at the application layer. Movement of an IP address to a different server could be confusing to the client, unless the new server also is aware of the state of the client work.

UDP applications whose interactions consist of atomic interactions (a single request followed by one or more responses, with no state information maintained at the server between requests) can use dynamic VIPAs in the multiple application-instance scenario. However, if the server application maintains state information between interactions (for example, NFS), then moving a dynamic VIPA to another server might not work unless the client/server application protocol can detect the discontinuity. In that case, the unique application-instance scenario might apply, which would require the restart of the server instance on another TCP/IP.

In addition, the following types of work are not appropriate for distribution with distributed dynamic VIPA:

- Applications that establish affinity with a particular TCP/IP stack, such as SNMP.
- FTP servers that receive the PASV or EPSV command for a distributed DVIPA that did not specify SYSPLEXPORTS. The PASV and EPSV commands are supported when SYSPLEXPORTS was specified on the VIPADISTRIBUTE statement of the distributed DVIPA that is the destination IP address being used by the FTP server. These commands request that the FTP server bind() on a data port that is not the default data port, or the one specified on the VIPADISTRIBUTE statement, and to wait for a connection rather than initiate one on receipt of a transfer command (for example, RETR). Because SYSPLEXPORTS cannot be specified for a non-distributed dynamic VIPA, passive mode FTP connections made to a non-distributed dynamic VIPA cannot be recovered when the VIPA is moved through a planned takeover. The control connection remains on the original stack, but attempts to create new data connections for control connections that existed prior to the move will fail. When SYSPLEXPORTS is used with a distributed dynamic VIPA, new data connections

are always sent to the same stack containing their control connection. For more information on using the SYSPLEXPORTS parameter in the VIPADYNAMIC block, refer to *z/OS Communications Server: IP Configuration Reference*.

- FTP servers that accept connections on an IPv6 distributed DVIPA that does not have SYSPLEXPORTS specified. The FTP server expects an EPSV command for data transfers to or from an IPv6 address, and use of the EPSV command is supported only when SYSPLEXPORTS was specified on the VIPADISTRIBUTE statement of the distributed DVIPA that is the destination IP address being used by the FTP server.

Configuring VIPAs for activation with VIPABACKUP

When adding a dynamic VIPA to a functioning sysplex, the normal order is to add VIPABACKUP statements to backup stacks, and then add a VIPADEFINE statement to the TCP/IP stack where the VIPA should normally reside, at which time the VIPA would be activated and ready for use. A VIPADISTRIBUTE statement could be added at the same time, along with the VIPASMPARMS statement if the SERVICEMGR parameter is coded on the VIPADEFINE statement. The VIPA would only be activated when the VIPADEFINE was processed on the TCP/IP that would normally own it, and not when a VIPABACKUP is processed on another stack. This is, in part, because parameters on the VIPADEFINE that are needed to activate the VIPA were not found on the VIPABACKUP statement.

However, on the rare occasion that a disaster occurs, it might be necessary to IPL all of the systems in a sysplex. Assuming that many dynamic VIPAs are in use and the VIPADEFINE statements are spread across the available TCP/IP stacks in the sysplex, most of the dynamic VIPAs have a lengthy wait before the owning operating system, TCP/IP, and application are started and fully operational. The intent of dynamic VIPAs defined with VIPADEFINE and VIPABACKUP is to move the dynamic VIPAs under a functioning application as soon as possible. Therefore, optional parameters have been added to the VIPABACKUP statement to allow the dynamic VIPA to be activated when the VIPABACKUP is processed at TCP/IP initialization or VARY TCPIP,,OBEYFILE command processing. If the MOVEABLE IMMEDIATE or MOVEABLE WHENIDLE parameter is specified on a VIPABACKUP profile statement, along with required address mask and optional SERVICEMGR parameters for an IPv4 DVIPA, the dynamic VIPA will be activated when the profile statement is processed, if it is not active elsewhere in the sysplex already. The following notes apply to this case:

- If the MOVEABLE parameter is specified, the address mask must also be specified. For an IPv4 DVIPA, the SERVICEMGR parameter on a VIPABACKUP statement is optional. The address mask and SERVICEMGR parameters can only be specified on a VIPABACKUP statement if the MOVEABLE parameter has also been specified.
- The first started VIPABACKUP TCP/IP stack that is configured to start a particular dynamic VIPA with VIPABACKUP will activate the dynamic VIPA and own it until the VIPADEFINE stack is started. As long as the first activating VIPABACKUP stack remains operational, starting another stack with a VIPABACKUP statement will not cause the dynamic VIPA to move, even if the second stack has a higher backup rank. Only the initialization of the VIPADEFINE stack will cause the VIPA to move automatically.
- When the VIPADEFINE is eventually processed, its parameters will override the values specified on the VIPABACKUP that activated the dynamic VIPA, if the VIPADEFINE has parameters that are different from those specified on the VIPABACKUP.

It is always a good idea, when able to specify a parameter on both primary and backup, to specify exactly the same values for each. When all the parameters common to VIPADEFINE and VIPABACKUP are specified the same on both for a particular DVIPA, the behavior exhibited when the VIPADEFINE stack comes up will be the same as when the DVIPA is activated first by a VIPABACKUP stack. However, for IPv4, there are several parameters that could be specified inconsistently on the VIPABACKUP and VIPADEFINE statements. If they are specified inconsistently, following are some of the implications by parameter:

- SERVICEMGR: If the Cisco forwarding agents are not configured to participate, then it does not matter whether or not SERVICEMGR is specified. If the Cisco forwarding agents are configured to participate, the following apply if SERVICEMGR is specified differently on participating stacks:
  - If SERVICEMGR is not specified on the VIPABACKUP stack, the Cisco forwarding agents will send all packets through normal IP forwarding to the stack owning the DVIPA (the VIPABACKUP stack until the VIPADEFINE stack comes up). When the VIPADEFINE stack comes up with SERVICEMGR, it will inform the forwarding agents and normal function will continue, even for existing connections. In other words, nothing fails, including existing connections, but optimal routing is not achieved unless all participating stacks code SERVICEMGR.
  - If SERVICEMGR is specified on the VIPABACKUP stack, things will operate normally for MNLB integration until the VIPADEFINE stack is activated. When that happens, no new updates will be sent to the Cisco forwarding agents, meaning that existing connections will be routed optimally for a while, but all packets for new connections will be sent to the VIPADEFINE stack for routing. After the 15-minute timeout for connection routes in the forwarding agents, the existing connections will have all packets sent to the VIPADEFINE stack. Once again, nothing fails, including existing connections, but optimal routing is not achieved unless all participating stacks code SERVICEMGR.
- MOVEABLE: If there is any doubt, this parameter should be coded as MOVEABLE IMMEDIATE on the VIPABACKUP stack, so that the stack will maintain connection information to be sent to the VIPADEFINE stack when it comes up. Subsequent operation in the sysplex will be determined by the coding (or defaulting) of MOVEABLE on the VIPADEFINE statement, regardless of what was coded on the VIPABACKUP of the stack that first activated the DVIPA. This is true even if that same stack later takes over the DVIPA from the VIPADEFINE or another stack. (This is the reason MOVEABLE was chosen as the way to activate the new function on VIPABACKUP stacks, rather than a new keyword, to force consideration of the setting of the MOVEABLE parameter for VIPABACKUP stacks.)
- address_mask: The address_mask is used to build BSDROUTINGPARMS statements. If the attached routing daemon is OMPROUTE and there is no corresponding interface definition with the subnet mask parameter in OMPROUTE configuration, OMPROUTE might send a different address mask from the VIPADEFINE stack than it will for the VIPABACKUP stack, which might confuse the routing network. The address_mask should be the same on VIPABACKUP and VIPADEFINE statements for a particular DVIPA.

  Result: Incorrect OMPROUTE configuration, such as missing interface definitions for RIPv1, RIPv2, or OSPF, can lead to unexpected results, especially if OMPROUTE uses defaults for the address mask.

Assuming that dynamic VIPAs are spread across the sysplex, and that critical applications are as well, the first stack to be activated might activate most, if not

all, DVIPAs, and might therefore assume a considerable amount of the workload. For any particular application, all of the workload is directed to the first stack supporting that application to be activated, regardless of what other workload will be executing on that same stack or LPAR.

Recommendation: Consider the planned sysplex TCP/IP activation sequence, and have only those DVIPAs for critical applications activating on TCP/IP stacks early in the sequence, so that less important work does not interfere with the business-critical applications during sysplex startups.

Example of configuring dynamic and distributed VIPAs

The TCP/IP profiles needed to implement dynamic VIPA (DVIPA) on multiple systems in a sysplex are shown in the following examples. The VIPADEFINE and VIPABACKUP statements allow automatic dynamic VIPA takeover to occur if needed (see "Configuring the multiple application-instance scenario" on page 309), and the VIPARANGE statements allow dynamic VIPAs to be dynamically created by an application or by the MODDVIPA utility (see "Configuring the unique application-instance scenario" on page 310). The VIPADISTRIBUTE statements allow a single VIPA to be shared among several TCP/IPs. Including the SOURCEVIPA and TCPSTACKSOURCEVIPA parameters on the IPCONFIG and IPCONFIG6 statements, on each target stack with the same dynamic VIPA specified, enables a single DVIPA address to be used as a sysplex-wide source DVIPA address for outbound TCP connections. The following examples show both IPv4 and IPv6 DVIPAs, and the output is shown in the IPv6-enabled, or long, format.

```
TCPCS
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.1 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0001 INTFID 6:7:8:9
 SOURCEVIPAINT SVIPA1 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE 255.255.255.240 201.2.10.11 201.2.10.12
  VIPADEFINE 255.255.255.240 201.2.10.14 201.2.10.15
| VIPADEFINE 255.255.255.240 201.2.10.23
| VIPADISTRIBUTE SYSPLEXPORTS DISTM SERVERWLM 201.2.10.11
   PORT 20 21 DESTIP ALL
  VIPADISTRIBUTE 201.2.10.12 PORT 20 21 DESTIP 193.9.200.2
  VIPADISTRIBUTE DISTMETHOD ROUNDROBIN 201.2.10.14 DESTIP 193.9.200.2
  VIPADISTRIBUTE TIMEDAFF 30 201.2.10.15 PORT 23 DESTIP ALL
| VIPADISTRIBUTE 201.2.10.23 PORT 4000 DESTIP ALL
  VIPABACKUP 100 201.2.10.13
  VIPABACKUP 80 201.2.10.21 201.2.10.22
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA1 2001:0DB8:1::1
| VIPADISTRIBUTE SYSPLEXPORTS DISTMETHOD SERVERWLM DVIPA1 DESTIP ALL
  VIPADEFINE DVIPA2 2001:0DB8:2::2
  VIPADISTRIBUTE TIMEDAFF 45 DISTM ROUNDROBIN DVIPA2 PORT 23 DESTIP ALL
  VIPARANGE VRANGE1 2001:0DB8:3::1/100
ENDVIPADYNAMIC
```

```
TCPCS2
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.2 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0002
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE 255.255.255.192 201.2.10.13
  VIPABACKUP 100 201.2.10.11 201.2.10.21
  VIPABACKUP 75 201.2.10.12 201.2.10.22
```

```
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA3 2001:0DB8:3::3
  VIPABACKUP 200 DVIPA2
  VIPABACKUP 220 DVIPA1
  VIPADISTRIBUTE SYSPLEXPORTS DVIPA1 PORT 23 DESTIP ALL
  VIPABACKUP 10  DVIPA4
ENDVIPADYNAMIC
```

```
TCPCS3
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.3 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0003
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE MOVE IMMED 255.255.255.192 201.2.10.21 201.2.10.22
  VIPABACKUP 10 201.2.10.11 201.2.10.12 201.2.10.13
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA4 2001:0DB8:4::4
  VIPABACKUP 110 DVIPA2
  VIPABACKUP 100 DVIPA1
ENDVIPADYNAMIC
```

```
TCPCS6
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.6 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0006
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

TCPCS6 does not have dynamic VIPAs defined so it does not
contain a VIPADYNAMIC definition.  It has DYNAMICXCF specified
for IPv4 and IPv6 to enable XCF dynamic support and to allow
TCPCS6 to be a target for dynamic VIPA distribution.
```

Start TCP/IP on each system as shown above.

- On system1, start TCPCS and TCPCS2.
- On system2, start TCPCS3, on system3 start TCPCS6.
- On system1, run the MODDVIPA utility to define the DVIPA 201.2.10.193.

  ```
  //TCPDVP   PROC
  //*
  //*
  //TCPDVP  EXEC PGM=MODDVIPA ,REGION=0K,TIME=1440,                      x
  //             PARM='POSIX(ON) ALL31(ON)/-p TCPCS -c 201.2.10.193'
  //SYSPRINT DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
  //SYSERR   DD SYSOUT=*
  //SYSERROR DD SYSOUT=*
  //SYSDEBUG DD SYSOUT=*
  //SYSUDUMP DD SYSOUT=A
  //SYSABEND DD SYSOUT=*
  //*
  //*Run program here
  //*
  //TCPDVP  EXEC PGM=MODDVIPA ,REGION=0K,TIME=1440,                      x
  //             PARM='POSIX(ON) ALL31(ON)/-p TCPCS -d 201.2.10.193'
  ```

  The PARM field can be -c for create or -d for delete. The example above will create DVIPA 201.2.10.193 for the TCP/IP named TCPCS. After intermediate program has completed (and the comment character is removed), the DVIPA will be deleted.

- On system 1, deactivate DVIPA 201.2.10.23 by issuing the following command on the MVS console:

  ```
  V TCPIP,TCPCS,SYSPLEX,DEACTIVATE,DVIPA=201.2.10.23
  ```

Verifying the DVIPAs in a sysplex

The following display command can be used to display dynamic VIPAs in a sysplex:

```
d tcpip,tcpname,sysplex,vipadyn
```

In the following example taken from stack TCPCS, the ORIGIN lines show that 201.2.10.11, 201.2.10.12, 201.2.10.14, 201.2.10.15, DVIPA1, and DVIPA2 were all created by VIPADEFINE on this stack. 201.2.10.13, 201.2.10.21, and 201.2.10.22 were created by VIPABACKUP statements. (Note that the deactivated DVIPA 201.2.10.23 does not appear in this display.)

The ORIGIN line indicates how the DVIPA is configured on the stack specified by tcpname. Each stack (TCPNAME) for each system (MVSNAME) is shown with its status (STATUS). Two other status values not shown in the following example are:

QUIESCING
: This DVIPA has been deactivated, or it was a target for distribution and has been removed as a target. However, it is still servicing one or more connections for this DVIPA. The DVIPA will be removed when all connections complete.

MOVING
: This DVIPA was active on this stack and has been moved to another stack. Connections on this stack for this DVIPA prior to the move will still be serviced by this stack until completion.

The rank (RANK) indicates which of the stacks is eligible to take over if the stack on which the DVIPA is active stops. The stack with the highest rank is the one that will take over the DVIPA.

```
d tcpip,tcpcs,sysplex,vipadyn
EZZ8260I SYSPLEX CS V1R7 711
VIPA DYNAMIC DISPLAY FROM TCPCS    AT MVS004
LINKNAME: VIPLC9020A0B
IPADDR/PREFIXLEN: 201.2.10.11/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
LINKNAME: VIPLC9020A0C
IPADDR/PREFIXLEN: 201.2.10.12/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.13
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
LINKNAME: VIPLC9020A0E
IPADDR/PREFIXLEN: 201.2.10.14/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
```

```
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
LINKNAME: VIPLC9020A0F
IPADDR/PREFIXLEN: 201.2.10.15/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

TCPCS2, TCPCS3, and TCPCS6 all display the same information about all the DVIPAs. ORIGIN fields are displayed for the DVIPAs that are configured on this stack.

```
d tcpip,tcpcs2,sys,vipad
EZZ8260I SYSPLEX CS V1R7 714
VIPA DYNAMIC DISPLAY FROM TCPCS2   AT MVS004
IPADDR: 201.2.10.11
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
```

```
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
IPADDR: 201.2.10.12
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
LINKNAME: VIPLC9020A0D
IPADDR/PREFIXLEN: 201.2.10.13/26
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.14
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.15
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
```

```
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

In the following example, TCPCS6 knows about the DVIPAs on the other stacks. There are no DVIPAs configured on TCPCS6, thus, no ORIGIN fields displayed.

```
d tcpip,tcpcs6,sys,vipad
EZZ8260I SYSPLEX CS V1R7 904
VIPA DYNAMIC DISPLAY FROM TCPCS6   AT MVS005
IPADDR: 201.2.10.11
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
IPADDR: 201.2.10.12
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.13
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.14
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.15
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
```

```
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

Using Netstat support to verify dynamic VIPA configuration

The Netstat commands (TSO NETSTAT, z/OS UNIX netstat, and MVS console DISPLAY TCPIP,,NETSTAT) have a VIPADCFG (-F) report and a VIPADYN (-v) report. These reports show the dynamic VIPA configuration for a particular TCP/IP.

Note: Use the CONFIG (-f) report to verify the rest of the stack's configuration, including SOURCEVIPA and TCPSTACKSOURCEVIPA, and use the SRCIP(-J) report to verify the SRCIP and ENDSRCIP profile statement block for particular jobs.

The dynamic VIPA information section is only displayed when there are DVIPAs configured on this stack. The VIPA Range section, displayed only if a VIPARANGE statement was processed in this stack's initial profile (or in a data set referenced by a VARY TCPIP,,OBEYFILE command), indicates only that a range was configured. It does not indicate whether any ioctl or BIND has actually created a DVIPA in the specified range. The VIPA distribute section is displayed only if there are VIPADISTRIBUTE statements configured on this stack. The deactivated dynamic VIPA information section is displayed only when there are DVIPAs that have been deactivated on this stack.

On stack TCPCS, using netstat on OMVS:

```
# netstat -p tcpcs -F
MVS TCP/IP NETSTAT CS V1R7       TCPIP Name: TCPCS           18:23:46
Dynamic VIPA Information:

  VIPA Backup:
    IpAddr/PrefixLen: 201.2.10.13
      Rank: 000100  Moveable:            SrvMgr:
    IpAddr/PrefixLen: 201.2.10.21
      Rank: 000080  Moveable:            SrvMgr:
    IpAddr/PrefixLen: 201.2.10.22
      Rank: 000080  Moveable:            SrvMgr:

  VIPA Define:
    IpAddr/PrefixLen: 201.2.10.11/28
      Moveable: Immediate  SrvMgr: No
    IpAddr/PrefixLen: 201.2.10.12/28
      Moveable: Immediate  SrvMgr: No
    IpAddr/PrefixLen: 201.2.10.14/28
      Moveable: Immediate  SrvMgr: No
    IpAddr/PrefixLen: 201.2.10.15/28
      Moveable: Immediate  SrvMgr: No
    IntfName: DVIPA1
      IpAddr: 2001:0DB8:1::1
        Moveable: Immediate  SrvMgr: n/a
```

```
    IntfName: DVIPA2
      IpAddr: 2001:0DB8:2::2
        Moveable: Immediate  SrvMgr: n/a

  VIPA Range:
    IpAddr/PrefixLen: 201.2.10.192/26
      Moveable: NonDisr
    IntfName: VRANGE1
      IpAddr/PrefixLen: 2001:0DB8:3::1/100
        Moveable: NonDisr

  VIPA Distribute:
    Dest:        201.2.10.11..20
      DestXCF:   ALL
        SysPt:   Yes  TimAff: No    Flg: ServerWLM
    Dest:        201.2.10.11..21
      DestXCF:   ALL
        SysPt:   Yes  TimAff: No    Flg: ServerWLM
    Dest:        201.2.10.12..20
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: BaseWLM
    Dest:        201.2.10.12..21
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: BaseWLM
    Dest:        201.2.10.14..n/a
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: Roundrobin
    Dest:        201.2.10.15..23
      DestXCF:   ALL
        SysPt:   No   TimAff: 30    Flg: BaseWLM
    DestIntf:    DVIPA1
      Dest:      2001:0DB8:1::1..n/a
        DestXCF: ALL
          SysPt: Yes  TimAff: No    Flg: BaseWLM
    DestIntf:    DVIPA2
      Dest:      2001:0DB8:2::2..23
        DestXCF: ALL
          SysPt: No   TimAff: 45    Flg: Roundrobin

Deactivated Dynamic VIPA Information:

  VIPA Define:
    IpAddr/PrefixLen: 201.2.10.23/28
      Moveable: Immediate  SrvMgr: No

  VIPA Distribute:
    Dest:        201.2.10.23..4000
      DestXCF:   ALL
        SysPt:   No   TimAff: No    Flg: BaseWLM
```

On stack TCPCS2 from the console:

```
d tcpip,tcpcs2,net,vipadcfg
EZD0101I NETSTAT CS V1R7 TCPCS2 721
DYNAMIC VIPA INFORMATION:
  VIPA BACKUP:
    IPADDR/PREFIXLEN: 201.2.10.11
      RANK: 000100  MOVEABLE:            SRVMGR:
    IPADDR/PREFIXLEN: 201.2.10.12
      RANK: 000075  MOVEABLE:            SRVMGR:
    IPADDR/PREFIXLEN: 201.2.10.21
      RANK: 000100  MOVEABLE:            SRVMGR:
    IPADDR/PREFIXLEN: 201.2.10.22
      RANK: 000075  MOVEABLE:            SRVMGR:
  VIPA DEFINE:
    IPADDR/PREFIXLEN: 201.2.10.13/26
      MOVEABLE: IMMEDIATE  SRVMGR: NO
```

```
    INTFNAME: DVIPA3
      IPADDR: 2001:0DB8:3::3
        MOVEABLE: IMMEDIATE  SRVMGR: N/A
  VIPA RANGE:
    IPADDR/PREFIXLEN: 201.2.10.192/26
      MOVEABLE: NONDISR
```

On stack TCPCS3 from the console:

```
d tcpip,tcpcs3,net,vipadcfg
EZD0101I NETSTAT CS V1R7 TCPCS3 907
DYNAMIC VIPA INFORMATION:
  VIPA BACKUP:
    IPADDR/PREFIXLEN: 201.2.10.11
      RANK: 000010  MOVEABLE:            SRVMGR:
    IPADDR/PREFIXLEN: 201.2.10.12
      RANK: 000010  MOVEABLE:            SRVMGR:
    IPADDR/PREFIXLEN: 201.2.10.13
      RANK: 000010  MOVEABLE:            SRVMGR:
  VIPA DEFINE:
    IPADDR/PREFIXLEN: 201.2.10.21/26
      MOVEABLE: IMMEDIATE  SRVMGR: NO
    IPADDR/PREFIXLEN: 201.2.10.22/26
      MOVEABLE: IMMEDIATE  SRVMGR: NO
    INTFNAME: DVIPA4
      IPADDR: 2001:0DB8:4::4
        MOVEABLE: IMMEDIATE  SRVMGR: N/A
  VIPA RANGE:
    IPADDR/PREFIXLEN: 201.2.10.192/26
      MOVEABLE: NONDISR
```

On stack TCPCS6 from the console:

```
d tcpip,tcpcs6,net,vipadcfg
EZD0101I NETSTAT CS V1R7 TCPCS6 910
```

The VIPADYN (-v) report displays all the dynamic VIPAs available to this stack, as shown in the following examples. (Note that deactivated DVIPAs do not appear in this report.)

On stack TCPCS using netstat on OMVS:

```
# netstat -p tcpcs -v
MVS TCP/IP NETSTAT CS V1R7       TCPIP Name: TCPCS           18:32:26
IpAddr/PrefixLen: 201.2.10.11/28
  Status: Active     Origin: VIPADefine       DistStat: Dist/Dest
IpAddr/PrefixLen: 201.2.10.12/28
  Status: Active     Origin: VIPADefine       DistStat: Dist
IpAddr/PrefixLen: 201.2.10.13/26
  Status: Backup     Origin: VIPABackup       DistStat:
IpAddr/PrefixLen: 201.2.10.14/28
  Status: Active     Origin: VIPADefine       DistStat: Dist
IpAddr/PrefixLen: 201.2.10.15/28
  Status: Active     Origin: VIPADefine       DistStat: Dist/Dest
IpAddr/PrefixLen: 201.2.10.21/26
  Status: Backup     Origin: VIPABackup       DistStat:
IpAddr/PrefixLen: 201.2.10.22/26
  Status: Backup     Origin: VIPABackup       DistStat:
IntfName: DVIPA1
  IpAddr: 2001:0DB8:1::1
    Status: Active     Origin: VIPADefine       DistStat: Dist/Dest
IntfName: DVIPA2
  IpAddr: 2001:0DB8:2::2
    Status: Active     Origin: VIPADefine       DistStat: Dist/Dest
```

On stack TCPCS2 from the console:

```
d tcpip,tcpcs2,net,vipadyn
EZD0101I NETSTAT CS V1R7 TCPCS2 731
IPADDR/PREFIXLEN: 201.2.10.11/28
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.12/28
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.13/26
  STATUS: ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
IPADDR/PREFIXLEN: 201.2.10.14/28
  STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.15/28
  STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.21/26
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
IPADDR/PREFIXLEN: 201.2.10.22/26
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
INTFNAME: DVIPA1
  IPADDR: 2001:0DB8:1::1
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
INTFNAME: DVIPA2
  IPADDR: 2001:0DB8:2::2
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
INTFNAME: DVIPA3
  IPADDR: 2001:0DB8:3::3
    STATUS: ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
10 OF 10 RECORDS DISPLAYED
```

On stack TCPCS3 from the console:

```
d tcpip,tcpcs3,net,vipadyn
EZD0101I NETSTAT CS V1R7 TCPCS3 913
IPADDR/PREFIXLEN: 201.2.10.11/28
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.12/28
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
IPADDR/PREFIXLEN: 201.2.10.13/26
  STATUS: BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
IPADDR/PREFIXLEN: 201.2.10.15/28
  STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.21/26
  STATUS: ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
IPADDR/PREFIXLEN: 201.2.10.22/26
  STATUS: ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
INTFNAME: DVIPA1
  IPADDR: 2001:0DB8:1::1
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
INTFNAME: DVIPA2
  IPADDR: 2001:0DB8:2::2
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
INTFNAME: DVIPA4
  IPADDR: 2001:0DB8:4::4
    STATUS: ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
9 OF 9 RECORDS DISPLAYED
```

On stack TCPCS6 from the console:

```
d tcpip,tcpcs6,net,vipadyn
EZD0101I NETSTAT CS V1R7 TCPCS6 916
IPADDR/PREFIXLEN: 201.2.10.11/28
  STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
IPADDR/PREFIXLEN: 201.2.10.15/28
  STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
INTFNAME: DVIPA1
   IPADDR: 2001:0DB8:1::1
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
```

```
INTFNAME: DVIPA2
  IPADDR: 2001:0DB8:2::2
    STATUS: ACTIVE     ORIGIN:                  DISTSTAT: DEST
4 OF 4 RECORDS DISPLAYED
```

Verifying Sysplex Distributor workload

The Netstat commands (TSO NETSTAT, z/OS UNIX netstat, and MVS console DISPLAY TCPIP,,NETSTAT) have a VDPT (-O) report and a VCRT (-V) report. Refer to *z/OS Communications Server: IP System Administrator's Commands* for more information on these commands.

Run Netstat VDPT/-O on the distributing stack to confirm that there are target stacks available with server applications ready. This display will only show target stacks that are currently up and have joined the sysplex. The RDY field indicates how many, if any, applications the target TCP/IP, identified by its DestXCF Addr, has bound to DPort. If none, then this target TCP/IP will not receive any connection workload. The TotalConn field indicates how many connections this distributing TCP/IP has forwarded to the target TCP/IP.

Note: TotalConn is a historical count and will wrap.

The following netstat display command on the distributing stack, TCPCS, shows which target stacks are available with the server applications ready. The target stack is identified by its dynamic XCF address (DESTXCF). The RDY field indicates how many applications on that target stack have bound to the DEST port. TOTALCONN is the number of all connections the distributing stack, TCPCS, has routed to the target stack. WLM is the workload manager weight value for the target TCP/IP stack. FLG is a flag field indicating how the connection was distributed. For example, the value DYNAMIC below indicates this DVIPA is being distributed by the dynamic port assignment function. The target server responsiveness (TSR) field indicates how well a target server is accepting TCP connection setup requests. A value of 100 indicates that the target server is accepting all TCP connection setup requests successfully. A value of 0 indicates that the target server is accepting no TCP connection setup requests successfully. A value between 100 and 0 indicates the relative success rate of TCP connection setup requests for this target server. For more details, refer to *z/OS Communications Server: IP System Administrator's Commands*.

```
d tcpip,tcpcs,net,vdpt
EZD0101I NETSTAT CS V1R7 TCPCS 734
DYNAMIC VIPA DESTINATION PORT TABLE:
DEST:        201.2.10.11..20
  DESTXCF:   193.9.200.1
  TOTALCONN: 0000000050  RDY: 001  WLM: 15  TSR: 100
  FLG: SERVERWLM
DEST:        201.2.10.11..20
  DESTXCF:   193.9.200.2
  TOTALCONN: 0000000024  RDY: 001  WLM: 15  TSR: 100
  FLG: SERVERWLM
DEST:        201.2.10.11..20
  DESTXCF:   193.9.200.3
  TOTALCONN: 0000000069  RDY: 001  WLM: 15  TSR: 100
  FLG: SERVERWLM
DEST:        201.2.10.11..20
  DESTXCF:   193.9.200.6
  TOTALCONN: 0000000077  RDY: 001  WLM: 15  TSR: 100
  FLG: SERVERWLM
DEST:        201.2.10.11..21
  DESTXCF:   193.9.200.1
  TOTALCONN: 0000000009  RDY: 001  WLM: 15  TSR: 100
```

```
|                         FLG: SERVERWLM
                        DEST:        201.2.10.11..21
                          DESTXCF:   193.9.200.2
|                         TOTALCONN: 0000000015  RDY: 001  WLM: 15  TSR: 100
|                         FLG: SERVERWLM
                        DEST:        201.2.10.11..21
                          DESTXCF:   193.9.200.3
|                         TOTALCONN: 0000000023  RDY: 001  WLM: 15  TSR: 100
|                         FLG: SERVERWLM
                        DEST:        201.2.10.11..21
                          DESTXCF:   193.9.200.6
|                         TOTALCONN: 0000000052  RDY: 001  WLM: 15  TSR: 100
|                         FLG: SERVERWLM
                        DEST:        201.2.10.12..20
                          DESTXCF:   193.9.200.2
|                         TOTALCONN: 0000000047  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
                        DEST:        201.2.10.12..21
                          DESTXCF:   193.9.200.2
|                         TOTALCONN: 0000000016  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
                        DEST:        201.2.10.14..6000
                          DESTXCF:   193.9.200.2
|                         TOTALCONN: 0000000093  RDY: 001  WLM: 01  TSR: 100
                          FLG: DYNAMIC, ROUNDROBIN
                        DEST:        201.2.10.15..23
                          DESTXCF:   193.9.200.1
|                         TOTALCONN: 0000000088  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
                        DEST:        201.2.10.15..23
                          DESTXCF:   193.9.200.2
|                         TOTALCONN: 0000000022  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
                        DEST:        201.2.10.15..23
                          DESTXCF:   193.9.200.3
|                         TOTALCONN: 0000000069  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
                        DEST:        201.2.10.15..23
                          DESTXCF:   193.9.200.6
|                         TOTALCONN: 0000000068  RDY: 001  WLM: 01  TSR: 100
                          FLG: BASEWLM
|                       DESTINTF:      DVIPA1
|                         DEST:        2001:0DB8:1::1..23
|                           DESTXCF:   2001:0DB8::151:1
|                           TOTALCONN: 0000000072  RDY: 001  WLM: 15  TSR: 100
|                           FLG: SERVERWLM
|                       DESTINTF:      DVIPA1
|                         DEST:        2001:0DB8:1::1..23
|                           DESTXCF:   2001:0DB8::151:2
|                           TOTALCONN: 0000000023  RDY: 001  WLM: 15  TSR: 100
|                           FLG: SERVERWLM
|                       DESTINTF:      DVIPA1
|                         DEST:        2001:0DB8:1::1..23
|                           DESTXCF:   2001:0DB8::151:3
|                           TOTALCONN: 0000000018  RDY: 001  WLM: 15  TSR: 100
|                           FLG: SERVERWLM
|                       DESTINTF:      DVIPA1
|                         DEST:        2001:0DB8:1::1..23
|                           DESTXCF:   2001:0DB8::151:6
|                           TOTALCONN: 0000000039  RDY: 001  WLM: 15  TSR: 100
|                           FLG: SERVERWLM
                        DESTINTF:      DVIPA2
                          DEST:        2001:0DB8:2::2..23
                            DESTXCF:   2001:0DB8::151:1
|                           TOTALCONN: 0000000076  RDY: 001  WLM: 01  TSR: 100
                            FLG: ROUNDROBIN
                        DESTINTF:      DVIPA2
```

```
  DEST:        2001:0DB8:2::2..23
    DESTXCF:   2001:0DB8::151:2
    TOTALCONN: 0000000021  RDY: 001  WLM: 01  TSR: 100
    FLG: ROUNDROBIN
DESTINTF:      DVIPA2
  DEST:        2001:0DB8:2::2..23
    DESTXCF:   2001:0DB8::151:3
    TOTALCONN: 0000000008  RDY: 001  WLM: 01  TSR: 100
    FLG: ROUNDROBIN
DESTINTF:      DVIPA2
  DEST:        2001:0DB8:2::2..23
    DESTXCF:   2001:0DB8::151:6
    TOTALCONN: 0000000033  RDY: 001  WLM: 01  TSR: 100
    FLG: ROUNDROBIN
23 OF 23 RECORDS DISPLAYED
```

The following Netstat display command on the distributing stack displays all current connections being distributed by TCPCS.

```
d tcpip,tcpcs,net,vcrt
EZD0101I NETSTAT CS V1R7 TCPCS 844
DYNAMIC VIPA CONNECTION ROUTING TABLE:
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.5..1029
  DESTXCF: 193.9.200.1
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.8..1050
  DESTXCF: 193.9.200.2
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.11..1079
  DESTXCF: 193.9.200.3
DEST:      201.2.10.12..21
  SOURCE:  193.9.200.9..1030
  DESTXCF: 193.9.200.2
DEST:      2001:0DB8:1::1..21
  SOURCE:  2001:0DB8::151:0006..1038
  DESTXCF: 2001:0DB8::151:0003
```

Dynamic VIPAs and routing protocols

With dynamic VIPAs, IP addresses may move from one stack to another. These changes need to be communicated to the network. Therefore, dynamic routing should be implemented when dynamic VIPAs are being used.

IPv4 considerations for OMPROUTE

The names of dynamic VIPA interfaces are assigned dynamically by the stack when a dynamic VIPA interface is created. Therefore, the Name field set on the Interface or OSPF_Interface statement for a dynamic VIPA will be ignored by OMPROUTE.

It is recommended that a host have an Interface or OSPF_Interface definition for every dynamic VIPA address which that host might own. Because this could be a large number of interfaces, additional wildcard capabilities have been added to OMPROUTE, for dynamic VIPA interfaces only.

Ranges of dynamic VIPA interfaces can be defined using the subnet mask parameter on the OSPF_Interface or Interface statement. The range defined will be all the IP addresses that fall within the subnet defined by the mask and the IP address. The IP address parameter must be the subnet number of the range being defined, not a host address within that range. The following example defines a range of dynamic VIPA addresses from 10.138.165.81 to 10.138.165.94:

```
OSPF_Interface
   IP_address = 10.138.165.80
   Name = dummy_name (see note)
   Subnet_mask = 255.255.255.240;
```

Tips:

- The *Name* parameter is required and must be unique, but it is not actually used for dynamic VIPAs.
- When defining ranges, it is not necessary or desirable to code a destination address. OMPROUTE will automatically set the destination address of a dynamic VIPA to its IP address.
- There is nothing in the interface definition statements that informs OMPROUTE that a particular interface definition statement is for a dynamic VIPA or a range of dynamic VIPAs. Rather, OMPROUTE learns this information from the stack when these interfaces are created or taken over.
- Since dynamic VIPAs can move between z/OS hosts within a sysplex, it is highly recommended that a ROUTERID be configured, and for it to specify a physical interface or a static VIPA and not a dynamic VIPA.

The MTU size defined on OSPF_INTERFACE statements limits the size of advertisements that can be sent or received over OMPROUTE interfaces. OMPROUTE cannot build an advertisement whose size would exceed the largest MTU size of all its interfaces. Also, OMPROUTE cannot receive an advertisement that is larger than the largest MTU size defined for all its interfaces. In either of these cases, you will see the following message:

```
EZZ7967I ADVERTISEMENT DISCARDED, OVERFLOWS BUFFER: LS
   TYPE x ID x.x.x.x ORG y.y.y.y
```

When this happens on an originating host, that host will not be able to send router Link State Advertisements (LSAs), and therefore other hosts will not be able to calculate routes to any destinations (for example, VIPAs) owned by the originating host. OMPROUTE will terminate if it encounters this condition. If it cannot send its router LSA, it is useless as a router. When this happens on a receiving host, that host will not be able to compute routes to any destinations advertised in the discarded LSA. Also note that other OSPF implementations might have similar or stricter limitations, in which case they would be unable to receive or propagate large router LSAs received from OMPROUTE. These scenarios can severely affect network connectivity and routing capability. If large numbers of VIPA interfaces are going to be used, we recommend you examine OSPF MTU sizes throughout your network to ensure that large router LSAs can be propagated.

Normally, large router LSAs would not be a problem, as LSAs seldom exceed their allowed MTU sizes. However, if a large number of VIPA or dynamic VIPA interfaces are defined on a host, this can become a consideration. The size of the router LSA will include 52 bytes for headers, plus the number of bytes required to advertise the host's owned interfaces. The number of bytes required for each interface is:

VIPA	12 bytes, plus 12 bytes for each VIPA subnet (see the following example)
Point-to-point	24 bytes
Point-to-multipoint	12 bytes, plus 12 bytes for each neighbor on the interface
All other types	12 bytes

For owned VIPA interfaces, OMPROUTE normally advertises both host and subnet routes. The size of router LSAs required can be minimized by careful subnet planning. For example, assume the following definition exists in the OMPROUTE configuration file:

```
OSPF_Interface
   IP_Address=3.3.3.*
   Name = VIPA1A
   Subnet_Mask=255.255.255.252
   Attaches_To_Area=1.1.1.1
   MTU=1024
   Cost0 = 1;
```

If 101 VIPA interfaces, numbered 3.3.3.1 to 3.3.3.101, are activated, in addition to the headers and any other owned interfaces, OMPROUTE would need 1512 bytes to advertise 126 links in its router LSA (1 host route to each of the VIPAs, plus 25 subnet routes since each subnet contains only four addresses).

By contrast, assume the following definition exists in the OMPROUTE configuration file:

```
OSPF_Interface
   IP_Address=3.3.3.*
   Name = VIPA1A
   Subnet_Mask=255.255.255.0
   Attaches_To_Area=1.1.1.1
   MTU=1024
   Cost0 = 1;
```

If the same 101 VIPA interfaces are activated, OMPROUTE would advertise 102 links in its router LSA (1 host route to each VIPA, plus 1 subnet route since all the VIPAs are in the same subnet). This would only require 1224 bytes to advertise the VIPAs. If the MTU size on the network is 1500, this can make the difference between being able to send or receive a router LSA or not being able to send or receive a router LSA. This limitiation can further be circumvented by suppressing VIPA host routes by coding the Advertise_VIPA_Routes or Subnet parameters on the OSPF_INTERFACE statement for the VIPA interfaces. However, there are limits on when this can be done. For details, see the *z/OS Communications Server: IP Configuration Reference*.

Ensure that all members of a sysplex are defined to be in the same OSPF area. Failure to do this causes routing problems during the transition of ownership of a DVIPA from a member in one area to a member in another area. It can also disrupt Sysplex Distributor operation for clients within the sysplex.

IPv4 considerations for RIP (Routing Information Protocol)

If using RIP services and Host Route advertising is not supported by adjacent routers (that is, inability to learn host routes), the following restrictions for VIPA addresses must be applied to benefit from fault tolerance support:

- If you use subnetting and VIPA addresses are in the same network as the physical IP addresses, the subnetwork portion of any VIPA addresses must not be the subnetwork portion of any physical IP addresses in the network. In this case, assign a new subnetwork for the VIPA address.
- If subnetting is not used on any physical interface, the network portion of any VIPA addresses must not be the network portion of any physical IP addresses in the network. In this case, assign a new network for the VIPA address, preferably a class C network address.

If using RIP services and Host Route advertising is supported by adjacent routers, the network or subnetwork portions of VIPA addresses can be the same across multiple z/OS TCP/IP stacks in the network. To enable Host Route advertising in OMPROUTE, configure RIP_Interface Send_Host_Routes=YES.

IPv6 considerations

For IPv6, all interfaces are defined to OMPROUTE by name only, and name wildcards are supported. This greatly simplifies definition considerations for OMPROUTE and VIPA interfaces.

To define IPv6 interfaces to OMPROUTE, you must know their names. Unlike IPv4, in which VIPA link names are generated by the stack, you specify link names to TCP/IP when defining IPv6 dynamic VIPA interfaces. These link names should be used when defining the same interfaces to OMPROUTE.

Tip: Using consistent naming conventions and name wildcards can make this task easier. For example, if you require that all IPv6 VIPA link names start with the letter V, and do not allow other interface types to have link names starting with V, a single definition can define all IPv6 VIPA interfaces to OMPROUTE as follows:

```
IPV6_OSPF_INTERFACE
      NAME=V*;
```

Unlike IPv4, there are no size limitations on the number of IPv6 dynamic VIPA addresses that OMPROUTE can advertise. In IPv6 OSPF, OMPROUTE will advertise all host and prefix addresses associated with owned interfaces, including VIPAs and dynamic VIPAs. In IPv6 RIP, sending and receiving of host routes should be turned on to allow host addresses associated with VIPAs to be advertised.

Chapter 8. TCP/IP in a sysplex

The increasing demands of network servers, and in particular z/Series servers, has led to the creation of different techniques to address performance requirements when a single server is not capable of providing the availability and scalability demands placed on it by its clients. Specifically, network solutions make use of what is referred to as the clustering technique, whereby multiple servers are associated together into a cluster to provide sufficient processing power and availability characteristics to handle the demands of the clients.

In the scope of this chapter, this cluster functionality is provided by the sysplex. That is, the sysplex provides the necessary capability to cluster together a number of z/Series servers that cooperate with one another to deliver the processing power needed to service the demands required of a particular service environment.

Solutions utilizing the clustering approach to increase server availability and processing capability attempt to provide mechanisms by which they ensure the viability of the cluster in an environment containing a large number of clients generating a potentially high number of requests. To do so, the cluster technique can provide for two main objectives, high availability and load balancing. In some cases, clustering techniques address only high availability, as is the case with dynamic VIPA that provides for availability in spite of potential TCP/IP stack or z/OS image failures. In other cases, the intent is to provide for both high availability and load balancing, as is done by the Domain Name System/Workload Manager solution (DNS/WLM) and Sysplex Distributor.

In general, load balancing refers to the ability to utilize different systems within the cluster simultaneously, thereby taking advantage of the additional computational function of each. Further, clustering techniques addressing load balancing lead to other system requirements, such as that of a single systemwide image (one identity by which clients access the system), horizontal growth, and ease of management.

The traditional view of a single server has been primarily a single machine with perhaps a few network interfaces (IP addresses). This tends to lead to many potential points of failure within the server: the machine itself (hardware), the operating system (including TCP/IP stack) kernel executing on the machine, or a network interface (and the IP address associated with it). Static Virtual IP Addresses (VIPAs) exclude the network interface as a point of failure while dynamic VIPAs additionally aid with server (image) or kernel failure. In this way, high availability is seen as the availability of the entire server cluster and the service it provides. Further, VIPAs can be used in conjunction with the load balancing solutions discussed in this document, DNS/WLM and Sysplex Distributor.

Clustering techniques that address the load balancing of connections requests also typically provide for some high availability. That is, these techniques dispatch connections to target servers and can exclude failed servers from the list of target servers that can receive connections. In this way, the dispatching function avoids routing connections and requests to a server incapable of satisfying such requests.

Load balancing is the ability for a cluster to spread workload evenly (or based on some policy) to target servers comprising the cluster. Usually, this load balancing is

measured by some notion of perceived load on each of the target servers. This chapter describes two techniques that provide load balancing: DNS/WLM and Sysplex Distributor. Each identifies the target zSeries servers willing to receive client connections based on some specification.

By providing load balancing, clustering techniques must also provide for other system requirements in addition to the dispatching of connections. These include the ability to advertise some single systemwide image or identity so that clients can uniquely and easily identify the service. Additionally, clustering techniques should also provide for horizontal growth of the system and ease of management.

There is an excellent description of sysplexes in *z/OS Parallel Sysplex Overview*. Refer to the Redbook, *TCP/IP in a Sysplex*, for more detailed information on implementing load balancing and availability in your sysplex.

Note: This chapter applies to both IPv4 and IPv6, unless otherwise noted.

Connectivity in a sysplex

With dynamic VIPAs, IP addresses may move from one stack to another. These changes need to be communicated to the network. Therefore, dynamic routing should be implemented when dynamic VIPAs are being used. Refer to "Dynamic VIPAs and routing protocols" on page 358 for more detailed information.

Dynamic XCF

Tip: Dynamic XCF support is available for both IPv4 and IPv6, enabled with the IPCONFIG DYNAMICXCF and IPCONFIG6 DYNAMICXCF statements, respectively. Though not specifically mentioned throughout this section, unless otherwise noted, this section applies to IPCONFIG6 DYNAMICXCF as well as IPCONFIG DYNAMICXCF.

Use the DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statements to create trusted, internal links to other stacks within a sysplex. Dynamic XCF creates a single IP address by which all other stacks in the sysplex can reach the stack. Most point-to-point links have their own unique IP address. A unique pair of IP addresses is needed for each stack within a sysplex using normal point-to-point links. This tends to use more IP addresses. IP addresses can be saved by using dynamic XCF. Additionally, dynamic XCF automatically generates the appropriate DEVICE, LINK, HOME, INTERFACE, BSDROUTINGPARMS, and BEGINROUTES definitions (as described below) and activates the devices to enable the stack to communicate with other stacks in the sysplex.

Dynamic XCF devices and links, when activated, appear to the stack as though they had been defined in the TCP/IP profile. They can be displayed using standard commands.

Dynamic XCF can be used to generate dynamic definitions for TCP/IP stacks that reside on another z/OS host in a sysplex and for additional TCP/IP stacks that reside on the same z/OS host.

The minimum requirements in order for TCP/IP stacks to utilize XCF dynamics differ based on whether same host or inter-host communication is being used. In order to generate definitions for two TCP/IP stacks that reside on different MVS hosts:

- Both MVS hosts must belong to the same sysplex.

- VTAM must have XCF communications enabled by specifying XCFINIT=YES or XCFINIT=DEFINE as a startup parameter or by activating the VTAM XCF local SNA major node, ISTLSXCF. For details about configuration, refer to *z/OS Communications Server: SNA Network Implementation Guide*.
- DYNAMICXCF must be specified in the TCP/IP profile of each stack.

With this configuration, both same host and inter-host communication can be performed using dynamic XCF.

In order to generate definitions for two TCP/IP stacks that reside on the same MVS host, you are required to specify DYNAMICXCF in the TCP/IP profile of each stack.

At initialization, each TCP/IP stack configured for XCF joins a well-known XCF group. When other stacks in the group discover the new stack, the definitions are created automatically, the links are activated, and the remote IP address for each link is added to the routing table. After the remote IP address has been added, IP traffic proceeds as usual.

In VTAM, you must activate the XCF local SNA major node. You can do this using the start option XCFINIT=YES or XCFINIT=DEFINE. If dynamically defined XCF definitions have been created for another VTAM in the sysplex that has since stopped and restarted with a different CPName, dynamic XCF recognizes this situation and automatically modifies existing definitions to accommodate the CPName change. If the XCF local SNA major node is inactive when TCP/IP is started and is not activated until after TCP/IP has finished initialization, TCP/IP will not generate any dynamic definitions for other TCP/IP hosts already started in the sysplex until either:

- A new TCP/IP host is detected
- A profile related operator command is issued (such as the VARY TCPIP,,OBEYFILE, VARY TCPIP,,START, or VARY TCPIP,,STOP commands)

For IPv4, to request dynamics for XCF or same host connections, enter the following in the IPCONFIG statement:

```
DYNAMICXCF IPAddress SubnetMask CostMetric
```

For IPv6, to request dynamics for XCF or same host connections, enter the following in the IPCONFIG6 statement:

```
DYNAMICXCF IP6Address
```

If the following are true:

- TCP/IP detects another instance of TCP/IP on the same z/OS
- No device exists with the name IUTSAMEH
- No link exists with the name EZASAMEMVS (IPv4) or EZ6SAMEMVS (IPv6)

Then, internal definitions equivalent to the following are created:

IPv4

DEVICE IUTSAMEH MPCPTP AUTORESTART
: Device definition to obtain the most efficient stack-to-stack communications within the same MVS image.

LINK EZASAMEMVS MPCPTP IUTSAMEH
: Link definition for the IUTSAMEH device.

HOME IPAddress EZASAMEMVS
Associates the IP address with the IUTSAMEH link.

BSDROUTINGPARMS EZASAMEMVS 65535 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for EZASAMEMVS.

Note: The DestIPAddress is always 0.

START IUTSAMEH
Starts the IUTSAMEH device.

IPv6

INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR IP6Address
Interface definition for EZ6SAMEMVS.

START EZ6SAMEMVS
Starts the EZ6SAMEMVS interface.

If the following are true:

- TCP/IP detects another instance of TCP/IP in the sysplex
- No device with the name of the CPName of the remote VTAM exists
- No HiperSockets connectivity between the two images exists
- No link exists with the name EZAXCF*nn* (IPv4) or EZ6XCF*nn* (IPv6), where *nn* is the value of the MVS system symbol (SYSCLONE) for the MVS hosting the VTAM with the device name

Then, internal definitions equivalent to the following are created:

IPv4

DEVICE CPName MPCPTP AUTORESTART
Device definition to communicate with TCP/IP stacks hosted by the remote VTAM.

LINK EZAXCF*nn* MPCPTP CPName
Link definition for the device, where *nn* is the SYSCLONE value for the remote VTAM and MVS.

HOME IPAddress EZAXCF*nn*
Associates the IP address with the dynamic XCF link.

BSDROUTINGPARMS EZAXCF*nn* 55296 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for EZAXCF*nn*.

START CPName
Starts the specified device.

IPv6

INTERFACE EZ6XCF*nn* DEFINE MPCPTP6 TRLENAME CPName IPADDR IP6Address
Interface definition, where *nn* is the SYSCLONE value for the remote VTAM and MVS.

START EZ6XCF*nn*
Starts the specified interface.

Notes:

1. If EZAXCF*nn* (IPv4) or EZ6XCF*nn* (IPv6) is already defined as a link name, or the CPName is already defined as a device name, dynamic XCF definitions will not be generated for discovery of another stack in the same MVS image.
2. The DestIPAddress is always zero.

If the following are true:

- TCP/IP detects another instance of TCP/IP in the sysplex
- The images reside on the same CPC
- HiperSockets connectivity between the two images exists
- The host processor supports HiperSockets and z/OS Communications Server is properly configured
- No link exists with the name IQDIOLNK*nnnnnnnn* (where *nnnnnnnn* is the hexadecimal representation of the IP address specified on the IPCONFIG DYNAMICXCF statement)

Then, internal definitions equivalent to the following are created for IPv4:

DEVICE IUTIQDIO MPCIPA AUTORESTART
Device definition to communicate with TCP/IP stacks hosted by the remote VTAM.

LINK IQDIOLNK*nnnnnnnn* IPAQIDIO IUTIQDIO
Link definition for the device, where *nnnnnnnn* is the hexadecimal representation of the IP address specified on the IPCONFIG DYNAMICXCF statement (that is, IPAddress).

HOME IPAddress IQDIOLNK*nnnnnnnn*
Associates the IP address with the dynamic XCF link.

BSDROUTINGPARMS IQDIOLNK*nnnnnnnn* 57344 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for IQDIOLNK*nnnnnnnn*.

START IUTIQDIO
Starts the specified device.

Notes:

1. If IQDIOLNK*nnnnnnnn* is already defined as a link name or IUTIQDIO is already defined as a device name, dynamic XCF definitions will not be generated for discovery of another stack in the same MVS image.
2. The DestIPAddress is always zero.

For details about these XCF-related statements, refer to *z/OS Communications Server: IP Configuration Reference*. For information about changes to Netstat displays of dynamic XCF settings, refer to *z/OS Communications Server: IP System Administrator's Commands*.

IUTSAMEH

Communications Server provides internal links between TCP/IP stacks that are running within the same MVS image. This support is referred to as a Same Host (IUTSAMEH) link. If DYNAMICXCF is defined, TCP/IP always creates and activates a same host (IUTSAMEH) device and link (unless a static IUTSAMEH device is already defined) even if this is the only stack on the MVS image. When TCP/IP activates the IUTSAMEH device, VTAM dynamically builds the IUTSAMEH TRLE. The generated device name is IUTSAMEH, the generated link name is EZASAMEMVS (IPv4), and the generated interface name is

EZ6SAMEMVS (IPv6). As other stacks are brought up within the same MVS image, a host route is created to each of these stacks across the same host link. It is recommended that users do not configure a static device for IUTSAMEH (allow TCP/IP to dynamically create the device and link). Communications Server also uses the IUTSAMEH link for Enterprise Extender support.

The IUTSAMEH DEVICE and LINK (IPv4) or INTERFACE (IPv6) will not become active until either another TCP/IP stack or Enterprise Extender is activated on this MVS image.

```
DEVNAME: IUTSAMEH        DEVTYPE: MPCPTP
DEVSTATUS: SENT SETUP
LNKNAME: EZASAMEMVS      LNKTYPE: MPCPTP      LNKSTATUS: NOT ACTIVE
⋮
INTFNAME: EZ6SAMEMVS     INTFTYPE: MPCPTP6     INTFSTATUS: NOT ACTIVE
```

In the case where IUTSAMEH is active solely because of Enterprise Extender and the IUTSAMEH is subsequently stopped, to restart the IUTSAMEH for Enterprise Extender, the XCA major node must be recycled. Otherwise, it might require manual reactivation of the Enterprise Extender LINEs and PUs, and manual redials.

XCF

When a subsequent stack within the sysplex is started that is not within the same MVS image, TCP/IP creates and activates an XCF device and link (unless an XCF device is already defined). The XCF links connect using the sysplex coupling facility or CTC links. A new device and link are created for every other VTAM node in the sysplex with at least one TCP/IP stack active on the same system with DYNAMICXCF specified. The generated device name is the CP name (for APPN) or SSCP name (for subarea-only nodes) of the remote VTAM. The generated link name is EZAXCF*nn* (IPv4), and the generated interface name is EZ6XCF*nn* (IPv6), where *nn* is the 2-character &SYSCLONE value. A host route across the XCF link is created when the XCF link is successfully activated.

Examples of definitions generated by dynamic XCF

This section contains examples of definitions generated by dynamic XCF in both IPv4 and IPv6 environments. The notes following the examples pertain to both environments.

IPv4 example 1:

This configuration consists of two MVS systems (MVS1,MVS2) that are members of the same sysplex. Each MVS host has one TCP/IP stack (TCPIP1 and TCPIP2, respectively). From the syntax descriptions described above, the following information is needed to generate the dynamic definitions:

- MVS sysclone value
- VTAM CPName
- Status of XCF in VTAM
- The values specified on the IPCONFIG DYNAMICXCF keyword

Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3

MVS2:
```

```
Sysclone = B2 VTAM
Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.2 255.255.255.248 2
```

After both TCPIP1 and TCPIP2 have been started, the following definitions will be generated.

TCPIP1 will generate the equivalent of these definitions:.

```
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.1 EZAXCFB2
BSDROUTINGPARMS EZAXCFB2 55296 3 255.255.255.248 0
START VTAM2
```

TCPIP2 will generate:

```
DEVICE VTAM1 MPCPTP AUTORESTART
LINK EZAXCFA1 MPCPTP VTAM1
HOME 9.1.1.2 EZAXCFA1
BSDROUTINGPARMS EZAXCFA1 55296 2 255.255.255.248 0
START VTAM1
```

When an XCF link becomes active, each TCPIP will generate a route to the other TCPIP over the XCF link. In this example, when the XCF link becomes active, TCPIP1 will generate a route to TCPIP2 over the XCF link and vice versa.

IPv4 example 2:

The configuration is the same as Example 1 except a second TCP/IP stack (TCPIP1A) was added to MVS1.

Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3
TCPIP1A: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0

MVS2:
Sysclone = B2
VTAM Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.2 255.255.255.248 2
```

After both TCPIP1 and TCPIP2 have been started, the following definitions will be generated, as in Example 1.

TCPIP1 will generate the equivalent of these definitions:

```
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.1 EZAXCFB2
BSDROUTINGPARMS EZAXCFB2 55296 3 255.255.255.248 0
START VTAM2
```

TCPIP2 will generate:

```
DEVICE VTAM1 MPCPTP AUTORESTART
LINK EZAXCFA1 MPCPTP VTAM1
HOME 9.1.1.2 EZAXCFA1
BSDROUTINGPARMS EZAXCFA1 55296 2 255.255.255.248 0
START VTAM1
```

Now, TCPIP1A is started. TCPIP1 and TCPIP2 recognize that TCPIP1A has started. TCPIP1A will generate definitions for both TCPIP1 and TCPIP2. TCPIP1 will generate IUTSAMEH definitions for TCPIP1A. However, TCPIP2 does not need to generate and will not generate any new definitions except for routing information for TCPIP1A. New definitions do not need to be created because the DEVICE and LINK definitions are based on the discovery of a new VTAM node in the sysplex. (The DEVICE name is the VTAM CPName.)

TCPIP1 will generate the equivalent of these definitions:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.1 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 3 255.255.255.248 0
START IUTSAMEH
```

When the IUTSAMEH connection becomes active, each TCPIP will generate a route to the other TCPIP over the IUTSAMEH connection.

TCPIP2 does not generate anything.

TCPIP1A will generate:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.3 EZAXCFB2
HOME 9.1.1.3 EZASAMEMVS
BSDROUTINGPARMS EZAXCFB2 55296 0 255.255.255.248 0
BSDROUTINGPARMS EZASAMEMVS 65535 0 255.255.255.248 0
START IUTSAMEH
START VTAM2
```

When the IUTSAMEH connection becomes active, each TCPIP will generate a route to the other TCPIP over the IUTSAMEH connection.

IPv4 example 3:

To continue Example 2, add another MVS host (MVS3) with a VTAM node (VTAM3) with one TCP/IP stack (TCPIP3).

```
MVS3:
Sysclone = C3
VTAM Cpname = VTAM3
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP3: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0
```

In this example, the previously active TCP/IP stacks will generate definitions for TCPIP3 because a new VTAM stack has become active in the sysplex. TCPIP3 will generate definitions for definitions for TCPIP1/TCPIP1A and TCPIP2.

TCPIP1 will generate the equivalent of these definitions:

```
DEVICE VTAM3 MPCPTP AUTORESTART
LINK EZAXCFC3 MPCPTP VTAM3
HOME 9.1.1.1 EZAXCFC3
BSDROUTINGPARMS EZAXCFC3 55296 3 255.255.255.248 0
START VTAM3
```

TCPIP2 will generate:

```
DEVICE VTAM3 MPCPTP AUTORESTART
LINK EZAXCFC3 MPCPTP VTAM3
HOME 9.1.1.2 EZAXCFC3
BSDROUTINGPARMS EZAXCFC3 55296 2 255.255.255.248 0
START VTAM3
```

TCPIP1A will generate:

```
DEVICE VTAM3 MPCPTP AUTORESTART
LINK EZAXCFC3 MPCPTP VTAM3
HOME 9.1.1.3 EZAXCFC3
BSDROUTINGPARMS EZAXCFC3 55296 0 255.255.255.248 0
START VTAM3
```

TCPIP3 will generate:

```
DEVICE VTAM1 MPCPTP AUTORESTART
LINK EZAXCFA1 MPCPTP VTAM1
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.3 EZAXCFA1
HOME 9.1.1.3 EZAXCFB2
BSDROUTINGPARMS EZAXCFA1 55296 0 255.255.255.248 0
BSDROUTINGPARMS EZAXCFB2 55296 0 255.255.255.248 0
START VTAM1
START VTAM2
```

IPv4 example 4:

This example illustrates how dynamic XCF can generate IUTSAMEH definitions without VTAM having its XCF enabled.

```
MVS1:
Sysclone = A1  (not used in this example)
VTAM Cpname = VTAM1  (not used in this example)
VTAM has XCFINIT=NO specified and has not activated the major node ISTLSXCF.
TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3
TCPIP1A: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0
```

TCPIP1 will generate the equivalent of these definitions:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.1 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 3 255.255.255.248 0
START IUTSAMEH
```

TCPIP1A will generate:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.3 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 0 255.255.255.248 0
START IUTSAMEH
```

IPv6 example 1:

This configuration consists of two MVS systems (MVS1,MVS2) that are members of the same sysplex. Each MVS host has one TCP/IP stack (TCPIP1 and TCPIP2, respectively). From the syntax descriptions described above, the following information is needed to generate the dynamic definitions:

- MVS sysclone value
- VTAM CPName
- Status of XCF in VTAM

- The values specified on the IPCONFIG6 DYNAMICXCF keyword

Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
MVS2:
Sysclone = B2
Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::222:f001
```

After both TCPIP1 and TCPIP2 are started, the following definitions are generated.

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f001
START EZ6XCFB2
```

TCPIP2 generates:

```
INTERFACE  EZ6XCFA1 DEFINE MPCPTP6 TRLENAME VTAM1 IPADDR 2001:0DB8::222:f001
START EZ6XCFA1
```

The INTERFACE statement combines the definitions of DEVICE, LINK, and HOME into a single statement for IPv6. When an XCF interface becomes active, each TCP/IP generates a route to the other TCP/IP over the XCF interface. In this example, when the XCF interface becomes active, TCPIP1 generates a route to TCPIP2 over the XCF interface, and TCPIP2 generates a route to TCPIP1 over the XCF interface.

IPv6 example 2:

The configuration is the same as Example 1 except a second TCP/IP stack (TCPIP1A) is added to MVS1.

Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
TCPIP1A: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f002
MVS2:
Sysclone = B2
VTAM Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::222:f001
```

After both TCPIP1 and TCPIP2 are started, the following definitions are generated, as in Example 1.

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f001
START EZ6XCFB2
```

TCPIP2 generates:

```
INTERFACE  EZ6XCFA1 DEFINE MPCPTP6 TRLENAME  VTAM1 IPADDR 2001:0DB8::222:f001
START EZ6XCFA1
```

Now, TCPIP1A is started. TCPIP1 and TCPIP2 recognize that TCPIP1A has started. TCPIP1A generates definitions for both TCPIP1 and TCPIP2. TCPIP1 generates IUTSAMEH definitions for TCPIP1A. However, TCPIP2 does not need to generate and does not generate any new definitions, except for routing information, for TCPIP1A. New definitions do not need to be created because the INTERFACE definition is based on the discovery of a new VTAM node in the sysplex. (The TRLENAME is the VTAM CPName.)

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME  IUTSAMEH IPADDR 2001:0DB8::111:f001
START EZ6SAMEMVS
```

When the IUTSAMEH connection becomes active, each TCP/IP generates a route to the other TCP/IP over the IUTSAMEH connection.

TCPIP2 does not generate anything.

TCPIP1A generates:

```
INTERFACE EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f002
START EZ6XCFB2
INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f002
START EZ6SAMEMVS
```

When an XCF interface becomes active, each TCP/IP generates a route to the other TCP/IP over the XCF interface. In this example, when the XCF interface becomes active, TCPIP1A generates a route to TCPIP2 over the XCF interface, and TCPIP2 generates a route to TCPIP1A over the XCF interface. When the IUTSAMEH connection becomes active, each TCP/IP generates a route to the other TCP/IP over the IUTSAMEH connection.

IPv6 example 3:

To continue Example 2, add another MVS host (MVS3) with a VTAM node (VTAM3) with one TCP/IP stack (TCPIP3).

```
MVS3:
Sysclone = C3
VTAM Cpname = VTAM3
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP3: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::333:f001
```

In this example, the previously active TCP/IP stacks generate definitions for TCPIP3 because a new VTAM stack has become active in the sysplex. TCPIP3 generates definitions for TCPIP1, TCPIP1A, and TCPIP2.

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::111:f001
START EZ6XCFC3
```

TCPIP2 generates:

```
 INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::222:f001
 START EZ6XCFC3
```

TCPIP1A generates:

```
INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::111:f002
START EZ6XCFC3
```

TCPIP3 generates:

```
INTERFACE EZ6XCFA1 DEFINE MPCPTP6 TRLENAME VTAM1 IPADDR 2001:0DB8::333:f001
START EZ6XCFA1
INTERFACE EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::333:f001
START EZ6XCFB2
```

IPv6 example 4:

This example illustrates how dynamic XCF can generate IUTSAMEH definitions without VTAM having its XCF enabled.

```
MVS1:
Sysclone = A1  (not used in this example)
VTAM Cpname = VTAM1  (not used in this example)
VTAM has XCFINIT=NO specified and has not activated the major node ISTLSXCF.
TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
TCPIP1A: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f002
```

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f001
START EZ6SAMEMVS
```

TCPIP1A generates:

```
 INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f002
 START EZ6SAMEMVS
```

Notes:

1. Because the interfaces generated by dynamic XCF use a single IP address, the output of the SIOCGIFCONF ioctl() contains multiple entries with the same IP address. If an application is using the SIOCGIFCONF output to issue bind() to all the entries returned, the application could receive EADDRINUSE on a bind() if there are multiple XCF devices defined by dynamic XCF in the list.
2. If you want to define a static route to a link which is generated by dynamic XCF, you must wait until the dynamic devices are started and then use the VARY TCPIP,,OBEYFILE command. For IPv4, the GATEWAY or BEGINROUTES statement that refers to a dynamically defined linkname must be in a separate data set from the data set used to define the dynamic devices (either the initial profile data set or the data set referenced by a VARY TCPIP,,OBEYFILE command). For IPv6, the ROUTE statement that refers to a dynamically defined interface name must be in a separate data set from the data set used to define the dynamic interfaces (either the initial profile data set or the data set referenced by a VARY TCPIP,,OBEYFILE command).
3. Even though the HOME, BSDROUTINGPARMS and BEGINROUTES definitions are full replacement statements, the definitions generated by dynamic XCF will not replace any existing definitions. Likewise, user-defined HOME, BSDROUTINGPARMS and BEGINROUTES definitions will not affect existing or future definitions generated by dynamic XCF.
4. A mix of static and dynamic IPv4 and IPv6 definitions for a device is not allowed. For example, if a static IUTSAMEH IPv4 device and link is defined, an IPv6 dynamic definition for IUTSAMEH will not be created. If a static IUTSAMEH IPv6 interface is defined, an IPv4 dynamic definition for IUTSAMEH will not be created. The same logic also applies for XCF links; a mix of static and dynamic IPv4 and IPv6 definitions is not allowed for an XCF link.
5. The address specified on the IPCONFIG6 DYNAMICXCF statement cannot be an existing, statically defined, interface address. If a profile contains an IPv6 interface statement for an address that is also used in the IPCONFIG6 DYNAMICXCF statement, the IPCONFIG6 DYNAMICXCF statement is ignored.

Deleting dynamically defined XCF devices: You can delete dynamically defined XCF devices and links by first stopping the devices to be deleted and then issuing a VARY TCPIP,,OBEYFILE command that references a data set containing a DELETE LINK EZAXCF*nn* and DELETE DEVICE. Because the HOME statement processing does not affect dynamically defined XCF HOME list entries, the HOME *xx.xx.xx.xx* EZAXCF*nn* entry is automatically deleted by DELETE LINK.

For IPv6, you can delete dynamically defined XCF interfaces by first stopping the interface previously defined in an INTERFACE statement and then issuing a VARY TCPIP,,OBEYFILE command that references a data set containing an INTERFACE *interfacename* DELETE statement.

HiperSockets

HiperSockets is a zSeries hardware feature that provides high performance internal communications between LPARs within the same CPC without the use of any additional or external hardware equipment (for example, channel adapters, LANs, and so on).

If the host processor supports HiperSockets and Communications Server is properly configured, Communications Server will attempt to create XCF connectivity between same-CPC LPARs using a HiperSockets link or interface. In cases in which the HiperSockets link or interface cannot be activated, TCP/IP creates a normal XCF link or interface.

The HiperSockets for DYNAMICXCF device and link or interface are dynamically built and the device or interface is started during TCP/IP DYNAMICXCF stack initialization. The HiperSockets for DYNAMICXCF device and link and interface cannot be configured. The generated device name is IUTIQDIO. The generated IPv4 link name is IQDIOLNK*nnnnnnnn*, where *nnnnnnnn* is the character representation of the hexadecimal version of the DYNAMICXCF IP address. The generated IPv6 interface name is IQDIOINTF6. In general, where an XCF link or interface would normally have been used (for intra-CPC), a HiperSockets link or interface will be used.

Similar to IUTSAMEH, VTAM will dynamically build the TRLE for IUTIQDIO when the IUTIQDIO device is started. The TRLE statement is not configured (defined) by the user.

Although HiperSockets for DYNAMICXCF is not configured with TCP/IP DEVICE and LINK or INTERFACE statements, and the TRLE is not defined by the user, the following steps must be taken to define the HiperSockets subchannel devices and IQD CHPID:

1. Using HCD or IOCP, the system administrator must define (create the IOCDS) the HiperSockets IQD CHPID (channel path ID) and subchannel devices to the applicable LPARs. To dynamically build the HiperSockets TRLE, VTAM requires a minimum of 3 subchannel devices configured with each IQD CHPID within HCD. The maximum number of subchannel devices that VTAM will use (associate with each TRLE or MPC group) is 10. For additional details regarding configuring the HiperSockets subchannel devices and IQD CHPID, refer to *z/OS HCD Planning* and Appendix D, “Using HCD,” on page 1203.
2. When more than one IQD CHPID is configured to a specific LPAR, VTAM start option IQDCHPID must be used to specify which specific IQD CHPID this LPAR should use. The VTAM start option controls which IQD CHPID (and related subchannel devices) VTAM selects to include in the HiperSockets MPC group (IUTIQDIO) when it is dynamically built for HiperSockets for DYNAMICXCF connectivity. Start option IQDCHPID controls the VTAM IQD

CHPID selection for the HiperSockets for DYNAMICXCF device IUTIQDIO (MPC group) only. It does not control IQD CHPID selection for a user defined HiperSockets device (MPCIPA). However, a user defined HiperSockets device (IQD CHPID) cannot use (conflict with) the same IQD CHPID that the HiperSockets for DYNAMICXCF device is currently using.

For example, if IQD CHPID 'FE'x is currently in use by DYNAMICXCF due to one of the following:

a. VTAM start option IQDCHPID=FE is currently specified
b. VTAM start option IQDCHPID=ANY is currently specified, but the HiperSockets for DYNAMICXCF device IUTIQDIO is currently using the 'FE' CHPID

then an attempt to configure and start a user defined HiperSockets device IUTIQDFE will not be allowed (IQD CHPIDs conflict). This option can also be modified with a VTAM modify command. In most cases, the default setting will be sufficient. For additional details regarding this start option refer to the *z/OS Communications Server: SNA Resource Definition Reference*.

For additional details regarding HiperSockets, refer to "HiperSockets concepts and connectivity" on page 92.

Sysplex problem detection and recovery

Each sysplex member monitors itself and can automatically leave the sysplex if it determines that it is no longer able to function correctly as a router, target, backup, or owner of a DVIPA. Through a variety of methods, it monitors:

- Internal resources and conditions, such as timely execution of sysplex functions, available private storage, and common storage
- External resources, such as availability of VTAM, or OMPROUTE if it is being used

As long as the TCP/IP stack is a member of the TCP/IP sysplex group (EZBTCPCS), the sysplex monitor gets control periodically. The time period is determined from the TIMERSECS value specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in PROFILE.TCPIP. The default TIMERSECS value is 60 seconds.

Once a problem is detected, further actions depend on whether RECOVERY or NORECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. NORECOVERY is the default value.

Joining EZBTCPCS can be delayed until OMPROUTE is active by using the DELAYJOIN option on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. No sysplex-related definitions within the TCP/IP profile (that is, VIPADYNAMIC and DYAMICXCF statements) will be processed until the sysplex group is joined. During a planned or unplanned outage, the DVIPAs and distributed DVIPAs for a TCP/IP stack are taken over by backup TCP/IP stacks. When the primary TCP/IP stack is restarted, the DVIPAs and distributed DVIPAs are taken back from the backup TCP/IP stacks. If dynamic routing is used to advertise routes to these DVIPAs, and OMPROUTE is not yet active on the primary TCP/IP stack, existing connections to these DVIPAs might be reset and new connect requests to these DVIPAs might fail. By using the GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN configuration statement in the TCP/IP profile on the primary TCP/IP stack, it is possible to delay taking back the DVIPAs and distributed DVIPAs until OMPROUTE is started and active. Thus,

new and existing connections continue to be serviced by the backup TCP/IP stacks until OMPROUTE is active on the primary TCP/IP stack.

For more information on the GLOBALCONFIG statement and its parameters, refer to *z/OS Communications Server: IP Configuration Reference*.

Problem detection

If this stack is not the only member of the TCP/IP sysplex group and it is advertising DVIPAs (owns a DVIPA or is the primary routing node for a Distributed DVIPA), the following resource checks are made by the monitor:

- VTAM address space availability

  If the VTAM address space is not currently active and the elapsed time since VTAM was last detected as active has exceeded the TIMERSECS value, message EZZ9671E is issued.
- Route availability

  If there are no routes available to all partners, given the configured method chosen for forwarding data (VIPAROUTE statement or dynamic XCF interfaces), there are at least two other MVS systems in the sysplex, and the elapsed time since an active route was detected has exceeded the TIMERSECS value, message EZZ9673E or EZD1172E is issued.

If OMPROUTE was successfully initialized, it periodically sends a heartbeat to the stack, so the monitor can always make the following resource checks:

- If the elapsed time since a heartbeat was received has exceeded half of the TIMERSECS value, message EZZ9672E is issued as a warning and no other actions occur.
- If the stack is not the only member of a TCP/IP sysplex group, is advertising DVIPAs, and the elapsed time since a heartbeat was received has exceeded the TIMERSECS value, message EZZ9678E is issued.

Ensure that the WLM policy for the OMPROUTE address space receives sufficient priority in relationship to other work managed by WLM on the system, so that OMPROUTE receives the CPU cycles necessary for this task. For more information on this topic, see "Sysplex autonomics" on page 239.

If this stack is advertising DVIPAs or is a Sysplex Distributor target, the following checks are made:

- The monitor checks for CSM storage availability. If CSM storage is continuously critical for greater than the TIMERSECS value, message EZZ9679E is issued.
- If TCP/IP ECSA storage limits are defined on the ECSALIMIT parameter of the GLOBALCONFIG statement, monitoring of this storage is similar to CSM monitoring. If TCP/IP ECSA storage is continuously critical for a time greater than the TIMERSECS value, message EZD1187E is issued. If there are no TCP/IP ECSA storage limits defined on the GLOBALCONFIG statement and a storage request cannot be satisfied, message EZD1170E is issued. These messages indicate that there is a TCP/IP problem with ECSA storage.
- If TCP/IP private storage limits are defined on the POOLLIMIT parameter of the GLOBALCONFIG statement, monitoring of this storage is similar to CSM monitoring. If TCP/IP private storage is continuously critical for a time greater than the TIMERSECS value, message EZD1187E is issued. If there are no TCP/IP private storage limits defined on the GLOBALCONFIG statement and a storage request cannot be satisfied, message EZD1170E is issued. These messages indicate that there is a TCP/IP problem with private storage.

Note, however, that if all DVIPAs on this stack have a status of MOVING and the stack is not a DVIPA target, the checks above are not made.

In addition to the sysplex monitor, when the stack joins the TCP/IP sysplex group it requests that XCF monitor the TCP/IP sysplex component on the local stack. The XCF component performs this function by monitoring a status field updated by the TCP/IP sysplex component. If XCF detects that the sysplex functions have not been responsive for the TIMERSECS value, it schedules a TCP/IP routine that issues message EZZ9674E.

If TCP/IP encounters a nonrecoverable sysplex error, it issues the eventual action message EZZ9670E.

All of the above messages are eventual action messages. If the detected problem condition is corrected (for example, VTAM is started), the eventual action message is cleared. For more information on these eventual action messages, refer to *z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM)*.

Recovery

Recovery actions occur if RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. If the default setting, SYSPLEXMONITOR NORECOVERY, is active, other than issuing the message, no further actions occur if the problem is not corrected.

The VARY TCPIP,,SYSPLEX,LEAVEGROUP command can be used to manually force the sysplex member to leave the TCP/IP sysplex group. As a stack leaves the TCP/IP sysplex group, message EZZ9670E is cleared, as well as message EZD1170E. All other outstanding eventual action messages are cleared when the condition is cleared (for example, starting VTAM). For information on the VARY TCPIP,,SYSPLEX command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in the TCP/IP profile, and this stack is not the only member of the TCP/IP sysplex group, the stack leaves the TCP/IP sysplex group when one of the messages is issued. The one exception to this is EZZ9672E, which is only issued as an OMPROUTE warning message. No actions occur unless the corresponding EZZ9678E OMPROUTE message is subsequently issued.

To determine if the stack is currently joined to the TCP/IP sysplex group, issue the DISPLAY TCPIP,,SYSPLEX,VIPAD command. If the stack is not currently joined to the TCP/IP sysplex group, this command displays the following message:

```
EZZ8269I tcpstackname mvsname IS NOT A MEMBER OF THE SYSPLEX GROUP
```

If you are not currently logged on to the stack, use the D XCF,GROUP,EZBTCPCS command to see the systems currently in the sysplex.

If the RECOVERY option is specified and a TCP/IP stack initiates an automated recovery action by leaving the TCP/IP sysplex group, all local DVIPAs are inactivated and all the VIPADYNAMIC block definitions are saved. Any applications bound to dynamically created DVIPAs (VIPARANGE or MODDVIPA) will receive an asynchronous error, `EUNATCH (3448) - the protocol required to support the address family is unavailable`.

If internal problems prevent the removal of these resources, eventual action message EZZ9675E is issued, and restarting the stack is necessary to be able to

become part of the TCP/IP sysplex group. If all DVIPAs are successfully removed, eventual action message EZZ9676E is issued, indicating that sysplex problem detection cleanup has succeeded. There are two ways for the stack to rejoin the sysplex group and clear this message after a successful cleanup has occurred:

- If AUTOREJOIN was configured on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, the stack automatically rejoins the group and reprocesses its saved VIPADYNAMIC configuration when all detected problems have been relieved. The AUTOREJOIN option is the recommended setting when the RECOVERY option is configured.
- Issue the VARY TCPIP,,SYSPLEX,JOINGROUP command to cause the stack to rejoin the group, and reprocess its VIPADYNAMIC configuration.

Recovery is the preferred method of operation since this allows other members of the TCP/IP sysplex to automatically take over the functions of a member with no actions needed by an operator. There are, however, some environments and scenarios where this automated recovery action might not be desirable and perhaps should not be enabled:

- DVIPAs or Distributed DVIPAs are defined, but no backup TCP/IP stacks are identified or no provisions are made to move the DVIPAs in cases of failure.

  The basic premise of the automated recovery actions is that one or more other TCP/IP stacks in the system can pick up ownership responsibilities for any DVIPAs owned by the failing TCP/IP stack. If this is not the case, it is suggested that you carefully evaluate the benefits of designating backup TCP/IP stacks and implement a configuration that includes backup capabilities. If this is not possible or desirable, RECOVERY should not be specified. If RECOVERY is not specified, automated recovery actions are disabled by default.

  For example, one such configuration is if you are only using DVIPAs that are always bound to a specific TCP/IP stack (that is, in lieu of static VIPAs). In this scenario, since there is no possibility of having ownership of these DVIPAs transferred automatically, there is no value in triggering the automated recovery action and you should consider not enabling the automated recovery function or using static VIPAs (since static VIPAs are not affected by the automated recovery actions).
- Test environments where individual system images have very limited resources (CPU, storage, and so on).

  This can include environments where you are running z/OS as a second level guest under z/VM, or in LPARs with shared processors and very limited resources. Not enabling the automated recovery actions in these environments can help prevent unwanted recovery actions that are triggered by false positive conditions, such as scenarios where artificial severe resource shortages are detected.
- Environments where VTAM or OMPROUTE are stopped for intervals longer than the TIMERSECS value specified on the SYSPLEXMONITOR parameter.

  If your current operations procedures have provisions for stopping VTAM or OMPROUTE for extended periods of time, you should consider disabling and re-enabling the automated recovery processing around the periods of time where you stop and restart these components. This can be accomplished using the VARY TCPIP,,OBEYFILE command.

  An alternative solution could be to increase the TIMERSECS value to accommodate the longest period of time you would expect VTAM or OMPROUTE to be inactive during normal operating procedures. One potential drawback of this approach is that the monitoring of other conditions and triggering of automatic recovery functions is less responsive.

Setting TIMERSECS

As discussed, the responsiveness of the self-monitoring and automated recovery actions is governed by the TIMERSECS value specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. The default value is 60 seconds. You should evaluate whether this default value is appropriate in your environment and, if not, specify a value that is better suited to your environment. Following are some of the considerations that should be evaluated in selecting an appropriate value:

- Specifying a smaller TIMERSECS value increases the responsiveness of the self-monitoring functions.

  The monitor periodically performs its checks for problem conditions, every 60 seconds or every quarter of the TIMERSECS interval (TIMERSECS/4), whichever is less. Therefore, when a smaller TIMERSECS value is specified, operator message alerts are issued sooner if a problem condition is detected. Automated recovery actions are also triggered sooner if the RECOVERY option is in effect.

  While this might be desirable, care should be taken to not specify a timer interval value that is small enough to trigger recovery actions for conditions that are transient in nature. For example, the value specified should probably be larger than the spin loop timeout interval for the system (SPINTIME parameter in the EXSPAT*xx* parmlib member). Otherwise, a spin loop timeout condition that might be recoverable could trigger an unnecessary TCP/IP sysplex recovery action, as a result of the TCP/IP sysplex monitor not being able to be dispatched temporarily before any spin loop timeout recovery actions can take place. A good rule of thumb is to specify a TIMERSECS value that is at least two times larger than the spin loop timeout interval in effect for the system.
- Specifying a longer TIMERSECS value has the opposite effect, and the self-monitoring and automated recovery actions are less responsive.

  In addition, a longer TIMERSECS value can cause disruptions to existing TCP connections when problem conditions are detected, even if automated recovery actions are initiated. For example, if the problem condition persists for a long enough interval and it impacts the delivery of data on existing TCP connections, the remote TCP/IP hosts might terminate these connections before a recovery action is initiated.
- Review your current settings for the sysplex failure detection interval (INTERVAL keyword in COUPLE*xx*) and the ISOLATETIME value specified in your SFM policy, if defined. These values indicate how responsive the XCF component should be in removing systems from the sysplex when a status update missing condition occurs. As a result, these settings can provide a reasonable reference point for the setting of the TIMERSECS value. For more information on these intervals, refer to *z/OS MVS Setting Up a Sysplex*.

Summary of problems monitored and actions taken

When the sysplex monitor detects a problem, an operator message is issued. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in the TCP/IP profile, the stack leaves the sysplex group, deletes all its dynamic VIPAs, and inactivates all its VIPADYNAMIC definitions. When the detected problem is relieved, the operator message is deleted. If AUTOREJOIN was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, the stack might or might not rejoin the sysplex group, depending on the severity of the problem.

Table 18 on page 381 summarizes the various problems that are monitored and the specific actions that are taken for each detected problem.

Table 18. Sysplex problem monitoring

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
VTAM address space is not active	DVIPAs advertised and VTAM address space has been inactive greater than TIMERSECS interval	1. Issue EZZ9671E, VTAM inactive for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when VTAM is started.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
Dynamic or static XCF route, or VIPAROUTE	DVIPAs advertised, acting as forwarder for a DVIPA, two other MVS systems in the sysplex, and either XCF routes (when VIPAROUTE is not specified) or all IP routes (when VIPAROUTE is enabled) to all sysplex partners are not available for TIMERSECS interval	1. If VIPAROUTE is not enabled, issue message EZZ9673E, dynamic XCF connectivity to all partners not available for at least TIMERSECS interval. 2. If VIPAROUTE is enabled, issue message EZD1172E , dynamic XCF connectivity and IP routes to all partners are not available for at least TIMERSECS interval. 3. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 4. Delete the operator message when any VIPAROUTE or XCF route becomes active.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
OMPROUTE	First heartbeat received, and no heartbeat received for half of the TIMERSECS interval	1. Issue message EZZ9672E, the OMPROUTE not responsive warning message. 2. Delete the operator message when heartbeat received, or when message EZZ9678E is issued due to OMPROUTE not responding for at least TIMERSECS interval.	Not applicable

Table 18. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
OMPROUTE	First heartbeat received, DVIPAs advertised, and no heartbeat received for TIMERSECS interval	1. Issue message EZZ9678E, OMPROUTE not responsive for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. If a last heartbeat was not received (OMPROUTE did not terminate properly), abend and dump the TCP/IP and OMPROUTE address space. 3. Delete the operator message when heartbeat received.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
TCP/IP private storage, when POOLLIMIT is not coded on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP private storage allocation fails	1. Issue message EZD1170E, TCP/IP private storage allocation failed message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP private allocation requests are successful for a full monitor interval (1/4 of the TIMERSECS interval or 1 minute, whichever is less), or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.

Table 18. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
TCP/IP private storage, when POOLLIMIT values are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP private pool is critical for TIMERSECS interval. If TCP/IP private storage allocation request from MVS fails before the TCP/IP private pool is critical for TIMERSECS interval, the problem is treated as if no POOLLIMIT values were specified.	1. Issue message EZD1187E, TCP/IP private storage is critical. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP private pool storage exits the critical state.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
TCP/IP ECSA storage limits are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP ECSA is critical for TIMERSECS interval. If TCP/IP ECSA storage allocation request from MVS fails before TCP/IP is critical for TIMERSECS interval, the problem is treated as if no TCP/IP limits were specified.	1. Issue message EZD1187E, TCP/IP ECSA storage is critical. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP ECSA critical state is exited.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Table 18. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
No TCP/IP ECSA storage limits are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP ECSA storage allocation request fails	1. Issue message EZD1170E, TCP/IP ECSA storage allocation failed message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP ECSA allocation requests are successful for a full monitor interval (1/4 of the TIMERSECS interval or 1 minute, whichever is less), or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.
CSM storage critical	DVIPAs advertised or this is a DVIPA target, and CSM critical for greater than TIMERSECS interval	1. Issue message EZZ9679E, CSM critical message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when CSM critical state exited.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Table 18. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
XCF status	Status not updated for TIMERSECS interval	1. Issue message EZZ9674E, sysplex processing not responsive message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when the monitor is able to run, or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.
Abend in sysplex code		1. Issue message EZZ9670E, sysplex processing encountered nonrecoverable error. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave sysplex group, inactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.

Target server connection setup responsiveness monitoring

Sysplex Distributor measures the responsiveness of target servers in accepting new TCP connection requests at intervals of approximately 1 minute. Target server responsiveness (TSR) values calculated from these measurements are used to modify the WLM recommendation when using WLM system weight or server-specific WLM distribution, to favor those servers that are more successfully accepting new TCP connection setup requests. The TSRs are displayed as a percentage. Thus, a value of 100 indicates full responsiveness and a value of 0 indicates no responsiveness in setting up new TCP connections. A value of 0 for the TSR causes this target server to be bypassed when new TCP connection setup requests are distributed to target servers for a DVIPA and port, even if the

distribution method is round-robin. Periodically, the distributor sends a new connection request to a target with a TSR of 0, to check whether the responsiveness of that target has improved.

The calculated TSR values are applied only to the WLM weights if there are no stacks participating in the distribution of connections for this DVIPA that are at a release level earlier than z/OS V1R7. The TSRs are calculated by combining two factors, the target connectivity success rate (TCSR) and the Servers' accept Efficiency Fraction (SEF). These factors are defined as follows:

TCSR The target connectivity success rate is the measure of how many new TCP connection setup requests (SYNs) sent by the distributing stack to a target stack are received by the target stack. A low value could indicate a problem with the connectivity between the distributing stack and the target stack.

SEF The Servers' accept Efficiency Fraction measures how well a target server is accepting new TCP connection setup requests and managing its backlog queue.

The connection establishment rate (CER) is a subcomponent of the SEF, and measures how many TCP connection setup requests received by the target stack become established (that is, how many enter the ESTAB state). A low value could indicate a problem with the target server establishing connections with the client.

Workload balancing

Load balancing is the ability for a cluster to spread workload evenly (or based on some policy) to target servers comprising the cluster. Usually, this load balancing is calculated based on the perceived load on each of the target servers. Various techniques are available to perform IP load balancing for a zSeries sysplex. These techniques typically fall into the following categories:

- Internal load balancing solutions

  These solutions typically rely on a component executing within the z/OS sysplex to perform the load balancing with very few dependencies on the external network. Because they are executing within the sysplex environment, these solutions typically have access to information that can be used to optimize load balancing, such as system capacity and current load information. Examples of this technique include the DNS/WLM and Sysplex Distributor solutions.
- Sysplex-aware external load balancing solutions

  These solutions are typically comprised of an external load balancer that relies on components inside the z/OS sysplex for advice on how workload should be distributed. As a result, these solutions usually provide for load balancing that is optimized for a z/OS sysplex environment. Examples of these solutions include the IBM Network Dispatcher and any external load balancing solution that supports the Server/Application State Protocol (SASP) provided by the z/OS Load Balancing Advisor (for example, the CISCO Content Switching Module).
- External load balancing solutions

  These solutions are usually comprised of load balancing components that execute outside of the z/OS sysplex, typically on one or more hosts or routers in the network, with little or no specific awareness of the z/OS sysplex environment. Several vendors provide such load balancing solutions.

This section describes and compares these three load balancing techniques and selected associated solutions. Each identifies the target z/Series servers willing to receive client connections based on some specification.

By providing load balancing, clustering techniques must also provide for other system requirements in addition to the dispatching of connections. These include the ability to advertise some single systemwide image or identity so that clients can uniquely and easily identify the service. Additionally, clustering techniques should also provide for horizontal growth of the system and ease of management.

Single systemwide image

Clients connecting to a cluster should not be aware of the internal makeup of a cluster. More specifically, clients should not even be aware that the service they are requesting is actually being serviced by a collection or cluster of servers. Instead, clients must be provided with some single image identifier to be used when connecting to the service. For example, most of the load balancing solutions (internal or external) discussed in this section (other than DNS/WLM) use a single IP address to represent a cluster of servers to clients. The clients simply use this IP address to establish connections and are not aware that the load balancing solution directs requests to a specific instance of the server.

Horizontal growth

As the clients' demands on the service increase, clusters must provide a way to expand the cluster of servers to accommodate for such growing demand. Put in another way, the cluster must provide a mechanism by which to add servers without disrupting the operation of the cluster. To this end, the service is made available to clients at all times and can grow horizontally to accommodate for increased demand placed on the cluster by the clients.

Ease of management

The administrative burden associated with the cluster should not increase as you add servers to the cluster. It is desirable to use the same configurations for many systems in the cluster (sysplex). Within a sysplex, servers are homogenous, since a sysplex is comprised solely of z/Series servers. As such, many of the configurations can be shared among the different z/Series servers, thereby reducing the administrative burden associated with the sysplex. Additionally, as the size of the cluster increases, the administrative overhead in adding systems to the cluster should be as low as possible.

Administrative activities required to maintain a load balancing solution are highly dependent on the type of load balancing solution deployed. For example, administrative tasks associated with the maintenance of internal load balancing solutions typically consist of administrative tasks within the z/OS sysplex, such as changing z/OS configuration files. External load balancing solutions typically require administrative tasks performed on the external load balancing components, while sysplex-aware external load balancing solutions might require administrative tasks on both internal and external load balancing components. Depending on your environment and organizational structure, the administrative scope required by the various load balancing solutions can play a key role in your selection process. For example, do you need to deploy a solution that requires only administrative tasks on z/OS, or can you make required updates to network components as necessary?

Internal load balancing solutions

Internal load balancing solutions typically rely on components that reside within the z/OS sysplex to perform the load balancing. While other components might be required in the network, after the solution is configured, minimal changes should be needed from network components outside of the sysplex. Examples of internal load balancing solutions include the Sysplex Distributor and the DNS/WLM solutions.

Sysplex Distributor

Sysplex Distributor extends the notion of dynamic VIPA and automatic VIPA takeover to allow for load distribution among target servers within the sysplex. It extends the capabilities of dynamic VIPAs to enable distribution of incoming TCP connections to ensure high availability of a particular service within the sysplex.

The functionality of Sysplex Distributor is that one IP entity advertises ownership of an IP address by which a particular service is known. In this fashion, the single system image of Sysplex Distributor is that of a special IP address. This IP address is called a distributed DVIPA. Further, in Sysplex Distributor, the IP entity advertising the distributed DVIPA and dispatching connections destined for it is itself a system image within the sysplex, referred to as the distributing stack.

Sysplex Distributor makes use of Workload Manager (WLM) and its ability to gauge server load and provide a WLM recommendation. In this paradigm, WLM provides the distributing stack with a WLM recommendation (a WLM system weight) for each target system, or the target stacks provide the distributing stack with a WLM recommendation (a server-specific WLM weight) for each target server. The distributing stack uses this information to optimally distribute incoming connection requests between a set of available servers. Additionally, Sysplex Distributor has the ability to specify certain policies within the Policy Agent so that it can use QoS information from target stacks to further modify the WLM recommendation. Further, these policies can specify which target stacks are candidates for clients in particular subnetworks.

Sysplex Distributor also measures the responsiveness of target servers in accepting new TCP connection setup requests, favoring those servers that are more successfully accepting new requests. For more information, see "Target server connection setup responsiveness monitoring" on page 385.

The Workload Manager weight is a relative value. The WLM system weight is an indication of the target system's capacity for additional work, and the server-specific WLM weight is a more granular indication of the specific target server's capacity for additional work. Higher numbers indicate a target with comparatively greater capacity. The Netstat VDPT/-O display shows these values. WLM system weights are indicated by a `B` beside the weight, and WLM server-specific weights are indicated by an `S` beside the weight.

Distribution using WLM system weights is the default distribution method. This can be specified by using the BASEWLM parameter on the VIPADISTRIBUTE statement. Distribution using server-specific WLM weights can be specified by using the SERVERWLM parameter on the VIPADISTRIBUTE statement. If the SERVERWLM parameter is used and all stacks are able to provide server-specific WLM weights for that VIPA/port, server-specific WLM weights are used. Otherwise, WLM system weights are used.

When determining a WLM system weight recommendation, WLM assigns a relative weight to each system in the sysplex, with the highest weight going to the

system with the most available capacity. If all systems in the sysplex are running at or near 100%, WLM assigns the highest weights to the systems with the largest amount of lower importance work. In this way, new connection requests are distributed to the systems with the highest displaceable capacity. However, this assumes that the new work is of a high enough importance level to displace this work. A system can be so loaded that only higher importance work is running on that system, in which case the new work request is not able to meet the goals specified in the WLM policy for that server. For more information on how displaceable capacity is calculated and using the Sysplex Routing Services, refer to *z/OS MVS Programming: Workload Management Services*.

When determining a server-specific WLM recommendation, WLM determines the service class of the server's address space, and then assigns a weight based on the following:

- How well that server is actually meeting the goals of its service class
- The amount of displaceable workload available, given the importance of the service class

For most applications, the server-specific WLM recommendations provide a more accurate way to distribute workload to the servers. However, when a server acts as an access point to applications that run in other address spaces (and therefore in a different service class), WLM system weights might be the preferred distribution method. For example:

- The TN3270 server is a communication gateway function that enables clients to access SNA applications over an IP network. As a result, most of the actual work associated with a TN3270 workload takes place in the SNA application, which in all likelihood is classified to a different WLM service class than the TN3270 server and probably at a lower importance level. Therefore, while the server-specific WLM recommendation would provide an accurate reflection of how well an individual TN3270 server is performing, it would not necessarily provide an accurate reflection of the available capacity required by the back-end SNA applications that the client is accessing. As a result, WLM system weight is probably a more appropriate distribution method for this server.
- The INET daemon also provides access to other applications that are probably associated with service classes of a lower importance level (for example, z/OS UNIX Telnet, REXECD, and RSHD). As a result, similar considerations apply, and WLM system weights are the more appropriate distribution method for this server.
- The FTP daemon address space typically performs very little processing on behalf of new FTP sessions. After accepting a new connection, it performs a fork() for an FTP server process that will service the new FTP session. The FTP server process is classified again by WLM, possibly resulting in a different service class than that of the FTP daemon.
  - If the FTP servers run in a different service class than the FTP daemon, then WLM system weights should be used.
  - If WLM policies are set up so that the FTP servers are classified in the same service class as the FTP daemon, server-specific WLM weights should be the distribution method. Although the WLM recommendation will not take into account how well the actual FTP server is meeting its goal, the recommendation will more accurately reflect the amount of displaceable capacity because it is based on the importance of the service class.

Specifying the SHAREPORT parameter on the PORT statement in the TCP/IP profile enables a group of servers to listen on the same port, and thereby share the

incoming workload. As new connections are received, the number of active connections is evenly balanced across the available servers using a weighted, round-robin distribution based on the Servers' accept Efficiency Fractions (SEFs). Specifying the SHAREPORTWLM parameter on the PORT statement enables connections to be distributed in a weighted, round-robin fashion based on the WLM server-specific recommendations modified by the SEFs. For more information on SEFs, see "Target server connection setup responsiveness monitoring" on page 385.

Use the same considerations concerning application type to determine the type of port sharing distribution to use. If the shared port is a sysplex distributed port and server-specific WLM weight is the distribution method that is being used by the distributor, code the SHAREPORTWLM paremeter on each target's PORT statement to take advantage of the WLM server-specific recommendations when connections are received at the target. Otherwise, use the SHAREPORT parameter on the PORT statement.

In the following Netstat VDPT/-O example, the weights represent normalized weights. That is, the original raw weights received from WLM are proportionally reduced for use by the distribution algorithm. Connections are distributed to these servers in a weighted, round-robin fashion using the normalized weights. In this example, the target server responsiveness (TSR) values are all 100, indicating that all servers are fully responsive to new connection requests. A value of 100 is also displayed for stacks that are at a level prior to z/OS V1R7. In that case, no target server responsiveness calculations are applied to the WLM values.

The SERVERWLM parameter was not coded for DVIPA 201.2.10.11, port 245, so it is using WLM system weights.

```
MVS TCP/IP NETSTAT CS V1R7       TCPIP NAME: TCPCS            12:19:18
Dynamic VIPA Distribution Port Table:
Dest IPaddr     DPort DestXCF Addr    Rdy TotalConn  WLM  TSR Flg
-----------     ----- ------------    --- ---------  ---  --- ---
201.2.10.11     00245 201.1.10.10     001 0000000000 12   100 DB
201.2.10.11     00245 201.1.10.15     001 0000000000 04   100 B
201.2.10.12     04011 201.1.10.10     001 0000000000 04   100 S
201.2.10.12     04011 201.1.10.15     001 0000000000 08   100 S
201.2.10.12     04011 201.1.10.40     001 0000000000 16   100 S
```

If the BASEWLM parameter or the SERVERWLM parameter is specified and WLM weights are not available, incoming connections are distributed among all available servers using round-robin distribution.

Now consider the same example, as shown in the following Netstat VDPT/-O display, but with TSR values indicating that some of the servers are not accepting new connection setup requests productively. The displayed WLM weights have been modified by the TSR values:

```
MVS TCP/IP NETSTAT CS V1R7       TCPIP NAME: TCPCS            12:19:18
Dynamic VIPA Distribution Port Table:
Dest IPaddr     DPort DestXCF Addr    Rdy TotalConn  WLM  TSR Flg
-----------     ----- ------------    --- ---------  ---  --- ---
201.2.10.11     00245 201.1.10.10     001 0000000000 09   075 DB
201.2.10.11     00245 201.1.10.15     001 0000000000 04   100 B
201.2.10.12     04011 201.1.10.10     001 0000000000 03   090 S
201.2.10.12     04011 201.1.10.15     001 0000000000 06   075 S
201.2.10.12     04011 201.1.10.40     001 0000000000 03   020 S
```

The SERVERWLM parameter was coded for DVIPA 201.2.10.12, port 4011. For the server on destination 201.1.10.10, a TSR value of 90 indicates that the server is 90%

responsive in accepting new connection requests. For the server on 201.1.10.15, a TSR value of 75 indicates that this server is 75% responsive in handling new connection requests. Likewise, for the server on 201.1.10.40, a TSR value of 20 indicates that it is only 20% effective in accepting new connection requests. These factors are used to modify the server-specific WLM weights, and the modified weights are normalized. As a result, the server on destination XCF 201.1.10.10 has a normalized WLM value of 3, the server on XCF 201.1.10.40 has a WLM value of 3, and the server on destination XCF 201.1.10.15 has a WLM value of 6. So, the server on destination XCF 201.1.10.15 is now favored over the other servers, and the server at 201.1.10.40 is now equal to the server at 201.1.10.10. Connections are distributed to these servers in a weighted, round-robin fashion using these normalized weights.

The SERVERWLM parameter was not coded for DVIPA 201.2.10.11 port 245, so it is using WLM system weights. The server at 201.1.10.10 has a TSR of 75, so the normalized weight is 9, three-quarters of what it was in the previous example. The server at 201.1.10.15 has a TSR of 100, so the normalized weight is the same as it was in the previous example.

In certain scenarios, a round-robin distribution might be appropriate for specific server applications. You can select round-robin distribution among DVIPA/port targets for a particular application, rather than distribution based on WLM recommendations and policy information, by specifying the ROUNDROBIN parameter on the VIPADISTRIBUTE statement.

This distribution value does not have any affect on incoming connection requests that have an active affinity established to a specific server instance (through the TIMEDAFFINITY parameter). When an affinity exists, it has priority over the distribution method setting.

As with Network Dispatcher, connection requests are directed to the distributed stack of Sysplex Distributor. The stack selects which target server is the best candidate to receive an individual request and routes the request to it. All inbound traffic for these connections must be routed to the distributing stack. The distributing stack maintains state so that it can forward data packets associated with this connection to the correct stack. Additionally, data sent from servers within the sysplex need not travel through the distributing stack.

Tip: In networks that have CISCO routers, it might be possible to remove the requirement that all inbound traffic needs to traverse the distributing stack. For more details, see “Optimized connection load balancing using Sysplex Distributor in a network with CISCO routers (IPv4 only)” on page 393.

Sysplex Distributor also enhances the dynamic VIPA and automatic VIPA takeover functions. The enhancements allow a DVIPA to move nondisruptively to another stack. That is, in the past, a DVIPA was only allowed to be active on one single stack in the sysplex. This led to potential disruptions in service when connections existed on one stack, yet the intent was to move the DVIPA to another stack. With Sysplex Distributor, the movement of DVIPAs can now occur without disrupting existing connections on the original DVIPA owning stack.

Refer to “Configuring distributed DVIPAs — Sysplex Distributor” on page 316 for more information.

Policy interactions: The Policy Agent interacts with the Sysplex Distributor to assist with workload balancing. There will be one Policy Agent running on an

LPAR regardless of how many stacks are configured. First, the Policy Agent can be configured to collect network performance statistics for applications being distributed on target stacks. These network performance statistics are then used to modify the overall WLM weight assigned to a target server. In this way, processor performance, server performance, and application network performance are taken into account when distributing work. Second, policies established on the distributing stack can be configured to restrict the set of target stacks to be considered for any given inbound connection request. In this way, the total set of target stacks can be partitioned among different groups of users or applications requesting connections to distributed applications.

Previously, the QoS performance data was collected by the Policy Agent on the target for each DVIPA and port or application. After collecting the QoS information, the Policy Agent on the target stack pushed this information down to the stack sysplex function which then forwarded it to the stack sysplex function on the distributing stack. There are two significant additions to Policy Agent and sysplex interaction:

- The Policy Agent at each target will collect information with an additional level of granularity; the QoS performance data will be collected for each service level that a target's DVIPA port or application supports.
- The Policy Agent on the distributing stack drives the collection of this information by pulling it from the Policy Agents on the target stacks:
  - The Policy Agent on the distributor opens up one or two TCP connections to each of the Policy Agents on the target stacks. The distributor can be configured to distribute connections to IPv4 DVIPAs, IPv6 DVIPAs, or both. The Policy Agent opens an IPv4 connection to a target stack's IPv4 XCF address if it is configured to distribute to an IPv4 DVIPA on that target, and likewise opens an IPv6 connection to a target stack's IPv6 XCF address if it is configured to distribute to an IPv6 DVIPA on that target. As a result, if the routing stack is distributing connections to both IPv4 and IPv6 DVIPAs on a given target, then two connections to that target are opened.

    For more information on how the Sysplex Distributor determines its targets, refer to "Configuring distributed DVIPAs — Sysplex Distributor" on page 316.
  - The Policy Agent on the distributing stack will send across a list of QoS service level names to the Policy Agent on each target.
  - The Policy Agent on each target will send back a QoS Policy Action weight fraction for each requested service level that each target DVIPA port/application supports. A specific Policy Action weight fraction will not be sent unless the distributing stack's Policy Agent requests it. Only weight fractions for IPv4 DVIPA ports and applications flow on the IPv4 Policy Agent connection, and similarly, only IPv6 weight fractions flow on the IPv6 Policy Agent connection.
  - Upon receiving the QoS Policy Action weight fraction, the Policy Agent on the distributing stack will pass this information down to the sysplex distribution function on the stack. If two connections to a given target Policy Agent are active, weight fractions for IPv4 DVIPA ports or applications and IPv6 DVIPA ports or applications are passed to the sysplex distribution function separately. The stack sysplex distribution function uses this additional information when it is selecting targets for incoming connections. If it does not have a QoS Policy Action weight fraction, then it uses the existing weight fraction described above to make the load distribution decision instead.

Steps for enabling Policy Agent load distribution functions: Perform the following steps to enable Policy Agent load distribution functions:

1. Define the PolicyPerfMonitorForSDR statement in the PAGENT configuration file to enable the policy performance monitor function. This function must be active on the target and distributing stacks.

   ---

2. z/OS Communications Server load distribution needs to be specifically enabled for each service level; a Policy Action with the same service level name needs to be defined on each of the appropriate target stacks and also on the distributing stack for these targets. Note that it is reasonable to have a subset of key service level names defined to the distributing stack. Traffic mapping to those service level names that are defined to the distributing stack will receive z/OS Communications Server load distribution by service level. All other traffic will receive Communications Server V2R10 load distribution.

   ---

3. A backup distributing stack must have the same Policy Action configuration definitions as the active distributing stack for the corresponding DVIPA targets which it is backing up, if it is desired that the Policy Action behavior stay the same when the backup distributing stack takes ownership of the DVIPA. It will also need to have the Policy Agent performance monitor function active.

   ---

4. Common PAGENT port numbers will be used by the listener (pagentQosListener) and the collector (pagentQosCollector) They are part of the /etc/services install file. If PAGENT is running on an LPAR containing a target stack, it will open a listening connection using the pagentQosListener port number. PAGENT running on an LPAR containing a distributing stack will establish a TCP connection with each PAGENT listener using the pagentQosCollector as the source port and the pagentQosListener as the destination port. The listener will fail a connect request if the source/destination port does not match the defined collector/listener port. The /etc/services file on all LPARs in the sysplex must be updated to contain these port numbers.

   ---

5. Define the two port numbers mentioned above as reserved ports for PAGENT via the PORT statement in the PROFILE.TCPIP data set.

   ---

6. Define the DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statements in PROFILE.TCPIP. The PAGENT TCP connections use the XCF IP addresses.

For more information on setting up polices, refer to "Sysplex Distributor policy performance monitoring configuration" on page 755, "Sysplex Distributor policy example" on page 761, or "Sysplex Distributor routing policy example" on page 769.

Optimized connection load balancing using Sysplex Distributor in a network with CISCO routers (IPv4 only): The IBM Sysplex Distributor function provides a workload balancing function within a parallel sysplex. The Sysplex Distributor consists of a primary distributor stack (denoted by a dynamic VIPA) and a set of target stacks. An inbound packet destined for that DVIPA flows through the

primary distributor stack which then forwards the packet over an internal link (XCF, IUTSAMEH, or HiperSockets) to the selected target stack.

The Cisco Multi-Node Load Balancer (MNLB) provides a workload balancing function which distributes traffic through Cisco routers across multiple destination TCP/IP stacks. The MNLB consists of a service manager (the Cisco local director which is denoted by a cluster IP address) and a set of forwarding agents (Cisco routers). For a TCP connection to the cluster IP address, the forwarding agent sends the SYN packet to the service manager, which then selects a target stack and notifies the forwarding agent of this decision. The forwarding agent then sends all future packets for that TCP connection directly to the target stack.

A solution is available to enable you to use a combination of the Sysplex Distributor and the MNLB to provide workload balancing.

The scope of a cluster IP address managed by Sysplex Distributor is still a single sysplex, and integration with Cisco forwarding agents merely allows the Sysplex Distributor routing stack to be bypassed for inbound traffic. If workload balancing for a single cluster IP address across nodes in multiple clusters (sysplexes) is desired, MNLB using Cisco local director as the service manager will continue to be used. Sysplex Distributor will continue to advertise network ownership of the cluster IP address with any attached routing daemon so that Sysplex Distributor appearance and behavior toward the attached routing network is unchanged except for its new relationship with Cisco forwarding agents.

This solution allows the choice of providing the workload distribution inside the sysplex, outside the sysplex, or a combination of both.

Steps for setting up Sysplex Distributor to be the service manager for the Cisco MNLB (IPv4 only): Perform the following steps to set up Sysplex Distributor to be the service manager for the Cisco MNLB:

1. The Cisco router must be configured as a forwarding agent. The IP CASA control address (which is NOT the interface address to the forwarding agent) must be advertised by the Cisco routing daemons. This is not automatically done by Cisco and must be enabled by a Cisco command. For more information on the commands, refer to online documentation for Cisco at:

   `http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120t/120t5/ipclus.htm`

   ---

2. Specify the SERVICEMGR keyword on the VIPADEFine statement in the TCPIP profile.

   ---

3. Specify the VIPASMparms statement in the TCPIP profile. Specify the same multicast group and UDP port on the VIPASMparms statement in the TCPIP profile as are configured in the MNLB.

   ---

4. Optionally, use MD5 authentication:

   Specify the same password (MD5 key) on the VIPASMparms statement in the TCPIP profile as is configured on the Cisco routers which will communicate with the Sysplex Distributor. If a password (SMPASSword) is specified, then the Sysplex Distributor will perform MD5 authentication for all communications with the Cisco forwarding agents. For both the Cisco forwarding agent and the Sysplex Distributor, the password is treated simply

as ASCII characters. No translation or conversion is performed. For more information on MD5 authentication, refer to RFC 1321.

---

5. If using the Cisco MNLB in a configuration where there is an OSA adapter between a Cisco router and the destination TCP/IP stacks such that multiple stacks are sharing the OSA, then configure GRE tunnels on the Cisco routers. The Sysplex Distributor stack is the only stack that registers the dynamic VIPA to OSA. Therefore, if GRE tunnels are not configured, OSA will send all packets destined for the DVIPA to the Sysplex Distributor stack (or to the default router stack if the OSA is not shared with the Sysplex Distributor stack). You must configure the GRE tunnels on the Cisco router such that any packets destined for a DVIPA that are routed by the forwarding agents directly to TCP/IP target stacks are encapsulated to the target stack's dynamic XCF address or, if specified in the distributor, to the VIPAROUTE target IP address. If the primary and backup TCP/IP stacks specify different VIPAROUTE statements for a particular target, you must define GRE tunnels for each target IP address that might be used. For more detailed information on configuring GRE tunnels, refer to the Cisco router publications located at http://www.cisco.com/univercd/cc/td/doc/product/core/index.htm.

---

6. If using the Cisco MNLB in a mixed environment where V2R10 targets exist, routing must be configured so that the selected route from the R10 target to the routing stack is not through the Cisco MNLB service manager for this distributed DVIPA.

---

7. Special consideration must be made for each target stack that will receive data from an OSA that is not shared with the distributor stack. Connection load balanced IP packets routed to target stacks that do not use GRE tunnels will arrive with a destination address of the dynamic VIPA address. Only the OSA associated with the distributor stack is aware of the dynamic VIPA address. If the OSA is not the primary router, it will discard the IP packet. In this case, you must either configure GRE tunnels on the Cisco router or configure the target stack to be the primary router for the OSA. If configuring GRE tunnels on the Cisco router, ensure that any packets destined for a DVIPA that are routed by the forwarding agents directly to TCP/IP target stacks are encapsulated to the target stack's dynamic XCF address or, if specified in the distributor, to the VIPAROUTE target IP address.

---

8. Verification:

   The Netstat VIPADCFG/-F report may be used to verify the configuration. Refer to the *z/OS Communications Server: IP System Administrator's Commands* for more information on this command.

   Cisco's ***show ip casa*** commands may be used to display MNLB information. For more detailed information on these commands, refer to Cisco's online documentation at:

   `http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120t/120t5/ipclus.htm`

Following is a sample VIPADYNAMIC statement:

```
VIPADYNAMIC
   VIPADEFINE  MOVEABLE IMMED SERVICEMGR 255.255.255.0 197.11.221.1
   VIPASMPARMS SMMCAST 224.0.1.2 SMPORT 1637 SMPASS ABCD
```

```
    VIPADIST 197.11.221.1 PORT 80 20 21 23
    DESTIP 199.11.87.104
           199.11.87.105
           199.11.87.106
           199.11.87.108
           199.11.87.109
           199.11.87.110
ENDVIPADYNAMIC
```

For more information on the VIPADYNAMIC statement, refer to the *z/OS Communications Server: IP Configuration Reference*.

DNS/WLM

The DNS solution is based on the DNS name server and the z/OS Workload Manager. This solution is only available with the BIND 4.9.3 name server and not with the BIND 9 name server. Intelligent sysplex distribution of connections is provided through cooperation between WLM and DNS. For customers who elect to place a name server in a z/OS sysplex, the name server can utilize WLM to determine the best system to service a given client request.

In general, DNS/WLM relies on the hostname to IP address resolution for the mechanism by which to distribute load among target servers. Hence, the single system image provided by DNS/WLM is that of a specific hostname. Note that the system most suitable to receive an incoming client connection is determined only at connection setup time. Once the connection is made, the system being used cannot be changed without restarting the connection.

The DNS approach works only in a sysplex environment, because the Workload Manager requires it. If the server applications are not all in the same sysplex, then there can be no single WLM policy and no meaningful coordination between WLM and DNS.

Sysplex-aware external load balancing solutions

Sysplex-aware external load balancing solutions are typically comprised of an external load balancing solution that communicates with components within the z/OS sysplex, to optimize load balancing based on the current conditions and workloads in the sysplex environment. Examples of these solutions include:

- External load balancers that use the Server/Application State Protocol (SASP) to obtain recommendations and topology information related to applications and systems within a sysplex environment.
- The IBM Network Dispatcher can obtain WLM recommendations from z/OS, and uses these recommendations to determine how workload requests are routed.

External load balancers can obtain detailed information regarding the state of target z/OS applications and systems by communicating with the z/OS Load Balancing Advisor using the SASP protocol. Using SASP, external load balancers can also obtain detailed recommendations on how workload should be distributed within the sysplex environment.

A key component of these recommendations is derived from the z/OS Workload Manager (WLM). The WLM information reflects the current available system capacity for target systems, and can also reflect how well individual server applications are performing compared to the WLM policy goals specified for that application. These recommendations are also influenced by how well the applications are performing from a TCP/IP perspective. For example, are the

applications processing new requests quickly enough? Or are new requests filling up the backlog queues that TCP/IP maintains for each application?

Using SASP, external load balancers can perform load balancing that is optimized for your sysplex environment, based on the current configuration and workload conditions. For more information on SASP and the z/OS Load Balancing Advisor, see Chapter 19, "z/OS Load Balancing Advisor," on page 1009.

External IP workload balancing solutions

IBM's Network Dispatcher (part of WebSphere Edge Server) and Cisco's Content Switching Module (CSM) are examples of external IP workload balancing solutions. Such solutions exist outside the sysplex, but can direct work into the sysplex. External IP workload balancing solutions typically define a single IP address representing all instances of the server, and then balance new work requests (for example, new TCP connection requests) among available servers.

External load balancers typically use a cluster IP address to represent the set of applications being load balanced. Client applications use this cluster IP address as the destination IP address for all their requests. External load balancers might be capable of using different methods for forwarding packets to their destinations. Various load balancing solutions might use different terms when describing these methods. In this discussion, dispatch mode and directed mode are discussed as two examples of these methods.

When a load balancer uses dispatch mode, the destination IP addresses for incoming IP packets is not changed. Instead, the load balancer forwards the packet to a target z/OS system by using the MAC address of a network adapter on that system. This technique works well when the target systems and the load balancer are attached to the same network (that is, there are no intervening routers). The receiving z/OS system inspects the destination IP address of the packet, and if it matches one of the IP addresses in its HOME list, it accepts the packet. As a result, with dispatch mode, all target z/OS systems must have the load balancer's cluster IP address defined in their HOME list. It is important, however, that these addresses are not advertised externally through dynamic routing protocols. One way to accomplish this is by defining these IP addresses as loopback addresses on z/OS.

Dispatch mode also has special considerations when the load balancer is more than one hop away from the target systems (that is, a packet must be routed to the target), or when multiple z/OS target systems share the same OSA-Express adapter. In these scenarios, techniques such as generic routing encapsulation (GRE) tunnels can be used to route the packets to the proper target system using an IP address that is exclusively owned by the target system.

On the other hand, when a load balancer uses directed mode, the load balancer converts the destination IP address (that is, the cluster IP address) to an IP address owned by the target z/OS system, using technologies such as network address translation (NAT). When IP packets for these connections are sent back to clients, the load balancer converts the source IP address (that is, the target z/OS system's IP address) back to the cluster IP address that the application had used on its request.

Some load balancer configurations might also perform NAT on the client's IP address. The client's IP address can be replaced with an IP address owned by the load balancing device. This might be necessary to ensure that all outgoing packets

from a target system traverse the load balancing device, so that NAT can be performed to change the server's IP address back to the cluster IP address that the client had originally used. Therefore, it is important to note that with directed mode solutions, the IP addresses of load balanced connections reaching the sysplex might not reflect the real IP addresses of the clients making the requests. This can be an important consideration if any definitions or configuration within the sysplex rely on the client's IP address being visible on incoming connections.

External load balancing solutions might provide several configuration options that influence how workload is distributed, such as round-robin, weighted round-robin, and so on. Some of the solutions might also obtain recommendations from components inside the z/OS sysplex that might affect their workload balancing decisions. For more information, see "Sysplex-aware external load balancing solutions" on page 396.

Additional Considerations

As described in the previous sections, several load balancing solutions are available for a z/OS environment. Which solution is the best one to deploy in your environment? The answer depends on several factors that are related to your environment and the workload being load balanced, including:

- What type of workload needs to be load balanced, and does a particular load balancing solution support this type of workload? For example, does the workload have affinities to specific servers? Can these affinities be determined only by inspecting data content? If so, an external load balancing solution that supports content inspection and affinities might be the most appropriate solution.
- Is it important that the solution is primarily configured and maintained within the z/OS sysplex, with minimal interactions with network components? In these scenarios, an internal load balancing solution might be the preferred answer.
- Do you already have an external load balancing solution in the network? If so, can it be used to load balance z/OS workloads? Does it support SASP, so that it can perform more optimal load balancing decisions?
- Does the load balancing solution have the needed level of availability? For example, if the primary load balancer fails, can a secondary load balancer take over transparently? Can it do so without disrupting existing connections? What if target applications or systems fail? Does the load balancer recognize these conditions?

While a single load balancing solution might be desirable for all workloads in some environments, you can select multiple solutions based on your specific requirements. For example, you might select an external load balancing solution for a particular workload, while an internal load balancing solution might be enabled for another workload. Table 19 on page 399 lists some of the key attributes of the various load balancing techniques and specific solutions that were discussed in this chapter. It can be used as a quick reference for comparing these solutions.

Table 19. Load balancing solution quick reference

Features or considerations	DNS/WLM	Sysplex Distributor	External load balancers with SASP	External load balancers
How is the solution configured and administered?	Initial setup and configuration requires DNS updates on z/OS and in the network. Ongoing administration (adding and removing target server applications and systems) typically confined within z/OS systems.	Initial setup might require some interaction with the network (dynamic routing protocols, DNS updates for dynamic VIPAs, and so on). Ongoing administration (adding and removing target server applications and systems) typically confined within z/OS systems.	Initial setup and configuration on load balancer and on z/OS. Ongoing administration might need to be performed on both the load balancer and the z/OS systems.	Initial setup and configuration on load balancer, some configuration on z/OS might be required. Ongoing administration should be mostly confined to the load balancer, although z/OS configuration might be necessary, such as when adding new target systems.
When is the server instance decision made?	When the host name is resolved by the client. Depends on client (or network DNS) not caching the responses.	Connection setup (in line Syn segment)	Connection setup (in line Syn segment)	Connection setup (in line Syn segment)
Support for TCP and UDP applications?	Yes	TCP only	Depends on the load balancer implementation; SASP supports both TCP and UDP	Depends on the load balancer implementation
Extra network flows?	Requires DNS flow per TCP connection (or per set of related UDP datagrams that constitute a request or session) for optimal distribution and maximum availability.	No, for outbound traffic. Yes, for inbound traffic. Inbound traffic must traverse the Sysplex Distributor node. If Sysplex Distributor is configured as service manager for CISCO routers, the inbound traffic can flow directly to the target application.	Depends on the load balancer implementation; can be avoided if the load balancer is implemented as part of a router or switch.	Depends on the load balancer implementation; can be avoided if the load balancer is implemented as part of a router or switch.
Support for affinities between TCP connection requests based on data content?	No, but client code can establish affinity by reusing the same IP address for duration of the affinity	No, but support does exist for timer-based affinities	Depends on implementation; some support affinities for HTTP and HTTPS requests by inspecting data content (correlating cookies, jsessionid)	Depends on implementation; some support affinities for HTTP and HTTPS requests by inspecting data content (correlating cookies, jsessionid)
Network address translation?	Not needed; client and server IP addresses are not modified	Not needed; client and server IP addresses are not modified	Might be required by some implementations; client and server IP addresses might be translated	Might be required by some implementations; client and server IP addresses might be translated

Table 19. Load balancing solution quick reference (continued)

Features or considerations	DNS/WLM	Sysplex Distributor	External load balancers with SASP	External load balancers
Support for IPv6?	No, not planned	Yes	Depends on the load balancer implementation; SASP supports both IPv4 and IPv6	Depends on the load balancer implementation
z/OS WLM recommendations?	Yes, system level only	Yes, system level and server level	Yes, system level and server level	Depends on the load balancer implementation
z/OS network QoS recommendations?	No	Yes, based on z/OS QoS policy	No	No
z/OS TCP/IP server health information?	No	Yes	Yes	No
Detection of target application and target system state changes, active or inactive?	Yes. Application state changes require application registration with WLM. State changes are recognized within a configurable time period, 60 seconds by default. How quickly this information becomes available to clients depends on the time-to-live (TTL) values associated with the sysplex zone.	Yes, application and system state changes are detected in near real-time fashion.	Yes, the z/OS Load Balancing Advisor and Agents detect application and system state changes within a configurable time period, 60 seconds by default. How quickly external load balancers become aware of these changes depends on several factors: • If the load balancer is using a push model with SASP, the Load Balancing Advisor sends a notification of a state change as soon as it is detected. • If the load balancer is using a poll model with SASP, it depends on the load balancer's polling interval. • The load balancer might also have additional mechanisms for detecting application and system state changes, which might provide for faster detection of these changes.	Depends on the load balancer implementation.

Table 19. Load balancing solution quick reference (continued)

Features or considerations	DNS/WLM	Sysplex Distributor	External load balancers with SASP	External load balancers
High availability solution, load balancing continues even if the primary load balancing component becomes unavailable?	Yes, additional DNS/WLM instances can be activated and configured for backup use in cases where the primary DNS/WLM instance, or system it is executing on, fails.	Yes, one or more backups can be configured to enable dynamic takeover in cases where the TCP/IP stack, or system that is acting as the distributor, fails.	For failures to the load balancer, it depends on the load balancer implementation. Some solutions provide for backup load balancers that can dynamically take over load balancing responsibilities in cases of failures. The z/OS Load Balancing Advisor and Agents can be configured for high availability to minimize the impact of an Advisor, Agent, or system failure.	For failures to the load balancer, it depends on the load balancer implementation. Some solutions provide for backup load balancers that can dynamically take over load balancing responsibilities in cases of failures.
Caching issues?	Caching of IP addresses by client applications can produce suboptimal performance and lengthen recovery times in cases of failures.	No, every new connection request is eligible for load balancing.	No, every new connection request is eligible for load balancing.	No, every new connection request is eligible for load balancing.

Part 2. Server applications

Chapter 9. Network connectivity with an SNA network

The objective of this chapter is to guide you through the steps required to implement:

- SNALINK LU0
- SNALINK LU6.2
- X.25 NPSI
- NCPROUTE

Before you configure:

Read and understand Chapter 2, "Configuration overview," on page 11. It covers important information about data set naming and search sequences.

SNALINK LU0 environment

SNALINK allows TCP/IP to send and receive packets using SNA sessions instead of dedicating physical network hardware (such as a channel-to-channel adapter or channel connection to a 3745/46 Communication Controller).

Prior to NCP V7R3, NCP did not support cross-channel native IP transmission of the transport PDUs associated with RIP traffic. NCP expects these PDUs to be carried in SNA frames. SNALINK is therefore still required for installations where dynamic routing is performed with the NCP (via NCPROUTE). See *z/OS Communications Server: IP Configuration Reference* for more information.

SNALINK allows an installation to multiplex SNA and IP traffic over the same I/O subchannels, rather than requiring separate subchannels dedicated to VTAM and TCP/IP. While such multiplexing capability may be desirable at some installations, the native TCP/IP CTC and 3745/46 device drivers will likely outperform SNALINK connections. Interaction with the SNALINK address space is very CPU-intensive, and is not required with the native TCP/IP CTC and 3745/46 device drivers. (See the *z/OS Communications Server: IP Configuration Reference* for configuration information.) It is therefore important to weigh the multiplexing capability that SNALINK provides against its performance cost, in determining whether to use SNALINK or the native TCP/IP CTC or 3745/46 device drivers.

Understanding the SNALINK environment

The SNALINK environment interfaces between the TCP/IP environment's SNAIUCV driver and the customer's SNA network. SNALINK communicates with one or more instances of SNALINK at remote nodes, using the SNA LU type 0 protocol. See Figure 42 on page 406 for a description of the SNALINK environment interfaces.

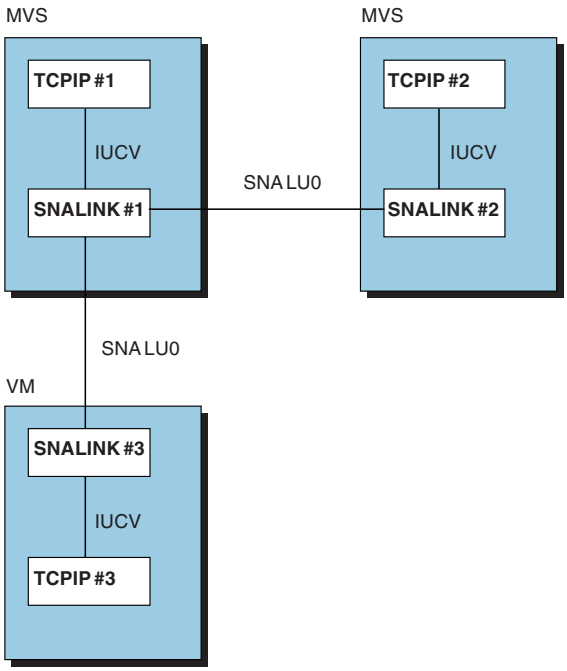


Figure 42. SNALINK environment interfaces

Each SNALINK environment can communicate with up to 9999 SNALINKs simultaneously. The number of connections is determined by the parameters you pass to the SNALINK cataloged procedure. The default is 6 sessions running in dual mode for a total of 3 SNALINKs.

- When operating in single mode, SNALINK opens one full duplex session.
- When operating in dual mode, SNALINK opens two System Network Architecture (SNA) sessions for each remote logical unit (LU) with which it communicates, one for sending and one for receiving.

Configuring SNALINK LU0

Steps to configure SNALINK LU0:

1. Specify configuration statements in *hlq*.PROFILE.TCPIP.
2. Update the SNALINK cataloged procedure.
3. Define the SNALINK application to VTAM.
4. Configure the program properties table (PPT) for SNALINK LU0.

Step 1: Specify configuration statements in *hlq*.PROFILE.TCPIP

The following sections describe the changes you must make to your TCPIP address space configuration data set (*hlq*.PROFILE.TCPIP).

Defining SNA DLC links: SNA DLC links are point-to-point and require DEVICE and LINK statements in the configuration data set. The DLC link constitutes a separate network, even though it includes only two hosts. To define a link, each host to which the DLC link is attached requires:

- A pair of SNA LU0 DEVICE and LINK statements
- A HOME statement
- A BSDROUTINGPARMS or GATEWAY or BEGINROUTES statement

SNA DLC links are defined in one of two ways:

- By unique network or subnetwork numbers, if the hosts to which they connect are not attached to other networks.
- By the IP address of the hosts to which they connect, if the hosts are attached to other networks.

You usually have to assign a unique network or subnetwork number to the SNALINK. If the link connects 2 hosts that also have other networks attached to them, the DLC link does not need its own subnetwork number. Figure 43 illustrates how to define an SNA DLC link if the 2 hosts are connected to other networks in the following way:
- Host A and Host B are connected by SNA DLC
- Host A is also connected to a token ring, 193.1.1
- Host B is also connected to a token ring, 193.1.2
- Host A's home address on its token ring is 193.1.1.1
- Host B's home address on its token ring is 193.1.2.1

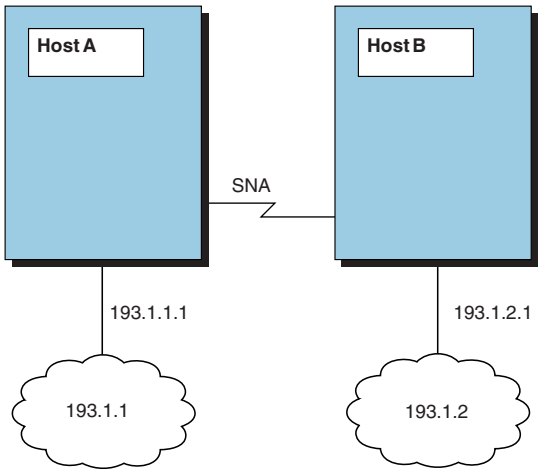


Figure 43. SNA DLC link

Host A's *hlq*.PROFILE.TCPIP could contain:

```
DEVICE  LCS1 LCS BA0
LINK  TR1  IBMTR 0 LCS1
DEVICE SNALU0 SNAIUCV SNALINK LU000000 SNALINKA
LINK SNAIUCV1 SAMEHOST  1 SNALU0
HOME
    193.1.1.1  TR1
    193.1.1.2  SNAIUCV1

GATEWAY
;  Network         First hop  Link    Packet size    Subnet mask
    193.1.1.0      =          TR1         2000        0
    193.1.2.0      =          SNAIUCV1    2000        0
```

Host B's *hlq*.PROFILE.TCPIP could contain:

```
DEVICE LCS2 LCS BE0
LINK  TR1 IBMTR 0 LCS2
DEVICE SNALU0 SNAIUCV SNALINK LU000001 SNALINKA
LINK SNAIUCV1 SAMEHOST  1 SNALU0
HOME
    193.1.2.1  TR1
    193.1.2.2  SNAIUCV1

GATEWAY
```

```
;  Network         First hop  Link    Packet size    Subnet mask
    193.1.2.0      =          TR1         2000        0
    193.1.1.0      =          SNAIUCV1    2000        0
```

Notes:

1. The *lu_name* must be different on each host. In the example, the *lu_name* for Host A is LU000000. The *lu_name* for Host B is LU000001.
2. In the example, the *lu_name* is the remote or partner LU.

Hosts A and B are addressed by their token-ring home addresses, even if the packets reach them through the SNA DLC link.

If Host B had no other network attached to it, you would have to assign a separate subnetwork number to the SNA DLC link. Even in this case, Host A does not need a separate home address for its SNA link, because it can be addressed by its token-ring home address. Host B's only home address is the home address for the SNA link.

Note: If you plan to run a network-monitoring protocol that requires each subnet to have its own subnet number, you can assign a separate subnet network number to the DLC link.

Defining NCPROUTE and 3745 LAN attachments: If your TCP/IP configuration supports NCPROUTE or 3745 Communications Controller Ethernet or token-ring links, you must do the following:

- Match the *lu_name* on the DEVICE statement to the LU statement in NCST section of your NCP generation.

  The following example shows the LU name A04TOLU1 defined in the *hlq*.PROFILE.TCPIP DEVICE statements and in the NCP generation.

```
DEVICE  SNA1LINK SNAIUCV SNALINK A04TOLU1 SNAL1STC
LINK  SNALINK SAMEHOST 1 SNA1LINK

HOME
    9.67.116.66  SNALINK

GATEWAY
;  Network         First hop  Link    Packet size    Subnet mask
    9.67.116.65   =          SNALINK     2000        HOST

START SNA1LINK
```

```
*************************************************************************
*               NCST IP INTERFACES**
*************************************************************************
A04NCSTG GROUP NCST=IP,LNCTL=SDLC,VIRTUAL=YES
A04NCSTL LINE LINEFVT=CXSXFVT,PUFVT=CXSXFVT,LUFVT=(CXSXFVT,CXSXFVT),LIN*
              ECB=CXSXLNK
A04NCSTP PU VPACING=0,PUTYPE=2,PUCB=CXSP0000S
*
A04TOLU1 LU INTFACE=(NCSTALU1,1492),REMLU=SNALKLU1,LUCB=(CXSXL0000,CXSS0*
              000),LOCADDR=1
*************************************************************************
```

- Match the remote LU name SNALKLU1 in the NCP generation to the APPLID in the SNALINK cataloged procedure parameters and in the VTAM APPL definition.

```
//SNALINK  PROC MODULE=SNALINK,TCPID='TCPV3',APPLID='SNALKLU1'
//SNALINK  EXEC PGM=&MODULE<REGION=$4096K,TIME=1440,
           PARM='&TCPID &APPLID C7 6 0003 SINGLE'
```

For additional information on configuring these links, see *z/OS Communications Server: IP Configuration Reference*.

Step 2: Update the SNALINK cataloged procedure

Update the SNALINK cataloged procedure by copying the sample in SEZAINST(SNALPROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Specify SNALINK parameters and change the DD statements, as required. Refer to *z/OS Communications Server: IP Configuration Reference* for more information about the SNALINK cataloged procedure.

Step 3: Define the SNALINK application to VTAM

In dual mode, SNALINK opens 2 SNA sessions for each remote logical unit with which it communicates: one for sending and one for receiving. In single mode, SNALINK opens one full-duplex session.

Figure 44 is an example of a typical VTAM APPL statement for SNALINK. The application identifier (SNALKB03 in this example) must match the APPLID specified in the SNALINK cataloged procedure parameters.

```
SNALKB03  APPL ACBNAME=SNALKB03,                                  X
               AUTH=(ACQ,VPACE),                                  X
               SRBEXIT=YES,                                       X
               EAS=12,                                            X
               PARSESS=YES,                                       X
               SONSCIP=YES,                                       X
               VPACING=0
```

Figure 44. APPL statement for SNALINK

Note: ***SRBEXIT must be YES.***

Refer to *z/OS Communications Server: SNA Resource Definition Reference* more information about defining VTAM applications.

VTAM considerations:

- Each connection requires 100KB of virtual storage.
- SNALINK provides its own BIND parameters, so it does not assume or require any particular LOGMODE entries.
- The EAS value should be two times the number of maximum sessions passed to the SNALINK cataloged procedure.
- SRBEXIT=YES.
- You might have to specify pacing values (VPACING). Consult your VTAM network administrator for further details.
- For *max_ru_size*, be sure to consider the size of the TH, RH, and RU portions. If the maximum size PIU exceeds MAXRU, the NCP issues a negative response with sense 800A0000 (PIU too long). The definition used in NCP and SNALINK must be such that MAXRU is at least 29 bytes less than MAXDATA. Refer to *z/OS Communications Server: SNA Network Implementation Guide* for more information on defining the MAXDATA, MAXBFRU, and UNITSZ operands.

Step 4: Configure PPT for SNALINK LU0

To avoid startup problems, ensure that the PPT entry for SNALINK LU0 (PGM=SNALINK) is configured with the required program properties. For more information, refer to *z/OS MVS Initialization and Tuning Reference*.

Stopping and starting SNALINK

If necessary, you can immediately retry a session that is waiting for the retry delay to expire by stopping and starting the SNALINK LU0 interface.

To stop SNALINK and close all connections, use the STOP command on the operator's console. For example, if SNALPROC was the name of the cataloged procedure used to start SNALINK, you enter:

```
STOP SNALPROC
```

You can also stop SNALINK with the HALT parameter on the MODIFY command. See "Controlling the SNALINK LU0 interface with the MODIFY command" on page 412.

SNALINK can be started by:

- Restarting the TCPIP address space if you have included the SNALINK procedure in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.
- Issuing START *procname* at the command console (where *procname* is the name of the cataloged procedure used to start the SNALINK LU0 interface).

For example, to restart SNALPROC, enter

```
START SNALPROC
```

Sample console

The example in Figure 45 on page 411 and the accompanying information illustrate SNALINK operation.

The line number notations in the example have been added for clarity. They do not appear in the console output.

```
Line 1              | Init complete, APPLID SNALKB03, TCPIP id TCPIPB
Line 2              | Maximum RU size is 00000600
Line 3     SNALKC04 | DLC path 00000001 pending
Line 4     SNALKC04 | Ready to accept bind from remote LU
Line 5     SNALKA04 | DLC path 00000002 pending
Line 6     SNALKA04 | Sending BIND request for SNA send session
Line 7     SNALKA04 | OPNDST   CHECK err. R15 00000004 R0 00000010 RTNCD 00000010 FDBK2 00000000
Line 8     SNALKA04 | OPNDST   sense: SSENSEI,SSENSMI,USENSEI: 00000000
Line 9     SNALKA04 | DLC path 00000002 pending
Line 10    SNALKA04 | Sending BIND request for SNA send session
Line 11    SNALKA04 | OPNDST   CHECK err. R15 00000004 R0 00000010 RTNCD 00000010 FDBK2 00000000
Line 12    SNALKA04 | OPNDST   sense: SSENSEI,SSENSMI,USENSEI: 00000000
Line 13    SNALKC04 | Received BIND request for SNA receive session
Line 14    SNALKC04 | Sending BIND request for SNA send session
Line 15    SNALKC04 | SNA receive session established
Line 16    SNALKC04 | SNA send session established
Line 17    SNALKC04 | Accepting DLC path 00000001
Line 18    SNALKA04 | DLC path 00000002 pending
Line 19    SNALKA04 | Sending BIND request for SNA send session
Line 20    SNALKA04 | SNA send session established
Line 21    SNALKA04 | Accepting DLC path 00000002
Line 22    SNALKA04 | Received BIND request for SNA receive session
Line 23    SNALKA04 | SNA receive session established
Line 24    SNALKC04 | NSEXIT CLEANUP request for receive session
Line 25    SNALKC04 | RECEIVE  CHECK err. R15 00000004 R0 0000000C RTNCD 0000000C FDBK2 0000000B
Line 26    SNALKC04 | RECEIVE  sense: SSENSEI,SSENSMI,USENSEI: 00000000
Line 27    SNALKC04 | DLC path 00000001 pending
Line 28    SNALKC04 | Ready to accept bind from remote LU
Line 29    STOP SNALINK
Line 30             | Received STOP command, shutting down
```

Figure 45. SNALINK console example

Line number	Description
Lines 1 and 2	SNALINK displays its startup information from its command line parameters, which are customized as described in *z/OS Communications Server: IP Configuration Reference*. The maximum RU size and all other values are displayed in hexadecimal.
Lines 3 and 4	The TCPIP address space, TCPIPB, issues a DLC CONNECT to establish a session with the remote LU SNALKC04. SNALKC04 is higher in the collating sequence than the local LU name SNALKB03. Consequently, SNALKB03 takes the passive role in connecting to SNALKC04, and waits for SNALKC04 to establish a session.
Lines 5 and 6	TCP/IP issues another DLC CONNECT to establish a session with SNALKA04. In this case, SNALKA04 is lower in the collating sequence. Consequently, SNALKB03 takes an active role in connecting to SNALKA04.
Lines 7 and 8	The session establishment attempt to SNALKA04 has failed, as indicated by the (nonzero) return code and the sense information printed.
Lines 9 through 12	Thirty seconds later, TCP/IP again tries to connect to SNALKA04.
Lines 13 and 14	SNALINK receives a BIND request from SNALKC04. SNALINK calls the resulting session

	the receive session, because it is used only to send data from SNALKC04. Now that the active end has initiated communication, SNALKB03 as the passive end, sends a BIND request to establish a send session.
Lines 15 through 17	The send and receive sessions are fully established. Establishment of the send session causes SNALINK to accept the corresponding DLC path.
Lines 18 through 23	TCP/IP again tries to connect to SNALKA04. This time it is successful (success is indicated by no nonzero return codes).
Lines 24 through 26	SNALKC04 terminates its sessions, and various error messages result.
Lines 27 and 28	Thirty seconds later, TCP/IP again tries to establish communication with SNALKC04. As in lines 13 and 14, SNALKB03 is the passive partner.
Lines 29 and 30	The operator issues a STOP SNALINK command, which causes SNALINK to stop. All DLC paths and SNA sessions are ended.

Verifying connection status using NETSTAT DEVLINKS

The DLC connect protocol between TCP/IP and SNALINK causes the status of the SNAIUCV device, reported by NETSTAT DEVLINKS, to reflect the status of the SNA sessions to the remote LU. Refer to *z/OS Communications Server: IP System Administrator's Commands* for more information on the NETSTAT command.

Controlling the SNALINK LU0 interface with the MODIFY command

The following command would pass parameters to a SNALINK LU0 address space started with a procedure named SNLK12.TCPSETUP.

```
MODIFY SNLK12.TCPSETUP,HALT
```

SNALINK LU6.2

The SNALINK LU6.2 cataloged procedure runs a VTAM application program called SNALNK62, which is an interface between the TCPIP address space and the SNA network. SNALNK62 uses SNA LU type 6.2 sessions to pass the TCP/IP data to or from SNALNK62 devices running on other hosts. Examples of SNALNK62 devices include an OS/2® workstation running TCP/IP for OS/2 or a host running TCP/IP for MVS.

Configuring SNALINK LU6.2

Steps to configure SNALINK LU6.2:

1. Specify DEVICE and LINK statements in *hlq*.PROFILE.TCPIP.
2. Update the SNALINK LU6.2 cataloged procedure.
3. Define the SNALINK LU6.2 application to VTAM.
4. Update the SNALINK LU6.2 configuration data set.

Step 1: Specify DEVICE and LINK statements in *hlq*.PROFILE.TCPIP

You must update the *hlq*.PROFILE.TCPIP data set to include a DEVICE and LINK statement for each DLC connection to be established between the main TCPIP address space and the SNALINK LU6.2 address space.

Step 2: Update the SNALINK LU6.2 cataloged procedure

Update the SNALINK LU6.2 cataloged procedure by copying the sample in SEZAINST(LU62PROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. No system parameters are required for the SNALINK LU6.2 address space.

The DD statements in the cataloged procedure should be defined as follows:

DD Name	Description
SYSTCPD	TCPIP.DATA configuration data set
LU62CFG	SNALINK LU6.2 configuration data set
SYSPRINT	Runtime diagnostic or trace output
SYSUDUMP	User abend dump output (optional)

Refer to “Resolver configuration files” on page 40 for information on data set search sequences.

Step 3: Define the SNALINK LU6.2 application to VTAM

SNALINK LU6.2 opens two SNA LU type 6.2 sessions with each destination node; one for sending and one for receiving. If a destination node supports parallel SNA LU type 6.2 sessions (PARSESS=YES), the two sessions use the same remote logical unit; otherwise, two remote logical units are used. In either case, SNALINK LU6.2 uses a single local logical unit that must support parallel sessions.

The SNALINK LU6.2 address space must be defined to VTAM as an SNA LU type 6.2 application program. The following APPL statement defines a SNALINK LU6.2 application to VTAM.

```
LU62APPL  APPL ACBNAME=LU62APPL,                                *
               PRTCT=QWERTY,                                    *
               AUTH=(ACQ,VPACE),                                *
               SRBEXIT=NO,                                      *
               EAS=12,                                          *
               PARSESS=YES,                                     *
               SONSCIP=YES,                                     *
               APPC=YES,                                        *
               DLOGMOD=LU62MODE,                                *
               VPACING=0
```

Figure 46. APPL statement for SNALINK LU6.2

Note: ***SRBEXIT must be NO.***

See *z/OS Communications Server: SNA Resource Definition Reference* for further information about defining VTAM applications.

The LOGMODE table entry specified by the APPL DLOGMOD parameter should have the following form:

```
LU62MODE MODEENT LOGMODE=LU62MODE,FMPROF=X'13',TSPROF=X'07',          *
               PRIPROT=X'B0',SECPROT=X'B0',COMPROT=X'D0B1',           *
               RUSIZES=X'8585',ENCR=B'0000',                          *
               PSERVIC=X'060200000000000000000300'
```

See *z/OS Communications Server: SNA Customization* for more information about defining log mode tables and *z/OS Communications Server: SNA Programming* for information on PSERVIC values.

Step 4: Update the SNALINK LU6.2 configuration data set

Customize the SNALINK LU6.2 configuration data set by copying the sample provided in SEZAINST(LU62CFG) to your system or recognized PROCLIB and modifying it to suit your local conditions. Add or change the configuration statements as required. Be sure the //LU62CFG statement in the cataloged procedure points to this data set. Refer to *z/OS Communications Server: IP Configuration Reference* for more information about parameters.

Sample console

The example in Figure 47 shows the messages that are expected when the SNALINK LU6.2 address space is started and a network connection is established.

```
S SNAL621A
$HASP100 SNAL621A ON STCINRDR
$HASP373 SNAL621A STARTED
 1 IEF403I SNAL621A - STARTED - TIME=15.26.03
 2 EZA5927I LU62CFG : NO ERRORS DETECTED - INITIALIZATION WILL CONTINUE
 3 EZA5932I INITIALIZATION COMPLETE - APPLID: SNAL621A  TCP/IP: TCPCS
 4 EZA5935I SEND CONVERSATION ALLOCATED FOR 9.67.22.2
 5 EZA5933I LINK SNALU62L OPENED
EZZ4313I INITIALIZATION COMPLETE FOR DEVICE SNALU621
 4 EZA5936I RECEIVE CONVERSATION ALLOCATED FOR 9.67.22.2
```

Figure 47. Sample MVS system console messages on SNALINK LU6.2 address space startup

The following list explains the MVS system console messages on SNALINK LU6.2 address space startup as shown in Figure 47.

1 The SNAL621A address space has been started.

2 The SNALINK LU6.2 configuration data set for the SNAL621A address space has been successfully parsed.

3 The SNAL621A address space displays its local VTAM application LU and the TCP/IP address space name to which it will connect.

4 The SNAL621A address space establishes a network connection through the VTAM API.

5 The SNAL621A address space establishes a DLC connection with its TCP/IP address space.

X.25 NCP Packet Switching Interface (NPSI)

The X.25 NPSI server runs a VTAM application program called XNX25IPI, which is the interface between the TCPIP address space's DLC driver and your X.25 network. XNX25IPI communicates with the X.25 NCP Packet Switching Interface in a front-end IBM 37xx Communications Controller.

Large scale X.25 network applications often require multiple physical lines to the network switch for increased capacity and reliability. You can configure the X.25 NPSI server to support multiple lines as a group, rather than individually. In this configuration, the collection of lines is assigned a single address called a *hunt group*

address. Incoming X.25 calls are distributed among the lines in either rotary or traffic balancing fashion, depending on the services offered by the X.25 network provider.

For information about improving the performance of the X.25 NPSI network, see the options on the PORT statement and GATEWAY statement in the *hlq*.PROFILE.TCPIP and the explanation provided in the *TCP/IP: Performance Tuning Guide*.

Configuring X.25 NPSI

This section describes how to configure the X.25 NPSI server.

Steps to configure the X.25 NPSI server:

1. Specify X.25 configuration statements in *hlq*.PROFILE.TCPIP.
2. Update the X.25 NPSI cataloged procedure.
3. Update the X.25 NPSI server configuration data set.
4. Define the X.25 NPSI configuration.
5. Define the X.25 NPSI application to VTAM.
6. Define VTAM Switched Circuits.

If you want to run the X.25 NPSI cataloged procedure in a different domain than the X.25 NPSI communication controller, see *z/OS Communications Server: IP Configuration Reference*.

For information about operating the X.25 NPSI server with the MODIFY command, see *z/OS Communications Server: IP Configuration Reference*.

Step 1: Specify X.25 configuration statements in *hlq*.PROFILE.TCPIP

To configure the *hlq*.PROFILE.TCPIP data set for X.25 NPSI, include appropriate DEVICE, LINK, HOME, GATEWAY, and START statements. The following example shows the statements that would correspond with the other X.25 samples in this chapter.

```
;
DEVICE X25DEV X25NPSI TCPIPX25
LINK X25LINK SAMEHOST 1 X25DEV
;
HOME
   199.005.058.23     X25LINK
;
GATEWAY
;
; Network  First hop   Link name Packet size  Subnet mask  Subnet value
  192.005      =        X25LINK     2000       0.0.255.0    0.0.58.0
;
START X25DEV
;
```

Note: Only one DEVICE and LINK statement per TCPIPX25 address space is allowed.

Step 2: Update the X.25 NPSI cataloged procedure

Update the X.25 NPSI cataloged procedure by copying the sample provided in SEZAINST(X25PROC) to your system or recognized PROCLIB and modifying it to suit your local conditions.

Change the data set names as needed:

- Refer to "Resolver configuration files" on page 40 for data set search sequence information.
- Modify the //X25IPI DD statement to point to your X.25 configuration data set.

Step 3: Update the X.25 NPSI server configuration data set

A sample configuration data set provided in SEZAINST(X25CONF) gives examples of how to define a public network connection, a Defense Data Network connection, and private point-to-point connection to a router. Copy this sample to the data set pointed to by the //X25IPI DD statement in your X.25 NPSI cataloged procedure. Update this sample to define your X.25 connections using the statements listed in the *z/OS Communications Server: IP Configuration Reference*.

Each connection must have a LINK and at least one DEST statement. You can optionally define hunt groups, fast connects, and call handling options for each link, and global options such as trace levels, when to clear inactive connections, and the buffer size to use for IP datagrams. You can find complete syntax for each of these statements in *z/OS Communications Server: IP Configuration Reference*.

Step 4: Define the X.25 NPSI configuration

Define the X.25 NPSI configuration according to the information in *X.25 NPSI Planning and Installation*. The X.25 NPSI server supports use of the LOGAPPL operand on the X25.MCH definition in the X.25 NPSI configuration to allow automatic recovery. You can use either the Generalized Access to X.25 Transport Extension (GATE) or Dedicated Access to X.25 Transport Extension (DATE).

IBM recommends using the X.25 NPSI GATE configuration which allows sharing of an X.25 physical link and provides better error recovery. A sample is provided in SEZAINST(NPSIGATE). NPSI GATE requires that you include the OPTIONS GATE statement in the X.25 NPSI configuration data set after the LINK statement, as shown in this portion of the X25CONF sample:

```
*
*        NPSI MCH       DTE             Window Packet Logical
*        LU Name   DNIC Address         Size   Size   Channels
*        --------  ---- --------------- -      ----   ---
Link     XU024     PRIV 1               2      1024   2
Options  GATE
*
*        IP address      X.25 DTE addr       C.U.D.
*        --------------- ---------------     --------
Dest     192.5.57.2      2
```

Sites that need to use the X.25 NPSI DATE configuration can find a sample in PV SEZAINST(NPSIDATE). See *X.25 NPSI Host Programming* for information about the definitions and parameters used in these configurations.

The following example shows portions of the sample NPSI GATE configuration (NPSIGATE). Ellipses (....) indicate code that has been omitted.

```
*********************************************************************
         OPTIONS NEWDEFN=YES,USERGEN=X25NPSI
*********************************************************************
....
....
NPSIV32  BUILD ADDSESS=400,                                          +
               AUXADDR=800,                                          +
               ERLIMIT=16,                                           +
               NAMTAB=120,                                           +
               MAXSESS=250,                                          +
               USGTIER=5,                                            +
```

```
                BRANCH=8000,                                            +
                BFRS=104,           BUFFER SIZE TO BE GENED             +
                CATRACE=(YES,255),  CHAN.ADAPTER TRACE OPTION           +
                CSMSG=C3D9C9E340E2C9E340D4C5E2E2C1C7C540C6D6D940E2E24040+
                40C2C340E3C5D9D4C9D5C1D3,                               +
                CWALL=26,                                               +
                ENABLTO=30.0,                                           +
                ERASE=YES,                                              +
                LOADLIB=NCPLOAD,    TARGET OF FINAL LINKEDIT            +
                LTRACE=8,           LINES TRACED SIMULTANEOUSLY         +
                MAXSSCP=8,          NUMBER OF CONCURRENT SSCP'S         +
                MODEL=3745,                                             +
                VERSION=V5R2.1,                                         +
                NEWNAME=NPSITCP,    NAME OF NCP LOAD MODULE             +
                NUMHSAS=8,          HOST SA IN CONCURRENT COMMUNICATION +
                OLT=YES,            ONLINE TERMINAL TEST                +
                PWROFF=YES,                                             +
                BACKUP=500,                                             +
                SALIMIT=511,                                            +
                SLODOWN=12,         BUFFER SLOWDOWN THRESHOLD (PERCENT) +
                SUBAREA=03,                                             +
                TRACE=(YES,100),    ADDRESS TRACE OPTION IN CORE TABLE  +
                TYPGEN=NCP,                                             +
                TYPSYS=MVS,         NCP TO BE GENERATED ON MVS          +
                TWXID=(E8D6E4C3C1D3D311,C2C9C7D5C3D7C3C1D3D325),        +
                VRPOOL=30,                                              +
                TRANSFR=32,                                             +
                NETID=NETA,                                             +
                X25.USGTIER=5,                                          +
                X25.IDNUMH=01,                                          +
                X25.MCHCNT=4,                                           +
                X25.MAXPIU=64K
....
....
*************************************************************************
*
*                  NPSI DEFINITIONS
*
*************************************************************************
X25XXX   X25.NET CPHINDX=1,                                             +
               NETTYPE=1,                                               +
               DM=YES,                                                  +
               OUHINDX=1
         X25.VCCPT INDEX=1,                                             +
               MAXPKTL=128,                                             +
               VWINDOW=2
         X25.OUFT INDEX=1
*-----------------------------------------------------------------------*
*        Hunt group primary line 021 with fast connect                  *
*-----------------------------------------------------------------------*
HGRP01A  X25.MCH ADDRESS=21,                                            +
               FRMLGTH=131,         128 byte packet + 3 byte header     +
               PKTMODL=8,                                               +
               ANS=CONT,                                                +
               LCGDEF=(0,16),       16 logical channels in group 0      +
               MWINDOW=2,                                               +
               STATION=DTE,                                             +
               SPEED=9600,                                              +
               LCN0=NOTUSED,                                            +
               GATE=GENERAL,        GATE                                +
               LLCLIST=(LLC4),                                          +
               CONNECT=YES,         Fast connect                        +
               LOGAPPL=TCPIPX25,                                        +
               DBIT=NO,                                                 +
               DIRECT=NO,                                               +
               SUBADDR=NO
         X25.LCG LCGN=0
```

```
         X25.VC LCN=(1,16),                                          +
               MAXDATA=1034,       MAXDATA only with Fast connect!   +
               TYPE=SWITCHED,                                        +
               CALL=INOUT,                                           +
               OUFINDX=1,                                            +
               VCCINDX=1
*-------------------------------------------------------------------*
*        Hunt group secondary line 022 with fast connect            *
*-------------------------------------------------------------------*
HGRP01B  X25.MCH ADDRESS=22,                                         +
               FRMLGTH=131,        128 byte packet + 3 byte header   +
               PKTMODL=8,                                            +
               ANS=CONT,                                             +
               LCGDEF=(0,16),      16 logical channels in group 0    +
               MWINDOW=2,                                            +
               STATION=DTE,                                          +
               SPEED=9600,                                           +
               LCN0=NOTUSED,                                         +
               GATE=GENERAL,       GATE                              +
               LLCLIST=(LLC4),                                       +
               CONNECT=YES,        Fast connect                      +
               LOGAPPL=TCPIPX25,                                     +
               DBIT=NO,                                              +
               DIRECT=NO,                                            +
               SUBADDR=NO
         X25.LCG LCGN=0
         X25.VC LCN=(1,16),                                          +
               MAXDATA=1034,       MAXDATA only with Fast connect!   +
               TYPE=SWITCHED,                                        +
               CALL=INOUT,                                           +
               OUFINDX=1,                                            +
               VCCINDX=1
*-------------------------------------------------------------------*
*        DDN line 023                                               *
*-------------------------------------------------------------------*
DTE01    X25.MCH ADDRESS=23,                                         +
               FRMLGTH=131,        128 byte packet + 3 byte header   +
               PKTMODL=8,                                            +
               ANS=CONT,                                             +
               LCGDEF=(0,16),      16 logical channels in group 0    +
               MWINDOW=2,                                            +
               STATION=DTE,                                          +
               SPEED=9600,                                           +
               LCN0=NOTUSED,                                         +
               GATE=GENERAL,       GATE                              +
               LLCLIST=(LLC4),                                       +
               LOGAPPL=TCPIPX25,                                     +
               CTCP=(00),          paired with CUD list              +
               CUD0=(CC),          incoming CUD selects CTCP         +
               DBIT=NO,                                              +
               DIRECT=NO,                                            +
               SUBADDR=NO
         X25.LCG LCGN=0
         X25.VC LCN=(1,16),                                          +
               TYPE=SWITCHED,                                        +
               CALL=INOUT,                                           +
               OUFINDX=1,                                            +
               VCCINDX=1
*-------------------------------------------------------------------*
*        Private line 024: DCE station to router                    *
*-------------------------------------------------------------------*
DCE01    X25.MCH ADDRESS=24,       1024 byte packet + 3 byte header  +
               FRMLGTH=1027,                                         +
               PKTMODL=8,                                            +
               ANS=CONT,                                             +
               LCGDEF=(0,2),                                         +
               MWINDOW=2,                                            +
```

```
              STATION=DCE,                                          +
              SPEED=9600,                                           +
              LCN0=NOTUSED,                                         +
              GATE=GENERAL,                                         +
              LLCLIST=(LLC4),                                       +
              CTCP=(00),          paired with CUD list              +
              CUD0=(CC),          incoming CUD selects CTCP         +
              DBIT=NO,                                              +
              DIRECT=NO,                                            +
              SUBADDR=NO
         X25.LCG LCGN=0
         X25.VC LCN=(1,2),                                          +
              TYPE=SWITCHED,                                        +
              CALL=INOUT,                                           +
              OUFINDX=1,                                            +
              VCCINDX=1
         X25.END
*********************************************************************
....
....
GENEND   GENEND
```

Step 5: Define the X.25 NPSI application to VTAM

Define the X.25 NPSI VTAM application with an APPL statement in VTAMLST. Following is an example of a VTAM APPL statement for X.25 NPSI.

```
         VBUILD TYPE=APPL                                           +
TCPIPX25 APPL  ACBNAME=TCPIPX25,                                    +
              PRTCT=TCPX25,                                         +
              AUTH=(ACQ),                                           +
              PARSESS=YES,                                          +
              EAS=20
```

Step 6: Define VTAM switched circuits

X.25 NPSI switched virtual circuits (SVCs) appear to VTAM as switched links, [1] requiring a switched circuit definition of a physical unit (PU) and logical unit (LU) for each SVC. The sample provided in SEZAINST(X25VSVC) shows the definitions of a VTAM switched circuit corresponding to the sample X.25 NPSI GATE configuration.

The definitions are associated with the SVCs by identifying numbers (IDNUMs) created automatically during X.25 NPSI generation. The entries, in hexadecimal, run in steps of 2, by default, in the opposite order of the MCH and SVC definitions in the X.25 NPSI configuration.

Notes:

1. Permanent virtual circuits (PVCs) are not supported.
2. If you specify a local version of the z/OS UNIX table with the SSCPFM operand, the table must not have an entry for message 10 (the welcome message); otherwise, the X.25 NPSI server does not operate correctly.

Following is a sample SVC configuration data set (X25VSVC):

```
          VBUILD TYPE=SWNET,MAXGRP=1,MAXNO=1
*--------------------------------------------------------------------*
*   Switched circuits for DDN line 023 (16 VCs, IDNUMS 006-024)      *
*                                                                    *
*   COPYRIGHT = NONE.                                                *
*--------------------------------------------------------------------*
```

1. Except when using fast connect, where they appear as leased lines to VTAM. For more information, see *z/OS Communications Server: IP Configuration Reference*.

```
VP023001 PU    ADDR=23,IDBLK=003,IDNUM=01024,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023001 LU    LOCADDR=0
VP023002 PU    ADDR=23,IDBLK=003,IDNUM=01022,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023002 LU    LOCADDR=0
VP023003 PU    ADDR=23,IDBLK=003,IDNUM=01020,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023003 LU    LOCADDR=0
VP023004 PU    ADDR=23,IDBLK=003,IDNUM=0101E,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023004 LU    LOCADDR=0
VP023005 PU    ADDR=23,IDBLK=003,IDNUM=0101C,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023005 LU    LOCADDR=0
VP023006 PU    ADDR=23,IDBLK=003,IDNUM=0101A,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023006 LU    LOCADDR=0
VP023007 PU    ADDR=23,IDBLK=003,IDNUM=01018,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023007 LU    LOCADDR=0
VP023008 PU    ADDR=23,IDBLK=003,IDNUM=01016,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023008 LU    LOCADDR=0
VP023009 PU    ADDR=23,IDBLK=003,IDNUM=01014,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023009 LU    LOCADDR=0
VP023010 PU    ADDR=23,IDBLK=003,IDNUM=01012,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023010 LU    LOCADDR=0
VP023011 PU    ADDR=23,IDBLK=003,IDNUM=01010,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023011 LU    LOCADDR=0
VP023012 PU    ADDR=23,IDBLK=003,IDNUM=0100E,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023012 LU    LOCADDR=0
VP023013 PU    ADDR=23,IDBLK=003,IDNUM=0100C,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023013 LU    LOCADDR=0
VP023014 PU    ADDR=23,IDBLK=003,IDNUM=0100A,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023014 LU    LOCADDR=0
VP023015 PU    ADDR=23,IDBLK=003,IDNUM=01008,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023015 LU    LOCADDR=0
VP023016 PU    ADDR=23,IDBLK=003,IDNUM=01006,                          +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,         +
               SSCPFM=USSNTO
VL023016 LU    LOCADDR=0
```

```
*-------------------------------------------------------------------*
*   Switched circuits for private line 024 (2 VCs, IDNUMS 002-004)  *
*-------------------------------------------------------------------*
VP024001 PU    ADDR=24,IDBLK=003,IDNUM=01004,                      +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,     +
               SSCPFM=USSNTO
VL024001 LU    LOCADDR=0
VP024002 PU    ADDR=24,IDBLK=003,IDNUM=01002,                      +
               DISCNT=(YES,F),MAXDATA=1034,MAXPATH=1,PUTYPE=1,     +
               SSCPFM=USSNTO
VL024002 LU    LOCADDR=0
```

NCPROUTE

NCPROUTE is a server that provides an alternative to using the Network Control Program (NCP) as a static host-independent IP router. NCPROUTE has the following effects:

- NCP becomes an active RIP router on a TCP/IP network.
- NCP becomes responsive to SNMP route table queries.

Notes:

1. NCPROUTE requires NCP V7R1, or later.
2. NCPROUTE requires SNALINK LU0 when using NCP V7R3 or previous.
3. SNALINK and IP over CDLC is supported for ESCON, BCCA, and CADS channels.
4. IP over CDLC can be used instead of SNALINK when using NCP V7R4, or later.
5. If using RIP Version 2, NCPROUTE requires NCP V7R6, or later. Also, the NCP generation definition must have VSUBNETS=YES specified on the BUILD statement.
6. NCP versions V6R1 and V6R2 support static IP routing only. NCP uses these static route tables to deliver datagrams over connected TCP/IP networks. NCP V7R1 can be specified only as a host-dependent router and it requires the NCPROUTE server to function as a RIP router.
7. If using NCPROUTE with SNALINK, IP over CDLC channels, and OMPROUTE, you should customize the NCST interface metric on the NCP client side for the SNALINK NCST connection so the routes will be less preferred. This causes OMPROUTE to prefer routes from the IP over CDLC interface over the ones from the SNALINK interface. To customize the interface metric, see the *interface metric* option in "Step 8: Configure the NCPROUTE gateways data set (Optional)" on page 436. Do the same for the SNALINK interface on the MVS host side by customizing the metric in the BSDROUTINGPARMS statement. RIP traffic will be carried over the IP over CDLC interface, while transport PDUs (for example, Hello, Add Route Request, Delete Route Request) will be carried over the SNALINK interface.
8. NCPROUTE does not support zero subnets.

NCPROUTE provides dynamic route table updates for one or more NCP clients that have been generated as IP routers and have NCPROUTE specified as the NCPROUTE server. NCPROUTE tables are updated periodically in the NCP client based on updates sent by the NCPROUTE server. These updates reflect dynamic changes in route states.

An NCPROUTE server at the host uses the Routing Information Protocol (RIP), described in RFC 1058 (RIP version 1) and in RFC 1723 (RIP version 2). The same routing protocols are used by OMPROUTE. NCPROUTE is implemented as a RIP

server operating on an MVS host connected to a RIP client in the NCP. Together they provide the appearance to the TCP/IP network of an IP router using the RIP protocol. The same client/server pair also provides SNMP agent support for network management route table queries. RIP Versions 1 and 2 are currently supported by NCPROUTE. For a brief description of RIP (Versions 1 and 2), see Chapter 6, "Routing," on page 217.

Understanding the NCPROUTE environment

The NCPROUTE server:

- Supports multiple host-attached, link-attached, and remote link-attached NCP clients as illustrated in Figure 48
- Generates RIP datagrams for the NCP to send
- Maintains separate routing tables for each NCP client
- Generates SNMP route table responses for each NCP SNMP agent

The client NCP unit appears as an active router to other RIP routers on the network. Multiple NCP clients can connect to the same NCPROUTE server. Each NCP appears as an IP router to the rest of the network. Each NCP client must have one or more LU0 sessions established with SNALINK. One LU0 session per client is used as the primary session, with the remaining sessions serving as backups.

Figure 48 illustrates the different ways the NCPROUTE server can support NCP clients. NCP3 and NCP4 are host-attached NCP clients, NCP5 and NCP6 are link-attached NCP clients, and NCP1 and NCP2 are remote link-attached NCP clients.

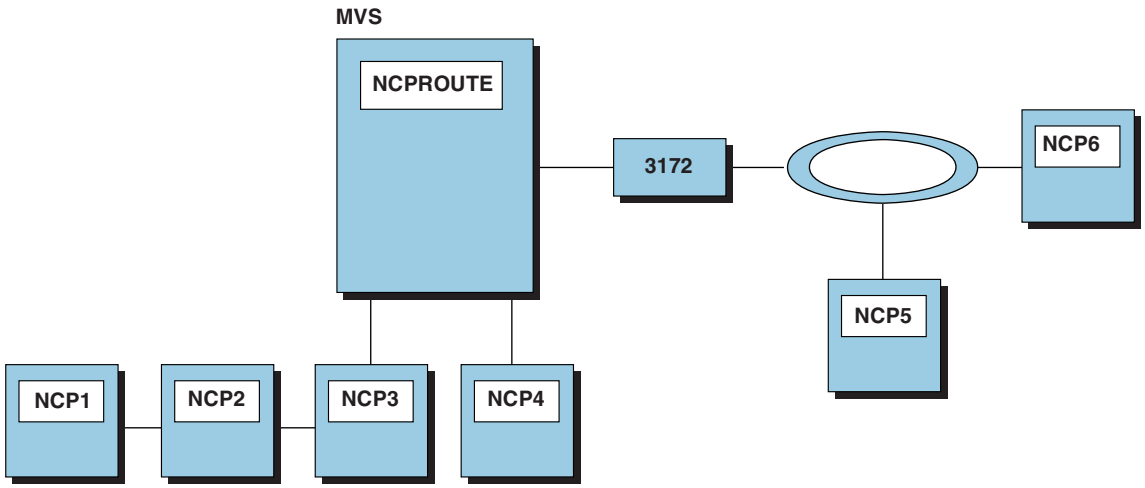


Figure 48. NCPROUTE environment

Server requirements

NCPROUTE processes RIP and SNMP datagrams addressed to all attached NCP units, generates datagrams for the NCP units, and maintains the state of each NCP unit's routing tables.

SNMP support is limited to route table queries. Queries are made to the NCP, which sends the request to the NCPROUTE server for processing.

NCPROUTE operation

An NCP's IPOWNER statement defines the controlling host and the interface this NCP client must use to reach the host. The NCP client initiates contact with NCPROUTE by sending a datagram, known as a "Hello" message, to the controlling host. It transmits this datagram on UDP port 580.

Note: The port number is generated in the NCP (using the UDPPORT keyword on the IPOWNER statement) and configured in NCPROUTE.

The "Hello" message identifies the client NCP and determines which member from the *hlq*.NCPROUTE.GATEWAYS partitioned data set to use for this NCP's route table. Any valid MVS data set name can be used for the gateways data set.

The NCP client then sends a list of its inactive links to NCPROUTE. NCPROUTE uses additional routes defined for this NCP in the NCPROUTE gateways data set, as defined in the NCPROUTE profile. It also uses the inactive links provided dynamically by the NCP to build the current route table for this NCP. The following process is repeated for each NCP that has been generated to act as a RIP router:

1. A RIP packet arrives at the NCP client from a foreign router.
2. The NCP client sends this datagram to the NCPROUTE server.
3. The NCPROUTE server processes the RIP packet.
4. The NCPROUTE server creates a RIP update for an NCP client.
5. This update is sent to the NCP client.
6. The NCP client transmits the datagram to the network.

NCPROUTE sends route table updates to each NCP client every 30 seconds. After a client has been activated, updates must be supplied over each of its interfaces every 30 seconds. The NCPROUTE server creates these updates and sends them to the NCP client along with the IP addresses of other RIP routers that the NCP client should send them to.

At the same time, adjacent RIP routers are providing periodic updates every 30 seconds to NCPROUTE. These updates are sent by the NCP client to the NCPROUTE server, where they are processed, and the results are reflected in future updates back to the NCP client.

The NCP client sends all SNMP and RIP datagrams to the NCPROUTE server for processing. The NCPROUTE server provides RIP packets and SNMP replies to the NCP client to send to their final destination.

NCPROUTE gateways:

Passive RIP route: Information about passive routes is put in NCP's and NCPROUTE's routing tables. A passive entry in NCPROUTE's routing table is used as a placeholder to prevent a route from being propagated and from being overwritten by a competing RIP route. With the exception of directly-connected passive routes, passive routes are not propagated; they are known only by this router.

Using passive routes can create routing loops, so care must be taken when creating them.

Do not define passive routes such as these:
- A to C is via B.
- B to C is via A.

Passive routes should be used when adding routes where the host or network is not running RIP. Passive routes should also be used when adding a default route, because this is the only way to prevent a route from timing out.

External RIP route: External routes are managed by other protocols, for example, the External Gateway Protocol (EGP). NCPROUTE needs to know not to interfere with these routes and not to delete them.

An external entry exists in the NCPROUTE routing table as a place holder to prevent a route from being overwritten by a competing RIP route. External routes are not propagated. NCPROUTE does not manage an external route. Therefore, NCPROUTE only knows that there is an existing route to the host or network and that is the route known by NCP.

External routes should be used when the local machine is running a non-RIP routing protocol that dynamically changes the TCP/IP routing tables. The remote machine does not need to run any routing protocol, since the only concerns are how to route traffic from the local machine to the remote machine, and how to prevent multiple routing protocols from interfering with each other.

RIP route advertising rules:

Note: RIPv1 and RIPv2 protocols are mutually exclusive; you cannot run RIPv1 and RIPv2 simultaneously.

Table 20 illustrates the differences between routing rules on the basis of RIP version.

Table 20. RIP route advertising rules

Version[2]	**Advertised destination route**[1]	**Same subnet as interface**	**Different network from interface with same subnet mask**	**Same network as interface regardless of subnet mask**	**Different network from interface**	**Same supernet as interface**	**Different supernet from interface**
RIPv1	Host	Yes[3]	Yes[3]	Yes[3]	Yes[3]		
	Subnet	No	Yes	No	No		
	Network			No	Yes		
	Supernet						
	Default				Yes[3]		
RIPv2	Host	Yes[3]	Yes[3]	Yes[3]	Yes[3]	Yes[3]	Yes[3]
	Subnet	No	Yes	Yes	Yes	Yes	Yes
	Network			No	Yes	No	Yes
	Supernet			No	Yes	No	Yes
	Default				Yes[5]		

Notes:

1. According to RIP design, route advertising relies on network-specific routes because they are the lowest common denominator. The network-specific routes consist of supernet, network, and subnet routes. The advertising of host specific routes is optional.
2. RIPv1 is the default setting for the RIP version. To set to RIPv2, specify the RIP2 parameter in NCPROUTE Profile and/or on interface options in the NCPROUTE Gateways data set.
3. The optional host specific routes are allowed to be advertised outside networks, and they are advertised in addition to the network specific routes. The option is enabled when the system -h parameter (or SUPPLY HOSTS option in NCPROUTE Gateways data set) is specified.
4. Although it is possible to advertise only the host specific routes using the RIP filters, doing so creates network unreachable problems when some routers in the network do not support the host specific routes. These routers rely on network-specific routes.
5. A default route has a network number of zero and is usually advertised over all network interfaces.
6. It does not matter whether the advertised route is VIPA or not. VIPA routes follow the same advertising rules as the non-VIPA routes.
7. Routes that are subjected to RIP filters may not be advertised at all over certain network interfaces.

NCPROUTE active gateways: Active gateways are treated as remote network interfaces. Active gateways are routers that are running RIP, but are reached through a medium that does not allow broadcasting or multicasting and is not point-to-point, for example, Hyperchannel. NCPROUTE normally requires that routers be reachable by broadcast or multicast for non-point-to-point links or by unicast addresses for point-to-point links. If the interface is neither, then an active gateway entry can add the gateway to NCPROUTE's interface list. NCPROUTE will treat the active gateway as a remote network interface. Note that the active gateway must be directly connected.

Active gateways should be used when the foreign router is reachable over a non-broadcast-capable, non-multicast-capable, and non-point-to-point network, and is directly connected to the local host.

NCPROUTE communicates with active routes by unicast transmissions to the gateway address. Routes are not added immediately to either NCPROUTE or the NCP routing table. They are added and propagated normally when route advertisements arrive from an active gateway. The sole effect of an active gateway statement is to bypass the requirement for broadcast communication on non-point-to-point links. Interfaces that are not broadcast, non-point-to-point, and are not active gateways are assumed to be loopback interfaces to the local machine. Also, while a route to an active gateway might timeout, the interface entry is never removed. If transmissions resume, then the new routes will still be available to the active gateways.

NCPROUTE gateways summary

Table 21 on page 426 provides a list of NCPROUTE gateways and their characteristics.

Table 21. NCPROUTE gateways summary

	Propagate	Kernel	NCPROUTE	Timeout
Dynamic (1)	Yes	Yes	Yes	Yes
Passive	No (2)	Yes	Yes	No
External	No	No	Yes	No
Active	Yes	Yes	Yes	Yes

1 Dynamic routing is provided by NCPROUTE.

2 Except directly-connected passive routes. Directly-connected passive routes are propagated to other network interfaces for network reachability. A directly-connected passive route is one where the gateway address is one of the local interfaces in an NCP client.

RIP input/output filters

The RIP input/output filters provide routing table manipulation and routing control. The filters are provided by NCPROUTE and consist of:

- Route receiving (unconditional and conditional)
- Route noreceiving
- Route forwarding (unconditional and conditional)
- Route noforwarding
- Interface Supply switch
- Interface RIP On/Off switch
- Gateway noreceiving

For more information on these filters, see "Step 8: Configure the NCPROUTE gateways data set (Optional)" on page 436.

Configuring NCPROUTE

Steps to configure NCPROUTE:

1. Specify configuration statements in *hlq*.PROFILE.TCPIP.
2. If using SNALINK, configure VTAM and SNALINK applications.
3. If using IP over CDLC, configure IP over CDLC DEVICE and LINK statements.
4. Update the NCPROUTE cataloged procedure.
5. Update *hlq*.ETC.SERVICES.
6. Configure the host-dependent NCP clients.
7. Configure the NCPROUTE profile data set for SNMP support and specification of an NCPROUTE gateways data set (optional).
8. Configure the NCPROUTE gateways data set for each NCP client (optional).
9. If OMPROUTE is not used, define a directly connected host route to each NCP client.

Figure 49 on page 427 shows the network addresses used in the configuration examples:

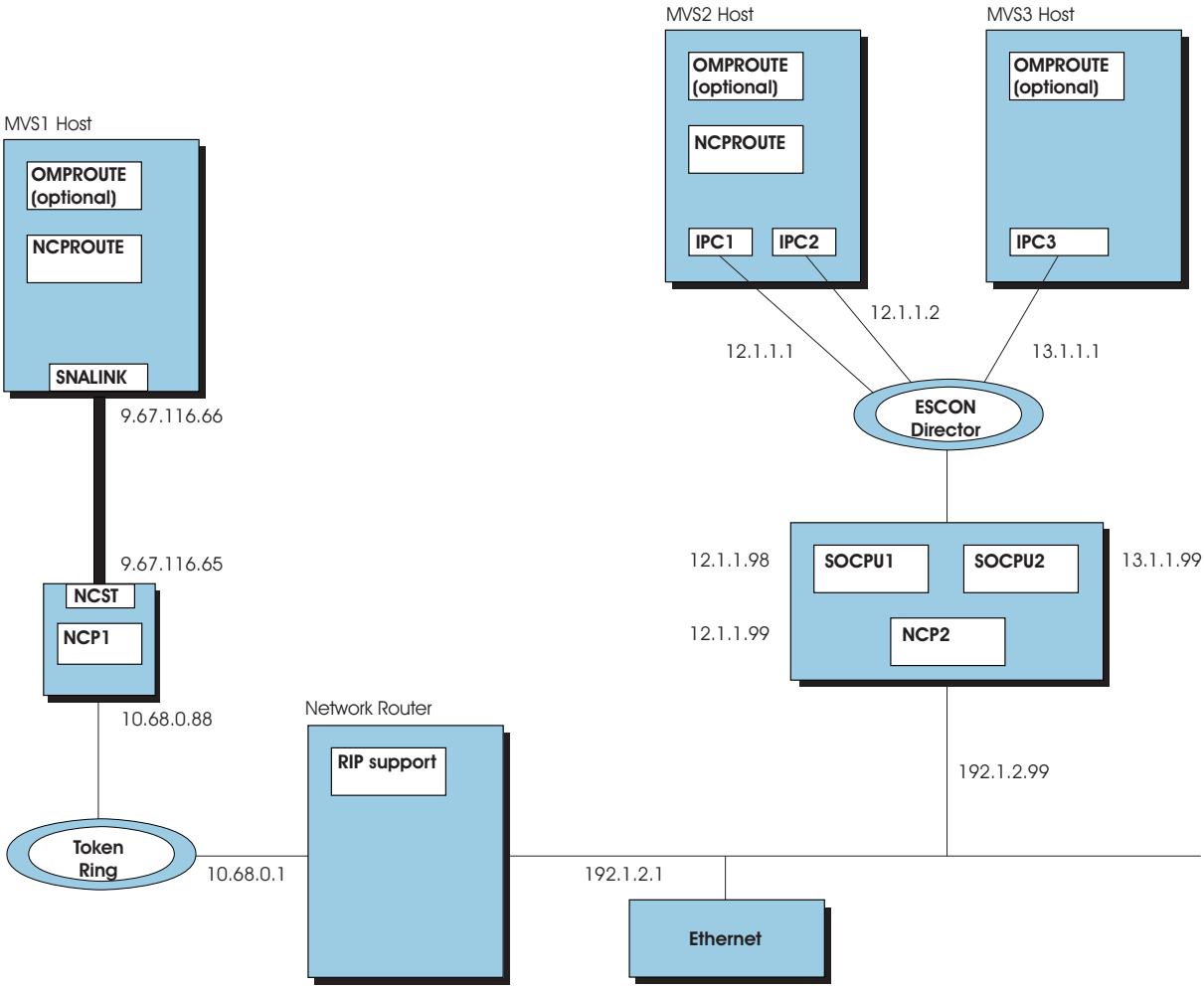


Figure 49. NCPROUTE example configuration

Step 1: Specify configuration statements in *hlq*.PROFILE.TCPIP

1. To have the NCPROUTE server started automatically when the TCPIP address space is started, include the name of the member containing the NCPROUTE cataloged procedure in the AUTOLOG statement in *hlq*.PROFILE.TCPIP:

   ```
   AUTOLOG
      NCPROUT
   ENDAUTOLOG
   ```

2. To ensure that UDP port 580 is reserved for the NCPROUTE server, also add the name of the member containing the NCPROUTE cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP:

   ```
   PORT
      580 UDP NCPROUT
   ```

   Note: This port number must match the one defined in the NCP generation definition (using the UDPPORT keyword on the IPOWNER statement) and assigned in *hlq*.ETC.SERVICES.

3. NCPROUTE also requires HOME and BSDROUTINGPARMS statements for the SNALINK type LU0 and IP over CDLC connections. For example, you would

use this HOME and BSDROUTINGPARMS statement and, optionally, the BEGINROUTES or GATEWAY statement for the configuration shown in Figure 49 on page 427:

```
MVS1:   HOME
           9.67.116.66  SNALINK
           BSDROUTINGPARMS false
             SNALINK 2000     0    255.255.240.0    9.67.116.65
           ENDBSDROUTINGPARMS
MVS2:   HOME
           12.1.1.1   IPC1
           12.1.1.2   IPC2
           BSDROUTINGPARMS false
             IPC1    1000     0    255.255.255.0   12.1.1.98
             IPC2    1000     0    255.255.255.0   12.1.1.99
           ENDBSDROUTINGPARMS
MVS3:   HOME
           13.1.1.1   IPC3
           BSDROUTINGPARMS false
             IPC3    1000     0    255.255.255.128   13.1.1.99
           ENDBSDROUTINGPARMS
```

Notes:

a. If you are not using OMPROUTE to manage the host routes, configure static routes to the NCP client or clients in the BEGINROUTES or GATEWAY statement in *hlq*.PROFILE.TCPIP. The BSDROUTINGPARMS statement is required to route transport PDUs over static routes. See "Step 9: Define a directly connected host route for the NCST session" on page 439 for sample definition. For more information on the BEGINROUTES or GATEWAY statement, see *z/OS Communications Server: IP Configuration Reference* for each NCP client.

b. If using NCPROUTE with OMPROUTE, the BSDROUTINGPARMS statement is required to route transport PDUs prior to OMPROUTE activation. Because the BSDROUTINGPARMS parameters are overridden by the interface parameters defined in the OMPROUTE configuration, ensure that the interface parameters for the SNALINK or IP/CDLC channel connections are identical in both the BSDROUTINGPARMS statement and the OMPROUTE configuration file.

You can find a complete explanation of these configuration statements in *z/OS Communications Server: IP Configuration Reference*.

Step 2: Configure VTAM and SNALINK applications

If you are using NCP V7R3 or previous, NCPROUTE requires SNALINK type LU0 to run. To use SNALINK LU0, verify that you have configured the SNALINK LU0 interface, defined it to VTAM with a VTAM APPL definition, and included the correct DEVICE and LINK statements in *hlq*.PROFILE.TCPIP. For NCP V7R4, or later, IP over CDLC can be used instead of SNALINK.

If you are using the Cross Domain Resource (CDRSC), verify that the cross-domain resource managers are configured in VTAM.

Following is an example of an appropriate VTAM APPL definition:

```
*********************************************************************
*       SNALINK VTAM APPL DEFINITION                                *
*********************************************************************
SNALKLU1 APPL AUTH=(ACQ,VPACE),ACBNAME=SNALKLU1,EAS=12,PARSESS=YES,   *
              SONSCIP=YES,VPACING=0,SRBEXIT=YES
```

Note: The application name (the ACBNAME value, SNALKLU1, in this example) must match the REMLU interface definition in the NCP clients generation program. See the example in "Step 6: Configure the host-dependent NCP clients" on page 430 for more information.

Following is an example of corresponding DEVICE and LINK statements:

```
;
;  DEVICE AND LINK DEFINITIONS FOR SNALINK LU0
;
DEVICE SNA1LINK SNAIUCV SNALINK A04TOLU1 SNALPROC
LINK SNALINK SAMEHOST 1 SNA1LINK
;
```

Note: The LU name on the DEVICE statement (A04TOLU1 in this example) must match the LU name of the NCST interface definition in the NCP clients generation program. See the example in "Step 6: Configure the host-dependent NCP clients" on page 430 for more information.

If you want the SNALINK device to start automatically, verify that you have a START statement for this device in *hlq*.PROFILE.TCPIP. For example, `START SNA1LINK`. Otherwise, you will have to start the device manually.

Step 3: Configure the IP over CDLC DEVICE and LINK statements

For NCPROUTE, IP over CDLC can be configured along with SNALINK for NCP V7R3, or later, or it can be used to replace SNALINK for NCP V7R4, or later.

Following is an example of corresponding DEVICE and LINK statements for the configuration shown in Figure 49 on page 427 for the MVS2 host:

```
;
;  DEVICE AND LINK DEFINITIONS FOR IP OVER CDLC
;
DEVICE IPC1NCP CDLC 013 40 40 1024 1024
LINK IPC1 CDLC 0 IPC1NCP
;
DEVICE IPC2NCP CDLC 014 40 40 1024 1024
LINK IPC2 CDLC 0 IPC2NCP
```

Note: If you want a CDLC device to start automatically, verify that you have a START statement for this device in *hlq*.PROFILE.TCPIP, for example, START IPC1NCP. Otherwise, you will have to start the device manually.

Step 4: Update the NCPROUTE cataloged procedure

Update the NCPROUTE cataloged procedure by copying the sample in SEZAINST(NCPROUT) to your system or recognized PROCLIB. Specify NCPROUTE parameters and change the data set names to suit your local configuration. See Figure 50 on page 435 for an illustration of NCPROUTE data set relationships.

Step 5: Update *hlq*.ETC.SERVICES

NCPROUTE uses the *hlq*.ETC.SERVICES data set to determine the port number on which to run. This data set can be used to define a port number other than the reserved well-known port for NCPROUTE. This data set must exist for NCPROUTE to run.

The ETC.SERVICES data set is dynamically allocated using the standard search sequence for data set names. This data set also can be explicitly allocated in the NCPROUTE cataloged procedure using the //SERVICES DD statement.

The entries in *hlq*.ETC.SERVICES are case and column sensitive. They must be in lowercase and begin in column 1.

Add the following lines to the *hlq*.ETC.SERVICES data set:

```
ncproute  580/udp
router    520/udp
```

Note: Verify that the NCPROUTE service port number is the port being used by the NCP clients. This number should match the port number defined in the NCP generation definition using the UDPPORT keyword on the IPOWNER statement. This port number does not necessarily have to match the reserved port number for NCPROUTE on the PORT statement in *hlq*.PROFILE.TCPIP.

The reserved router service port number is 520. It is required for the NCPROUTE transport of RIP packets to NCP clients which are responsible for broadcasting the packets to other RIP routers. It cannot be overridden.

If you want to use name aliases, refer to INFO APAR II08205 for information.

Step 6: Configure the host-dependent NCP clients

You should refer to the appropriate NCP documentation for more information about defining and generating the NCP and creating route information tables.

- For more information about defining IP, refer to *NCP, SSP, and EP Resource Definition Guide*.
- For more information about the IP Dynamics function, refer to *NCP and EP Reference*.
- For more information about NCP generation definitions for IP, refer to *NCP, SSP, and EP Resource Definition Reference*.
- For more information about generating NCP as an IP router, refer to *NCP, SSP, and EP Generation and Loading Guide*.

Note: See NCPROUTE notes in on page 421.

Generating the routing information tables: To support IP dynamics, NCP's Network Definition Facility (NDF) builds a routing information table (RIT) for networks and subnetworks for use by TCP/IP at NCP generation time.

The RIT consists of routing tables that are generated from the NCP IPROUTE and IPLOCAL statements. During NCP generation, the RIT is added as a member of the NCP load library partitioned data set ncp.v7r1.ncpload. You identify the member name of ncp.v7r1.ncpload that NCPROUTE uses at execution time with the NEWNAME parameter of the BUILD statement for each NCP client generation.

Determining the gateway route table name: There is one RIT in the ncp.v7r1.ncpload data set for each NCP client this server supports. The NCPROUTE server receives the NCP name from an NCP client in the "Hello" message. This name is used as the base to determine the member name in the ncp.v7r1.ncpload partitioned data set to use for the initial RIT for this NCP client. The RIT member name in the ncp.v7r1.ncpload data set is the NEWNAME parameter of the BUILD statement for the NCP generation with a suffix of **P** added. Specify a unique name on the NEWNAME parameter of the BUILD statement for each NCP client. This name is also used as the member name if the optional gateways data set (GATEWAYS_PDS) is specified in the NCPROUTE profile. The RIT is accessed by NCPROUTE from a //STEPLIB DD statement in the NCPROUTE cataloged procedure, LINKLST, or authorized library.

NCST session interface definition: The NCP Connectionless SNA Transport (NCST) interface is used to establish a session that can provide a connection to another IP node (NCP or z/OS) over a SNA network. Use this definition when using NCST PU interfaces to communicate with NCPROUTE using SNALINK devices with the MVS host. The NCST interface must be defined to match the SNALINK LU0 interface in VTAM so that an NCP client can establish a connection with NCPROUTE. The LU statement in the NCST interface definition tells VTAM which interface to use for the SNALINK application. The following are important keywords in this definition:

NCST
Specifies the protocol type. Must be coded as IP for internet protocol.

INTFACE
Specifies the name of the interface and the maximum transfer unit (MTU) size for the NCST session to the VTAM owner (IPOWNER).

REMLU
Specifies the name of the remote LU for the SNALINK LU0 VTAM connection. This name must match:
- The APPLID in the PROC statement of the SNALINK cataloged procedure
- The application name in the VTAM APPL definition

Note: If you define a backup NCST SNALINK session, the REMLU can specify the primary logical name for the remote LU or a different remote LU. Ensure that the MTU sizes are the same for the backup NCST sessions.

Following is an example of an NCST session interface definition:

```
*********************************************************************
*     NCST IP INTERFACES                                            *
*********************************************************************
A04NCSTG GROUP NCST=IP,LNCTL=SDLC,VIRTUAL=YES
A04NCSTL LINE LINEFVT=CXSXFVT,PUFVT=CXSXFVT,LUFVT=(CXSXFVT,CXSXFVT),   *
               LINECB=CXSXLNK
A04NCSTP PU VPACING=0,PUTYPE=2,PUCB=CXSP0000
*
A04TOLU1 LU INTFACE=(NCSTALU1,1492),REMLU=SNALKLU1,LUCB=(CXSL0000,CXSS0*
               000),LOCADDR=1
*
```

Note: The NCST LU name (A04TOLU1 in this example) must match the LU name on the SNALINK LU0 DEVICE statement in *hlq*.PROFILE.TCPIP. See the example in "Step 2: Configure VTAM and SNALINK applications" on page 428 for more information.

Channel PU interface definition: Use this definition with channel PU interfaces (ESCON, BCCA, or CADS) to communicate with NCPROUTE using IP over CDLC devices with the MVS host.

Following is an example of channel PU interface definition using the ESCON channel type:

```
*********************************************************************
*   PHYSICAL ESCON CHANNEL DEFINITIONS                              *
*********************************************************************
*
A04PSOC1 GROUP LNCTL=CA,MONLINK=NO,NPACOLL=NO,XMONLNK=YES
               SPEED=144000000,SRT=(32768,32768)
*
A04S2240 LINE  ADDRESS=2240
A04P2240 PU    ANS=CONTINUE,PUTYPE=1
```

```
*
*********************************************************************
*    LOGICAL ESCON CHANNEL DEFINITIONS                              *
*********************************************************************
*
A04PSOCB GROUP LNCTL=CA,PHYSRSC=A04P2240,NPACOLL=NO,
               DELAY=0.2,MAXPU=16,MODETAB=AMODETAB,
               DLOGMOD=INTERACT,SPEED=144000000,
               SRT=(21000,20000),PUDR=YES,
               TIMEOUT=150.0,CASDL=0.0
*
A04LSOC2 LINE  ADDRESS=NONE,HOSTLINK=1
A04L2S1  PU    ADDR=01,PUTYPE=1,ARPTAB=(10,,NOTCANON),
               INTFACE=SOCPU1
A04L2S2  PU    ADDR=02,PUTYPE=1,ARPTAB=(10,,NOTCANON),
               INTFACE=SOCPU2
```

NCP host interface definition: The IPOWNER statement in the NCP generation definition contains the TCP/IP host information and tells NCP which interface to use for NCPROUTE. The following are important keywords on the IPOWNER statement:

INTFACE
Specifies the name of the interface to the owning TCP/IP host that is running NCPROUTE.

HOSTADDR
Specifies the IP address of the owning TCP/IP. This address must match the IP address in the HOME statement in *hlq*.PROFILE.TCPIP data set for a SNALINK or IP over CDLC interface.

UDPPORT
Specifies the UDP port number for NCPROUTE. The default is 580. This port number must match the NCPROUTE service port number defined in the *hlq*.ETC.SERVICES data set. See "Step 5: Update ***hlq***.ETC.SERVICES" on page 429 for more information.

The IPLOCAL statement in the NCP generation definition contains the NCP routing information for the local attached routes. During NCP generation, this information gets included in the Routing Information Table (RIT) which NCPROUTE uses to build the interface and routing tables. IPLOCAL routes are predefined as permanent or static to prevent modification by NCPROUTE. The following are important keywords on the IPLOCAL:

INTFACE
Specifies the name of the locally attached interface.

LADDR
Specifies the IP address of the locally attached interface.

P2PDEST
For point-to-point interfaces only. Specifies the IP address of the remote end of the point-to-point link.

PROTOCOL
Specifies the type of protocol to be used for the interface. The default is RIP which indicates that the interface is RIP-managed by NCPROUTE.

SNETMASK
Specifies the subnetwork mask for a route to a network that is subnetted. Because RIP does not support variable subnetwork masking, this value must be equal to the subnetwork mask of the route's destination.

The IPROUTE statement in the NCP generation definition contains the NCP routing information for optional predefined routes. During NCP generation, this information gets included in RIT which NCPROUTE uses to add the routes to its routing tables. IPROUTE routes can be predefined as permanent or non-permanent for route management control by NCPROUTE. The following are important keywords on the IPROUTE statement:

INTFACE
Specifies the name of the locally attached interface for the route.

DESTADDR
Specifies the route's destination IP address.

DISP
Specifies the disposition for the route. A disposition of PERM indicates that this route is a permanent route and will not be modified by NCPROUTE. The default is NONPERM.

HOSTRT
Indicates whether this is a host route. The default is NO.

NEXTADDR
Specifies the IP address of the gateway through which the route can reach its destination. A value of 0 indicates that there is no gateway.

The following example shows typical NCP RIP router generation source statements.

```
*********************************************************************
*       IP ROUTING DEFINITIONS                                      *
*********************************************************************
*
        IPOWNER INTFACE=NCSTALU1,HOSTADDR=9.67.116.66,              *
                NUMROUTE=(100,100,100),MAXHELLO=25,UDPPORT=580
*
        IPLOCAL LADDR=9.67.116.65,INTFACE=NCSTALU1,METRIC=1,        *
                P2PDEST=9.67.116.66,PROTOCOL=RIP,SNETMASK=FFFFF000
        IPLOCAL LADDR=10.68.0.88,INTFACE=TR88,METRIC=1,             *
                SNETMASK=FFFFF000
        IPLOCAL LADDR=10.68.0.92,INTFACE=TR92,METRIC=1,             *
                SNETMASK=FFFFF000
*
        IPROUTE DESTADDR=11.0.0.1,NEXTADDR=0,INTFACE=TR88,METRIC=2, *
                DISP=PERM,HOSTRT=YES
        IPROUTE DESTADDR=12.0.0.0,NEXTADDR=13.0.0.1,INTFACE=TR92,   *
                METRIC=2,DISP=NONPERM
```

The following example shows IPOWNER and IPLOCAL statements for the ESCON channel PU interfaces in the configuration for NCP2 as shown in Figure 49 on page 427.

```
*************************************************************
*   IP ROUTING DEFINITIONS USING ESCON CHANNEL INTERFACES   *
*************************************************************
*
IPOWNER INTFACE=SOCPU1,UDPPORT=580,NUMROUTE=(110,120,130),
        HOSTADDR=12.1.1.1
*
IPLOCAL LADDR=12.1.1.98,INTFACE=SOCPU1,METRIC=1,
        P2PDEST=12.1.1.1,PROTOCOL=RIP,SUBNETMASK=FFFFFF00
*
IPLOCAL LADDR=12.1.1.99,INTFACE=SOCPU1,METRIC=1,
        P2PDEST=12.1.1.2,PROTOCOL=RIP,SUBNETMASK=FFFFFF00
```

```
*
IPLOCAL LADDR=13.1.1.99,INTFACE=SOCPU2,METRIC=1,
        P2PDEST=13.1.1.1,PROTOCOL=RIP,SUBNETMASK=FFFFFF80
```

Step 7: Configure the NCPROUTE profile data set (Optional)

To build the NCPROUTE profile, create a data set and specify its name in the //NCPRPROF DD statement in the NCPROUTE cataloged procedure. You can find a sample in SEZAINST(EZBNRPRF). Include configuration statements in this data set to define SNMP functions and to identify the NCPROUTE gateways data set. For more information, refer to *z/OS Communications Server: IP Configuration Reference*.

RIP_SUPPLY_CONTROL *supply_control*

Specifies one of the following options on a server-wide basis:

- RIP1—Unicast/Broadcast RIP Version 1 packets (Default)
- RIP2B—Unicast/Broadcast RIP Version 2 packets (Not Recommended)
- RIP2M—Unicast/Multicast/Broadcast RIP packets (Migration)
- RIP2—Unicast/Multicast RIP Version 2 packets
- NONE—Disables sending RIP packets

Note: If RIP2 is specified, the RIP Version 2 packets are multicast over multicast-capable interfaces only. No RIP packets are sent over multicast-incapable interfaces. For RIP2M, the RIP Version 2 packets are multicast over multicast-capable interfaces and RIP Version 1 packets over multicast-incapable interfaces. For RIP2B, the RIP Version 2 packets are unicast or broadcast; this option is not recommended since host route misinterpretations by adjacent routers running RIP Version 1 can occur. For this reason, RIP2B may become obsolete in a future release. For point-to-point interfaces that are non-broadcast and multicast-incapable, the RIP Version 2 packets are unicast.

RIP_RECEIVE_CONTROL *receive_control*

Specifies one of the following options on a server-wide basis:

- RIP1—Receive RIP Version 1 packets only
- RIP2—Receive RIP Version 2 packets only
- ANY—Receive any RIP Version 1 and 2 packets (Default)
- NONE—Disables receiving RIP packets

Note: If the client NCP does not support variable subnetting, the default of ANY is changed to RIP1.

RIP2_AUTHENTICATION_KEY *authentication_key*

Specifies a plain text password *authentication_key* containing up to 16 characters. The key is used on a router-wide basis and can contain mixed case and blank characters. Single quotes (') can be included as delimiters to include leading and trailing blanks. The key will be used to authenticate RIP Version 2 packets and be included in the RIP updates for authentication by adjacent routers running RIP Version 2. For maximum security, set RIP_SUPPLY_CONTROL and RIP_RECEIVE_CONTROL to RIP2. This will discard RIP1 and unauthenticated RIP2 packets. A blank key indicates that authentication is disabled. Following are examples of authentication passwords:

```
my password        (no leading or trailing blanks)
' my password '    (leading and trailing blanks)
''abc''            (single quotes part of password)
'     '            (5-character blanks)
```

SNMP_AGENT *host_name*
Specifies the host name or IP address of the host running an SNMP daemon. Only one NCPROUTE server can use a particular SNMP agent at a time.

SNMP_COMMUNITY *community_name*
Specifies a community name that SNMP applications must use to access data that the agent manages. Protect this information accordingly.

GATEWAY_PDS *dsname*
Specifies the optional partitioned data set that contains GATEWAY information for each client NCP. Quotation marks are not needed when specifying *dsname*. One member for each NCP client of this data set must be configured to match the NCP NEWNAME parameter with the **P** suffix which is the same as the NCP's RIT member name. See "Step 8: Configure the NCPROUTE gateways data set (Optional)" on page 436 for information on defining the statements necessary for the members of this data set.

Note: You can use a semicolon in column 1 to permit comments in the profile. Blank lines are also permitted.

Figure 50 shows the relationship between the data set names specified in the NCPROUTE cataloged procedure and the NCPROUTE profile, as well as the relationship between the members of the gateways PDS and the ncpload PDS.

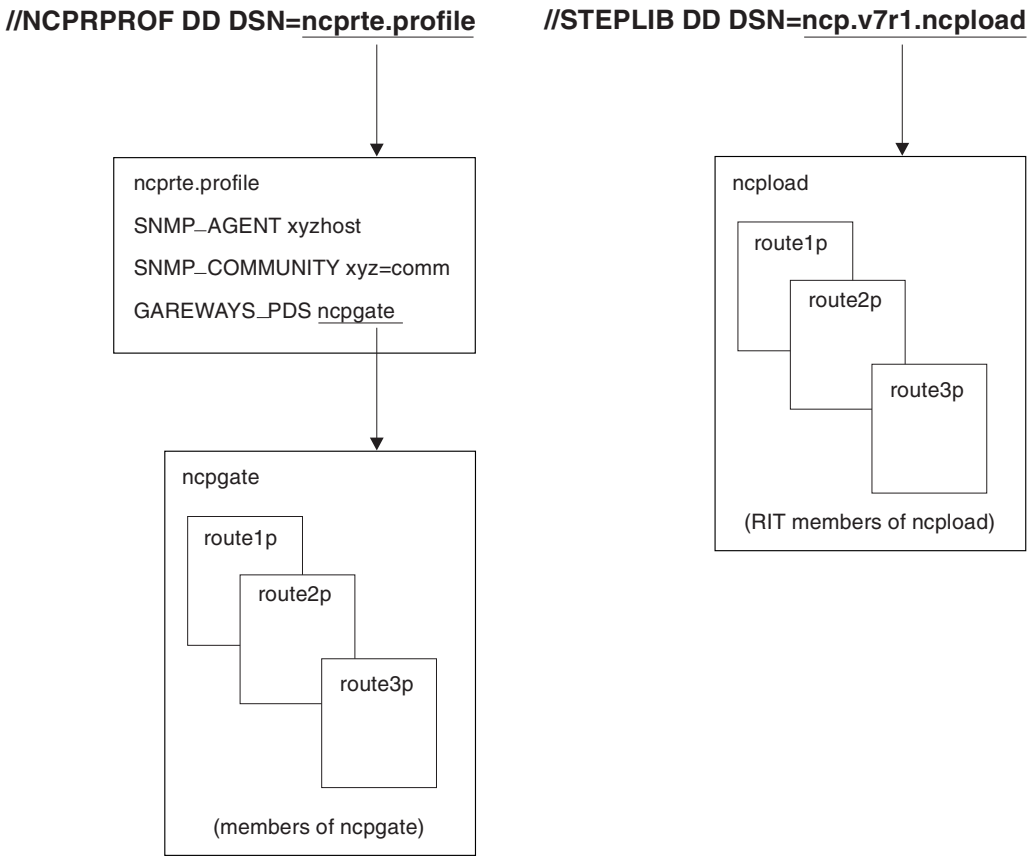


Figure 50. NCPROUTE data sets relationship

Step 8: Configure the NCPROUTE gateways data set (Optional)

The gateways data set is used to identify routes not defined in the NCP routing information table.

The NCPROUTE gateways data set is optional. However, if you use it, you must include the GATEWAY_PDS statement in the NCPROUTE profile to specify the gateway data set name. The NCPROUTE server queries the gateways data set for static routing information. It also dynamically receives routing information from the NCP client portion of this RIP router.

Allocate the gateways data set with partitioned organization (PO), a fixed block format (FB), a logical record length of 80 (LRECL), and any valid block size value for a fixed block, such as 3120.

A `passive` entry in the gateways data set is used to add a route to a part of the network that does not support RIP. An `external` entry in the gateways data set indicates a route that should never be added to the routing tables. If another RIP server offers this route to your host, the route is discarded and not added to the routing tables. An `active` entry indicates a gateway that can only be reached through a network that does not allow or support link-level broadcasting or multicasting.

Note: The gateways data set is not related to the GATEWAY statement used in *hlq*.PROFILE.TCPIP data set.

To configure NCPROUTE, add an entry to the gateways data set for each route not defined in the NCP RIT. Use the options statement to define the characteristics of the routes in this member of the PDS.

Configuring a passive route: Figure 51 illustrates an NCPROUTE configuration using NCP as the destination hosts. In other configurations, destination hosts might not necessarily be NCPs.

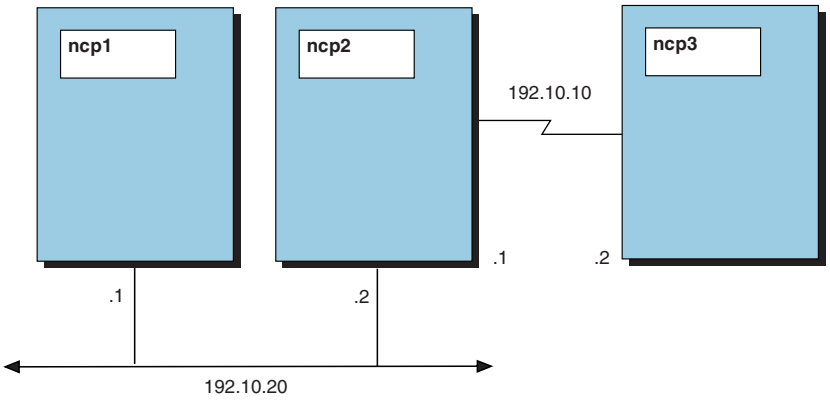


Figure 51. NCPROUTE configuration example of a passive route

Using Figure 51, assume that your NCP client `ncp1` is channel-attached to an MVS host running an NCPROUTE server. The other two NCP clients, `ncp2` and `ncp3`, are not running a RIP server. Also assume that permanent routes to `ncp2` and `ncp3` are not defined with the IPROUTE definitions in the NCP generation definition for `ncp1`. Your NCPROUTE server cannot learn a route to `ncp3`, because `ncp2` is not running a RIP server. Your NCPROUTE server sends routing updates to `ncp3` over the link to `ncp2`, but never receives a routing update from `ncp2`. After 180 seconds, your NCPROUTE server deletes the route to `ncp2`. This problem is inherent to the

RIP protocol and cannot be prevented. Therefore, you need to add a passive route to ncp3 in the NCPROUTE gateways data set for the NCP client ncp1. This is the data set defined by the GATEWAYS_PDS statement in the NCPROUTE profile.

You can use either of the following gateway statements:

```
host  ncp3         gateway  ncp2         metric 2  passive

host  192.10.10.2  gateway  192.10.20.2  metric 2  passive
```

Similarly, because ncp2 is not running an RIP server supported by NCPROUTE, you need to add a directly-connected passive route as follows:

```
host  ncp2         gateway  ncp1         metric 1  passive
```

A directly-connected passive route is one where the gateway address or name is one of the local interfaces in the NCP generation.

Assume that your NCP client is now ncp2 and is running an NCPROUTE server. ncp1 is also running a RIP server, but ncp3 is not. Your NCPROUTE server sends routing information updates to ncp3 over the link to ncp3 but never receives a routing update from ncp3. After 180 seconds, your NCPROUTE server deletes the route to ncp3.

You should add a passive route to ncp3 as follows:

```
   host  ncp3        gateway  ncp2        metric 1  passive
```

ncp1 cannot reach ncp3 unless a passive routing entry is added to ncp1. For example:

```
host   ncp3    gateway  ncp2  metric  2  passive
```

or

```
host  192.10.10.2  gateway  192.10.20.2  metric  2  passive
```

Configuring an external route: Using Figure 51, assume that your NCP client ncp1 is channel-attached to an MVS host running an NCPROUTE server. The other two NCP clients, ncp2 and ncp3, are also running a RIP server. Your NCPROUTE server normally learns a route to ncp3 from ncp2, because ncp2 is running a RIP server. You might not want ncp1 to route to ncp3 for security reasons. For example, a university might want to prevent student hosts from routing to administrative hosts.

To prevent your NCPROUTE server from adding a route to ncp3, add an external route to the NCPROUTE gateways data set. This is the data set defined by the GATEWAYS_PDS statement in the NCPROUTE profile. You can use either of the following gateway statements:

```
host  ncp3         gateway   ncp2        metric 2  external

host  192.10.10.2  gateway  192.10.20.2  metric 2  external
```

Configuring an active gateway:

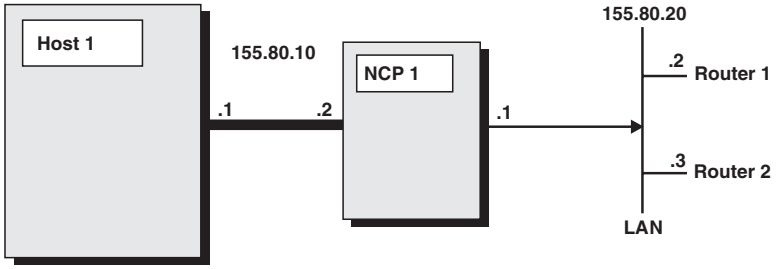


Figure 52. Configuring an active gateway

As shown in Figure 52, assume that your NCP client is `ncp1`, which is channel-attached to an MVS host running an NCPROUTE server and that it has a network attachment adapter that does not support link-level broadcasting or one that does not support ARP processing. Also, assume that there are routers `Router1` and `Router2` on the local area network. Because the IP addresses `Router1` and `Router2` are unknown by `ncp1`, they have to be manually configured in NCPROUTE for NCPROUTE to communicate with them. Configuring active gateways for `Router1` and `Router2` as remote network interfaces enables NCPROUTE to send RIP responses to the target addresses.

1. Specify IP addresses for each network adapter (without link-level broadcasting or ARP support) attached to the local network in the NCP client according to the NCP generation definition. For example, 155.80.20.1 is the IP address for the local network adapter attachment in `ncp1`.
2. Define active gateways for the remote routers in the NCPROUTE gateways data set specified on the GATEWAYS_PDS statement in the NCPROUTE profile:

```
active active gateway 155.80.20.2 metric 1 active
active active gateway 155.80.20.3 metric 1 active
```

NCPROUTE will use these active gateway addresses as the destination addresses to send RIP responses to the remote routers. In addition, NCPROUTE will continue to receive RIP responses from the active gateways over the NCP client.

Configuring a default route: A default route is typically used on a gateway or router to an internet, or on a gateway or router that uses another routing protocol, whose routes are not reported to other local gateways or routers.

To configure a route to a default destination, add a default route using the IPROUTE statement in the NCP generation definition. For example, if the default destination router has a gateway address 9.67.112.1, an IPROUTE statement might look like:

```
IPROUTE  DESTADDR=0.0.0.0,NEXTADDR=9.67.112.1,INTFACE=TR88,
         METRIC=1,DISP=PERM
```

An easier way would be to use the passive route definition specified in the NCPROUTE gateways data set for the NCP client. For example, the gateways entry would look like:

```
net 0.0.0.0 gateway 9.67.112.1 metric 1 passive
```

Only one default route to a destination gateway or router can be specified. For an NCP client, NCPROUTE currently does not support multiple default routes.

Configuration examples: The following example shows the contents of an NCPROUTE gateways data set containing multiple entries:

```
options default.router no trace.level 4 supply on
net testnet gateway 9.0.0.100 metric 1 passive
net 2.0.0.2 gateway 9.0.0.101 metric 2 external
host 2.0.0.3 gateway 9.0.0.102 metric 3 passive
host 2.0.0.4 gateway 9.0.0.103 metric 2 external
active active gateway 2.0.1.1 metric 1 active
```

In the second entry, the route indicates that NCPROUTE can reach network `testnet` through the gateway `9.0.0.100`, and that it is one hop away. This passive route is not broadcast to other RIP routers.

In the third entry, the route indicates that NCPROUTE can reach network `2.0.0.2` through the gateway `9.0.0.101`, and that it is two hops away. Because this route is external, NCPROUTE should not add routes for this network to the routing tables and routes received from other RIP routers for this network should not be accepted.

In the fourth entry, the route indicates that NCPROUTE can reach host `2.0.0.3` through gateway `9.0.0.102`, and that it is one hop away. This passive route is not broadcast to other RIP routers.

In the fifth entry, the route indicates that NCPROUTE can reach host `2.0.0.4` through gateway `9.0.0.103`, and that it is two hops away. Because this route is external, NCPROUTE should not add routes for this network to the routing tables, and routes received from other RIP routers for this network should not be accepted.

The sixth entry shows an active gateway. Note that it is specified as the last entry in the data set.

Note: If a default route is to be defined to a destination gateway or router, configure a default route in this gateways data set (if the default route is not defined in a NCP client's generation definition).

Step 9: Define a directly connected host route for the NCST session

If you are not using OMPROUTE, you need to configure a directly connected static host route using the BEGINROUTES or GATEWAY statement in *hlq*.PROFILE.TCPIP. For example, if you are using SNALINK as the host route and have the IP addresses shown in Figure 49 on page 427, the GATEWAY statement might look like this:

```
GATEWAY
; net_number first_hop link_name packet_size     subnet_mask  subnet_value
  9.67.116.65     =    SNALINK     32758           HOST
```

See *z/OS Communications Server: IP Configuration Reference* and the GATEWAY syntax information in "Step 8: Configure the NCPROUTE gateways data set (Optional)" on page 436 for more information about configuring GATEWAYS statements.

Note: The host routes on the MVS host are managed by TCP/IP, as defined on the BEGINROUTES or GATEWAY statement or as added dynamically by

OMPROUTE. NCPROUTE does not manage the host routes on the MVS host. It only manages the routes on the NCP clients.

Controlling the NCPROUTE address space with the MODIFY command

You can control most of the functions of the NCPROUTE address space from the operator's console using the MODIFY command.

For information about modifying the NCPROUTE address space with the MODIFY command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Chapter 10. Accessing remote hosts using Telnet

Telnet is a terminal emulation protocol that allows end users to log on to remote host applications as though they were directly attached to that host. Telnet protocol requires that the end user have a Telnet client emulating a type of terminal the host application will understand. The client connects to a Telnet server, which communicates with the host application. The Telnet server acts as an interface between the client and host application. A PC can support several clients simultaneously, each with its own connection to any Telnet server. This chapter describes how to set up and use the following:

TN3270E Telnet server
: Provides access to z/OS VTAM SNA applications on the MVS host using Telnet TN3270E, TN3270, or linemode protocol.

z/OS UNIX Telnet server
: Provides access to z/OS UNIX shell applications on the MVS host using Telnet linemode protocol.

It is possible to use the same port for both Telnet servers. For an overview of port management, see "Port management overview" on page 73. For more specific information on the PORT BIND statement, refer to "Setting up reserved port number definitions in PROFILE.TCPIP" on page 203.

TN3270E Telnet server

The TN3270E Telnet server acts as an interface between IP and SNA networks. End users in an IP network connect to the server which is also a VTAM application. The server activates one SNA application minor node logical unit (LU) to represent each Telnet IP client. The Telnet application LU establishes a session with a VTAM host application (for example, CICS), simulating a terminal (LU0 or LU2) or a printer (LU1 or LU3). To enable connections, the VTAM and either the TCP/IP or Telnet configuration data sets must be modified with Telnet statements. These statements describe the Telnet LUs, a listening port, and the characteristics of that port. After Telnet is started, VARY and DISPLAY commands specifically related to the Telnet server can be used to alter the state of Telnet or display information about Telnet. For more information about these command sets, refer to *z/OS Communications Server: IP System Administrator's Commands*.

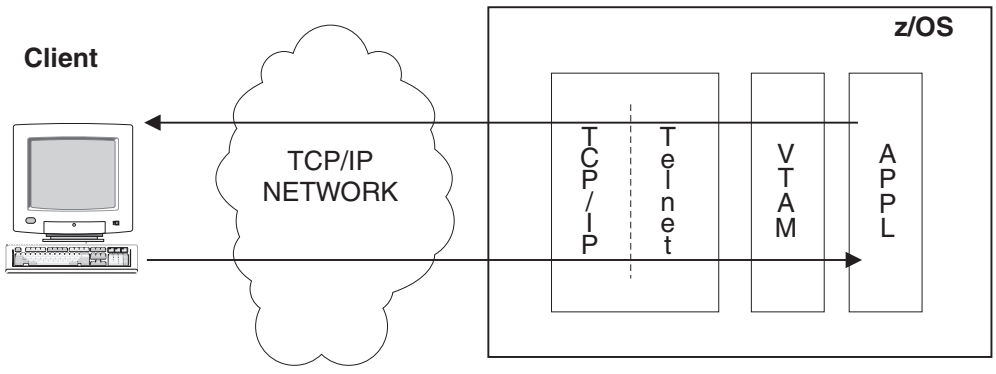


Figure 53. Telnet connectivity

Getting started

Whether Telnet is running within the TCP/IP stack or in its own address space, the following steps must be taken to start a Telnet session:

1. Set up VTAM LUs for use by Telnet.
2. Start Telnet, either with TCP/IP or in its own address space.
3. Start a client emulator and log on to a SNA application.

Set up VTAM LUs

Telnet must be able to activate VTAM LUs to represent the client. Ensure an APPL major node for Telnet LUs is in the concatenation of data sets specified on the VTAMLST DD statement in the procedure used to start VTAM. The LUs need to be in a connectable state. To do this, activate the APPL major node that represents the Telnet LUs.

Start Telnet

When Telnet runs as part of the TCP/IP stack, a Telnet port is activated if Telnet profile statements are specified in the initial profile, or in a data set referenced by a VARY TCPIP,,OBEYFILE command.

When Telnet runs as its own application running in its own address space, set up for the job is necessary. For details, see "Telnet in its own address space" on page 443.

Start a client emulator

A Telnet client emulator is provided for the TSO environment. To start the Telnet client emulator, issue the following TSO command, where *ipaddr* is the TCP/IP loopback address:

```
TELNET ipaddr
```

The command defaults to port 23. The Telnet server uses the first Telnet LU available to establish a session with TSO. The sample profiles define the default application to be TSO. When you connect in, Telnet immediately initiates a session with TSO, which then sends a prompt for a user ID. Enter any valid TSO user ID.

Telnet in the TCP/IP address space

Telnet running within the TCP/IP stack is part of the initial verification procedure (IVP). With the IVP, an end user can:

- Start the TSO Telnet client. For guidance on using the client component, refer to *z/OS Communications Server: IP User's Guide and Commands*.
- Establish a Telnet connection to the Telnet server using loopback.
- Begin a session with another TSO user ID on the same host.

Establishing a new session confirms that the Telnet server is working properly. The VTAM configuration data set for Telnet is SEZAINST(IVPLU), and the Telnet configuration statements are part of the SEZAINST(SAMPPROF) TCP/IP configuration data set. For more information on IVP, refer to *z/OS Communications Server: IP Programmer's Guide and Reference*.

Starting a Telnet session: You can use the IVP sample configuration data sets to start a simple Telnet. Follow these steps to establish a TSO session over a Telnet connection:

1. Ensure that the sample SEZAINST(IVPLU) member is in the concatenation of data sets specified on the VTAMLST DD statement in the procedure used to start VTAM, and activate the major node TCPABC to put the LUs in the necessary connectable state.
2. When the TCP/IP stack is started using SEZAINST(SAMPPROF), message EZZ6003I is displayed. This indicates that Telnet is ready to accept connections.
3. Start a TSO Telnet client emulator by issuing the following TSO command:
   `TELNET ipaddr`
4. Enter any valid TSO user ID.

This sample is designed to verify that the basic Telnet requirements are in place and functioning. Several features are not included in this sample, but you should consider them before using Telnet in production. Examples of Telnet features are LU and application mapping, security, and connection persistence. To learn more about Telnet capabilities and better understand TN3270, read the remainder of this chapter.

Telnet in its own address space

Telnet can be started as part of the TCP/IP stack, or it can run in its own address space separate from the stack. Advantages of running Telnet in its own address space as its own application include the following:

- Telnet priority can be set to a different priority than that of TCP/IP.
- Telnet can be stopped and restarted without stopping TCP/IP.
- When the TCP/IP stack is stopped, Telnet remains active.
- Separating Telnet and TCP/IP makes problem diagnosis easier.
- You can start multiple instances of Telnet.
- In a common INET environment, Telnet can be associated with multiple stacks, or have affinity to a single stack by using the TCPIPJOBNAME statement in TELNETGLOBALS.

Command processing: Commands are entered exactly as they are for the Telnet server running within the TCP/IP address space. When Telnet runs in its own address space, the procedure name must be specified on all commands. Otherwise, the command is processed by the default TCP/IP stack instead of the Telnet procedure. For example, assuming the Telnet procedure name is TN3270, the profile display command is as follows:

```
D TCPIP,TN3270,TELNET,PROFILE
```

If the procedure does not exist or the TCPIP keyword is typed incorrectly, the command is assumed to be a TCP/IP command and a TCP/IP error message is displayed.

Association with a TCP/IP stack: For some Telnet functions to work properly, Telnet must be associated with a single stack. In other words, Telnet must have affinity to a particular stack. When Telnet runs as part of the TCP/IP stack, Telnet is automatically associated with and has affinity to that stack.

When Telnet runs as its own procedure, association depends on whether or not the TCPIPJOBNAME statement in TELNETGLOBALS is used. Even in an INET environment where only a single TCP/IP stack can run, TCPIPJOBNAME must be specified for Telnet to be associated with that stack for some functions. Telnet connections automatically associate with the active stack, but the functions listed below will not work if TCPIPJOBNAME is not explicitly specified. In a common INET (CINET) environment, Telnet is associated with all running stacks. If another

stack is started while Telnet is active, the current LISTEN for the port is cancelled and reissued automatically to include the new stack. If association with one stack is desired for control purposes or for functionality support, specify TCPIPJOBNAME.

The following functions are affected by whether or not TCPIPJOBNAME is specified:

- Telnet SNMP subagent activation requires specification of a stack name to register with the agent. Without TCPIPJOBNAME, Telnet blocks the subagent activation request.
- WLM registration requires specification of a stack name for successful registration. Without TCPIPJOBNAME, a profile DEBUG message is issued if WLM registration is attempted.
- Netstat displays might not show all Telnet connections if multiple stacks are supporting the Telnet server. Only those connections supported by the stack where the command is issued are shown.

Telnet SNMP subagent limitation: The Telnet SNMP subagent can only register with one agent, and each agent can support only one Telnet subagent. If running Telnet in its own address space, in addition to the required affinity, you must be careful to plan for one agent per Telnet subagent, including the Telnet subagent that might be running as part of the TCP/IP stack. If multiple Telnet SNMP subagents initialize to the same agent, the agent forwards all data requests to the first subagent that connected, and all other initializations are queued. If the first subagent ends, the next subagent in the queue then receives all data requests.

SMF address space name: When running Telnet as part of the TCP/IP stack, the SMF 119 address space name of the writer or the SMF 118 started task name is always the same name, the name of the TCP/IP stack. When running Telnet as its own procedure, the address space name or started task name is the name of the Telnet procedure.

IPv6 and INET environments: Telnet does not support changing either environment without recycling the Telnet procedure. Telnet checks for environment change at every port activation. If a change is detected, the port is not activated. Undesired results can occur on existing ports if a change between IPv4 and IPv6 or INET and CINET is attempted.

Port reservation: Reserving a Telnet port using INTCLIEN does not work when Telnet is running as its own procedure. The port is reserved for Telnet running as part of the TCP/IP stack. The stack rejects the BIND request from Telnet running as its own procedure.

Multilevel security compliance: Telnet running as its own procedure is not multilevel security compliant. Telnet must be run as part of the TCP/IP stack to be multilevel security compliant.

CTRACE set up: Telnet uses the TCP/IP component name, SYSTCPIP. Only a subset of trace options are available in the component when it is being set up for Telnet. For trace option details, see "Tracing" on page 528. The component trace options can be changed in the JCL by specifying a new parmlib member in the form CTIEZB*xx*. A sample parmlib member, CTIEZBTN, is in SYS1.PARMLIB. For additional CTRACE set up information, refer to the TCP/IP services traces and IPCS support chapter in *z/OS Communications Server: IP Diagnosis Guide*.

Resolver search order: The default resolver search order is probably the search order best suited for Telnet's needs. However, if there are known unique parameters for the resolver when called by Telnet, those unique parameters can be defined in a data set specified on the SYSTCPD DD statement in the Telnet procedure JCL. The use of this optional statement is not recommended unless there are clearly understood reasons for its use.

Setting up the Telnet procedure: When Telnet is started as its own procedure running in its own address space, set up for the job is necessary. Before Telnet can be started, security for the procedure name and the associated user ID must be defined. This discussion assumes RACF is the security subsystem being used. If another security product is used, refer to its manuals for equivalent set up instructions.

The procedure name must be added to the RACF STARTED class and have a user ID associated with it as follows:

```
RDEFINE STARTED TELNET*.*  STDATA(USER(TN3270))
SETROPTS RACLIST(STARTED) REFRESH
```

Use an existing superuser ID, or define a superuser ID to associate with the job name by adding a user ID to RACF and altering it to superuser status as follows:

```
ADDUSER TN3270
ALTUSER TN3270  OMVS(UID(0) PROGRAM ('/bin/sh') HOME('/'))
```

In this example the user ID name is TN3270, but any name can be used. These two RACF commands can be combined into one command by putting the OMVS parameter on the ADDUSER command line. The add and alter commands are done separately in case the user ID already exists. If the add fails, the alter still succeeds.

If setting up a superuser ID is not desirable, you can instead permit the user ID to the BPX.SUPERUSER class using the following steps:

1. Add the user to RACF:

   ```
   ADDUSER TN3270
   ```

2. Permit the user ID:

   a. Create a BPX.SUPERUSER FACILITY class profile:

      ```
      RDEFINE FACILITY BPX.SUPERUSER
      ```

   b. If this is the first class profile, activate the FACILITY class:

      ```
      SETROPTS CLASSACT(FACILITY)
      SETROPTS RACLIST(FACILITY)
      ```

   c. Permit the user to the class:

      ```
      ALTUSER TN3270  OMVS(UID(23) PROGRAM ('/bin/sh') HOME('/'))
      PERMIT BPX.SUPERUSER CLASS(FACILITY) ID(TN3270) ACCESS(READ)
      ```

      In this example, the user ID is TN3270 and the UID is 23. The UID can be any nonzero number. UID 23 was used to match the well-known Telnet port number.

   d. Refresh the FACILITY class:

      ```
      SETROPTS RACLIST(FACILITY) REFRESH
      ```

Tip: If you run the TN3270E server without a UID value of 0, the server is not able to increase the MAXFILEPROC value. This could result in the TN3270 server refusing connections.

If hardware encryption is to be used, be sure the server user ID has read access to the RACF CSFSERV class resources. For details, see "Encryption algorithms" on page 1171.

Sample statements for adding the procedure to the STARTED class and defining a superuser ID are in SEZAINST(EZARACF). For more detail about these functions, refer to *z/OS UNIX System Services Planning*, *z/OS Security Server RACF Security Administrator's Guide*, and *z/OS Security Server RACF Command Language Reference*.

The MVS default program properties table (PPT) has the Telnet module set up as privileged, non-swappable, non-cancelable, running in key 6, and system task. These settings give Telnet the same priority as the TCP/IP stack. Either privileged or system task cause the started job to be assigned to the SYSSTC service class. The priority can be changed by assigning the job name to another service class within the STC subsystem.

Tip: The default PPT entry sets the TN3270 server to non-cancelable. As a non-cancelable application, the TN3270 server should not be started automatically by a TCP/IP stack using the AUTOLOG function. If the TCP/IP stack is recycled, the following messages will be issued repeatedly:

```
N 0140000 SA6I     2005147 04:59:27.69 STC07087 00000084  EZZ0621I AUTOLOG FORCING IBMTNSI0, REASON: TCP/IP HAS BEEN RESTARTED
NR0000000 SA6I     2005147 04:59:27.71 STC07087 00000080  IEE838I IBMTNSI0         NON-CANCELABLE - ISSUE FORCE ARM
```

To prevent this, specify NOAUTOLOG on the PORT reservation statement as follows:

```
PORT      23 TCP TN3270D  NOAUTOLOG  ; TN3270 server
```

Sample JCL can be found in SEZAINST(EZBTNPRC). The only valid parameter that can be passed in from the JCL is the component trace options parmlib member name. Specify the profile data set name on the PROFILE DD entry in the JCL. The data set must be fixed and blocked with a record length between 56 and 256. The block size must be evenly divisible by the record length.

Customizing the VTAM configuration data sets

Telnet uses application LUs that are defined in VTAM application (APPL) major nodes to represent clients. Their definition members must be made available to VTAM by being in the list of data sets specified on the VTAMLST DD statement in the procedure used to start VTAM. This member contains the Telnet application LUs, and ensures that these LUs will be available for activation after VTAM is started. To automatically activate the application definition deck, include it in ATCCONxx. If multiple Telnet servers are used, for example multiple TCP/IP stacks in a Sysplex Distributor environment, ensure each server uses unique LU names. Otherwise, the second server that uses the same LU name will be unable to establish a session. Either the OPEN ACB request will fail, or the cross-domain session request will fail. A sample VTAM configuration data set can be found in SEZAINST(IVPLU).

Telnet LUs, representing either terminal or printer emulators, can be defined in the VTAM configuration data set using a wildcard character instead of coding individual Telnet application LU statements. The Model Application Names function allows system administrators to code a model APPL name with an asterisk (*) or question mark (?). Refer to *z/OS Communications Server: SNA Resource Definition Reference* for detailed information. Use * as a wildcard character to replace a character string at the same position anywhere within the minor node name. This can produce a significant administration savings. For example, assume Telnet LUs are needed in the range of TCPABC01 through TCPABC99. The sample

configuration data set has a single VTAM application definition statement with a Telnet application minor node (Telnet LU) name of TCPABC* which supports all 99 LUs.

Because Telnet server LUs do not support multiple concurrent sessions, VTAM will automatically set SESSLIM=YES for Telnet LUs defined to VTAM.

CLOSEACB can hang waiting on UNBIND RESPONSE if the target (PLU) application does not issue CLSDST when its LOSTERM exit is driven. To avoid this hang, code LOSTERM=IMMED on all target (PLU) applications that will have a SNA session with TELNET.

Code EAS=1 to minimize Common Service Area (CSA) storage use. If EAS is allowed to default, excessive CSA storage will be used.

Parallel sessions should not be used with Telnet LUs. The default VTAM value for non-APPC applications is PARSESS=NO. Do not code PARSESS=YES.

Customizing the TCP/IP configuration data sets

Telnet configuration statements are processed during initialization of the TCP/IP address space or when using the VARY TCPIP,,OBEYFILE command. The purpose of the Telnet configuration statements is to:

1. Define connection characteristics
2. Facilitate session setup with MVS host VTAM applications
3. Assign an LU name to represent the client

Telnet configuration uses the following statement blocks:

TELNETGLOBALS/ENDTELNETGLOBALS
An *optional* statement block containing Telnet parameter statements. The parameters define connection characteristics for all ports.

TELNETPARMS/ENDTELNETPARMS
A *required* statement block containing Telnet parameter statements. The parameters define connection characteristics for the specified port.

BEGINVTAM/ENDVTAM
A *required* statement block containing Telnet mapping statements. The mapping statements define how applications and LU names are mapped (assigned) to clients.

PARMSGROUP/ENDPARMSGROUP
An *optional* parameter group statement within BEGINVTAM containing Telnet parameter statements. The parameters define connection characteristics for the mapped clients.

Refer to *z/OS Communications Server: IP Configuration Reference* for exact syntax rules.

TELNETPARMS and BEGINVTAM blocks are required for each port started or modified by the VARY TCPIP,,OBEYFILE command. See "Complete profile replacement" on page 453 for details on complete profile processing. It is recommended, but not required, that you reserve the Telnet port or ports by using the stand-alone PORT *num* INTCLIEN statement in the TCP/IP startup profile. If you do not code the PORT *num* INTCLIEN statement, another application might use the port before the Telnet application can claim it.

Use a separate profile member with only Telnet statements (TELNETGLOBALS, TELNETPARMS and BEGINVTAM blocks) to keep TCP/IP configuration more organized and allow for easy Telnet updates with the VARY TCPIP,,OBEYFILE command. To validate a Telnet profile without applying the profile, specify TESTMODE (a TELNETPARMS-only parameter). When no errors are reported, remove the TESTMODE statement and use the INCLUDE statement to add it to the TCP/IP startup profile. A sample TCP/IP configuration data set can be found in SEZAINST(SAMPPROF). For the Telnet procedure, a sample configuration can be found in SEZAINST(TNPROF).

Telnet LUs can be defined to the Telnet configuration data set using several different wildcard notations instead of coding individual LU names. The DEFAULTLUS statement can be used to define terminal emulator LUs. The sample profile defines Telnet application LUs TCPABC01 through TCPABC99 to represent clients connecting to port 23 of the stack IP address. The last two fields are defined to be numeric variables. If multiple Telnet servers are used, for example multiple TCP/IP stacks in a Sysplex Distributor environment, ensure each server uses unique LU names. Otherwise, the second server that uses the same LU name will be unable to establish a session. Either the OPEN ACB request will fail, or the cross-domain session request will fail. After connecting, the DEFAULTAPPL statement maps the TSO application to the client and causes the Telnet server to immediately initiate a session between the Telnet LU and TSO. If an error occurs during session initiation, the MSG07 statement allows an error message to be sent to the client. If MSG07 is not coded, the connection is dropped. The sample profile contains additional statements that are included as comments. These statements provide examples of advanced functions. Many of these statements are installation-specific and will require modification.

Connection characteristic parameters: Telnet initially sets all connection parameters to default values. Parameters can then be changed at three levels. Each level provides a different scope of coverage. Parameters merge down as shown in Figure 54 on page 449. Parameters coded in TELNETGLOBALS will be applied to all connections on all ports, unless overridden by parameters in either TELNETPARMS or PARMSGROUP. Parameters coded in TELNETPARMS will be applied to all connections on the specified port, unless overridden by parameters in PARMSGROUP. Parameters coded in PARMSGROUP will be applied to connections whose clients are mapped to that PARMSGROUP. Telnet parameters are described throughout this chapter where appropriate. For a complete list, see *z/OS Communications Server: IP Configuration Reference*.

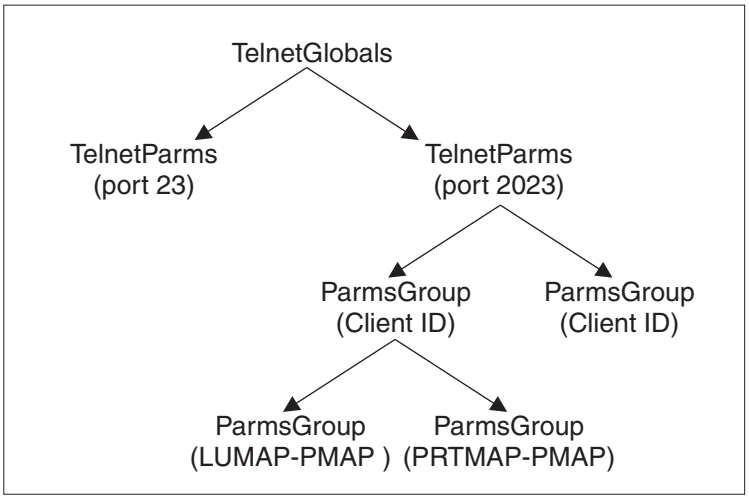


Figure 54. Telnet parameter placement

Session setup and LU assignment: There are 11 mapping statements available in the BEGINVTAM block, which map Objects to specified Client Identifiers within the port. Five are application setup related, four are LU name assignment related, one maps connection parameters, and one maps monitoring rules. Shortly after a connection request is accepted, the mapping statements are used by Telnet to map, or assign, as many of the 11 Objects to the client as possible. This set of Objects is used for the duration of the connection. For more detailed information, refer to "Mapping Objects to Client Identifiers" on page 467.

Managing the Telnet server

Commands

Many networking products (such as VTAM) use VARY commands to change the state of a device and DISPLAY commands to show information. The Telnet server also uses VARY and DISPLAY commands to change and monitor Telnet functions and debug problems.

Note that when a VARY TCPIP,,OBEYFILE command is issued, TELNETPARMS and BEGINVTAM blocks are both required for each port started or modified.

The following commands help manage the Telnet server. Syntax and examples can be found in *z/OS Communications Server: IP System Administrator's Commands*.

- Telnet VARY commands allow the operator to change the state of Telnet ports, enable or disable the use of certain Telnet LU names, and manage diagnostic tools. These commands include:
  - VARY TCPIP,,TELNET,QUIESCE a port to block any new connection requests but allow existing connections to continue activity.
  - VARY TCPIP,,TELNET,RESUME a port to end the QUIESCEd state and allow new connection requests.
  - VARY TCPIP,,TELNET,STOP a port to end connections on the port and close the port.
  - VARY TCPIP,,OBEYFILE, to start, restart, or change a port by updating the Telnet profile. The VARY TCPIP,,TELNET,STOP and VARY TCPIP,,OBEYFILE commands can be used to stop a Telnet port and then restart that port or a new port without stopping the TCP/IP stack.

- VARY TCPIP,,TELNET,ACT and VARY TCPIP,,TELNET,INACT LUs for use by the Telnet server. If an LU is already in use, the INACT command fails. Specify the name ALL to activate all inactive LUs with one command. These commands have no effect on the VTAM state of the LU.
- VARY TCPIP,,TELNET,DEBUG,OFF turns off all debug activity that might have been turned on to debug a problem.
- VARY TCPIP,,TELNET,ABENDTRAP sets an abend trap based on unique Telnet return codes set in Telnet modules.

WLM registration is affected by port changes. See "WorkLoad Manager for Telnet (WLM)" on page 529 for more information.

- Telnet DISPLAY commands are discussed in "Telnet diagnostics" on page 521.

Qualified ports

In some cases all clients need to use the same port number, but the Telnet parameters need to be differentiated by destination IP address or destination linkname. The destination IP address can be either an IPv4 or IPv6 IP address.

For example, two stacks are running with a Telnet server on each stack, and the two stacks are going to be merged into one stack. Currently, stack 1 has a home address of 1.1.1.1 and has Telnet running with a set of definitions for port 23. Stack 2 has a home address of 2.2.2.2 and has Telnet running with a different set of definitions for port 23. Before the stacks are merged into one, end users connect to either 1.1.1.1,port 23 or 2.2.2.2,port 23, depending on which Telnet services are desired. The sample definition statements are:

```
Stack 1

 TelnetParms
  Port 23
  Inactive 600      ; Drop after 10 minutes of no activity
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC01..TCPABC49 EndDefaultLus
  DefaultAppl TSO
 EndVTAM

Stack 2

 TelnetParms
  Port 23
  Inactive 0        ; Never drop
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC50..TCPABC99 EndDefaultLus
  DefaultAppl CICS
 EndVTAM
```

After the merge, both home addresses exist in a single stack. One way to keep the Telnet definitions separate would be to change the port number in one of the definition sets. For instance, the port 23 definitions associated with the old stack 2 could be changed to be port 2023. The end result is 1 TCPIP stack and 1 Telnet server with port 23 and port 2023, where port 23 has the definitions used in the old stack 1 and port 2023 has the definitions used in the old stack 2. The definitions are still separate. However, all the end users who were connecting to 2.2.2.2 port 23 now have to change their clients to port 2023. The sample definition statements would be changed to:

```
Merged Stack

 TelnetParms
  Port 23
  Inactive 600      ; Drop after 10 minutes of no activity
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC01..TCPABC49 EndDefaultLus
  DefaultAppl TSO
 EndVTAM

 TelnetParms
  Port 2023
  Inactive 0        ; Never drop
 EndTelnetParms

 BeginVTAM
  Port 2023
  DefaultLus TCPABC50..TCPABC99 EndDefaultLus
  DefaultAppl CICS
 EndVTAM
```

With port qualification, the system administrator can qualify the port number with the destination IP address or linkname to keep the Telnet services separate. In this case, the destination IP address is used. The qualified port allows the users of either old stack to connect without making any changes to their client. The sample definition statements would be changed to:

```
Merged Stack

 TelnetParms
  Port 23,1.1.1.1
  Inactive 600      ; Drop after 10 minutes of no activity
 EndTelnetParms

 BeginVTAM
  Port 23,1.1.1.1
  DefaultLus TCPABC01..TCPABC49 EndDefaultLus
  DefaultAppl TSO
 EndVTAM

 TelnetParms
  Port 23,2.2.2.2
  Inactive 0        ; Never drop
 EndTelnetParms

 BeginVTAM
  Port 23,2.2.2.2
  DefaultLus TCPABC50..TCPABC99 EndDefaultLus
  DefaultAppl CICS
 EndVTAM
```

You cannot QUIESCE, RESUME, or STOP a qualified portion of a port. If the port has several qualified port profiles, the VARY TCPIP,,QUIESCE, the VARY TCPIP,,RESUME, and the VARY TCPIP,,STOP commands affect all qualified port profiles associated with the port being quiesced, resumed, or stopped. In the example above, V TCPIP,,T,STOP,PORT=23 will stop port 23,1.1.1.1 and port 23,2.2.2.2. It is not possible to stop port 23,1.1.1.1 or port 23,2.2.2.2 individually. All display commands that allow port specification allow you to specify a qualified port. If just the port number is specified, only the unqualified port, if it exists, is

displayed. The qualified port profiles are not displayed. DBCSTRANSFORM can be active on only one port, but can be active on one, some, or all of the qualified profiles of that port.

Maximum connections supported

For each port, Telnet uses setrlimit() to automatically set the MaxFileProc value to the maximum allowed by UNIX System Services, currently 131 072. Each Telnet port will support that number of connections. Be sure that MaxSockets is large enough to support the anticipated number of sockets used by the system. If your system is IPv6 enabled, Telnet listening sockets are IPv6. Be sure to set IPv6 MaxSockets appropriately.

Tip: If you run the TN3270E server without a UID value of 0, the server is not able to increase the MAXFILEPROC value. This could result in the TN3270 server refusing connections.

Multiple ports

Telnet supports up to 255 ports on one TCP/IP stack. A unique TELNETPARMS block must be created for each port or qualified port. Telnet allows the use of the same BEGINVTAM block for all ports, some ports, or a unique BEGINVTAM block for each port. Both TELNETPARMS and BEGINVTAM blocks are required for each port started or modified by a VARY TCPIP,,OBEYFILE command. One or more PORT *num* TCP INTCLIEN reservation statements can be specified. The first PORT *num* TCP INTCLIEN specified will generate a default TELNETPARMS block and combine with a single BEGINVTAM block to start Telnet. It is recommended that every port be defined with explicit TELNETPARMS blocks to avoid confusion. There are several reasons more than one Telnet port or qualified port might be needed. The two most common reasons are discussed in the following sections.

Assigning a single application to a port simplifies the setup of clients on the workstation and the logon process. Workstation clients can be labeled with the associated application name and then be set up to connect to the appropriate port or qualified port. With a client per application on the workstation, the end user can select the needed client, connect, and be immediately in session with the application defined on the DEFAULTAPPL statement in BEGINVTAM. This implementation requires a unique BEGINVTAM block for each port due to the unique DEFAULTAPPL statements. The example below shows how to set up TSO, IMS, and CICS on ports 23, 2023, and 4023, respectively. The same LU names are used in each BEGINVTAM block. Telnet maintains a master LU "in-use" registry across all ports so that the same LU name will not be used by two different ports.

```
TELNETPARMS
      PORT  23
ENDTELNETPARMS
TELNETPARMS
      PORT  2023
ENDTELNETPARMS
TELNETPARMS
      PORT  4023
ENDTELNETPARMS

BEGINVTAM
      PORT  23
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   TSO
ENDVTAM
BEGINVTAM
      PORT  2023
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   IMS
```

```
ENDVTAM
BEGINVTAM
      PORT  4023
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   CICS
ENDVTAM
```

Assigning different security levels to different ports is an easy way to differentiate client security needs. External connections might require SSL security, while internal connections do not. Other than that difference, all other aspects of the Telnet profile can be the same. For example, external clients can connect to port 23 of a firewall that converts the request to the Telnet secure port 992. Internal clients would connect directly to the Telnet basic port 23. The statements below show how two ports allow implementation of different security levels. Note the same BEGINVTAM block is used for both ports, which can significantly reduce profile maintenance complexity. The PORT statement in BEGINVTAM links the BEGINVTAM block to the multiple TELNETPARMS blocks defined.

```
TELNETPARMS
      PORT  23
ENDTELNETPARMS
TELNETPARMS
      SECUREPORT  992
      KEYRING hfs /use/keyring/tcpcs.kdb
ENDTELNETPARMS
BEGINVTAM
      PORT  23 992
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      ALLOWAPPL   *
ENDVTAM
```

If a profile that contains a new port number is processed, it is treated as an additional port, and the VARY TCPIP,,OBEYFILE command request will succeed if all parameters for the new port are correctly specified. Existing, non-referenced ports remain active and unchanged. You can use the VARY TCPIP,,TELNET,STOP command to stop a port.

For multiport considerations with WLM, see "WorkLoad Manager for Telnet (WLM)" on page 529.

Complete profile replacement

When using the VARY TCPIP,,OBEYFILE command to update the Telnet configuration, new profile statements create a completely new configuration that is used by all connections accepted after the update. Existing connections continue to use the configuration in affect when those connections were accepted. For a successful configuration creation, both TELNETPARMS and BEGINVTAM blocks are required for each port started or modified. The new configuration create is *not* a cumulative update from the previous profile. If only one change is needed in the new profile, change the old profile or copy the profile to another data set member and make the change. After VARY TCPIP,,OBEYFILE command processing completes, the new profile is labeled the CURRent profile, and the replaced profile becomes profile 0001. If another update is done, the new update becomes the current profile and the replaced profile becomes profile 0002. If the profile update is for a subset of the active ports, the ports not being updated remain unchanged. Profile debug messages can be suppressed by coding DEBUG OFF or DEBUG SUMMARY in TELNETGLOBALS and placing it before all other Telnet statement blocks. The structural layout of the profiles and how connections are associated with profiles are shown in the following figure.

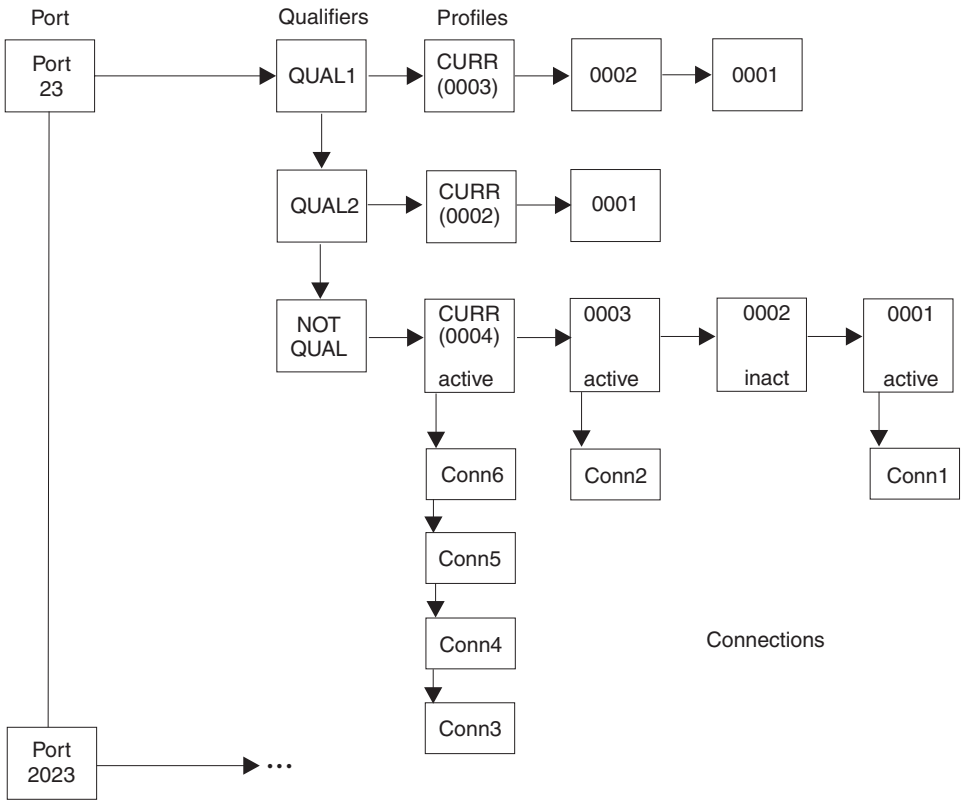


Figure 55. Telnet profiles and connections

Connection association

New connections are associated with the current profile and use the mappings and parameters defined by that profile. Even if a VARY TCPIP,,OBEYFILE command updates the port, existing connections remain associated with the same profile. The statements of non-current profiles remain in effect and continue to support all connections that were established when the non-current profile was the CURRent profile. When all connections associated with a non-current profile have ended, the storage for the non-current profile mapping rules is freed and the profile is considered INACTIVE.

Connection mode choices

The TN3270 Telnet server supports several connection types. The negotiation process is hierarchical in the order listed below:

- TN3270 Enhanced (TN3270E)
- TN3270
- Linemode
  - Standard
  - Binary
  - Transform

TN3270E is the default connection mode for the TN3270 Telnet server. If the client refuses TN3270E mode, the server tries TN3270 mode. If the client refuses TN3270 mode, the server then tries Linemode. The TN3270 Telnet Server does not support

Network Virtual Terminal (NVT) mode, except to allow the negotiation of TN3270E, TN3270, or linemode connections.

Note: The Type of Service (ToS) byte, also known as the Differentiated Services field, is not managed directly by the Telnet server. If you want to use Differentiated Services for Telnet, use the Quality of Service (QoS) support discussed in Chapter 15, "Quality of service (QoS)," on page 751.

TN3270E and TN3270 are very similar. If the TN3270E functions described in the following sections are not needed, the end user does not notice any difference between TN3270E and TN3270 connections. In some cases, older clients do not properly refuse the server request for a TN3270E connection, and the connection is dropped. In these unusual cases, use the NOTN3270E parameter to disable the TN3270E function for those clients. Similarly, use the NOSNAEXT parameter for any client that does not properly negotiate the extension functions (Contention Resolution and SNA Sense). TN3270E/NOTN3270E and SNAEXT/NOSNAEXT parameters can be coded at all three parameter block levels for different levels of granularity.

TN3270E and TN3270 clients can receive a Telnet solicitor panel to submit an application name, User ID, and password to the server. The cursor is positioned on the application line unless the OLDSOLICITOR parameter is specified which causes the cursor to be positioned on the user line. Refer to "Using the Telnet Solicitor or USS logon panel" on page 514 for detailed information.

The ATTN key function is supported over TN3270, TN3270E, and Transform Linemode connections. It is not supported over Standard or Binary Linemode. Default LOGMODEs for TN3270E connections are SNA, and default LOGMODEs for TN3270 and Transform connections are non-SNA. Telnet processes the ATTN key differently for SNA and non-SNA LOGMODEs. In addition, the Telnet server can be configured to handle double ATTNs sent by some clients by specifying SINGLEATTN. See "Device types and logmode considerations" on page 512 for more information.

For TN3270E, LU assignment is done during connection negotiation. For TN3270, LU assignment is done at application selection time. To delay LU assignment until application selection time for TN3270E, specify the SIMCLIENTLU parameter. Refer to "LU mapping by application name" on page 499 and "LU mapping selection rules" on page 501 for details.

You might experience unexpected results if you start a Telnet session from within an application that is already connected using Telnet. For example, if you start a new Telnet session from within a TSO session that was established on a TN3270E connection, the keyboard will unlock when it seems it should not. This happens when an unlock keyboard intended for only the original, first session is sent from Telnet. The second session should remain locked but does not. An unlock keyboard intended for only the first session has the affect of unlocking the keyboard for both the first and second session since both are represented by the same client.

Some host applications send 3270 read commands (for example, X'F2' read buffer) to the client during the course of a session. Telnet sends an unlock keyboard sequence (that is, X'F1C2') before the read command is sent to the client. This is the default behavior or can be specified by coding UNLOCKKEYBOARD BEFOREREAD. In some cases, a problem can arise if the keyboard is unlocked prior to the read command being forwarded to the client. The unlock keyboard sequence allows transmission of buffered keyboard data to the host application.

The buffered keyboard data is not expected in response to a read command. The UNLOCKKEYBOARD AFTERREAD parameter can be used to send the unlock keyboard sequence after the read command rather than before. In most cases, the default value will suffice and there is no need to code or change the setting of this parameter. Certain applications, however, will issue error messages when buffered keyboard data is unexpectedly received from the client. In these cases, UNLOCKKEYBOARD AFTERREAD can be coded to resolve the application error.

Some applications expect the end user to initiate session data traffic. If a USSMSG10 screen or solicitor panel was used to initiate the session, the keyboard is locked. A BIND flows to the client on a TN3270E connection alerting the client to unlock the keyboard. A non-TN3270E connection does not support sending a BIND to the client. Therefore, when a BIND is received from the application, Telnet sends an unlock keyboard to the client on a non-TN3270E connection to ensure the end user can initiate data traffic if necessary. This behavior is the default or can be specified by coding UNLOCKKEYBOARD TN3270BIND. In some cases, the unlock keyboard might not be correctly interpreted by an older client. If this is the case, UNLOCKKEYBOARD NOTN3270BIND can be coded to stop Telnet from sending an unlock keyboard when a BIND is received.

The two unlock keyboard functions must be specified on a single UNLOCKKEYBOARD statement. If only one is specified, it is assumed the other is the default value. The UNLOCKKEYBOARD parameter can be coded at all three parameter block levels for different levels of granularity.

TN3270 Enhanced (TN3270E)

TN3270E connections support full-screen 3270 emulation that is sometimes referred to as TN3270 Extended. Do not confuse TN3270E function with the IBM 327x device types that end in -E (for example, 3278-2-E). In these cases, the *E* indicates that the terminal supports Extended field attributes such as color and highlighting and is not related to Telnet functions.

Telnet is often used as the primary method of connection between client workstations and the SNA mainframe environment. To make this form of remote connection as seamless as possible, Telnet terminal emulation simulates actual SNA terminals as closely as possible. To accomplish this, RFC1647 and RFC2355 (both known as TN3270E) add the ability to specify device names at connection time, add support for printer devices, and add additional SNA functions. An Internet draft, RFC 2355 Extensions, adds Contention Resolution and SNA Sense code support.

Device name specification
The Telnet server assigns LUs based on the LU mapping statements supplied. Clients are assigned a device name (Telnet LU name) based on those statements. However, a TN3270E client can optionally specify that a particular device name be assigned, or it can specify that a device name from a pool of LUs be assigned. If the specified device name is allowed for this client based on the LU mapping statements and the LU is available, the server assigns the specified device name. If the specified device pool is allowed for this client based on the LU mapping statements and an LU within the pool is available, the server assigns a device name from the specified pool. Otherwise, the request is rejected with an appropriate reason code, and the connection is dropped. See “Mapping Objects to Client Identifiers” on page 467 for additional LU mapping information.

328x printer support
Many Telnet clients emulate 328x class printers (device type IBM-3287-1).

Most support both SNA Character Stream (SCS) as an LU1 and 3270 data stream as an LU3. The support of each is negotiated at connection time. When connected in TN3270E mode, the Telnet server supports these emulators in a manner similar to terminal LUs. Telnet can be configured to initiate a session at connection time or simply open an ACB to let the application initiate the session. The bind initiating each session is sent to the client, and the bind informs the emulator which data stream to expect. The VTAM application perceives the Telnet LU to be an an actual 3287-class printer and sends the SCS or 3270 data to the Telnet LU. Telnet passes the data on to the client, which prints the data. Telnet printer support allows you to use a single product, Telnet, to control both SNA terminals and SNA printers.

Some Telnet client printer emulators can request to be associated with a terminal device name by specifying the terminal device name during connection negotiation. Using printer association, end users can connect their Telnet terminals to an application and then have Telnet assign an associated printer device name based on the terminal name. To associate printers with terminals, Telnet must have a printer device pool of LUs defined and a terminal device pool of LUs defined with each having the same number of device names. For more information, see "Associated printer function" on page 496.

Additional negotiated TN3270E support

Responses and SysReq functions are supported by most clients that support TN3270E connections. Contention Resolution and SNA Sense support are newer and less prevalent.

- Responses - The client or host VTAM application can request that it receive definite, exception, or no response. Client responses to application requests provide more accurate response information for application-based monitoring tools compared to TN3270 connections. For TN3270 connections, the server must intercept response requests from the host and respond on behalf of the client, incorrectly reducing measured response time. Telnet monitoring will provide accurate response time information for either TN3270E or TN3270 connections.
- SysReq function - The end user can request that a current session be dropped by entering LOGOFF (in upper, lower, or mixed case) after pressing the SysReq key. If LUSESSIONPEND is not mapped to the client, the connection will be dropped. Otherwise, a USSMSG10 screen is sent to the client. If, instead of entering LOGOFF, the SysReq key is pressed a second time and if the application supports LUSTAT 082B (presentation screen is lost), the previous screen is resent to the client emulator.
- Contention Resolution - Improves communication between the client and host VTAM application regarding which owns the send state. Contention Resolution includes the following:
  - Keyboard Restore Indicator (KRI) - Whenever the host VTAM application sends End Bracket (EB), the server sends a KRI to the client. This directly notifies the client that the keyboard can be unlocked. Without the KRI indicator, Telnet would have to make sure a WCC with the unlock keyboard flag set is sent to the client. The KRI flag is set whether or not the keyboard restore flag in the WCC byte is set.

  - Start Data Indicator (SDI) - When the host VTAM application sends change direction or end bracket, the server sends the SDI to the client. This allows the client to know exactly when data can be sent to the server.
  - BID - A BID sent from the host VTAM application is forwarded to the client instead of being intercepted and handled by the server. This allows the client to manage the BID process itself.
  - Signal Indicator - A signal received from the host VTAM application is forwarded to the client. When the client responds to the signal, the server sends a change direction indicator to the host VTAM application.
- SNA Sense Support - Allows the client to include SNA sense codes in a response message. The client retains the option of letting the server map the errors to an appropriate sense code by not turning on the SNA-Sense-Code indicator in the response message.

TN3270

TN3270 connections support full-screen 3270 emulation. TN3270 connections do not support:

- Device name or pool name specification
- Printers
- Client involvement with responses, SysReq, Start Data Indicator, BID, Signal or SNA Sense data.

RFC1646 defines device name specification and printer support for TN3270 connections. However, this RFC is not supported on the TN3270 Telnet server. If either of these requests is received on a TN3270 connection, the server will drop the connection.

Linemode

In some cases, the client or the application does not support full-screen presentation, or the end user needs to work in a linemode environment. For these reasons, most emulators support linemode. Linemode supports a *go-ahead* function to simulate a half-duplex format. With *go-ahead* negotiated, the partner cannot send data until it receives a *go-ahead* from the current sender of data. In most cases, sessions are naturally half-duplex and the *go-ahead* adds unneeded transmissions. Therefore, the Telnet default is to Suppress Go Ahead (SGA). If *go-ahead* is needed to maintain a half-duplex format, use the NOSGA parameter. SGA or NOSGA can be coded at all three parameter block levels for different levels of granularity.

Telnet supports the following types of Linemode connections:

- Standard
- Binary
- Transform

Standard Linemode is assumed if neither DBCS transform nor BINARY linemode parameters are specified, or if the device type is not supported by transform. Standard Linemode is the only connection mode that requires translation by Telnet. Telnet supports NLS for standard Linemode connections. ASCII and EBCDIC code pages are the basis for translation. Telnet makes use of the National Language Support ICONV services available in the C runtime library. For custom code page information, refer to the ICONV services in *z/OS XL C/C++ Programming Guide*. When ASCII and EBCDIC code pages are specified, a conversion descriptor will be

given to Telnet. Telnet creates ASCII-EBCDIC and EBCDIC-ASCII translation tables based on the conversion descriptor. The CODEPAGE parameter is used to specify the code page names. For example:

```
CODEPAGE ISO8859-1 IBM1047
```

The possible results from CODEPAGE processing are:

- If a conversion descriptor is not returned, CODEPAGE is not coded, or there is an error in the syntax, a default code page of ISO8859-1 will be used for ASCII, and the language environment code page taken from locale information will be used as the EBCDIC code page.
- If a conversion descriptor is not returned again, a default code page of IBM-1047 will be used for EBCDIC.
- If a conversion descriptor is not returned again, predefined translation tables within Telnet will be used. These tables are similar, but not exactly the same as the tables which would have been generated if ISO8859-1 and IBM-1047 had worked. Some of the differences are noted below:

```
EBCDIC           ASCII
x'0D25' -----> x'0D0085' using ISO8859-1/IBM-1047
x'0D25' -----> x'0D0A'   using internal tables
x'15'   <----  x'0A'     using ISO8859-1/IBM-1047
x'25'   <----  x'0A'     using internal tables
```

No message is issued to the console if the first conversion succeeds. If there is any conversion failure a message is issued. If one of the later conversions succeeds, a message is issued indicating success.

If your Linemode connection does not perform correctly, the default translation tables may be causing the problem. Try the internal Telnet translation tables by specifying TNSTD for both ASCII and EBCDIC choices. For example:

```
CodePage TNSTD TNSTD
```

The internal code pages must be used together. If only one of the two internal tables is specified, then the other internal table will also be used.

CODEPAGE can be coded at all three parameter block levels for different levels of granularity.

Binary Linemode is set using the BINARYLINEMODE parameter. It indicates that Telnet should not do translation. The ASCII data from the client should be passed as-is to the VTAM application. BINARYLINEMODE or NOBINARYLINEMODE can be coded at all three parameter block levels for different levels of granularity.

Transform Linemode is set using the DBCSTRANSFORM parameter. When coded, all data that passes through Telnet will be *transformed* from DBCS or SBCS ASCII full screen to 3270 full screen for all supported device types. If the device type is not supported, Standard or Binary Linemode is used. DBCSTRANSFORM can be coded in TELNETPARMS or PARMSGROUP for different levels of granularity. It cannot be coded in TELNETGLOBALS. A unique logmode for transform can be set using TELNETDEVICE with a device type of TRANSFORM. Any logmode used must not support extended graphics.

Note: Transform can be used by only one port when multiple ports are active on one TCP/IP stack. DBCSTRANSFORM supports a maximum of 250 concurrent connections.

DBCSTRANSFORM can be used for either the VT100 single-byte character set (SBCS) or VT282 double-byte character set (DBCS) transform mode. When DBCSTRANSFORM is specified and the TCP/IP procedure JCL has been modified as shown below, ASCII-based terminal emulators (VT100 or VT282) will appear as full-screen 3270 terminals. The Telnet server receives ASCII data from the client and transforms it into SBCS or DBCS EBCDIC data, depending on the terminal type. Telnet adds appropriate SNA control bytes to give the appearance that the data is coming from a 3270 terminal. The Telnet server receives EBCDIC data from the host application and transforms the SNA control bytes and data into appropriate ASCII control bytes and data. The data is sent to the ASCII-based terminal where it is displayed in 3270 full-screen emulation. DBCSTRANSFORM requires additional special Data Definition (DD) statements in the TCP/IP procedure.

You must add the following three DD statements to the TCP/IP procedure JCL to support Transform:

```
//TNDBCSCN DD DSN=TCPIP.SEZAINST(TNDBCSCN),DISP=SHR
//TNDBCSXL DD DSN=TCPIP.SEZAXLD2,DISP=SHR
//TNDBCSER DD SYSOUT=*
```

- The TNDBCSCN DD statement must point to the configuration data set for 3270 DBCS transform mode. This configuration data set specifies the default DBCS conversion mode that will take effect at initialization time. Specify the CODEKIND and CHARMODE parameters according to the required DBCS code page. If CODEKIND and CHARMODE are not specified, or if the TNDBCSCN DD statement is not added, CODEKIND defaults to SJISKANJI and CHARMODE defaults to ALPHABET. A sample can be found in SEZAINST(TNDBCSCN).
- The TNDBCSXL DD statement must point to the data set containing binary translation table code files for 3270 DBCS transform mode. The installation data set, SEZAXLD2, contains the default binary translation table code files. The binary translation table code files for 3270 Transform can be customized by using the CONVXLAT command. Refer to *z/OS Communications Server: IP Configuration Reference* for more information about customizing translation table code files. If the TNDBCSXL DD statement is not added, the following message will appear and an abend will occur:

  ```
  IEC130I PASCAL01 DD STATEMENT MISSING
  ```

- The TNDBCSER DD statement defines where Transform-specific error messages are recorded. This DD statement can specify an output data set or SYSOUT=*. If the TNDBCSER DD statement is not added, transform initialization will fail.

Specifying the DBCSTRACE parameter sends detailed trace output from 3270 Transform to the location specified in the SYSPRINT output DD statement. Additional detailed trace output is also sent to TNDBCSER. Both data sets will contain detailed trace data. DBCSTRACE or NODBCSTRACE can be coded in TELNETPARMS or PARMSGROUP for different levels of granularity. They cannot be coded in TELNETGLOBALS.

Connection security

This section describes data overrun security, transport layer security (TLS), and Network Access Control.

Data overrun security

MAXRECEIVE: This parameter limits the number of bytes received from a client without an End Of Record (EOR) being received. If the data received exceeds the

limit, the connection is dropped. This parameter protects against a client stuck in a send-data loop. In general, large file transfers will not be affected because the sending client typically divides the file into smaller records that are sent. The receiving application rebuilds the file as the smaller records are received.

MAXVTAMSENDQ: This parameter limits the number of data segments (RPLs) queued to be sent to VTAM. If the queue size exceeds the limit, the connection is dropped. This parameter protects against using up large amounts of storage to hold data destined for a host application that is not receiving data.

MAXREQSESS: This parameter limits the number of session requests received by Telnet in a 10-second period. For this parameter, a BIND received by Telnet defines a session request. If the number of BINDs received in a 10-second period exceeds the limit, an error is reported. This parameter protects against session logon loops that are possibly created by an automatic CLSDST-PASS to an inactive session. This parameter cannot protect against logon loops caused by an inactive default application and a client using auto-reconnect.

MAXRUCHAIN: This parameter limits the number of chained RUs that can be received over a given session from a host application without a corresponding end chain RU. If the number of RUs exceeds the limit, the session is dropped. This parameter protects against using up large amounts of storage to hold data destined for a client in session with the host application.

The MAXRECEIVE, MAXVTAMSENDQ, MAXREQSESS, and MAXRUCHAIN parameters can be coded at all three parameter block levels for different levels of granularity.

Auto-reconnect loop: Without MSG07 coded a client connection error causes Telnet to drop the connection. The error may be an inactive DEFAULTAPPL or an LU assignment error. If the client has AUTO-RECONNECT specified, a continuous loop of retries occurs. The best protection against this is to code the MSG07 parameter which keeps the client from being disconnected. However, other applications can be chosen from the error screen returned to the end user. To block end users from other applications, use the DEFONLY parameter.

Transport layer security

TN3270 transport layer security overview: The TN3270 server provides the ability to secure Telnet connections with the transport layer security (TLS) or secure sockets layer (SSL) protocol. References to RACF apply to any other SAF-compliant security products which contain the required support. In this chapter, a port that is configured to use the TLS/SSL protocol is referred to as a secure port or SECUREPORT. A port that does not use the TLS/SSL protocol is referred to as a basic port. Connections are also either secure or basic. The flows between Telnet and VTAM are unchanged.

The Internet Engineering Task Force (IETF) TLS-based Telnet Security Draft is supported. This Draft allows a TN3270 negotiation to determine if the client wants or supports TLS/SSL prior to beginning the handshake. The default action that the TN3270 server will take for a secure port is to first attempt a TLS/SSL handshake. If the client does not start the handshake within the time specified by SSLTIMEOUT, an attempt will be made to negotiate TLS/SSL as defined by the TLS-based Telnet Security Draft. If the client responds that a secure connection is desired, the handshake is started; if the client rejects TLS/SSL, the connection will be closed. This allows installations to support both types of secure clients without

knowing which protocol the client is using. The default action can be changed by specifying the CONNTYPE parameter described later in this section.

Telnet server authentication and client authentication are described in Appendix B, "TLS/SSL security," on page 1167. The Telnet server supports level 1, level 2 and level 3 client authentication. Client authentication is done with the CLIENTAUTH parameter. Level 2 and level 3 client authentication use RACF services to translate the client certificate to an associated user ID. That user ID is used as a client identifier.

The Telnet server supports limiting and ordering the encryption algorithms. Use the ENCRYPTION parameter to define the number and order of encryption algorithms. The ENCRYPTION parameter can be coded in TELNETGLOBALS, TELNETPARMS or PARMSGROUP, providing a high level of granularity.

System SSL environment initialization occurs when the first secure port is activated; initialization is not done again unless all secure ports are stopped (V TCPIP,,T,STOP,PORT=SECURE). Whether or not hardware encryption is used is based on its availability at the time of Telnet initialization. For ICSF to be used by Telnet, ICSF must be available to System SSL when Telnet is started. The availability status can be changed only by completely stopping the Telnet procedure (or TCP/IP if Telnet is running as part of TCP/IP). A crypto assist status message is issued to sysout when the first Telnet secure port is activated. For more information on hardware encryption, see Appendix B, "TLS/SSL security," on page 1167.

Configuring the TN3270 server to support TLS/SSL connections: To implement secure connections, TCP must have APF authorized access to the System SSL DLLs. The System SSL DLLs are located in SYS1.SIEALNKE by default. System SSL uses the C runtime library (SCEERUN) and the C/C++ Standard Library, which must also be accessible to TCP. To access these libraries, either add them to the linklist or specify them in the TCP procedure's STEPLIB. If accessed through the linklist, the linklist must be authorized (LNKAUTH=LNKLST specified in the IEASYSxx parmlib member) or the libraries explicitly APF authorized. If accessed through a STEPLIB, the libraries must be APF authorized and DISP=SHR specified. The TCP/IP profile must also be updated. An overview of the SSL related profile parameters follows. For a detailed description of the parameters, refer to *z/OS Communications Server: IP Configuration Reference*.

The two essential parameters that must be specified are:

- SECUREPORT – All TLS/SSL enabled TN3270 ports must be defined by specifying a TELNETPARMS block for each port. The SECUREPORT port designation statement in the TELNETPARMS block indicates the port is capable of handling SSL connections.
- KEYRING – As mentioned in the overview section, a server certificate is required for the server authentication process defined by the SSL protocol. This certificate is stored in a key ring. The key ring type and location is specified in the KEYRING statement. Only one key ring can be used by the TN3270 server.

The key ring can be defined in the TELNETGLOBALS or TELNETPARMS block. TELNETGLOBALS is the preferred definition method because it ensures that the same key ring has been defined for all SECUREPORTs. If specified in TELNETPARMS, the same key ring type and file must be specified for each SECUREPORT. The first key ring file name read is considered the correct key ring file name. The TELNETGLOBALS key ring is read first and then the

TELNETPARMS key rings are read in reverse order. Any key ring that does not match the first is rejected and the port update fails.

The following steps are required to enable TLS/SSL support for Telnet, with server authentication.

1. Generate the Telnet server private key and server certificate.
2. Configure Telnet to include one or more TLS/SSL-enabled ports and specify the name of the key ring created in the step above in the TELNETGLOBALS block or the TELNETPARMS block. For example:
   - KEYRING HFS /usr/ssl/server.kdb (In this example two files, server.kdb and server.sth, were created using the gskkyman utility. The server's certificate is contained in the server.kdb file and designated as the default certificate.) The key database and the password stash file must reside in the same directory.
   - KEYRING MVS tcpip.mvs180.kdb (In this example 2 files, mvs180.kdb and mvs180.sth, were converted to MVS data sets from gskkyman files.)
   - KEYRING SAF serverkeyring (In this example, RACF is used to manage keys and certificates. The server certificate is connected to a key ring called SERVERKEYRING and designated as the default certificate.)
3. Restart TCP/IP or issue the VARY TCPIP,,OBEYFILE command with the updated configuration files.

Optional security parameters: Optional security parameters can only be specified for SECUREPORTs and can be specified in the TELNETGLOBALS, TELNETPARMS or PARMSGROUP blocks. The parameters specified in the PARMSGROUP block apply only to the clients mapped to the PARMSGROUP block by the PARMSMAP statement and override the parameters specified in the TELNETPARMS or TELNETGLOBALS block. The parameters specified in the TELNETPARMS block apply to any connection for that port if not overridden by a PARMSGROUP parameter. The parameters specified in the TELNETGLOBALS block apply to any connection for any port if not overridden by a TELNETPARMS or PARMSGROUP parameter.

The ENCRYPTION parameter is used to limit the encryption algorithms to only those included in the parameter statement. If this parameter is not specified, all encryption algorithms that can be specified will be used by Telnet. For the encryption algorithms that can be specified and the default order used by Telnet if the ENCRYPTION parameter is not used, refer to *z/OS Communications Server: IP Configuration Reference*. The following are some reasons to use this parameter:

- The applications supported on this port require a high level of security and the installation wants all data encrypted using a particular encryption algorithm
- Certain connections are local and the installation does not require encryption for local clients. NULL encryption can be specified for this subset of connections.

The SSLV2 parameter enables the use of the SSLV2 protocol. The default value is NOSSLV2, which prohibits the use of the SSLV2 protocol. The SSLV2 protocol might be necessary if SSLV3 or TLS is not supported by the client.

The CONNTYPE parameter sets the level of security for connections. If CONNTYPE is not specified, a SECUREPORT defaults to CONNTYPE SECURE and a basic port defaults to CONNTYPE BASIC.

Valid CONNTYPE options are:

- SECURE – Indicates that the TLS/SSL handshake will be used to start the connection. If the client does not start the handshake within the time specified by SSLTIMEOUT, an attempt will be made to do a negotiated TLS/SSL handshake (as defined by the IETF TLS-based Telnet Security Draft); if the client rejects TLS/SSL, the connection will be closed.
- NEGTSECURE – Indicates the client supports the IETF TLS-based Telnet Security Draft. A TN3270 negotiation with the client first determines if the client is willing to enter into a secure connection. If the client agrees, a TLS/SSL handshake is started and secure protocols will be used for all subsequent communication. If the client rejects TLS/SSL, the connection will be closed.

  If you know that the TN3270 secure clients connecting into the port are using the protocol defined by the TLS-based Telnet Security Draft, you should consider using this option. With this option the TLS/SSL handshake is not attempted until after a positive response to the TN3270 DO STARTTLS IAC is received. This avoids the timeout delay that can occur when a TLS/SSL handshake is immediately started (as done with CONNTYPE SECURE) but the client is expecting the protocol used by the TLS-based Telnet Security Draft.
- BASIC – Indicates that a basic (non-secure) connection will be used.
- ANY – Indicates that the connection can be either secure or basic. The TN3270 server will first try a standard TLS/SSL handshake. If the handshake times out, a negotiated TLS/SSL (see CONNTYPE NEGTSECURE) is attempted.
  - If the client is willing to enter into a secure connection, secure protocols will be used for all subsequent communication.
  - If the client is not willing to enter into a secure connection, a basic (non-secure) connection is used.
- NONE – Indicates that no connection is allowed and the connection will be closed. If this option is specified in TELNETPARMS, a PARMSMAP must cover every allowable connection and the related PARMSGROUP must specify the desired CONNTYPE.

The CLIENTAUTH parameter indicates that the client must send a client certificate to the server. If this parameter is not specified, a client certificate is not requested during the SSL handshake and no certificate based client authentication is done. The level of validation done depends on the option specified.

Valid CLIENTAUTH options are:

- SSLCERT (Level 1) – To pass authentication, the Certificate Authority (CA) that signed the client certificate must be considered trusted by the server (that is, a certificate for the CA that issued the client certificate is listed as trusted in the server's key ring).
- SAFCERT (Level 2 and 3) – The level 1 checking provided by SSLCERT is done and level 2 checking is done to verify that the certificate has been registered with RACF (or other SAF compliant security product that supports certificate registration). Additionally, if the SERVAUTH RACF class is active and a RACF resource has been defined for the port, level 3 client authentication is in effect and the connection is allowed only if the user ID associated with the client certificate has READ access to the RACF resource.
- NONE – No client certificate is requested.

The CRLLDAPSERVER parameter is specified in the TELNETGLOBALS block. It defines the name or IP address and port of the Certificate Revocation List (CRL) LDAP server. The CRL LDAP server is used only if client certificates are received (CLIENTAUTH specified). If CLIENTAUTH and the CRLLDAPSERVER have been

specified, the certificate revocation list is checked during client authentication. If the client's certificate is found on the certificate revocation list, the connection is closed. Only one CRL LDAP server can be defined to the Telnet server.

Changing the key ring (name, type, or contents) or the CRL LDAP server (name or location) cannot be done using the VARY TCPIP,,OBEYFILE command while secure ports are active. To change the key ring or the CRL LDAP server, all secure ports must first be stopped (V TCPIP,tcpname,T,STOP,PORT=S). The VARY TCPIP,,OBEYFILE command can then be used to restart the secure ports with a new key ring or CRL LDAP server. If the CRL LDAP server is stopped or connectivity is lost, System SSL may not recognize a subsequent reconnection. This situation must be handled like the CRL LDAP server change.

Using one port for both basic and SSL connections*:* SECUREPORT indicates that a port is capable of supporting SSL connections. The default connection type for SECUREPORT is SECURE. CONNTYPE can be used to modify connection types on a single port. Allowing a port to support both basic and secure connections assumes that either:

- The installation will allow the client to determine the connection type desired.
- A subset of the connections that should use a particular connection security type can be identified by Client Identifier.

In the first case, CONNTYPE ANY can be specified. If the port was defined as a SECUREPORT but the client wants a basic connection, there will be a slight delay before connection negotiation begins. This is because when CONNTYPE ANY is coded, the Telnet server will first attempt an SSL handshake to ensure the client is not requesting SSL support. It is only after the SSL handshake times out and negotiated security is rejected that the basic connection negotiation begins.

In the second case, the TELNETPARMS block should specify the default connection security type (see the CONNTYPE parameter). For connections with different connection security requirements:

- Identify the clients by Client Identifier.
- Create a PARMSGROUP with the alternate CONNTYPE definitions.
- Map the PARMSGROUP to the clients using the PARMSMAP statement.

Telnet profile example*:* The following example defines three ports with the characteristics discussed below.

- Port 23 allows only basic (non-secure) connections.
- Ports 992 and 1023 are enabled for secure connections and use the key ring defined in the TELNETGLOBALS block.
- Port 992 allows only secure connections. No client authentication is requested.
- Port 1023 allows both basic and secure connections. The installation desires the following characteristics for port 1023:
  - The system administrator is at IP address 9.37.88.1 and wants the capability to choose to connect with secure or non-secure connections.
  - Building A and B are local and do not need connection security. The clients in these buildings have identifiable host names. The installation wants only these clients to use basic connections to avoid the encryption overhead.
  - Connection security is desired on all other connections.
  - All secure connections require client authentication and will use the DES or triple DES encryption algorithms.

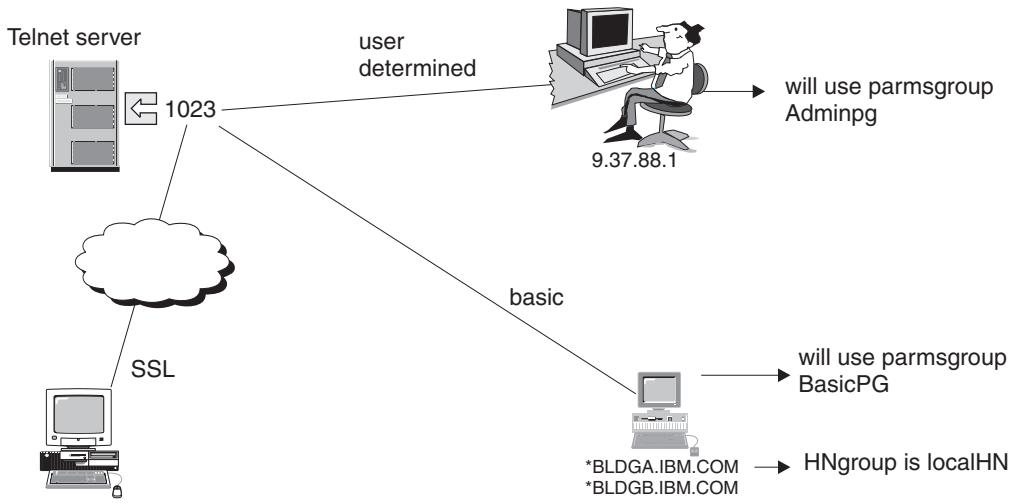


Figure 56. Port 1023 connection characteristics

Note: Only the definitions applicable to TLS/SSL connection security are shown; additional parameters might be needed.

```
TELNETGLOBALS
KEYRING hfs /usr/keyring/tcpcs.kdb ;keyring used by all SECUREPORTs
ENDTELNETGLOBALS


TELNETPARMS            ; basic port does not support secure connections
Port 23
ENDTELNETPARMS

TELNETPARMS            ; port that allows only secure connections
SECUREPORT 992         ; no client authentication requested
ENDTELNETPARMS         ; any supported encryption algorithm

TELNETPARMS            ; port that allows secure and BASIC connections.
SECUREPORT 1023        ; note: BEGINVTAM block has PARMSGROUP that may override CONNTYPE
CONNTYPE SECURE        ; SECURE is default
CLIENTAUTH SSLCERT     ; client certificate must be issued by a trusted CA
ENCRYPT SSL_DES_SHA SSL_3DES_SHA ENDENCRYPT ; only encrypt with DES or
ENDTELNETPARMS                              ;triple DES



BEGINVTAM
Port 1023
...                    ;Mapping statements
HNGROUP localHN
   *.BLDGA.IBM.COM
   *.BLDGB.IBM.COM
ENDHNGROUP

PARMSGROUP BasicPG     ; override telnetparms definitions
   CONNTYPE BASIC      ; support non-secure connections mapped to this group
ENDPARMSGROUP
PARMSGROUP AdminPG
   CONNTYPE ANY        ; connections mapped to this group allow any type of connection
ENDPARMSGROUP

PARMSMAP AdminPG 9.37.88.1 ; this ip address can use secure or non-secure connections
PARMSMAP BasicPG localHN   ; hosts defined in HNGROUP localHN,
                           ;will use non -SSL connections as defined in PARMSGROUP BasicPG
ENDVTAM


BEGINVTAM
Port 992 23
```

```
...                         ;Mapping statements
                            ;no PARMSGROUP defined for these ports
                            ;TELNETPARMS definitions used for all connections
ENDVTAM
```

Network Access Control

Network Access Control (NAC) limits user access to certain IP security zones defined by the NETACCESS statement. A security product, such as RACF, is used to check the permission of user IDs to send data to or receive data from these security zones. The NAC user ID is based on the application's address space information. TCP/IP is an exempt user and has access to all zones. Because Telnet runs in the TCP/IP address space, Telnet processes are also exempt from Network Access Control based on address space user ID information.

The NACUSERID parameter enables Network Access Control checking for the TN3270 server. This parameter is used to associate TN3270 server ports with a user ID defined to the security server. The user ID specified on the NACUSERID parameter must be a valid user ID defined to the security server. If not, the TN3270 port will fail initialization. NONACUSERID can be specified to disable Network Access Control. NACUSERID can be coded in TELNETGLOBALS to affect all ports or TELNETPARMS to affect a single port. NACUSERID cannot be coded in PARMSGROUP.

When Telnet is modified with a VARY TCPIP,,OBEYFILE command, the NACUSERIDs are reverified for the TN3270 ports defined in the data set referenced by the command. If a TN3270 port has NACUSERID *NAC_name_1*, you cannot use the VARY TCPIP,,OBEYFILE command to change that port's NACUSERID to *NAC_name_2*. The port must first be stopped, and then started with the new *NAC_name_2* using the VARY TCPIP,,OBEYFILE command.

On a restricted TCP/IP stack (not SYSMULTI), the NACUSERID will run under the SECLABEL of the TCP/IP stack. On an unrestricted TCP/IP stack (SYSMULTI), the NACUSERID will run under the default SECLABEL defined in the user profile. If no default SECLABEL is configured in the user profile, SYSLOW is used by default. The NACUSERID user profile must be permitted to the SECLABEL it is to run under or port activation will fail.

The NETACCESS statement in the TCP/IP profile is used to configure portions of your IP network into named security zones. Each defined security zone must have a SERVAUTH profile for the resource named EZB.NETACCESS.*sysname.tcpname.zonename*. The user ID associated with the TN3270 port must have READ access to the security zone that maps its bind address (0.0.0.0/32 for INADDR_ANY or ::/128 for IN6ADDR_ANY unless overridden by the PORT statement in the TCP/IP profile) and to every security zone that maps client IP addresses that TN3270 is to accept connections from on this port.

For more information, see "Network access control" on page 123.

Mapping Objects to Client Identifiers

The TN3270 Telnet server provides flexibility for mapping Objects to clients based on Client Identifiers. This section provides definitions, rules, and examples of many mapping methods. Examples start with simple concepts, then progress to more complicated concepts showing interaction between mapping statements. All mapping statements are specified in the BEGINVTAM block. Refer to *z/OS Communications Server: IP Configuration Reference* for statement rules not discussed here.

The general relationship of mapping statements is:

MAP OBJECTS to clients based on **CLIENT IDENTIFIER**

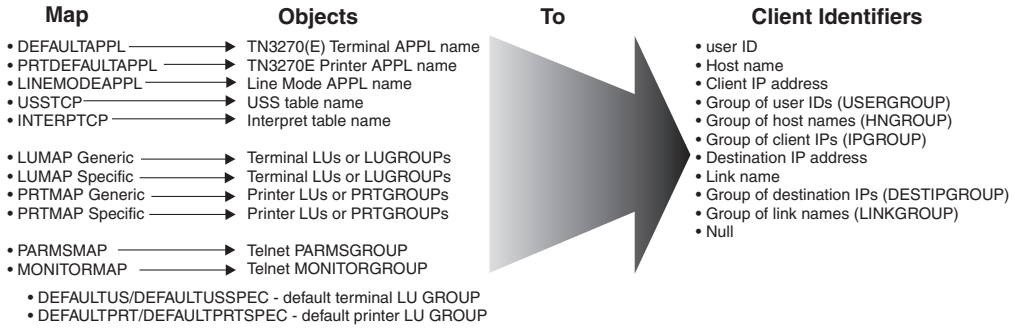


Figure 57. Mapping model

Telnet tries to assign all 11 Objects to a client based on the mapping statements when the connection is accepted. The search for Objects continues until all Objects are found or until all mapping statements are checked.

Objects

When a client connection request is made, Telnet must assign an LU name to represent the client. Optionally, a USS table, default application, unique parameters defined in the PARMSGROUP statement, or network monitoring can be assigned to the connection. See "Mapping methods" on page 478 for details about how these objects are mapped to clients. The complete list of object follows:

- TN3270(E) terminal application name – The DEFAULTAPPL mapping statement maps the TN3270(E) terminal application Object to a terminal client. When a TN3270 or TN3270E connection is negotiated, Telnet immediately initiates a session request to the VTAM application.
- TN3270E printer application name – The PRTDEFAULTAPPL mapping statement maps the TN3270E printer application Object to a printer client. When a TN3270E printer connection is negotiated, Telnet immediately initiates a session request to the VTAM application.
- Line Mode application name – The LINEMODEAPPL mapping statement maps the linemode application Object to a client. When a linemode connection is negotiated, Telnet will immediately initiate a session request to the VTAM application.
- USS table name – The USSTCP mapping statement maps the USS table Object to a client. When a TN3270 or a TN3270E connection is negotiated, Telnet will send a USSMSG10 screen to the client. A special case condition exists when an application name and a USS table are both mapped to the client by the exact same Client Identifier. In this case, Telnet will immediately initiate a session request to the VTAM application and use the USS table for error messages.
- Interpret table name – The INTERPTCP mapping statement maps the Interpret table Object to a client. When a TN3270 or a TN3270E connection is negotiated, Telnet uses the Interpret table to modify USS commands. The client must have a USS table mapped to it for the Interpret table to be used.
- Terminal LUs or LUGROUPs (Generic) – The Generic LUMAP mapping statement maps a single LU or LUGROUP Object to a client. For single LU mappings, Telnet will assign the LU name to the connection if the LU is available. For LUGROUP mappings, Telnet will assign an available LU from the group to the connection. An LU is required to represent the client when

initiating a VTAM session. DEFAULTLUS is a default terminal LUGROUP Object mapped Generically to the NULL Client Identifier.

- Terminal LUs or LUGROUPs (Specific) – The Specific LUMAP mapping statement maps a single LU or LUGROUP Object to a client. Unlike the Generic mapping in which Telnet assigns the LU, the Specific mapping requires the client to specify the LU name it wants. Telnet verifies the LU is mapped and available. The specified LU name can be either a mapped single LU, an LU within a mapped LUGROUP, or the LUGROUP name of the mapped LUGROUP. If the client specifies an LUGROUP name, Telnet assigns an available LU from within the group. DEFAULTLUSSPEC is a default terminal LUGROUP Object mapped Specifically to the NULL Client Identifier.
- Printer LUs or PRTGROUPs (Generic) – The Generic PRTMAP mapping statement maps a single LU or PRTGROUP Object to a client. For single LU mappings, Telnet will assign the LU name to the connection if the LU is available. For PRTGROUP mappings, Telnet will assign an available LU from the group to the connection. An LU is required to represent the client when initiating a VTAM session. DEFAULTPRT is a default printer PRTGROUP Object mapped Generically to the NULL Client Identifier.
- Printer LUs or PRTGROUPs (Specific) – The Specific PRTMAP mapping statement maps a single LU or PRTGROUP Object to a client. Unlike the Generic mapping in which Telnet assigns the LU, the Specific mapping requires the client to specify the LU name it wants. Telnet verifies the LU is mapped and available. The specified LU name can be either a mapped single LU, an LU within a mapped PRTGROUP, or the PRTGROUP name of the mapped PRTGROUP. If the client specifies a PRTGROUP name, Telnet assigns an available LU from within the group. DEFAULTPRTSPEC is a default printer PRTGROUP Object mapped Specifically to the NULL Client Identifier.
- Telnet PARMSGROUP – The PARMSMAP mapping statement maps the PARMSGROUP Object to a client. The parameters in the group override parameter values specified in either TELNETGLOBALS or TELNETPARMS.
- Telnet MONITORGROUP – The MONITORMAP mapping statement maps the MONITORGROUP Object to a client. The parameters in the group define what monitoring measurements will be done for a mapped client. For details, see "Connection monitoring mapping statement" on page 487.

The two following statements are not Objects but can affect application and LU Object usage:

- ALLOWAPPL – This statement allows client access to applications and optionally maps or confirms the mapping of an LU name to the client based on the application name chosen. DEFAULTAPPL application names are presumed allowed and do not require the ALLOWAPPL statement for Telnet acceptance. However, ALLOWAPPL may be used by default applications for LU assignment and other advanced functions.
- RESTRICTAPPL – This statement restricts Telnet acceptance of application names to only users that specify an acceptable User ID and password. It also optionally maps or confirms the mapping of an LU name to the client based on the application name and User ID chosen.

Client Identifiers

One client can be represented by many different Client Identifiers. For example, Telnet might assign an LU based on client host name, assign an application based on a client IP address, and assign a USS table based on connection link name. Refer to "Mapping methods" on page 478 for details about how these Client Identifiers are used to map Objects. In some cases, two different Client Identifiers

that represent the same client are used on mapping statements to map the same type of Object. In these cases, Telnet must determine which Client Identifier to use when assigning the Object. See "Client Identifier selection rules" on page 473 for more details. The complete list of Client Identifiers and mapping examples follow:

- User ID or USERGROUP name - If the CLIENTAUTH SAFCERT parameter is used with a secure connection, the client is required to send its client certificate to the Telnet server for client authentication. The SAFCERT option indicates that the client certificate can be translated to a User ID by a security product such as RACF. Telnet translates the certificate as soon as the SSL handshake is done. The resulting User ID is associated with the connection. Objects can be mapped to the connection based on an exact User ID, or Objects can be mapped to a USERGROUP name containing exact User IDs and wildcarded User IDs. For example, mobile employees need to be assigned a unique set of LU names and the manager must always be assigned LU name LUMOBL01. These employees are not within a secure network and always use client authenticated secure connections. Their certificates are translated to User IDs by Telnet.

  ```
  USERGROUP   USGMOBL1
     MOBL0002 MOBL0003
     MOBL1%%C
  ENDUSERGROUP
  LUGROUP LUGMOBL1
     LUMOBL02..LUMOBL20
  ENDLUGROUP
  LUMAP  LUMOBL01  USERID,MOBL0001   ; mgr mapping
  LUMAP  LUGMOBL1  USERGRP,USGMOBL1  ; employee mapping
  ```

  Rule: The specification of the Client Identifier type USERID is required on the mapping statement. If you do not specify this type, Telnet assumes that the name is a link name.

  Tip: The specification of the Client Identifier type USERGRP is optional. The following statement is equivalent to the last LUMAP statement in the previous example:

  ```
  LUMAP  LUGMOBL1  USGMOBL1
  ```

- Host name or HNGROUP name - If the network dynamically assigns IP addresses, the same client will not have the same IP address from one connection to the next. However, if Dynamic Domain Name System (DDNS) and Dynamic Host Configuration Protocol (DHCP) are used, the client host name can be constant. See Chapter 13, "Domain Name System (DNS)," on page 595 for more DDNS and DHCP information. With static host names, Objects can be mapped to clients based on their host name, or Objects can be mapped to HNGROUP names containing exact host names and wildcarded host names. For example, LUADMNM is mapped to exact host name ADMIN.DEPT1.GROUP1.COM, and application INVENTRY is mapped to HNGROUP name HNGINV.

  ```
  HNGROUP HNGINV
     INV1.DEPT1.GROUP1.COM
     *.DEPT3.GROUP1.COM
     **.GROUP3.COM
  ENDHNGROUP
  LUMAP LUADMNM  HOSTNAME,ADMIN.DEPT1.GROUP1.COM
  DEFAULTAPPL    INVENTRY  HNGRP,HNGINV
  ```

  Tip: The specification of the Client Identifier types HOSTNAME and HNGRP is optional. The following two mapping statements are equivalent to the last two statements in the previous example:

  ```
  LUMAP LUADMNM  ADMIN.DEPT1.GROUP1.COM
  DEFAULTAPPL    INVENTRY  HNGINV
  ```

Rule: Host name specification requires that Telnet resolve a host name from an IP address by using the resolver. To do this, you must specify GLOBALTCPIPDATA, DEFAULTTCPIPDATA, or a valid TCPIP.DATA data set. See Chapter 2, "Configuration overview," on page 11 for a description of how TCPIP.DATA is located.

Tip: Telnet uses the native MVS sockets search order to find a resolver. The Resolver_Config environment variable and the /etc/resolv.conf file are not used when searching for TCPIP.DATA.

Result: The most common reason for message `EZZ6011I, INIT_API failure` is that no resolver has been defined.

- Client (source) IP address or IPGROUP name - Client IP address is the most common method used to map Objects to the client. In a static network, Objects can be mapped to clients based on the exact IP address, or Objects can be mapped to IPGROUP names containing exact IP addresses and subnets. For example, LUADMN is mapped to exact IP address 1.1.1.1, and application PAYROLL is mapped to IPGROUP name IPGPAY.

```
IPGROUP IPGPAY
   1.1.2.2  1.1.2.3                 ;IPv4 addresses
   255.255.0.0:2.2.0.0              ;IPv4 subnet
   2001:0DB8:9:11:15:4              ;IPv6 address
   6C11:10::0/96                    ;IPv6 subnet
   6.1.3.4..6.1.3.8                 ;IPv4 range
   2AB0::12:5:1321..2AB0::12:5:1410  ;IPv6 range
ENDIPGROUP
LUMAP LUADMN IPADDR,1.1.1.1
DEFAULTAPPL  PAYROLL IPGRP,IPGPAY
```

  Tips:

  - The specification of the Client Identifier types IPADDR and IPGRP is optional. The following two mapping statements are equivalent to the last two statements in the previous example:

```
LUMAP LUADMN 1.1.1.1
DEFAULTAPPL  PAYROLL IPGPAY
```

  - The IP/subnet combinations of 0.0.0.0:0.0.0.0 (IPv4 only) and 0::0/0 (IPv4 and IPv6) are special cases that include all connections. This might be useful if you want to have a default mapping with a higher priority than the NULL client identifier.
  - The client IP address can be either an IPv4 or IPv6 IP address. IP address ranges can also be specified and are treated as if individual IP addresses were coded. An IPv4 range can vary in the last octet only. An IPv6 range can vary in the last two hexadecimal bytes only.

- Destination IP address or DESTIPGROUP name - A destination IP address is the host address that is the destination for a Telnet connection. Linkname can be used as a Client Identifier to map Objects to destination IP addresses when the linkname is static and defined in the profile. However, if the destination IP address is a dynamic virtual IP address (VIPA) , the linkname is not known before the VIPA is created. In this case, destination IP address is the ideal solution. In other cases, specifying the destination IP address in the Telnet profile might be more clear than specifying the linkname. For example, two TCP/IP stacks are backups for each other. Telnet connections to stack 1 (VIPA 5.5.5.1) default to logon manager application APPL1 and connections to stack 2 (VIPA 51CB:C3E4::9:4) default to logon manager application APPL2. If one of the stacks becomes unavailable, the other will take over and dynamically add the failing stack's VIPA. The dynamic linkname created is not easily predicted. Use the following statements in the profile of each stack to ensure users connecting

to 5.5.5.1 always get APPL1 and users connecting to 51CB:C3E4::9:4 always get APPL2 regardless of which stack is used.

```
DEFAULTAPPL  APPL1 DESTIP,5.5.5.1
DEFAULTAPPL  APPL2 DESTIP,51CB:C3E4::9:4
```

Rule: The specification of the Client Identifier type DESTIP is required on the mapping statement. If you do not specify this type, Telnet assumes that the IP addresses are client (source) IP addresses.

Tip: When the destination IP address is the IP address of a dynamic XCF address, multiple linkname values can be associated with the IP address. Telnet will use the first linkname associated with the IP address in the home list. If a dynamic XCF destination is used as a Client Identifier, it is recommended that DESTIP be used instead of linkname. Results can vary using linkname.

- Linkname or LINKGROUP name - A linkname is defined by the TCP/IP LINK statement. The linkname defines a host IP address that is a destination address for clients connecting to Telnet. Linkname can be useful in cases where Object assignment is dependent on the client destination IP address instead of the client source IP address. Several linknames may be defined and the same LU mapping or other Object mapping may be desired for several linknames. In this case, a LINKGROUP can be defined and used on a single mapping statement. For example, based on the statements below, a client connecting to LINK1 IP address will be assigned an LU from the LUGROUP name LUGLNKS and will establish a session with TPX1. A client connecting to LINK2 IP address will be assigned an LU from the LUGROUP name LUGLNKS and will establish a session with TPX2. Because LINK1 and LINK2 are not group names, host names, or IP addresses, they are assumed to be linknames. The Client Identifier type, LINKNAME, can be used for clarity but is not required.

  ```
  LINKGROUP    LNKGRP1
     LINK1 LINK2
  ENDLINKGROUP
  LUMAP        LUGLNKS  LINKGRP,LNKGRP1
  DEFAULTAPPL  TPX1     LINKNAME,LINK1
  DEFAULTAPPL  TPX2     LINKNAME,LINK2
  ```

  Tips:

  – The specification of the Client Identifier types LINKNAME and LNKGRP is optional. The following three mapping statements are equivalent to the last three statements in the previous example:

  ```
  LUMAP        LUGLNKS  LNKGRP1
  DEFAULTAPPL  TPX1     LINK1
  DEFAULTAPPL  TPX2     LINK2
  ```

  – When the destination IP address is the IP address of a dynamic XCF address, multiple linkname values can be associated with the IP address. Telnet will use the first linkname associated with the IP address in the home list. If a dynamic XCF destination is used as a Client Identifier, it is recommended that DESTIP be used instead of linkname. Results can vary using linkname.

- NULL (no Client Identifier) - The NULL Client Identifier type indicates that no Client Identifier was specified. The NULL Client Identifier is valid on the DEFAULTAPPL, LINEMODEAPPL, USSTCP, and INTERPTCP mapping statements. It is the implied Client Identifier for the DEFAULTLUS, DEFAULTLUSSPEC, DEFAULTPRT, and DEFAULTPRTSPEC Objects. ParmsGroup and MonitorGroup are the only Objects that cannot be mapped to the NULL Client Identifier. The NULL Client Identifier mapped Objects are the last Objects checked when assigning Objects to a client. For example, assume a client does not match any Client Identifier in the profile for DEFAULTAPPL or USSTCP. You can put the end user into session with a security application,

named SecAppl, that can verify the end user is authorized to use the company's system. The Client Identifier field is blank.

```
DEFAULTAPPL SECAPPL
```

Client Identifier selection rules

When Client Identifiers are used together, conflicts might occur. For example, host name NAME1.HOST1.COM may also be IP address 1.2.3.4. If the following DEFAULTAPPL statements exist, only one of the applications can be chosen.

```
DEFAULTAPPL  TSO  NAME1.HOST1.COM
DEFAULTAPPL  CICS 1.2.3.4
```

If USSTCP and DEFAULTAPPL have the same Client Identifier, DEFAULTAPPL will be used. For detailed information, refer to "Resolving DEFAULTAPPL and USS table conflicts" on page 484.

Telnet uses a very specific Client Identifier hierarchy when assigning Objects. The following order is used:

The mapping rule search order

- **Exact client identifier:**
  - 1) User ID, 2) hostname, 3) IP address
- **Exact client identifier in a group definition:**
  - 4) User group, 5) hostname group, 6) IP address group
- **Wildcard match for client identifier in a group definition:**
  - 7) User group, 8) hostname group, 9) IP address group
- **Exact destination:**
  - 10) destination IP address, 11) link name
- **Exact destination in a group definition:**
  - 12) destination IP address group, 13) link name group
- **Wild card match for destination in a group definition:**
  - 14) destination IP address group, 15) link name group
- **Null client ID**
  - 16) DEFAULTAPPL, LINEMODEAPPL, USSTCP, INTERPTCP, DEFAULTLUS, DEFAULTLUSSPEC, DEFAULTPRT, DEFAULTPRTSPEC

Examples:

- **Exact client identifier:**

```
1)  LUMAP   LU1     USERID,USER1
2)  LUMAP   LU2     NAME1.HOST1.COM
3)  LUMAP   LU3     1.2.3.4
```

  Client Identifier type USERID is required. If not specified, USER1 is assumed to be a link name.

- **Exact client identifier in a group definition:**

```
LUGROUP   LUGRP1    LU100..LU199   ENDLUGROUP
LUGROUP   LUGRP2    LU200..LU299   ENDLUGROUP
LUGROUP   LUGRP3    LU300..LU399   ENDLUGROUP

    USERGROUP USRGRP1
      USER1 USER2 USER3
    ENDUSERGROUP

    HNGROUP  HNGRP1
         NAME2.HOST1.COM  NAME2.HOST3.COM
    ENDHNGROUP
```

```
    IPGROUP  IPGRP1
      1.2.3.5     1.2.3.6
      1.3.4.7..1.3.4.E
    ENDIPGROUP
4)  LUMAP   LUGRP1   USRGRP1
5)  LUMAP   LUGRP2   HNGRP1
6)  LUMAP   LUGRP3   IPGRP1
```

- **Wild card match for client identifier in a group definition:**

```
    USERGROUP USRGRP2
      USER%%  TCPU*
    ENDUSERGROUP
    HNGROUP  HNGRP2
         *.HOST2.COM  **.HOST3.COM
    ENDHNGROUP
    IPGROUP  IPGRP2
      255.255.0.0:2.3.0.0
      2001:0DB8:3:274C::0/80
    ENDIPGROUP
7)  LUMAP   LUGRP1   USRGRP2
8)  LUMAP   LUGRP2   HNGRP2
9)  LUMAP   LUGRP3   IPGRP2
```

- **Exact destination:**

```
10) DEFAULTAPPL    TSO        DESTIP,1.2.3.4
11) USSTCP         USSTAB1    LINK1
```

Client Identifier type DESTIP is required. If not specified, destination IP address 1.2.3.4 is assumed to be a client IP address.

- **Exact destination in a group definition:**

```
    DESTIPGROUP DSTIPGRP1
      1.2.3.5  1.2.3.6
      79DA:10:3.4.9.5
      613D:10::9241..613D:10::C510
    ENDDESTIPGROUP
    LINKGROUP LINKGRP1
      LINK1 LINK2 LINK3
    ENDLINKGROUP
12) LUMAP LUGRP1 DSTIPGRP1
13) LUMAP LUGRP2 LNKGRP1
```

- **Wild card match for destination in a group definition:**

```
    DESTIPGROUP DSTIPGRP2
      255.255.0.0:1.4.0.0
    ENDDESTIPGROUP
    LINKGROUP LINKGRP2
      LINK* %LINK
    ENDLINKGROUP
14) LUMAP LUGRP1 DSTIPGRP2
15) LUMAP LUGRP2 LNKGRP2
```

- **Null client ID**

```
16) DEFAULTAPPL     TSO
    LINEMODEAPPL    CICS
    USSTCP          USSTAB1
    INTERPTCP       INTTAB1
    DEFAULTLUS
      LU01..LU99
    ENDDEFAULTLUS
```

NULL is a single Client Identifier. The order of the examples has no significance. If DEFAULTAPPL and USSTCP mapping statements both have the NULL Client

Identifier, the DEFAULTAPPL will be used regardless of order. For more information, refer to “Resolving DEFAULTAPPL and USS table conflicts” on page 484.

Object assignment examples

A client can be known by several different Client Identifiers. These Client Identifiers are used to assign as many Objects as possible to the connection based on the profile mapping statements. Telnet starts with the highest priority Client Identifier of the client and assigns all Objects mapped by that Client Identifier. If all 11 Objects are not assigned, Telnet uses the next highest priority Client Identifier (for prioritization details, see “Client Identifier selection rules” on page 473) and assigns all Objects mapped by that Client Identifier. This Object assignment process continues by using lower and lower priority Client Identifiers until all 11 Object types are found or until all of the matching Client Identifier mappings have been checked. If an Object is mapped by multiple Client Identifiers, only the Object mapped by the highest Client Identifier is used. It is unlikely all Objects are assigned to connections because not all Objects are always mapped. For example, many profiles do not contain PRTDEFAULTAPPL or INTERPTCP mapping statements. In this case, the printer default appl and Interpret table Objects will not be assigned.

Figure 58 on page 476 is a graphical representation of the following Telnet mapping statements. The numbered mapping statements correspond to the numbered buttons in the figure. The mappings that specify USERGROUP USGRP1 generate buttons 4 through 8 for exact user ID in a group and buttons 12 through 16 for wildcard user ID in a group.

```
LUGROUP      LUGRP1   LU01..LU10..FFNN    ENDLUGROUP
LUGROUP      LUGRP2   LU11..LU99..FFNN    ENDLUGROUP
PRTGROUP     PRTGRP1  PRT01..PRT10..FFFNN ENDPRTGROUP
PARMSGROUP   PGDBG    DEBUG DETAIL        ENDPARMSGROUP
PARMSGROUP   PGSCAN   SCANINTERVAL 10     ENDPARMSGROUP
PARMSGROUP   PGMTKO   TKOSPECLU  7        ENDPARMSGROUP
PARMSGROUP   PGALL    DEBUG DETAIL
                      SCANINTERVAL 10
                      TKOSPECLU  7        ENDPARMSGROUP
MONITORGROUP MONGRP1  NODYNAMICDR         ENDMONITORGROUP
USERGROUP    USGRP1   PAYUSR1   PAYUSR*   ENDUSERGROUP
HNGROUP      HNGRP1   USER1.GROUP3.COM
                      USER5.GROUP3.COM    ENDHNGROUP

   (1) PARMSMAP       PGALL    USERID,PAYUSR1
   (2) LINEMODEAPPL   TSO      9.9.9.9
   (3) PARMSMAP       PGDBG    9.9.9.9

(4,12) DEFAULTAPPL    PAYROLL  USGRP1
(5,13) PRTDEFAULTAPPL PAYPRT   USGRP1
(6,14) LUMAP          LUGRP1   USGRP1   SPECIFIC
(7,15) PRTMAP         PRTPGRP1 USGRP1   SPECIFIC
(8,16) PARMSMAP       PGTKO    USGRP1

   (9) USSTCP         USSTABHN HNGRP1
  (10) LUMAP          LUGRP2   HNGRP1   GENERIC
  (11) PARMSMAP       PGSCAN   HNGRP1

  (17) INTERPTCP      INTTAB1  LINK1
  (18) MONITORMAP     MONGRP1  LINK1
  (19) DEFAULTAPPL    TPX1
  (20) USSTCP         USSTAB1
```

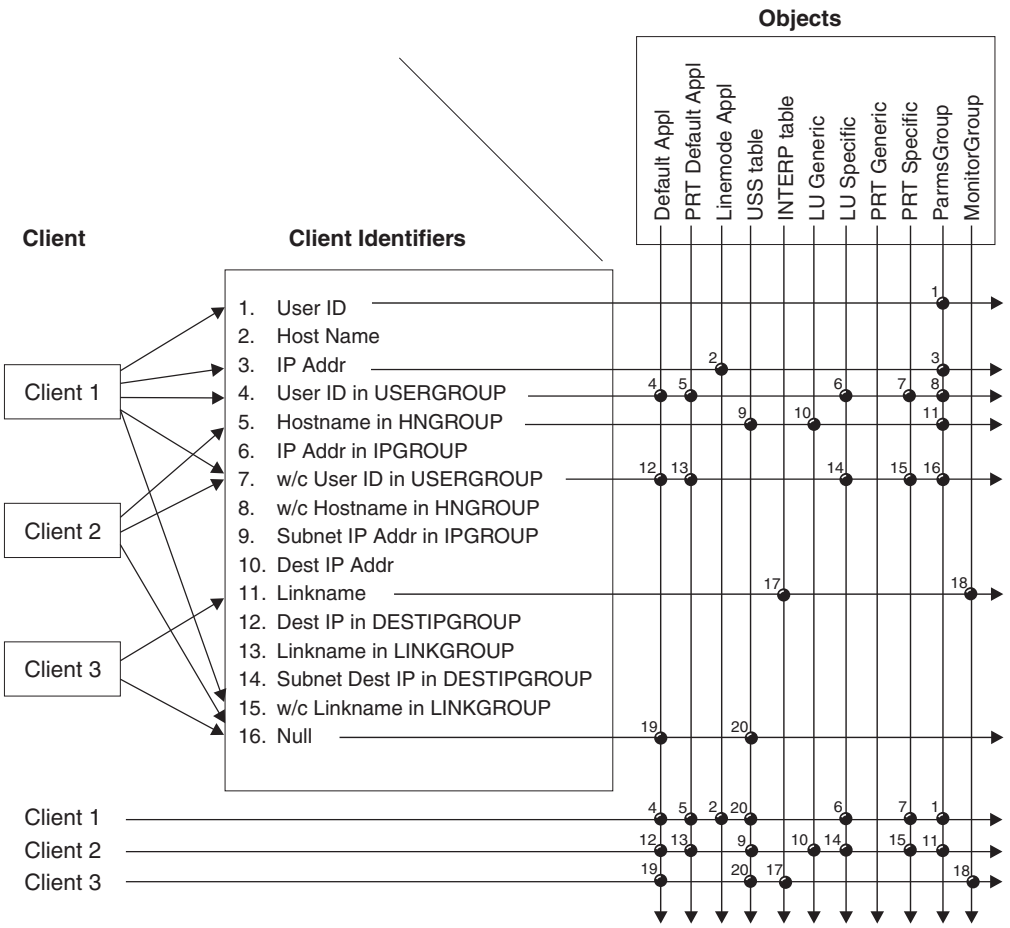


Figure 58. Search method

Client 1 example: Assume client 1 connects from IP address 9.9.9.9 using client authentication and is assigned PAYUSR1. The client does not have a host name ending in GROUP3.COM and does not have a linkname LINK1. Using Figure 58, the client will be assigned objects as shown in Table 22:

Table 22. Client 1 example

Button	Object type	Name	Action
(1)	ParmsGroup	PGALL	Assigned, exact user ID match
(2)	Linemode Appl	TSO	Assigned, exact IP address match
(3)	ParmsGroup		Not assigned, already assigned
(4)	TN3270(E) Appl	PAYROLL	Assigned, exact user ID match in group
(5)	Print Appl	PAYPRT	Assigned, exact user ID match in group
(6)	LUgroup-Spec	LUGRP1	Assigned, exact user ID match in group
(7)	PRTgroup-Spec	PRTGRP1	Assigned, exact user ID match in group
(8)	ParmsGroup		Not assigned, already assigned
(9)	USS table		Ignored, no exact hostname match in group
(10)	LUgroup-Gen		Ignored, no exact hostname match in group
(11)	ParmsGroup		Ignored, no exact hostname match in group

Table 22. Client 1 example (continued)

Button	Object type	Name	Action
(12)	TN3270 (E) Appl		Already covered, exact user ID in group
(13)	Print Appl		Already covered, exact user ID in group
(14)	LUgroup-Spec		Already covered, exact user ID in group
(15)	PRTgroup-Spec		Already covered, exact user ID in group
(16)	ParmsGroup		Not assigned, already assigned
(17)	Interp table		Ignored, no linkname match
(18)	MonitorGroup		Ignored, no linkname match
(19)	TN3270(E) Appl		Not assigned, already assigned
(20)	USS table	USSTAB1	Assigned, NULL Client Identifier match

Client 2 example: Assume client 2 connects from IP address 9.1.1.1 using client authentication and is assigned PAYUSR5 and has a host name of USER5.GROUP3.COM. The client does not have a linkname LINK1. Using Figure 58 on page 476, the client will be assigned objects as shown in Table 23:

Table 23. Client 2 example

Button	Object type	Name	Action
(1)	ParmsGroup		Ignored, no exact user ID match
(2)	Linemode Appl		Ignored, no exact IP address match
(3)	ParmsGroup		Ignored, no exact IP address match
(4)	TN3270(E) Appl		Ignored, no exact user ID match in group
(5)	Print Appl		Ignored, no exact user ID match in group
(6)	LUgroup-Spec		Ignored, no exact user ID match in group
(7)	PRTgroup-Spec		Ignored, no exact user ID match in group
(8)	ParmsGroup		Ignored, no exact user ID match in group
(9)	USS table	USSTABHN	Assigned, exact hostname match in group
(10)	LUgroup-Gen	LUGRP2	Assigned, exact hostname match in group
(11)	ParmsGroup	PGSCAN	Assigned, exact hostname match in group
(12)	TN3270 (E) Appl	PAYROLL	Assigned, wildcard user ID match in group
(13)	Print Appl	PAYPRT	Assigned, wildcard user ID match in group
(14)	LUgroup-Spec	LUGRP1	Assigned, wildcard user ID match in group
(15)	PRTgroup-Spec	PRTGRP1	Assigned, wildcard user ID match in group
(16)	ParmsGroup		Ignored, no wildcard user ID match in group
(17)	Interp table		Ignored, no linkname match
(18)	MonitorGroup		Ignored, no linkname match
(19)	TN3270(E) Appl		Not assigned, already assigned
(20)	USS table		Not assigned, already assigned

Client 3 example: Assume client 3 connects from IP address 9.2.2.2 without client authentication and has a host name of USER3.GROUP1.COM. The client connects to linkname LINK1. Using Figure 58 on page 476, the client will be assigned objects

as shown in Table 24:

Table 24. Client 3 example

Button	Object type	Name	Action
(1)	ParmsGroup		Ignored, no exact user ID match
(2)	Linemode Appl		Ignored, no exact IP address match
(3)	ParmsGroup		Ignored, no exact IP address match
(4)	TN3270(E) Appl		Ignored, no exact user ID match in group
(5)	Print Appl		Ignored, no exact user ID match in group
(6)	LUgroup-Spec		Ignored, no exact user ID match in group
(7)	PRTgroup-Spec		Ignored, no exact user ID match in group
(8)	ParmsGroup		Ignored, no exact user ID match in group
(9)	USS table		Ignored, no exact hostname match in group
(10)	LUgroup-Gen		Ignored, no exact hostname match in group
(11)	ParmsGroup		Ignored, no exact hostname match in group
(12)	TN3270 (E) Appl		Ignored, no wildcard user ID match in group
(13)	Print Appl		Ignored, no wildcard user ID match in group
(14)	LUgroup-Spec		Ignored, no wildcard user ID match in group
(15)	PRTgroup-Spec		Ignored, no wildcard user ID match in group
(16)	ParmsGroup		Ignored, no wildcard user ID match in group
(17)	Interp table	INTTAB1	Assigned, linkname match
(18)	MonitorGroup	MONGRP1	Assigned, linkname match
(19)	TN3270(E) Appl	TPX1	Assigned, NULL Client Identifier match
(20)	USS table	USSTAB1	Assigned, NULL Client Identifier match

Mapping methods

Once you have identified all the clients in your network, determine which clients need to remain exact Client Identifiers for special considerations and which clients can be combined or wildcarded into Client Identifier groups. Now map the application and LU Objects to these Client Identifiers.

1. Use LU name mapping statements to assign LU names to connections based on the Client Identifier. This step is required.
2. Use application mapping statements to facilitate session setup based on the Client Identifier.
3. Use the connection parameters mapping statement to change connection parameters based on specific Client Identifiers.
4. Consider the advanced topic features for additional Telnet functions.

LU name mapping statements

Every connection must be represented by an LU name before a session can be initiated. The time of LU assignment depends on the connection type. In general, for TN3270E clients, the LU name is assigned early during connection negotiation

before an application name is known. For all other types of clients, the LU name is assigned immediately after application name selection. For details and exceptions to this rule, see "Advanced LU mapping topics" on page 493. Mapping statements define which LU name is assigned to the connection.

DEFAULTLUS: The simplest way to assign LUs is to create a default LU group that Telnet can use for all terminal connections. DEFAULTLUS is a combination statement that defines the LUs in a default group and maps the group to the NULL Client Identifier. If the client's Client Identifiers do not match any LU mapping statements, the client is identified by the NULL Client Identifier and will be assigned LUs from the default group.

For example, use the following statement to create an LU group with a numeric range of LUG1001 to LUG1100. When Telnet assigns an LU to a terminal connection, it will assign the next available LU from that group of 100 LUs.

```
DEFAULTLUS  LUG1001..LUG1100..FFFFNNN  ENDDEFAULTLUS
```

By default, Telnet uses a sequential selection method to assign LUs from the LU group. No LU name will be reused until all the names in the group have been used. Specifying NOSEQUENTIALLU changes the selection process to always start at the beginning and find the first name available. If the range is large and a large number of LUs are already assigned, NOSEQUENTIALLU might degrade LU lookup performance.

DEFAULTPRT: The DEFAULTPRT statement is used to create a default LU pool that Telnet will use for all printer connections. For example, use the following statement to create an LU group with a numeric range of PRTG1001 to PRTG1100. When Telnet assigns an LU to a printer connection, it will assign the next available LU from that group.

```
DEFAULTPRT  PRTG1001..PRTG1100..FFFFFNNN  ENDDEFAULTPRT
```

LUMAP, PRTMAP, LUGROUP, PRTGROUP: The LUMAP and PRTMAP statements allow you to map LUs to connections based on the Client Identifier for terminal emulators and printer emulators, respectively. For example, use the following statements to map LU name LUT001 to any terminal client identified by the client IP address 1.1.1.1 and map LU name PRT001 to any printer client identified by client IP address 2.2.2.2.

```
LUMAP  LUT001  1.1.1.1
PRTMAP PRT001  2.2.2.2
```

An LU group can be used when it is not necessary to have an exact LU name to Client Identifier match. For example, use the following statements to create a terminal LU group and a printer LU group, and map both groups to the Client Identifier IPGPAY. When a terminal client connects, Telnet will assign an LU from LUGRP1. When a printer client connects, Telnet will assign an LU from PRTGRP1.

```
LUGROUP   LUGRP1    LUT101..LUT400..FFFXXX   ENDLUGROUP
PRTGROUP  PRTGRP1   PRT101..PRT400..FFFXXX   ENDPRTGROUP

IPGROUP IPGPAY   255.255.0.0:9.8.0.0  ENDIPGROUP

LUMAP    LUGRP1   IPGPAY
PRTMAP   PRTGRP1  IPGPAY
```

Once all 300 LUs are assigned, the next client connection request will fail. In this way, the LUGROUP Object can limit the number of clients connected at one time.

If a client connection is known by a Client Identifier that has an LU group mapping, only that mapping will be used to assign an LU name. The DEFAULTLUS group will not be used. It is used only in the case when no other LU mapping exists.

LU range specification: Telnet LU range rules allow for almost any type of LU range needed. Ranges can be alphabetic (A), numeric (N), alphanumeric (B), hexadecimal (X), or completely wildcarded (?) which includes alphanumeric and the three national characters (@,#,$). The range type can be different for each character position. Within the LU range, any character position can be fixed (F). To conform with VTAM LU naming convention, the first character must be alphabetic or a national character. If the first character is a range, only the alphabetic range can be used.

An LU range is created by specifying a starting LU name, an ending LU name, and the range rules to be used. For example, the following statement creates a range from TCPM1000 to TCPM1100.

```
TCPM1000..TCPM1100..FFFFFNNN
```

The three components are:

- Starting LU name (TCPM1000)
- Ending LU name (TCPM1100)
- Range rules (FFFFFNNN)

All three components must be the same length, the Starting LU name overall must be lower than the Ending LU name, and each character position value must be appropriate for the specified range rule.

Tip: In the above example, the character 1 following the character M is defined as fixed because it cannot change. The range rule cannot specify N even though it seems to be part of the number range.

The ascending order of characters is 0-9, A-Z, @, #, $.

If the range rule is omitted, Telnet assumes the following style, where LowerRange and UpperRange must be all numeric or all alphabetic:

```
LuBase+LowerRange..LuBase+UpperRange
```

Numeric values are lower than alphabetic values to facilitate the use of hexadecimal ranges. The range rules are:

```
Range         Rule       Characters
Numeric        N          0-9
Alphabetic     A          A-Z
AlphaNumeric   B          0-9,A-Z
Hexadecimal    X          0-9,A-F
Wildcard       ?          0-9,A-Z,@,#,$
```

The maximum number of LUs per range is 4294967295 and the maximum number of LUs per group is 4294967295.

The creation of LU name values from the range specification begins at the Starting LU and increments the rightmost variable position first, moving to the left as each variable position reaches its range maximum. The process is like an odometer, except that each position can have different basing instead of all positions being base 10. For example, the following statement has 223 LU name entries.

```
LU555..LU777..FFNNN
```

The breakdown of the range is:

```
LU555->LU559,                                                 5
LU560->LU569, LU570->LU579, LU580->LU589, LU590->LU599,      40
LU600->LU699,                                               100
LU700->LU769, LU770->LU777                                   78
                                                            ===
Total  ------------------------------------------------> 223
```

The LU names increment just as the numbers on an odometer would. A less intuitive case involves an alphabetic range of 1407 LU name entries.

```
LUCCC..LUEEE..FFAAA
```

The breakdown of the range is:

```
LUCCC->LUCCZ,                                                  24
LUCDA->LUCDZ, LUCEA->LUCEZ, LUCFA->LUCFZ, ... LUCZA->LUCZZ,   598
LUDAA->LUDZZ,                                                  676
LUEAA->LUEDZ, LUEEA->LUEEE                                     109
                                                              ====
Total  ---------------------------------------------------> 1407
```

It is important to realize that the ranges above do *not* break down in the following patterns:

```
LUCCC->LUCCE, LUCDC->LUCDE, LUCEC->LUCEE, ...
LU555->LU557, LU565->LU567, LU575->LU577, ...
```

It is an incorrect assumption that the LU name after LUCCE would be LUCDC. The correct LU name after LUCCE is LUCCF. The LU names increment to LUCCZ and the next name is LUCDA. When the rightmost position reaches the range maximum, the position to its left is incremented by one, and the rightmost position starts at the range beginning, not the character specified in the Starting LU name.

All range types are handled the same way. The position is incremented to its maximum value and then wraps to the beginning range value, not the specified Starting LU name value. By the same logic, the position is incremented to the ending range value and not the Ending LU name value.

All LU names increment the same way. A more complicated example mixes fixed and variable character positions with several different range types. The LU range has 39744 LUs.

```
LUAD1800..LUGD98FZ..FFAFNFXB
```

Calculating the number of LUs is easier if the fixed positions are removed. For purposes of calculating the number of LUs, the range is specified as follows:

```
A100..G9FZ..ANXB
```

This breaks down as follows:

```
A100->A10Z, D110->D11Z, ... A190->A19Z, A1A0->A1AZ ... A1F0->A1FZ    576
A200->A2FZ, A300->A3FZ, ... A900->A9FZ                              4608
B000->B9FZ, C000->C9FZ, ... F000->F9FZ                             28800
G000->G9FZ                                                          5760
                                                                   =====
Total ------------------------------------------------------------> 39744
```

SEQUENTIALLU: Telnet, by default, uses a sequential method to choose LUs from a group.

```
LUGROUP  LUGRP1
     LU001..LU120..FFNNN
     LU201..LU250..FFNNN
     LU240..LU280..FFNNN
     LU010..LU050..FFFNN
ENDLUGROUP
```

From the previous example, the first LU assigned is LU001, second is LU002, and so on. If five clients repeatedly connect and disconnect, they will be assigned new LUs farther into the range each time:

- When the end of the first range is reached, selection goes to the beginning of the second range.
- At the end of the second range, selection goes to the beginning of the third range.
- At the end of the third range, selection goes to the beginning of the fourth range.
- At the end of the fourth range, selection goes to the beginning of the first range again.

Telnet does not enforce an overall ascension in LU name selection. The selection process begins at the first name of the first range and progresses to the last name of the last range. In the example, after LU250 is assigned from range 2, LU240 from range 3 is attempted next. After LU280, LU010 is attempted. After LU050, the process starts over and LU001 is attempted.

The SEQUENTIALLU function can be turned off by coding NOSEQUENTIALLU. In this case, the five LUs that are repeatedly connecting and disconnecting would never use any LU names other than LU001, LU002, LU003, LU004, and LU005. NOSEQUENTIALLU might degradate LU lookup performance when a large range is specified and only LUs at the end of the range are available. Every connection has to relearn that most of the LUs are already in use. SEQUENTIALLU allows Telnet to start its search near the last chosen LU where LUs are more likely to be available. SEQUENTIALLU and NOSEQUENTIALLU parameters can be coded at all three parameter block levels for different levels of granularity.

If several clients are connecting at the same time, the order of LU assignment might not be in exactly the same order as the connection IDs due to process timing between connection ID assignment and LU name assignment.

If single LU names are in a group with LU ranges, the single LU names are selected before any LU range names are selected, regardless of their order. In the example below, LUAAA, LUBBB, LUCCC, and LUDDD are all processed before any of the range LU names.

```
 Profile LUGROUP                              LUGROUP as used by Telnet


LUGROUP  LUGRP2                                 LUGROUP  LUGRP2
     LUAAA                                           LUAAA
     LU001..LU120..FFNNN                             LUDDD
     LU201..LU250..FFFNN                             LUBBB
     LUDDD                                           LUCCC
     LUBBB                                           LU001..LU120..FFNNN
     LU240..LU280..FFFNN                             LU201..LU250..FFFNN
     LU010..LU050..FFFNN                             LU240..LU280..FFFNN
     LUCCC                                           LU010..LU050..FFFNN
ENDLUGROUP                                      ENDLUGROUP
```

Application mapping statements

When a client connects, Telnet either immediately initiates a session request to an MVS host VTAM application or solicits the end user for an application name.

DEFAULTAPPL: The DEFAULTAPPL mapping statement is used to assign an application name to the connection and immediately initiate a session with that application, and not solicit the end user for an application name. The DEFAULTAPPL statement applies only to terminal emulators connecting in TN3270, TN3270E, or DBCSTRANSFORM mode. For example, use the following statement to map the default application PAYROLL to any TN3270(E) terminal client identified by the IPGROUP IPGPAY. When a TN3270(E) client connects, Telnet will immediately initiate a session to the PAYROLL application.

```
DEFAULTAPPL PAYROLL IPGPAY
```

PRTDEFAULTAPPL and LINEMODEAPPL: The PRTDEFAULTAPPL mapping statement is used to assign an application to a printer emulator client connecting in TN3270E mode. The LINEMODEAPPL mapping statement is used to assign an application to a client connecting in standard or binary LINE mode. For example, use the following statements to map the default application PAYPRINT to any TN3270E printer client identified by the IPGROUP IPGPAY and to map the default application TSO to any linemode client identified by the linkname LINK1. When the printer client connects, Telnet will immediately initiate a session to the PAYPRINT application. When a linemode client connects, Telnet will immediately initiate a session to the TSO application.

```
PRTDEFAULTAPPL PAYPRINT IPGPAY
LINEMODEAPPL  TSO  LINK1
```

The DEFAULTAPPL, PRTDEFAULTAPPL, and LINEMODEAPPL statements imply a basic ALLOWAPPL statement for the application name if no ALLOWAPPL or RESTRICTAPPL is explicitly coded.

USSTCP: If the end user needs the ability to choose an application, custom solicitation panels can be created using unformatted system services (USS) message tables. These tables are mapped to clients using the USSTCP mapping statement. For example, use the following statement to map a USS table, USSTAB1, to any TN3270(E) client identified by any linkname that starts with LINK. When a TN3270(E) client connects, Telnet will immediately send a custom logon screen (USSMSG10) from the USS table.

```
LINKGROUP  LNKGRP1  LINK*   ENDLINKGROUP
USSTCP USSTAB1 LINKGRP1
```

Assembled USS tables used by VTAM can also be used by Telnet.

INTERPTCP: In some cases, the application name must be generated based on the name provided by the end user or the name might be dependent on the LU name representing the client. The INTERPRET table can provide this function. Telnet uses the input from the USSMSG10 screen as input to the INTERPRET table translation list or uses the USSMSG10 input and the LU name as input to one of the INTERPRET table user-written exits. Because USS logon data is required input to the INTERPRET process, any client with an INTERPRET table mapping must also have a USS table mapping. For example, use the following statement to map an INTERPRET table, INTTAB1, to any TN3270(E) client identified by the linkname LINK1. When a TN3270(E) client connects to LINK1, Telnet will immediately send a custom logon screen (USSMSG10) from the USS table. The end user responds

with a USS logon command. LINK1 client input is then processed through the INTTAB1 INTERPRET table to derive an application name. Telnet uses the derived name to initiate a session.

```
LINKGROUP LNKGRP1
     LINK*
ENDLINKGROUP
USSTCP    USSTAB1 LNKGRP1
INTERPTCP INTTAB1 LINK1
```

Assembled interpret tables used by VTAM can also be used by Telnet.

If neither a default application nor a USS table is mapped to the connection, the Telnet Solicitor panel is sent to the end user. For a detailed discussion of the Telnet Solicitor, USS table, and INTERPRET table, see "Using the Telnet Solicitor or USS logon panel" on page 514.

Resolving DEFAULTAPPL and USS table conflicts: If both a default application and a USS table are mapped to the same Client Identifier, Telnet will use the default application to immediately initiate a session. If each is mapped by a different Client Identifier, the Object mapped by the higher priority Client Identifier is used. In all cases, any error messages are sent using the USS table messages. For example, if CICS and USSTAB1 are both mapped to destination IP address 1.1.1.1, Telnet will initiate a session with CICS and use the USS messages for any session setup errors.

If CICS is mapped to USERID USER1 and USSTAB1 is mapped to client IP address 5.5.5.5, Telnet will initiate a session with CICS and use the USS messages for any session setup errors.

If CICS is mapped to linkname LINK1 and USSTAB1 is mapped to hostname TEST1.IBM.COM, Telnet will send a USSMSG10 logon panel to the end user. The USS messages will be used for any session setup errors. The default application mapping of CICS will never be used.

ALLOWAPPL: Telnet will not initiate a session for a solicited application name unless the name is allowed. The ALLOWAPPL statement is used to configure Telnet to allow the initiation request. For example, CICS01 and CICS02 are allowable names.

```
ALLOWAPPL CICS01
ALLOWAPPL CICS02
```

The ALLOWAPPL name can be wildcarded with an asterisk (*). For example, if there are no other CICS regions, the lines above could be reduced to the following:

```
ALLOWAPPL CICS*
```

All application names can be allowed by coding the following:

```
ALLOWAPPL *
```

Default application names do not need to be explicitly allowed. However, if the default application issues a CLSDST-PASS to another application name for the session, the second application must be in the ALLOWAPPL list. For example, TSO is the default application for the NULL Client Identifier. TSO typically passes the session to TSO00001, TSO00002, and so on. The following default application mapping will initiate a session with TSO, but when TSO issues a CLSDST-PASS the new bind to Telnet will have TSO00001 as the application name.

```
DEFAULTAPPL TSO
```

Telnet will fail this session request because TSO00001 is not allowed. Add an ALLOWAPPL statement to allow the TSO* names as follows:

```
DEFAULTAPPL TSO
ALLOWAPPL   TSO*
```

RESTRICTAPPL: In addition to the ALLOWAPPL statement, Telnet provides more restrictive access to applications. The RESTRICTAPPL statement requires the end user to enter a valid RACF user ID and password before the application name is used to initiate a session. This user ID is different than the Client Identifier user ID derived from the client certificate. This user ID is used only as a security check.

For example, use the following statement to allow users USER1, USER2, USER3, USER4, and USER5 access to the PAYROLL application. At the Solicitor panel, the end user enters USER1/password and the PAYROLL application name. Telnet verifies USER1/password is valid and then immediately initiates a session with PAYROLL.

```
RESTRICTAPPL PAYROLL
   USER USER1
   USER USER2
   USER USER3
   USER USER4
   USER USER5
```

Like ALLOWAPPL, the application name can be wildcarded with an asterisk (*). The USER value can also be wildcarded with an asterisk. The user ID/password combination is used by Telnet to verify the password given for that user ID. In no way is the user ID or password used by the application. No matter how the application name request arrived at the server (from DEFAULTAPPL or USSMSG10), Telnet uses the Solicitor panel to prompt for the user ID/password. Once the user ID is validated and a password is obtained, Telnet submits the user ID/password pair for authorization to a security program such as RACF. The user ID/password check authorizes the client to connect to the application through Telnet. The application itself might also ask for a user ID/password pair that can be completely different than the pair entered at the Telnet Solicitor panel. The user ID/password pair entered at the Telnet Solicitor panel is not in any way passed to the host application. The user ID/password pair is solicited only after an application name is entered on the Solicitor (or USSMSG10) panel. If a second application is reached through the original application using CLSDST-PASS, the second application is verified and Telnet will solicit a new user ID/password pair if necessary.

When searching for a match with the input application name, Telnet will find the most specific match whether it is on the ALLOWAPPL or RESTRICTAPPL statement. If each statement has the same name specified, the RESTRICTAPPL entry is used. For example, TSO has its own user ID/password requirement and probably does not need the additional Telnet security check. However, the Telnet security check may be needed for all other applications. This example can be supported with the following statements.

```
RESTRICTAPPL   *
  USER   *
ALLOWAPPL   TSO*
```

MSG07 and LUSESSIONPEND: MSG07 and LUSESSIONPEND are Telnet parameter statements that define what Telnet should do in case of a session setup error and after normal logoff when the client is emulating a terminal. These parameters do not affect a printer connection.

- Connection negotiation error - If any problems occur during negotiation nothing can be done to keep the connection. If appropriate, Telnet will send the client an error code to help inform the client why the connection was dropped and issue a CONN DROP DEBUG message at the console.
- Session setup error - If a problem occurs during session setup such as an application name that is not valid, session request failure, or a BIND error, Telnet will drop the connection and issue a CONN DROP DEBUG message. The end user cannot get to any application other than the default. No error messages are sent to the end user and auto-reconnect loops are possible. For these reasons it is recommended that MSG07 always be used. If the MSG07 parameter is coded, the connection will not be dropped and an error message will be sent to the end user. MSG07 function applies to any connection mode whether or not USS tables are mapped to the client. If a USS table is used, the end user can press the CLEAR key to return to the USSMSG10 screen. If the LUMAP-DEFAPPL or PRTMAP-DEFAPPL statement is coded and the default application is not available, an error screen will be sent to the client whether or not MSG07 is coded.
- Normal Session Logoff - When the end user logs off a session using a normal logoff, Telnet drops the connection. If the end user typically logs on to another application after logging off the first application, it might be more efficient if the user were presented another solicitor (or USSMSG10) panel or if Telnet initiated a new session with the default application after logoff. This can be accomplished by coding the LUSESSIONPEND parameter. Code LUSESSIONPEND to redrive the initial database lookup after session logoff. Later results will be identical to the first lookup. If a default application for the client exists, Telnet will immediately initiate another session request. Otherwise, a USSMSG10 screen or solicitor panel will be sent to the end user. When LUSESSIONPEND is coded, the connection remains active but terminal LU ACBs are closed.
- SYSREQ LOGOFF - When the end user logs off a session using a "SYSREQ LOGOFF" sequence (TN3270E connection supported) and LUSESSIONPEND is coded, Telnet does not drop the connection. Instead, the user is presented with a solicitor (or USSMSG10) panel. If DEFAULTAPPL is in effect, Telnet redrives the default application.
- USS LOGOFF - When the end user issues a LOGOFF command from the USSMSG10 panel, the connection is dropped whether or not the LUSESSIONPEND parameter is coded.

Connection parameters mapping statement

Connection parameters are typically defined once at the port level. Sometimes it is useful to have different connection parameters depending on the Client Identifier. The PARMSGROUP and PARMSMAP statements allow connection parameters to be mapped at the Client Identifier level. This level of granularity applies to almost all parameters. Refer to the *z/OS Communications Server: IP Configuration Reference* for a list of Telnet parameters allowed in the PARMSGROUP block.

Assume the PAYROLL department is assigned the highest level of security and connections are being monitored with summary debug messages, general users are assigned negotiable security, and inventory employees are experiencing intermittent problems with Telnet connections that require detailed debug messages for resolution. The following statements assign the security and debug levels to the areas needed and do not affect other areas. See "Transport layer security" on page 461 for security information and "Telnet diagnostics" on page 521 for debug information.

```
HNGROUP   HNGINV
  **.GROUP3.COM
ENDHNGROUP
IPGROUP   IPGPAY
  255.255.0.0:2.2.0.0
ENDIPGROUP
IPGROUP   IPGGEN
  255.0.0.0:2.0.0.0
ENDIPGROUP
PARMSGROUP   PRMGDBG
  DEBUG DETAIL
ENDPARMSGROUP
PARMSGROUP   PRMGSEC1
  CONNTYPE SECURE
  ENCRYPTION SSL_3DES_SHA  ENDENCRYPTION
  DEBUG SUMMARY
ENDPARMSGROUP
PARMSGROUP   PRMGSEC2
  CONNTYPE NEGT
  ENCRYPTION SSL_RC4_MD5   ENDENCRYPTION
ENDPARMSGROUP
PARMSMAP   PRMGDBG  HNGINV
PARMSMAP   PRMGSEC1 IPGPAY
PARMSMAP   PRMGSEC2 IPGGEN
```

Connection monitoring mapping statement

Telnet collects monitoring data for any client connection that is mapped to a MONITORGROUP Object. The collected data can be displayed using the D TCPIP,,TELNET,CONN,CONN=*connid* command, or a network management agent can request the data from the Telnet SNMP subagent once the subagent is started. For information on the TNSACONFIG statement and the parameters needed to start the subagent, refer to *z/OS Communications Server: IP Configuration Reference*. For details about configuring the SNMP agent, see Chapter 20, "Network management," on page 1049.

Tip: If Telnet is running as its own procedure, TCPIPJOBNAME must be specified for the subagent to connect with the agent.

Telnet is capable of collecting two types of response time data at the connection level.

- Response time statistics
  - Life-of-connection response time averages
  - Sliding-window response time averages
  - Sum of squares for variance and standard deviation calculations
- Response time counts by time bucket

The MONITORGROUP Object statement in BEGINVTAM is used to set the criteria for monitoring. Options allow the inclusion or exclusion of averages, counts by time bucket, and whether or not the IP network should be included in measurements. When a MONITORGROUP is mapped to Client Identifiers with the MONITORMAP mapping statement, the requested data will be collected for those connections.

Up to 255 MonitorGroups can be active at one time for all ports per Telnet instance. Once 255 MonitorGroups are active, no additional groups can be created until some groups are no longer in use by Telnet. A MonitorGroup is considered no longer in use when the group is not defined in a current profile and no connections are using the non-current profile where the MonitorGroup was defined. After being notified that no entries are available, it is up to the operator to

manage the removal of some entries by ending connections to free a profile. The operator does not get a notification when an entry becomes available.

Each MonitorGroup entry in the table is assigned an index number. The index number coincides with one of the 255 slots available in the MonitorGroup table. Each Telnet connection entry saves the index number of the MonitorGroup mapped to the connection. When the management application queries a connection for data, it can also query the MonitorGroup table to get the group name and configuration values based on the index found in the connection entry. With the data from the connection and the MonitorGroup criteria from the MonitorGroup table, a management application can generate several summary reports. If a management application is not being used to collect the connection monitoring data, the D TCPIP,,TELNET,CONN,CONN=*connid* command can be used to view all the collected data and MonitorGroup information for a single connection.

Collecting response time data: The typical Telnet data flow can be represented as shown in Figure 59.

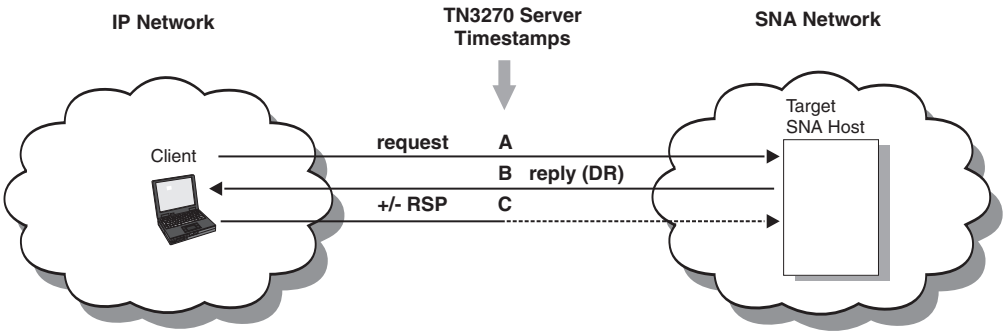


Figure 59. Typical Telnet data flow

The TN3270 server saves timestamp A when it sends a client data request to VTAM, saves timestamp B when the target SNA host application returns data, and saves timestamp C when the client responds to the definite response request that flowed with the reply.

There are many clients, mostly those not supporting TN3270E negotiation, that do not support the Definite Response (DR) function. In this case, Telnet approximates IP response time by appending a TIMEMARK request immediately after the data. A Telnet TIMEMARK acts as a synchronization mark or simply an “are you there” function for almost all clients. Almost all clients respond to the TIMEMARK request. Timestamps B and C are set based on when the TIMEMARK request is sent and a TIMEMARK response is received. In rare cases, a client does not respond to a TIMEMARK request. If Telnet does not receive a response to the TIMEMARK, a flag in the connection data indicates that IP transit time measurements were not attempted.

Some installations might not want to incur the additional network traffic of DRs or TIMEMARKs to measure IP transit time, and are only interested in the SNA application side transit time. In this case, turn off IP response measurements by specifying NOINCLUDEIP within MonitorGroup. Total response times will be the SNA response time only.

Specify INCLUDEIP within the MonitorGroup to include IP response measurements when monitoring a connection. It does not matter whether the SNA host application requested a definite response on its reply. Many applications save

processing time and IP transactions by not requesting Definite Response. If IP response time monitoring is desired and the client supports DR, specify DYNAMICDR within MonitorGroup to add the definite response request to ensure a response is received from the client. In this case, the TN3270 server does not forward the response to the target SNA host application, as indicated by the dotted line in Figure 59 on page 488. Specify NODYNAMICDR to turn off dynamic DR creation. If INCLUDEIP and NODYNAMICDR are specified and data from the application does not include a DR request, the TIMEMARK method is used.

If chained data is sent from the host application, the TN3270 server collects the entire chain before sending the complete data stream to the client. The server will save timestamp B when the entire data chain is sent.

It should be pointed out that the measured response time might not be exactly what the end user sees. The first timestamp, A, is recorded when the data passes from Telnet to VTAM on its way to the SNA VTAM application. For various SNA reasons, Telnet might have received the data much earlier and had to queue it before it could be sent to VTAM. The most common reason for this is the application might not have given Telnet direction (the ability to send data). SNA applications often use a Change Direction Indicator (CDI) to manage which side can send data. The current sender can send data until it sends a CDI, which gives the other side permission to send. There might be scenarios where a client data request comes in, Telnet does not have direction, and must queue the data until the application sends a CDI. The measured response time will not include this queue time.

Average response time data collection: Specify AVERAGE within MonitorGroup to collect average response time data. Specify NOAVERAGE to turn off average response time data collection. If average response time data is requested, the following data is available on a per connection basis:

- Life-of-connection response time averages
- Sliding-window response time averages
- Variance and standard deviation of response time averages

Life-of-connection response time averages: These averages are based on data collected since the beginning of the connection. Data collected for this average includes:

- Total transaction count.
- Sum of the round-trip response times. Round-trip response time is the time difference between timestamp C and timestamp A.
- Sum of the IP response times. IP response time is the time difference between timestamp C and timestamp B. Sum of the SNA response times can be derived by subtracting the IP sum from the round-trip sum.

With this data collected, the average round-trip time, average IP time, and average SNA time can be calculated by dividing each sum by the transaction count. The data accumulation values might wrap. It is up to the management application to detect this. The Telnet connection display will indicate the wrapped condition instead of displaying averages.

Sliding-window response time averages: These averages give higher significance to more recent data without ever completely losing the impact of earlier data. The sliding-window methodology calculates an average over an interval of time instead of over the life of the connection. An interval is made up of a specified number of equal time periods. Use AVGSAMPMULTIPLIER to specify how many periods are

in the interval. Use AVGSAMPPERIOD to specify the time length of a period. An interval should be long enough to have several data flows measured. Data collected during the period includes:

- Transaction count in the period (pTX).
- Sum of round-trip response times in the period (pRT).
- Sum of IP response times in the period (pIP). Sum of SNA response times can be derived.

In addition, there are three sliding-window variables that represent the interval and are used to calculate the sliding-window average. These variables are updated at the end of each period. They are:

- Sliding-window transaction count (swTX).
- Sliding-window sum of round-trip response times (swRT).
- Sliding-window sum of IP response times (swIP). Sliding-window sum of SNA response times can be derived.

When the interval has moved one period, subtracting the oldest period of collected data from the total and adding the newest period is one method for determining the new interval averages. However, this simple method loses all impact of earlier periods on the averages. Instead of dropping the collected data from the oldest period in the interval and completely losing its effect on the average, an average period of data is subtracted from the sliding-window totals. With this method, even the oldest period data continues to have a declining effect on the latest average calculations. And, by using an average period amount, the unique data for each period does not need to be retained. Only a total for the entire interval needs to be saved. Average period data is determined based on the fact that a period is a fraction of the total interval time. A period of data is $1/n$ of the current sliding-window totals, where n is the number of periods in the interval. The sliding-window average is derived from the new sliding-window values. At the end of each period the following calculations occur:

- Calculate the average period values:

```
avTX = swTX * (1/n)
avRT = swRT * (1/n)
avIP = swIP * (1/n)
```

- Remove an average period amount from the sliding-window totals and add the new period values:

```
swTX = swTX - avTX + pTX
swRT = swRT - avRT + pRT
swIP = swIP - avIP + pRT
```

- Calculate and report new sliding-window averages:

```
sliding-window round-trip average = swRT/swTX
sliding-window IP average = swIP/swTX
sliding-window SNA average = (swRT-swIP)/swTX
```

After the calculations are done, the period variables are reset and the next period of data collection begins. At the end of that period, the sliding-window variables are updated again with new information. The cycle continues for the life of the connection.

AVGSAMPMULTIPLIER 1 is a special case where 100% of the existing interval data is subtracted before the new period data is added. The result is an average for just that interval and the effect of older data is completely ignored.

AVGSAMPMULTIPLIER 0 is a special case that tells Telnet to include all data equally. The result is a life-of-connection average that is already available. A very large multiplier will have a similar effect.

A sliding-window round-trip average example follows. Assume an interval is made up of 5 periods. Each period is 2 minutes. Specify AVGSAMPMULTIPLIER 5 and AVGSAMPPERIOD 120. For illustrative purposes in this example, the average response time starts low and then increases to a steady state value. The data collected for each period is the sum of all response times in milliseconds (ms) and the number of transactions (tx).

Table 25. Sliding-window round-trip average example

Period	Sum of response times (ms)	Number of transactions (tx)	Average response time (ms/tx)
1	1000	10	100
2	1650	15	110
3	2800	20	140
4	3500	25	140
5	4200	30	140
6	4900	35	140
7	5600	40	140

The sliding-window average response time for period 1:

```
1000ms/10tx = 1000ms/10tx = 100 ms/tx
```

The sliding-window average response time for period 2:

```
(1000-(1000*(1/5))+1650)ms/ (10-(10*(1/5))+15)tx = 2450ms/23tx = 106.5 ms/tx
```

The sliding-window average response time for period 3:

```
(2450-(2450*(1/5))+2800)ms/ (23-(23*(1/5))+20)tx = 4760ms/38.4tx = 124.0 ms/tx
```

The sliding-window average response time for period 4:

```
7308ms/55.7tx  = 131.2 ms/tx
```

The sliding-window average response time for period 5:

```
10046.4ms/74.6tx = 134.7 ms/tx
```

The sliding-window average response time for period 6:

```
12937.1ms/94.7tx = 136.6 ms/tx
```

The sliding-window average response time for period 7:

```
15949.7ms/115.7tx = 137.9 ms/tx
```

The sliding-window method continues to include the effects of the earlier response time averages. The greater the number of periods (AVGSAMPMULTIPLIER), the longer older data will continue to impact new average calculations. Decreasing the number of periods (AVGSAMPMULTIPLIER) gives less emphasis to older data. The AVGSAMPMULTIPLIER gives the system administrator control over the emphasis placed on old data.

Each of the last five periods in the example had average response times of 140 ms/tx. Different multiplier values give the following results after seven periods. It is clear that the greater the number of periods in the interval, the greater the impact of the older data.

- 2 period interval - 139.7 ms/tx
- 5 period interval - 137.9 ms/tx
- 10 period interval - 136.6 ms/tx

Variance and standard deviation of response time averages: It is useful to know the dispersion of the response time data. The variance is a measure of how spread out the data is. The square root of the variance is the standard deviation. Assuming the distribution is a normal distribution, you can have a confidence level of the following:

- 68% that a response time value will fall within the average, plus or minus 1 standard deviation
- 95% that a response time value will fall within the average, plus or minus 2 standard deviations
- 99% that a response time value will fall within the average, plus or minus 3 standard deviations

The data saved for variance and standard deviation includes the data saved for life-of-connection averages, and the sum of each response time squared for the following:

- Complete round-trip.
- IP portion of the transaction.
- SNA portion of the transaction. The sum of squares cannot be derived in this case and must be separately maintained.

These values might wrap. It is up to the management application to detect this. The Telnet connection display will indicate the wrapped condition instead of displaying these values or the standard deviation. The typical formula used for variance is:

```
VARIANCE = SUM((Xi-Xm)**2)/(n-1)
```

However, that would require keeping all the individual response times to calculate the sum of the squares. The same formula can be expanded and rewritten to not need individual response times. The formula used by Telnet is:

```
VARIANCE = [SUM(Xi**2) - (SUM(Xi))**2/n]/(n-1)
```

The values needed in this formula are saved by Telnet for each connection requesting average response time monitoring.

Time buckets: Five time buckets, each defined by a maximum response time, are used. Every transaction has a total response time, and that time fits into one of the five configurable time buckets. The bucket boundaries are created by specifying the maximum response time for the first four buckets. The fifth bucket has an open-ended maximum value. The minimum response time boundary for each bucket is the maximum of the previous bucket. The first bucket has a minimum value of 0. A bucket transaction count is incremented by one when a response time is greater than the minimum and less than or equal to the maximum. Figure 60 on page 493 depicts the time buckets. For default values, refer to *z/OS Communications Server: IP Configuration Reference*.

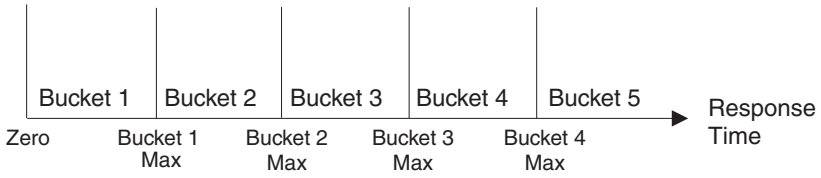


Figure 60. Time buckets

Advanced LU mapping topics

Beyond the basic LU mapping statements, there are several functions available to the advanced user. This section describes the following topics:

- Generic and Specific connection requests
- Mapping groups to Client Identifiers
- LU name assignment user exit
- Associated printer function
- Map default application and ParmsGroup by LU group
- Multiple LUMAP statements for one Client Identifier
- Keep LU for the Client Identifier
- LU group capacity warning
- LU mapping by application name
- LU mapping selection rules
- LU mapping with multilevel security active

Generic and Specific connection requests

There are three types of Telnet connection requests that dictate how the Telnet server chooses a name to represent the client. They are Generic requests, Specific requests, and associated printer requests. For details about associated printer requests, see "Associated printer function" on page 496. Most connection requests are Generic requests.

For Generic requests, the Telnet server has complete control over LU name assignment using the Generic mapping statements as a reference. All linemode and TN3270 connections use only Generic requests, and TN3270E terminal and printer emulators use Generic requests as their default request type. Specific mapping statements are ignored by Generic requests.

For Specific requests, the TN3270E client specifies the LU name to be used. The Telnet server validates the name using the Specific mapping statements as a reference. Requesting a Specific LU name allows a client to be assigned the same LU every time. This is important if the host application is LU name dependent, and the client does not have a constant Client Identifier to use for mapping an LU name. It is also important to block the server from assigning these LUs to Generic requests. That is why Specific mapping statements are ignored by Generic requests. If a Specific mapping does not find an LU match, Generic mapping statements are checked. The server confirms or denies the request during negotiation. If the LU mapping algorithms reject the client choice, the server sends a device type reject to the client. Most clients then notify the end user that the requested LU name is not valid or is already in use.

Default LU groups: DEFAULTLUS and DEFAULTPRT are default LU groups for Generic requests from terminal and printer emulators. DEFAULTLUSSPEC and DEFAULTPRTSPEC are default LU groups for Specific requests from terminal and printer emulators. DEFAULTLUS and DEFAULTPRT are explained in "LU name mapping statements" on page 478. Like the Generic pools, the Specific pools are checked only if there is no other LU mapping statement match. For example, use the following statements to create a terminal LU group with a numeric range of LUS1001 to LUS1100 and a printer LU group with a numeric range of PRTS1001 to PRTS1100. When Telnet receives a Specific connection request from a terminal, it will verify that the requested LU name is within the range specified.

```
DEFAULTLUSSPEC  LUS1001..LUS1100..FFFFNNN  ENDDEFAULTLUSSPEC
DEFAULTPRTSPEC  PRTS1001..PRTS1100..FFFFFNNN  ENDDEFAULTPRTSPEC
```

The sequential selection rules do not apply to Specific requests.

Mapping groups to Client Identifiers

The LUMAP and PRTMAP statements allow LUs to be mapped based on a Client Identifier. The LU group can be mapped Generically or Specifically. The default mapping is Generic. The keyword SPECIFIC must be coded to define a Specific mapping.

For example, use the following statements to create two LU groups. Map one group Generically to the IP group IPGPAY and map the other group Specifically to the same Client Identifier. When a Generic connection request is received, Telnet will assign the next available LU from LU group LUGRPGEN. When a Specific connection request is received, Telnet will verify the requested LU name is included in the LU group LUGRPSPC. If it is, Telnet will assign the LU name to the connection.

```
LUGROUP   LUGRPGEN  LUG101..LUG400..FFFXXX   ENDLUGROUP
LUGROUP   LUGRPSPC  LUS001..LUS100..FFFXXX   ENDLUGROUP

IPGROUP IPGPAY   255.255.0.0:9.8.0.0  ENDIPGROUP

LUMAP    LUGRPGEN   IPGPAY
LUMAP    LUGRPSPC   IPGPAY  SPECIFIC
```

Generic request connections can be assigned LUs only from Generically mapped LU groups. If no Generic mapping exists, the DEFAULTLUS group is checked. No Specific group is checked. This safeguards the Specific LU names from being used by Generic requests.

For Specific requests, Telnet first checks to see if the LU is in a Specifically mapped LU group. If the LU name is not found, the Generically mapped groups are searched. If neither LU group type contains the requested LU name, the connection request is rejected. The DEFAULTLUSSPEC group is *not* checked in this case because LU group mappings exist. If no LU group mappings exist, only the DEFAULTLUSSPEC group is checked. If the LU name is not found, the connection request is rejected. The Generic DEFAULTLUS group is *not* checked.

In addition to requesting an exact LU name, the client can request an LU group name. Telnet first searches within the mapped groups, assuming the name is an exact LU name. If that search fails, Telnet then checks the requested name against mapped Specific LU group names and then checks the name against mapped Generic LU group names. If the group name is found, the next available LU in the group is assigned using sequential LU selection unless it has been turned off.

The LU group itself is not defined as Generic or Specific. Rather, the LU group is *mapped* Generically or Specifically. It is possible to map the same LU group both Generically and Specifically. IBM recommends that you do not map the same group Generically and Specifically unless you are an advanced user.

LU name assignment user exit

Most LU assignment requirements can be satisfied using the Telnet LU group and LU mapping statements. However, there are cases when the LU assignment requirements are so specific that Telnet cannot satisfy them. In these cases, the LU name assignment user exit might be the solution. The LU name exit is defined like an LUGROUP and is mapped the same way LUGROUPs are mapped. The LUGROUP is defined as an exit by specifying *,EXIT* immediately after the LUGROUP name. For LUGROUPs, Telnet selects an LU from the group, verifies its availability, and assigns the LU to the connection. For LU name exits, Telnet calls the user-written assembler program passing a parameter list that contains client and other information. The program creates the LU name, places it in the parameter list, and returns control to Telnet. Telnet will then verify the LU name's availability and assign the LU to the connection. An LU name exit cannot be dynamically updated. Once it is loaded, it remains unchanged until the TCP/IP stack is recycled. To make a change without recycling TCP/IP, the exit name must be changed. The new name can then be added on a mapping statement. For a detailed description of the parameter list and coding requirements for the Telnet LU Exit, refer to *z/OS Communications Server: IP Configuration Reference*.

In addition to client information, the parameter list includes any LU names or ranges that were coded in the LUGROUP, and the requested application name if known. Telnet does not use the LU list. The LUs specified can be used as seed values if the LU name exit wants to use them. The LUGROUP can be defined without any LUs specified. However, if the exit is used with multilevel security, at least one LU must be specified so the LUGROUP can be assigned a security label. For details, see "LU mapping with multilevel security active" on page 503.

The LU name exit is called when the LU is assigned, when the LU is released, when the LU is inactivated, and when the LU is activated. A different function code is used for each type of call. If you do not need a certain function, like tracking inactivated LUs, the LU name exit can be written to ignore the function code. Telnet allows only one connection at a time to use the LU name exit, which serializes its use in case any local tables are maintained in the exit.

As an example, assume LU names are to be assigned based on client port number and application requested. The SIMCLIENTLU parameter is used to postpone TN3270E LU assignment until the application name is known. The parameter list includes the client port number and the requested application name. In this case, no seed LU names are needed. The LU name exit will create LU names based on the port number and application name in the parameter list.

```
LUGROUP  LUEXIT1,EXIT
ENDLUGROUP
```

In another example, assume the clients specify LUGROUP names that match up with default applications on the LUMAP statement. The LU names are to be created based on the last two numbers of the IP address and a prefix that identifies the application. For example, TSO, IMS, and CICS are three current applications. The prefix for each is TS, IM, and CI, respectively. The connection from 9.1.240.111 specifies LUGROUP LUGTSO and is assigned TS240111. The connection from 9.1.240.212 specifies LUGROUP LUGIMS and is assigned IM240212. The connection from 9.1.89.7 specifies LUGROUP LUGTSO and is assigned TS089007. Three LU

name exits are needed (LUGTSO, LUGIMS, LUGCICS), but they are all functionally equivalent. The LU name specified in the LUGROUP is passed to the LU name exit as part of the parameter list. That name is used by the exit as the prefix. The client IP address is also in the parameter list. The LU name exit combines the prefix with the last portions of the IP address to create an LU name. The following statements can be used to support this scenario.

```
IPGROUP  IPGRP1  0.0.0.0:0.0.0.0  ENDIPGROUP  ; Matches all connections


LUGROUP  LUGTSO,EXIT    TS       ENDLUGROUP
LUGROUP  LUGIMS,EXIT    IM       ENDLUGROUP
LUGROUP  LUGCICS,EXIT   CI       ENDLUGROUP

LUMAP  LUGTSO  IPGRP1  DEFAPPL  TSO
LUMAP  LUGIMS  IPGRP1  DEFAPPL  IMS
LUMAP  LUGCICS IPGRP1  DEFAPPL  CICS
```

Capacity checks cannot be performed since Telnet has no way of knowing how many total LUs are available in the LU name exit.

Associated printer function

The associated printer function allows a printer emulator to specify an active LU terminal name during connection negotiation. The server understands this special request and knows to assign a printer LU name that is associated with the requested terminal LU name. The association is established by linking a pool of terminal LUs (LUGROUP) with a pool of printer LUs (PRTGROUP). The groups are linked with the LUMAP statement. The printer LU group name is linked to the terminal LU group name by adding the PRTGROUP name on the LUMAP statement.

The two LU groups *must* have the same number of LUs defined so the LUs can be paired. The groups must have the same number of single LU names, the same number of LU ranges, and the same number of LU names in each range. If the groups do not have the same number of LUs defined, error messages will be produced during profile processing.

Once the groups are linked, Telnet assigns the *n*th printer LU to a printer connection that requests association with the *n*th terminal LU. For example, a CICS table might specify that if terminal LU1 is requesting printer function, the output should be routed to printer PRT1. Within CICS, LU1 and PRT1 are associated with each other. Use the following statements to set up printer association.

```
LUGROUP  LUGCICS   LU1..LU9           ENDLUGROUP
PRTGROUP PRTCICS   PRT1..PRT9         ENDPRTGROUP
IPGROUP  IPGRP9    255.0.0.0:9.0.0.0  ENDIPGROUP
LUMAP  LUGCICS  IPGRP9  GENERIC  PRTCICS
```

If the terminal connection matches a DEFAULTAPPL or LUMAP-DEFAPPL mapping statement, Telnet immediately initiates a session request for the specified application. If the printer connection matches a PRTDEFAULTAPPL mapping statement, Telnet immediately initiates a session request for the specified application. Building on the previous example, use the following statements to set CICS as the default application for both the terminal and printer connections:

```
LUMAP LUGCICS IPGRP9 GENERIC DEFAPPL CICS PRTCICS
PRTDEFAULTAPPL CICS
```

Neither the LU group nor the printer group can be an LU exit group. If either is an LU exit, the mapping statement will be rejected.

Drop the printer connection when dropping the terminal connection: In many cases, the associated printer connection should be dropped when the terminal connection is dropped. If you code the DROPASSOCPRINTER parameter, Telnet will monitor the terminal connection. When the terminal connection is dropped, Telnet will initiate the closing and dropping of the printer connection. The DROPASSOCPRINTER and NODROPASSOCPRINTER parameters can be coded at all three parameter block levels for different levels of granularity.

Map default application and ParmsGroup by LU group

The DEFAPPL option on the LUMAP statement allows a host VTAM application to be mapped with an LU name or LUGROUP name instead of using DEFAULTAPPL. The LUMAP-DEFAPPL combination is treated just like DEFAULTAPPL when a Client Identifier matches the LUMAP statement. The LUMAP-DEFAPPL combination also supports the LOGAPPL, FIRSTONLY, and DEFONLY parameters that are used by DEFAULTAPPL, PRTDEFAULTAPPL, and LINEMODEAPPL. The LUMAP-DEFAPPL combination is a powerful statement when used with multiple LUMAP statements for the same Client Identifier. If the LUMAP-DEFAPPL or PRTMAP-DEFAPPL statement is coded and the default application is not available, an error screen will be sent to the client whether or not MSG07 is coded.

The PMAP option on the LUMAP statement allows assignment of connection parameters based on LU or LU group name. When the LU is assigned, the parameter values specified in the PMAP PARMSGROUP will override the parameter value specified in TELNETGLOBALS, TELNETPARMS, or PARMSGROUP mapped to this connection's Client Identifier. For example, any client residing in subnet 9.0.0.0 that specifies the LU group LUGTSO will immediately have a session initiated to TSO with the LOGAPPL function, and the TIMEMARK time will be set for two hours instead of the default three hours.

```
IPGROUP    IPGRP9  255.0.0.0:9.0.0.0             ENDIPGROUP
LUGROUP    LUGTSO  TCPTSO01..TCPTSO99..FFFFFFNN  ENDLUGROUP
PARMSGROUP PGRPT2  TIMEMARK 7200                 ENDPARMSGROUP
LUMAP     LUGTSO  IPGRP9  DEFAPPL TSO LOGAPPL  PMAP PGRPT2
```

An LUMAP-DEFAPPL defined default application name is always used if specified, regardless of USSTCP mappings. LUMAP-DEFAPPL has a higher priority than any DEFAULTAPPL or USSTCP. The connection parameters assigned using LUMAP-PMAP will override any other setting of the parameters. However, not all parameters have a meaningful use by the time the LU is assigned. For example, NOTN3270E controls whether or not Telnet should negotiate for TN3270E. That negotiation is done before LU assignments. For information on which parameters can be properly applied with LUMAP-PMAP, see the parameter table in *z/OS Communications Server: IP Configuration Reference*.

The PRTMAP statement supports PRTMAP-DEFAPPL and PRTMAP-PMAP in the same manner as LUMAP-DEFAPPL and LUMAP-PMAP.

Multiple LUMAP statements

Another feature of Specific LU name requests is that the client can specify an LUGROUP name, and the server will assign an available LU from that pool. This capability is useful when different applications require different LU naming schemes, but each end user client does not need to use an exact LU name for each emulator. For example, an administrator can create three pools, one for each of three applications. Only three client emulators need to be set up. One for TSO which requests LU name LUTSO, one for CICS which requests LUCICS, and one for IMS which requests LUIMS. Assume the general users are in subnet 3.0.0.0.

Any client connecting with a Client Identifier of IPGGEN can be set up to issue a Specific request for LU pool LUTSO, LUCICS, or LUIMS, and will be assigned an LU from the appropriate pool.

After an LU is assigned, the DEFAPPL option will cause Telnet to immediately issue a session request for the appropriate application. If LOGAPPL is coded and the application is not active, VTAM will continue session initiation once the application is active.

In most cases, DEFAPPL on multiple Generic LUMAPs is not useful. LUs are assigned in order starting with the first LUMAP statement. One case that may be useful is if an application has a user limit but can be cloned. Assume the INVENTRY application can support only 20 users but can be cloned. Multiple LUMAPs with DEFAPPL will direct the first 20 HNGINV clients to INVENTRY, the next 20 HNGINV clients to INVENTR2, and the next 20 HNGINV clients to INVENTR3.

```
IPGROUP  IPGGEN
     255.0.0.0:3.0.0.0
ENDIPGROUP
LUGROUP  LUTSO   TSO00001..TSO00999  ENDLUGROUP
LUGROUP  LUCICS  CICS0001..CICS0999  ENDLUGROUP
LUGROUP  LUIMS   IMS00001..IMS00999  ENDLUGROUP
HNGROUP  HNGINV
  *.INVDEPT.COM
ENDHNGROUP
LUGROUP  LUGINV1 LUINV01..LUINV20    ENDLUGROUP
LUGROUP  LUGINV2 LUINV21..LUINV40    ENDLUGROUP
LUGROUP  LUGINV3 LUINV41..LUINV60    ENDLUGROUP
LUMAP   LUTSO     IPGGEN   SPECIFIC  DEFAPPL  TSO      LOGAPPL  FIRSTONLY
LUMAP   LUCICS    IPGGEN   SPECIFIC  DEFAPPL  CICS     LOGAPPL
LUMAP   LUIMS     IPGGEN   SPECIFIC  DEFAPPL  IMS      LOGAPPL
LUMAP   LUGINV1   HNGINV             DEFAPPL  INVENTRY LOGAPPL  DEFONLY
LUMAP   LUGINV2   HNGINV             DEFAPPL  INVENTR2 LOGAPPL  DEFONLY
LUMAP   LUGINV3   HNGINV             DEFAPPL  INVENTR3 LOGAPPL  DEFONLY
```

Pool name specification is a powerful mapping method because multiple LUMAP statements with different Objects can be used for a single Client Identifier.

Keep LU for the Client Identifier

An LU name can be kept (or reserved) for a period of time so no other client is assigned that name. Only the same Client Identifier reconnecting to Telnet within the specified time can be assigned that LU name. After the specified time, the LU name is again available for any connection. This function is useful when the application does not clean up session information quickly and a released LU is quickly reassigned to another end user by Telnet. The application thinks the new session is a continuation of the previous session but it is not. With KEEPLU, the LU will not be reassigned to a different Client Identifier for a period of time, long enough for the application to clean up its session information. The LU name is kept based on the highest Client Identifier by which the connection is known. It is either a User ID derived from a client certificate, a Hostname, or an IP address, respectively.

LU group capacity warning

An LU group capacity threshold can be specified on the LUGROUP, PRTGROUP, and default LU group statements. If specified, Telnet will check the number of LUs used in the group when an LU is assigned from the group. A message is issued when the group's in-use LU count is at or above the specified percentage of the total. Once the message is issued, no other message is issued until the in-use count has dropped below the threshold by 10% of the total. For example, an LU group

has 200 LUs with a capacity threshold of 80%. When the 160th LU is assigned, EZZ6007I is issued. Ten percent of the total in the group is 20. Therefore, after the number of in-use LUs has dropped to 140 or lower, another warning message will be issued when the in-use count rises to 160 again. If multiple LU groups have the same LU name, the only LU group checked is the group from which the LU is assigned to the client. The other LU groups might go over their capacity limits, but notification will not be issued until an LU is taken from the group. Below are examples for setting the capacity warning.

```
LUGROUP         LUGRP1,80%    TCPLU000..TCPLUF9F..FFFFFXNX
PRTGROUP        PRTGRP1,60%   TCPRT000..TCPRTFFF..FFFFFXXX
DEFAULTLUS      ,75%          LU000000..LU999999..FFNNNNNN
DEFAULTPRTSPEC  ,90%          PRTDEFS1..PRTDEFS9
```

Capacity checking cannot be done for LU groups that are defined as LU name exits. During VARY TCPIP,,OBEYFILE command processing, all LU groups are checked for in-use LU counts and a capacity warning message is issued if needed.

Tip: Be sure to leave a blank space between the default LU group statement and the capacity (,*nnn*%). Do not leave a blank space between a group name and the capacity.

LU mapping by application name

In some cases, only certain LU names are eligible to be in session with the host application. Or only certain LU names are eligible to represent user IDs. The LU and LUG parameters on the ALLOWAPPL and RESTRICTAPPL statements provide this checking function and allow some LU name mapping based on application name. The LUG parameter can represent either an LUGROUP, a PRTGROUP, or a group that is a mixture of terminal and printer LUs so that both terminal and printer emulators can access the application. If single LUs are specified, they are assumed to be terminal LUs.

For example, assume the only LUs eligible to use the inventory set of applications are the LUs in the inventory LU pools. A new LUGROUP pool named LUGINVT contains LUs from LUGINV1, LUGINV2, and LUGINV3. The ALLOWAPPL statement requires that any session request to the inventory applications have an LU name defined in LUGINVT. The LUG parameter must be used carefully. When specified, Telnet must match the LU using both the common mapping algorithms and the mapping by application. For RESTRICTAPPL, assume security authorization is required to get to the PAYROLL application, and each of the PAYxx user IDs must map to a certain LU.

```
LUGROUP   LUGINV1  LUINV01..LUINV20  ENDLUGROUP
LUGROUP   LUGINV2  LUINV21..LUINV40  ENDLUGROUP
LUGROUP   LUGINV3  LUINV41..LUINV60  ENDLUGROUP
LUGROUP   LUGINVT  LUINV01..LUINV60  ENDLUGROUP
ALLOWAPPL   INVENTR*  LUG   LUGINVT
RESTRICTAPPL   PAYROLL
  USER  PAY01     LU    LUPAY01
  USER  PAY02     LU    LUPAY02
        (user pay03 through pay20 not listed)
```

The LU group specified on the LUG parameter cannot be an LU exit. If it is, the ALLOWAPPL statement is rejected. Multiple LUs can be assigned individually using the LU keyword or a single LU group can be assigned using the LUG parameter. LU and LUG cannot be mixed on a single statement and only one LUG entry per statement is permitted. LU assignment based on application is a convenient way to limit the access to applications. However, this increases mapping complexity significantly when LU mapping statements and connection types are part of the overall mapping equation. Non-TN3270E connections or

TN3270E connections with NOTN3270E or SIMCLIENTLU specified do not keep the LU name assigned to the connection after a session is dropped. For these connection types, the end user can establish a session with different application names even if different LU names are mapped to the application names with the ALLOWAPPL or RESTRICTAPPL-USER statement. However, LU mapping based on application name does not work well with TN3270E connections because the LU is assigned during connection negotiation before the desired application name is known. In all CLSDST-PASS cases, the LU name cannot change when switching from the first application to the second because the LU's ACB is not closed during the switch. If the LU mapping by application name requires an LU name switch, the new session attempt will be failed by Telnet.

TN3270 connections do not assign an LU to represent the client until an application name is chosen. Therefore, the LU and LUG parameters can be used as sole LU mapping statements for TN3270 connections. For example, assume no other mapping statements exist (LUMAP or DEFAULTLUS), and either no TN3270E connections will be used or SIMCLIENTLU has been specified. The following ALLOWAPPL statements will map LUs to the appropriate application based on the application name chosen. The following RESTRICTAPPL statement will assign a single LU or LU pool to each user.

```
ALLOWAPPL   TSO*  LUG   LUGTSO
ALLOWAPPL   CICS  LUG   LUGCICS
ALLOWAPPL   IMS   LUG   LUGIMS
RESTRICTAPPL   APP*
  USER   USER1*   LUG  LUG10
  USER   USER01   LU   LU01
  USER   USER02   LU   LU02
```

Both of these assignment methods were very popular before TN3270E connections were introduced. TN3270E connections will likely achieve poor mapping results. An LU must be assigned during connection negotiation before the application name is known which will likely result in an LU mismatch later. TN3270E connections require that an LU mapping statement exist because an LU must be assigned to the connection during negotiations before an application name is known. Consider the following example:

```
DEFAULTLUS
  LU1 LU2 LU3 LU4
ENDDEFALTLUS
RESTRICTAPPL  APPL1
  USER  USER3  LU  LU3
ALLOWAPPL   APPL2  LU  LU4
```

Assume two TN3270 connections are started.

- Two solicitor screens appear.
- Specify APPL1, USER3, and a password. The server selects LU3 based on both the DEFAULTLUS and the RESTRICTAPPL statements.
- Specify APPL2. The server selects LU4 based on both the DEFAULTLUS and the ALLOWAPPL statements.

Assume two TN3270E connections are started.

- Two solicitor screens appear. The server assigns LU1 and LU2.
- Specify APPL1, USER3, and a password. The server fails the connection because of an LU mismatch.
- Specify APPL2. The server fails the connection because of an LU mismatch.

If LU name mapping by application name or user ID is desired with TN3270E clients, the following three solutions are available:

- If the same application or user ID is always used at the same client, individual LUMAP statements can be used to map the correct LU name to each client. Then every connection request will result in the correct LU assignment for that client. The assumptions are that the client keeps the same Client Identifier and only one client exists per Client Identifier.
- Map the NOTN3270E parameter to clients to disable all TN3270E function in the server so those connections will be TN3270, not TN3270E. The drawback is that all TN3270E function is disabled. This includes printer function, Generic/Specific function, and SNA function to the client. The TN3270E and NOTN3270E parameters can be coded at all three parameter block levels for different levels of granularity.
- Mapping the SIMCLIENTLU parameter is a less severe solution. This function will send a dummy LU name of EZBSIMLU to all TN3270E clients issuing Generic connection requests to satisfy the negotiation but will not assign a Telnet LU until an application name is chosen. This alternative preserves printer function, Specific requests, and SNA function to the client. The drawback is the name sent to the client is not the name Telnet ultimately uses to represent the client. Printer association will not work for these TN3270E Generic connections and any emulator programming that depends on the LU name will be using the dummy LU name. The SIMCLIENTLU and NOSIMCLIENTLU parameters can be coded at all three parameter block levels for different levels of granularity.

LU mapping selection rules

LU mapping selection can become complicated because of the many variations of mapping statements, TN3270E versus TN3270 connections, Generic versus Specific connection requests, printer association, and LU mappings based on application name. LU mapping is very different between TN3270E and TN3270 and will be discussed separately. But first, some general Mapping Rules for both TN3270E and TN3270 follow:

- If multiple LUMAP statements exist for a Client Identifier all Specific LUMAPs are searched (TN3270E only) and then all Generic LUMAPs are searched in the order they are listed in the profile.
- If the application is known during the LU lookup and the ALLOWAPPL or RESTRICTAPPL-USER statement has LUs listed, then the found LU must be in both the mapped LU group and in the application LU group.
- Once an LU match is found, the search stops.
- Telnet performs database lookup for Objects based on the Client Identifier. TN3270E connections require an Early Lookup so Telnet can give the client an LU name during connection negotiation. In all cases a Complete Lookup is done when the application name is known. Telnet performs an Early Lookup and a Complete Lookup for TN3270E connections. Telnet performs only a Complete Lookup for TN3270 and Linemode connections.

TN3270E LU mapping: TN3270E connections require an Early Lookup during connection negotiation. Telnet will use as much information as is available to assign an LU to the client. However, the eventual application is not known at this time unless an LUMAP-DEFAPPL or DEFAULTAPPL statement defines the application name. After connection negotiations are complete, Telnet will either send a logon solicitor (or USSMSG10) screen to the client or will perform a Complete Lookup using the application name obtained from the LUMAP-DEFAPPL or DEFAULTAPPL statement. If Complete Lookup is successful, Telnet will begin session initiation. If a solicitor (or USSMSG10) screen is sent to

the client, an application name must be entered, at which time Telnet will perform a Complete Lookup. If LU mapping is being done based on application name, a conflict might occur between the application LU mapping and the LU already assigned to the connection. For TN3270E, once an LU name is assigned during connection negotiation it can never change until the connection is dropped. The SIMCLIENTLU statement allows Telnet to assign LUs for TN3270E connections as though they were TN3270 connections. See "TN3270 LU mapping" on page 503 for mapping Generic TN3270E connection requests with SIMCLIENTLU. A request for a Specific LU from the Telnet Client will be treated as if SIMCLIENTLU were not specified. The exact lookup process for TN3270E (non-SIMCLIENTLU) is described below.

Early Lookup: An LU must be found during Early Lookup. LUMAP-DEFAPPL and DEFAULTAPPL statements are considered but not necessarily used. Possible lookup results are:

- An LU is found.
- An LU is not found, the connection is dropped.

Perform TN3270E Early Lookup in the following order. The process stops when LU lookup is successful. Printer connections use the same process, substituting PRTMAP and PRTDEFAULTAPPL.

- Check for LUMAP matches considering application lookup results and possible application-based LU mappings.
  1. For each Specific LUMAP used for Specific connection requests: If the Specific LUMAP has DEFAPPL, or DEFAULTAPPL was specified and the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
  2. For each Generic LUMAP used for Specific or Generic connection requests: If the Generic LUMAP has DEFAPPL, or DEFAULTAPPL was specified and the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
- Check for LUMAP matches without considering application lookup results.
  1. For each Specific LUMAP used for Specific connection requests: Ignore DEFAPPL and DEFAULTAPPL and perform LU lookup.
  2. For each Generic LUMAP used for Specific and Generic connection requests: Ignore DEFAPPL and DEFAULTAPPL and perform LU lookup.
- If LUMAP statements were *not checked* (different from *checked but no match*), use the appropriate Default LU pool considering application lookup results and possible application-based LU mappings. In this case the only relevant application is the DEFAULTAPPL, if specified. If the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
- If LUMAP statements were not checked, try the appropriate Default LU pool without considering application lookup results. Perform LU lookup.

Complete Lookup: An application name is required for Complete Lookup. The application name is obtained from one of three sources in the order specified.

1. Input from the USER or VTAM (via CLSDST with OPTCD=PASS)
2. DEFAPPL parameter on the LUMAP statement
3. DEFAULTAPPL statement

Use the application name and the previously found LU to perform Complete Lookup. Possible lookup results are:

- The application is not valid.
- The application is valid (return code OK or USER_REQUIRED) for the existing LU.
- The application-based LU map does not match the already chosen LU.

TN3270 LU mapping: TN3270 connections only perform Complete Lookup after all information is known. LU lookup is not done during connection negotiation. Telnet will either send a solicitor (or USSMSG10) screen to the client or will perform Complete Lookup using the application name known through the LUMAP-DEFAPPL or DEFAULTAPPL statement. If Complete Lookup is successful, Telnet will begin session initiation. If not successful, the solicitor (or USSMSG07) screen is sent to the client without an LU being assigned to the connection or the connection is dropped. The LU is not assigned until the application name is valid. If the application name is a RESTRICTAPPL, the LU is not assigned until a user ID is specified. Application-based LU mappings have a very good chance of success due to the late LU mapping aspect of TN3270 connections. When SIMCLIENTLU is coded, Generic TN3270E connections have this same characteristic.

Complete Lookup: An application name is required for Complete Lookup. The application name is obtained from one of three sources in the order specified.

1. Input from the USER or VTAM (CLSDST with OPTCD=PASS)
2. DEFAPPL parameter on the LUMAP statement
3. DEFAULTAPPL statement

Use the application name to perform Complete Lookup. Possible lookup results are:

- The application is not valid.
- The application is valid but an LU is not found.
- The application is valid (return code OK) and an LU is found.
- The application LU map does not match the Client Identifier LU map.

If the application is not valid, no LU is assigned to the connection and an error message is sent to the client. If the application is valid, continue the LU lookup in the following order.

- Check for LUMAP matches considering application-based LU lookup results. Only Generic LUMAPs are searched. If the application lookup return code is OK, then perform LU lookup.
- If no LUMAP statements were used, check for application-based LU mappings. If the application lookup return code is OK and LUs are defined on the application statement, perform LU lookup.
- If no LUMAP or application-based LU mapping statements were used, use the DEFAULTLUS pool considering application lookup results. If the application lookup return code is OK, then perform LU lookup.

LU mapping with multilevel security active

Telnet can be in a multilevel secure environment that uses security labels. For more information on preparing for TCP/IP networking in a multilevel secure environment, refer to Chapter 4, “Preparing for TCP/IP networking in a multilevel secure environment,” on page 145 and *z/OS Planning for Multilevel Security and the Common Criteria*. To ensure correct security label comparisons, Network Access Control (NAC) must also be active for Telnet. For more information about NAC, see “Network Access Control” on page 467.

Tip: If Telnet is running as its own procedure, it is not multilevel security compliant. Only Telnet running as part of the TCP/IP stack is multilevel security compliant.

If multilevel security is active, Telnet ensures the security label of the selected LU is compatible with the security label of the client.

- Telnet retrieves the security label of the client when the connection is accepted.
- Telnet assigns a security label to all LUGROUPs based on the first LU name in the group. The first single LU name in the group is used. If no single LU names exist, the first LU name within the first LU range is used.
  - If multilevel security is active, an LUGROUP EXIT is required to have at least one LU name in the group. The LU name is used to obtain a security label for the group. The name is passed to the exit in the parameter list and can be used or ignored by the exit.
  - A single LU name on a mapping statement is treated as an LUGROUP with one LU name. That LU name is used to obtain the security label for the LUGROUP created by Telnet.

When multilevel security is active, LU lookup uses the following process:

1. The security label of the client is compared with that of the mapped LUGROUP. If the group is compatible, Telnet searches for an available LU in the group. If not compatible, the LUGROUP is skipped.
2. Telnet retrieves the security label of the selected LU and compares it with the security label of the LUGROUP. If the selected LU is not compatible with the LUGROUP, the LU is inactivated and no other LU in the group is tried.
3. If the LUGROUP was not compatible or no LU was available, the steps are repeated for each mapped LUGROUP until an LU is found or all LUGROUPs are checked.

Advanced application topics

In addition to the basic function of facilitating session setup, Telnet supports several advanced functions such as:

- Session initiation management (LOGAPPL, QINIT, FIRSTONLY, and DEFONLY)
- Connection and session takeover
- Queuing sessions
- Disconnect on session error
- Bypass RESTRICTAPPL with CERTAUTH
- Allow printer sessions with RESTRICTAPPL
- Keeping the ACB open
- Express Logon Feature

Session initiation management (LOGAPPL, QINIT, FIRSTONLY, and DEFONLY)

The LOGAPPL, QINIT, FIRSTONLY, and DEFONLY options can be coded on DEFAULTAPPL, PRTDEFAULTAPPL, LINEMODEAPPL, LUMAP-DEFAPPL, or PRTMAP-DEFAPPL. For the remainder of this section, DEFAULTAPPL represents all the default application statements.

LOGAPPL, QINIT: The LOGAPPL or QINIT functions keep the Telnet LU active if a Request Session fails due to the host VTAM application not being active. In addition, VTAM remembers the attempted Request Session and will initiate a session request to the Telnet LU on behalf of the application when the application

becomes active. When the Request Session fails, Telnet sends the client a solicitor panel or USSMSG07 screen. The end user then has the option of logging on to a different host VTAM application (if DEFONLY is not coded). When this different session is started, VTAM drops the queued Request Session for the original session.

What happens at session logoff depends on whether or not LUSESSIONPEND and FIRSTONLY are coded and whether LOGAPPL or QINIT is coded. If LUSESSIONPEND is coded, the connection remains. Otherwise, it is dropped. If FIRSTONLY is coded, Telnet will send a USSMSG10 screen or Solicitor panel to the client. If FIRSTONLY is not coded, Telnet will initiate another session with the default application defined by the DEFAULTAPPL statement. LOGAPPL and QINIT have different results when logging off the original application. When LOGAPPL is specified, a USSMSG10 or Solicitor panel is sent to the client. When QINIT is specified, the default application is redriven.

FIRSTONLY, DEFONLY: Sometimes a default application is used at initial connection, but after LOGOFF a USSMSG10 or solicitor panel is more appropriate than redriving the default. In this case, code the FIRSTONLY parameter. This indicates the default should be used on the first session only. After a session has been established, any subsequent lookups will ignore the default and send the USSMSG10 screen or solicitor panel.

If MSG07, LUMAP-DEFAPPL, or PRTMAP-DEFAPPL is coded and the default application is inactive, an error screen will be sent to the client. The DEFONLY parameter will block a user-entered application choice if it is different than the default. This parameter prevents application choice while giving the end user error information.

The following table summarizes several possible session initiation failure scenarios.

- ReqSess OK - A Request Session to the default application succeeded or a Request Session to a second application from the USSMSG07 screen succeeded.
- ReqSess Fail - A Request Session to the default application failed.
- 2nd Appl Fail - A Request Session to a second application from the USSMSG07 screen failed.

Scenario / Mapping Statement	ReqSess OK	ReqSess Fail		2nd Appl Fail	
		MSG07	No MSG07	MSG07	No MSG07
DEFAULTAPPL name	1	2	5	2	N/A
DEFAULTAPPL name LOGAPPL DEFAULTAPPL name QINIT	1	3	3	4	4

Figure 61. Session initiation failures scenarios

FIRSTONLY is not a consideration because the first session has not been established.

1. In session.
2. Send USSMSG07 or Solicitor panel to client. Close the ACB.
3. Send USSMSG07 or Solicitor panel to client. Keep ACB open, queue original session request in VTAM.
4. Send USSMSG07 or Solicitor panel to client. Keep ACB open, keep the new queued session request in VTAM.

5. Drop connection.

The following table summarizes several possible session ending scenarios. The session is ending due to normal LOGOFF or session breakage (possibly caused by loss of the application).

- Original Application - User is in session with the original default application.
- CLSDST from Original Appl - User is in session with a second (or later) application after issuing a CLSDST-PASS from the original application.

Scenario / Mapping Statement	Original Application		2nd Appl or CLSDST from Orig	
	Logoff	Break	Logoff	Break
DEFAULTAPPL name DEFAULTAPPL name QINIT	1	1	1	1
DEFAULTAPPL name LOGAPPL	2	1	1	1
DEFAULTAPPL name FIRSTONLY LOGAPPL DEFAULTAPPL name FIRSTONLY QINIT	2	1	2	1

Figure 62. Session ending scenarios

In all cases, if LUSESSIONPEND is not coded the connection is dropped.

1. Redrive the default application.
2. Send USSMSG10 or Solicitor panel to client. Close the ACB.

Connection and session takeover

In some cases the route for a connection is lost without Telnet being notified. When this occurs, the end user can no longer make contact with the host application. The end user will typically disconnect the emulator and initiate another connection to the Telnet server over a difference route. Even though reconnected, in many cases the session cannot be reestablished. For example, assume the host application is TSO and the end user is in session with TSO user ID USER1. The route is lost, so the end user disconnects and establishes a new connection over a different route. Assume a Generic connection so Telnet will assign a different LU to represent the client. When the TSO logon to USER1 is attempted, TSO will fail the logon because USER1 is still in session with the original Telnet LU. There is no way for the end user to bring down the original connection. The end user has to wait for an inactivity timer in Telnet or TSO to bring down the original session. If a Specific connection is requested the connection is rejected sooner. The second Specific connection request will specify the original LU. Telnet will fail the request during Early Lookup indicating that the LU is already in use. The problem with both situations is that the original connection is still assumed active by Telnet.

The takeover function fixes this problem. When takeover is active for a connection, Telnet LU lookup attempts takeover whenever an LU name is specified upon entry to the LU lookup routines.

The TKOSPECLU and TKOSPECLURECON statements activate takeover for the connection and require that the end user specify the LU name. The statement name is derived from the function of connection takeover (TKO) for a Specific LU (SPECLU) connection request. Because the LU name is specified, LU lookup attempts to take over the original connection to which the LU was assigned. The

Specific LU request allows an end user to move to any client to take over a connection that is lost and provides some level of security because it requires the end user to know the LU name.

The TKOGENLU and TKOGENLURECON statements activate takeover for a connection without the end user specifying an LU name. The statement name is derived from the function of connection takeover (TKO) for a Generic LU (GENLU) connection request. During the original connection LU lookup, Telnet saves the LU name by Client Identifier. If another connection request is received from the same Client Identifier, Telnet assumes takeover should be attempted using the original LU name. If multiple connections are made from a single client, all additional connection setups will be delayed by the time it takes to attempt takeover of the first connection. When the takeover attempt fails, Telnet resumes normal Generic LU lookup for that connection. For Generic LU takeover to work, the Client Identifier must remain the same. The Client Identifier can be the client user ID derived from the client security certificate. If no user ID is available, the client hostname is used. If no hostname is available, the client IP address is used.

The TKOSPECLU and TKOGENLU statements cause the following events to occur. When the takeover connection request arrives, Telnet LU lookup discovers the LU name is in use and suspends the new request. Telnet sends a TIMEMARK request to the original client, which acts as an "are you there" message. The client is required to respond to the TIMEMARK. If no response is received by Telnet within the time specified on the takeover statement, Telnet drops the original session and connection. The original LU is reserved during the drop process. Once the original session and connection are dropped, Telnet resumes processing the new request. This time the LU is not in use, only reserved for takeover purposes, and is assigned to the new takeover session. The end user is essentially starting over. The original session has been dropped, allowing the end user to immediately log on to the same TSO user ID again.

The TKOSPECLURECON or TKOGENLURECON statements can be used to accomplish the same connection drop but avoid the session drop. When the original connection is dropped, the Telnet LU stays in session with the host application. The new connection is established and Telnet sends an LUSTAT to the host application indicating that Presentation Space Integrity was lost (X'082B'). Depending on the application, it will either end the session or resend the previous screen. By resending the previous screen, the end user is able save the original session and avoid the SNA session tear-down and restart process. At worst, if the application drops the session upon receipt of the LUSTAT, the end user is able to immediately log on again as if TKOSPECLU or TKOGENLU were coded.

In some cases, the original client TCP/IP stack may respond to the Timemark with a RESET. Telnet interprets this RESET as a client disconnect and assumes the end user disconnected the session. Telnet then drops the session. To keep the session in this case, add the KeepOnTmReset option to the TKOSPECLURECON or TKOGENLURECON statement. A security risk exists when using this parameter. An end user may actually disconnect just before the Timemark arrives due to an unauthorized takeover attempt. Telnet will interpret the disconnect as a response to the Timemark and allow the takeover without loss of the VTAM session.

Part of session initiation includes Telnet issuing a SETLOGON VTAM macro that contains control vectors. The control vectors include client information, such as IP address and host name if available. If the connection is capable of being taken over, a flag is set in the control vector. This information is passed to the application's logon exit for use by the application. If the connection is taken over, a

second SETLOGON macro is issued. However, the control vector information does not go beyond the VTAM that processed the SETLOGON. The application's logon exit is not redriven and is not aware of connection changes. Some applications require the IP address to remain constant. Since the application cannot be notified of changes, Telnet has the ability to only allow takeover by clients from the same IP address. Specify SAMEIPADDR on the TKOGENLURECON or TKOSPECLURECON statements to ensure takeover is done by a client from the same IP address.

TKOSPECLU, TKOSPECLURECON, TKOGENLU, and TKOGENLURECON can be coded at all three parameter block levels for different levels of granularity. Code NOTKO to turn off all takeover function at any of the three levels.

The new connection must have an equal or higher security level to take over the original connection. The order for connection security is:

1. Basic
2. Secure
3. Secure with CLIENTAUTH SSLCERT
4. Secure with CLIENTAUTH SAFCERT

If the original connection used SSL with CLIENTAUTH SAFCERT, either takeover method will verify that the new connection is using the same client certificate. Telnet does this by translating the new certificate to a user ID and comparing the new user ID to the user ID on the original connection.

Recommendation: If there is a chance the connection can be taken over by an unauthorized user, TKOSPECLURECON and TKOGENLURECON should not be used. Neither statement requires the end user to reverify user authenticity to the host application.

Tip: A time value of zero is permitted. In this case the server will always perform the takeover whether or not the original connection is still active. The zero value is intended for testing purposes rather than production use.

Sometimes a takeover attempt will not complete as expected. This may be due to one of the following factors:

- The profile of the original connection defines how the original connection can be taken over. Be sure the original connection supports the desired takeover method.
- The new connection request must specify the LU name of the connection being taken over if TKOSPECLU or TKOSPECLURECON is specified.
- TKOSPECLURECON and TKOGENLURECON will not preserve a session if the takeover is done from a different port. The ACB of the LU is associated with the original port and must be closed before it can be associated with the new port. The takeover will function as a TKOSPECLU and TKOGENLU takeover.
- TKOSPECLURECON and TKOGENLURECON might not preserve a session if the original client connection is ended before the TKOSPECLURECON or TKOGENLURECON timer expires. If the close reason is TIMEMARK or INACTIVE, the session will be preserved under the assumption that the inactivity is due to a lost connection. Any other close reason will cause the takeover to function as a TKOSPECLU or TKOGENLU takeover. This is done to protect users who disconnect their client as a means of logging off their session. These sessions will not be taken over. Instead, the end user will have to issue a new logon.

Takeover is also affected by where the new client resides and how the old client responds. There are several event scenarios and results will vary.

- Event 1

  New client connects from a different IP address or Port.

  Original client responds to TIMEMARK.

  Result - Takeover will not occur in this case because the original client is still responding. With a Specific LU request, the new client will receive an error indicating the LU is already in use. With a Generic LU request, Telnet assigns the next available LU to the connection.

- Event 2

  New client connects from a different IP address or Port.

  Original client does not respond to TIMEMARK.

  Result - Connection takeover will occur. If TKOSPECLURECON or TKOGENLURECON is mapped to the client, session takeover will occur. A likely scenario in this case is Telnet has lost connectivity to the old connection due to a failed router and the new connection is using a different route, the original machine lost power and has not reestablished connectivity, or the original machine lost power but reestablished connectivity with a different IP address.

- Event 3

  New client connects from a different IP address or Port.

  Original client stack responds with RESET.

  Result - Connection takeover will occur. However, even if TKOSPECLURECON or TKOGENLURECON is coded, session takeover will not occur because Telnet handles the RESET as a client disconnect. A likely scenario in this case is a PC lost power and then regained power. The takeover request is accepted from either a different PC or the same PC using a different port. After the new connection request is accepted, Telnet sends a TIMEMARK to the original client PC stack. The PC stack does not recognize the IP-port and responds with a RESET. If KeepOnTmReset is specified and the RESET is received by Telnet after the Timemark has been sent, Telnet will keep the session.

- Event 4

  New client connects from the same IP address and Port.

  Telnet stack rejects the request.

  Original client times out and sends a RESET.

  Result - The original session and connection are dropped. Takeover does not occur. The new client is able to connect on retry because the original connection and session were cleaned up. A likely scenario in this case is a PC has lost power and then regained power. The same PC is used to attempt takeover. The client is assigned the same port as before the power loss and has the same IP address.

Queuing sessions

Logon manager applications are very popular. Typically they are set up as a default application which sends a selection screen to the end user. Once the end user specifies the destination application choice, the logon manager typically issues a CLSDST macro with OPTCD=PASS to the destination application. A new session is started with the destination application. The logon manager session is closed with a special UNBIND sent to Telnet indicating that a new session BIND is forthcoming. Telnet receives that special UNBIND and then waits for the next BIND instead of cleaning up as it would when receiving a normal UNBIND. When the end user logs off the destination application, Telnet will either redrive the

initial database lookup process or drop the connection, depending on whether or not LUSESSIONPEND is coded. Logoff of the original application will cause Telnet to perform normal close function instead of leaving the LU ACB open.

Many logon managers were written to support real terminals, not Telnet, and have the added function of issuing a SIMLOGON Q for the logon manager session itself immediately after issuing the CLSDST-PASS. When SIMLOGON Q is coded, the logon request is added to a VTAM queue. The first application queued is the first application off the VTAM queue. Immediately after user logoff from the destination application, VTAM (on behalf of the logon manager) will request a session with the terminal LU (or Telnet LU representing the client). This works very well for real terminals, but causes timing problems for Telnet when the logon manager is a default application. In this case, Telnet and VTAM both end up requesting a session.

The QSESSION option on ALLOWAPPL or RESTRICTAPPL can be used to correct this timing problem. When coded for the logon manager, Telnet will *not* do normal close processing when the UNBIND from the destination application arrives. Telnet will leave the LU ACB open and wait for the BIND from the logon manager that is generated because of the Queued SIMLOGON. When the BIND does arrive, Telnet will verify that the application name is the original logon manager and finish session setup.

You must ensure that the application defined as the QSESSION application will issue a SimlogonQ so a BIND will be sent after logoff from the destination application. Otherwise, Telnet will leave the LU ACB open waiting for a BIND that is never coming. The connection will be essentially hung. Only a Telnet expiration timer or client disconnect will clean up the connection.

The BEGINVTAM statement, QUEUESESSION, provides the same function with less control than the QSESSION option and will be made obsolete in a future release. QUEUESESSION affects all defined DEFAULTAPPLs whether or not all are queue session applications. If you are using QUEUESESSION, you should update your profile to use the QSESSION option instead.

As an example of the QSESSION parameter, assume that APPL1, APPL2, and APPL3 are each defined in VTAM. APPL1 will issue a SimlogonQ after CLSDST-PASS. The following Telnet statements allow connections to access the applications and define which is a QSESSION application.

```
ALLOWAPPL APPL1 QSESSION
ALLOWAPPL APPL*
```

The client first logs on to APPL1 and Telnet saves the name in the first slot of a 10 slot Qsession table. A CLSDST-PASS to APPL2 and a SimlogonQ are issued. The APPL2 name is saved in slot two of the table. Finally, a CLSDST-PASS to APPL3 is done and the APPL3 name is saved in slot three of the table. When the APPL3 session is ended, the APPL3 slot is cleared and VTAM sends an APPL1 BIND to Telnet. Telnet will start at the end of the Qsession table (APPL2 since the APPL3 entry was just cleared) and check each slot to try to find an application name match. Eventually a name match is made with slot one. Now, all slots above slot one are cleared. A CLSDST-PASS to APPL4 will cause the APPL4 name to be put into slot 2. Because the slots for APPL2 and APPL3 were cleared earlier, Telnet will no longer accept a BIND from APPL2 or APPL3 after ending the session with APPL4.

The Qsession table is set up when a BIND is received for an application that is defined with the QSESSION parameter. If the first application does not have QSESSION coded and a CLSDST-PASS is issued to an application name that has QSESSION coded, it is the second application that will be in slot one of the Qsession table. When the session in slot one is ended, Telnet will clean up the LU and the connection status will depend on whether or not LUSESSIONPEND is coded.

Disconnect on session error

The DISCONNECTABLE option on either the ALLOWAPPL or RESTRICTAPPL statement determines what type of session termination to send to the host VTAM application when Telnet initiates session termination. If DISCONNECTABLE is coded, Telnet issues a TERMSESS UNBIND(0F). Otherwise, Telnet issues a TERMSESS UNCOND. For example, when DISCONNECTABLE is coded for the TSO application, an unexpected connection loss results in an UNBIND(0F) being sent to TSO putting it in a reconnectable state. The DISCONNECTABLE parameter has no effect on a session ended normally by the end user logging off the session. The QSESSION parameter can be coded with DISCONNECTABLE on either statement.

Bypass RESTRICTAPPL with CERTAUTH

CERTAUTH is an option on RESTRICTAPPL used in conjunction with client authenticated secure connections or Express Logon. In both cases the client certificate is used to derive a user ID. If the end user chooses an application that is a RESTRICTAPPL, the normal Telnet response is to request a valid user ID and password before allowing access to the application. However, if the end user has been authenticated with a client certificate it may not be necessary to require a user ID/password. With the CERTAUTH option on RESTRICTAPPL Telnet will use the derived user ID. If the user ID is valid (listed on the RESTRICTAPPL statement), Telnet will bypass the end user solicitation and immediately give access to the application. The derived user ID value depends on the type of connection. If Express Logon is being used, the user ID is derived from the latest Client Certificate/Applid combination received from the client. If Express Logon is not being used, the user ID is the Client Identifier user ID derived from the Client Certificate from the SSL handshake.

Allow printer sessions with RESTRICTAPPL

The ALLOWPRINTER option enables printers to establish a session with an application defined as a RESTRICTAPPL. The TN3270E Telnet server verifies all session requests before finishing session setup, and if the application name is restricted by the RESTRICTAPPL statement, Telnet requires the user to enter a user ID and password before session initiation begins. The user of a printer emulator has no way to provide a user ID and password, and you might want to allow printer sessions without the user ID and password verification, mimicking the behavior of the ALLOWAPPL statement. The ALLOWPRINTER option on RESTRICTAPPL gives you the flexibility to keep the terminal LU sessions restricted while allowing all printer LU sessions access to the application, without user ID and password. In most cases, printer sessions originate from the application and not the user, keeping security control in the application. However, use this option with care if you have coded PRTDEFAULTAPPL or PRTMAP-DEFAPPL. In this case, it is possible for users to originate a printer session.

Keeping the ACB open

Some host VTAM applications are set up to inquire if a secondary LU is active. If the LU is active, the application will initiate a session. For this to work, the secondary LU must be active. The LUMAP option, KEEPOPEN, will cause the ACB

to open when the LU is assigned to the connection and remain open for as long as the LU is assigned to the client by this mapping statement.

LU assignment is different for TN3270E and TN3270 connections. TN3270E connections have an LU assigned during connection negotiation. TN3270 connections do not have an LU assigned until the VTAM application name is known.

When the host VTAM application initiates a session with the secondary LU, the host must issue an INQUIRE to see if the LU is active with an OPEN ACB. This INQUIRE will fail for Telnet LUs because Telnet does not open the ACB until a session request is sent from Telnet to VTAM. When the end user gets the solicitor (or USSMSG10) panel, Telnet has not opened the ACB of the LU assigned to a TN3270E connection. TN3270 and LineMode connections do not have LUs assigned yet. If the KEEPOPEN parameter is coded on the LUMAP statement used by Telnet to assign an LU during Early Lookup for a terminal TN3270E connection, Telnet will open the ACB before sending the solicitor (or USSMSG10) panel. At that time an end user can either log on to an application as usual or wait for a host application to INQUIRE about the LU and initiate a session. TN3270 connections will not have an LU assigned until an application name is known. When the name is known, an LU is assigned to the connection, the ACB of the LU is opened, and a session request is issued. If the KEEPOPEN parameter is coded on the LUMAP statement used by Telnet to assign the LU, the LU stays assigned to the connection with the ACB open until the connection is dropped. If profile statements define LUs uniquely to different applications, a second logon to a different application might fail. When KEEPOPEN is mapped to a connection, the MSG07 and LUSESSIONPEND functions are in effect whether or not they were explicitly coded. When a session is ended, the connection remains and the ACB remains open. Only a client disconnect, a Telnet error, or the KEEPINACTIVE/INACTIVE timers will cause a KEEPOPEN connection to be dropped. The KEEPINACTIVE timer is used whenever the Telnet LU is not in session with a VTAM application. Otherwise, the INACTIVE timer is used. See “Timers” on page 519 for information about ending idle KEEPOPEN connections.

Express Logon Feature (ELF)

The Express Logon Feature allows an end user to connect to an MVS host VTAM application without explicitly entering a user ID or password. Telnet uses the client certificate to resolve the user ID and RACF generates a temporary password, a passticket. ELF requires a secure connection with level 2 client authentication, a client that supports ELF, and RACF passticket setup.

The ELF function is activated by specifying the EXPRESSLOGON parameter. The function can be inactivated by specifying NOEXPRESSLOGON. Either parameter can be coded in TELNETGLOBALS, TELNETPARMS or PARMSGROUP. For a detailed discussion of ELF, refer to Appendix C, “Express Logon Feature (ELF),” on page 1195.

Device types and logmode considerations

The VTAM logmode defines many characteristics of the session established between the Telnet LU representing the client and the host VTAM application. For example, the logmode defines response types, presentation style, and the type of LU Telnet is emulating. LU0 (non-SNA) and LU2 (SNA) represent terminal LU types. LU1 (SCS) and LU3 (3270 Data) represent printer LU types.

Telnet matches a VTAM logmode to each client as it connects based on the client device type, unless the end user specifies a logmode on the USSMSG10 screen or

logmode is configured in the USS table mapped to the connection. Refer to *z/OS Communications Server: IP Configuration Reference* for default device type and logmode table information. The default terminal logmodes are non-SNA for TN3270 connections and SNA for TN3270E connections. At session request time, Telnet indicates to VTAM the desired logmode based on device type. The host VTAM application usually honors the request and binds the session using the requested logmode. However, depending on VTAM statements, the application can override the requested logmode and bind the session using different characteristics than Telnet requested. For this reason, some screen sizes might not work correctly even though the logmode defined in Telnet is correct. If the KEEPOPEN function is used to allow session initiation by the host application, the desired logmode must be coded on the DLOGMOD parameter as part of the VTAM application definition statement that defines the Telnet LU. Otherwise, the application will choose its own logmode.

Telnet processes the ATTN KEY request differently for non-SNA and SNA sessions. For non-SNA sessions (BIND FM value 02), Telnet converts the ATTN KEY request to a '6C'x data byte and sends it to the application. For SNA sessions (BIND FM value 03), Telnet converts the ATTN KEY request into a SNA signal and sends it to the application as expedited data. Some clients send both an ATTN KEY function code and a '6C'x data byte to ensure the ATTN is seen by the application. Telnet converts the ATTN KEY function into either a '6C'x data byte or a SNA signal and also forwards the '6C'x data. Some applications give unexpected results or Telnet might appear to not support ATTN when two ATTNs are received. The SINGLEATTN parameter causes Telnet to drop the second ATTN if it immediately follows an ATTN. The SINGLEATTN and NOSINGLEATTN parameters can be coded at all three parameter block levels for different levels of granularity.

To change either the TN3270 or the TN3270E logmode for a device type, use the TELNETDEVICE parameter. Whenever Telnet initiates the session request, Telnet will request that the logmode specified on the TELNETDEVICE statement be used for the session. The application (the primary LU) does have the ability to override the requested logmode and use a completely different logmode. The TELNETDEVICE parameter can be coded in all three parameter block levels for different levels of granularity. Coding TELNETDEVICE in a PARMSGROUP that is mapped on the LUMAP-DEFAPPL-PMAP statement enables the logmode to be LU and application specific.

If the application initiates the session, the TELNETDEVICE logmode has no affect on the session. For example, printer sessions are initiated by the application unless a printer default application is specified. The LUMAP-KEEPOPEN parameter can be used to open a terminal ACB and wait for the primary application to initiate the session.

Transform Linemode connections can have a unique logmode by coding TELNETDEVICE with a device type of TRANSFORM. Any logmode used must not support extended graphics.

The special case logmode NONE can be specified indicating that Telnet should not send any logmode request when initiating the session.

In the examples that follow, the first line causes only the 3270 logmode to change from the default to SNX32705. The second line causes both the 3270 and 3270E logmodes to change from their defaults to SNX32705 and SNX32702. The third line causes only the 3270E logmode to change from the default to SNX32702.

```
TELNETDEVICE 3278-5-E SNX32705
TELNETDEVICE 3278-5-E SNX32705,SNX32702
TELNETDEVICE 3278-5-E ,SNX32702
```

Using the Telnet Solicitor or USS logon panel

This section describes the Telnet Solicitor panel and Telnet Unformatted System Services (USS) support. All information needed to establish a session can be entered on the Telnet Solicitor panel. However, Telnet is often used as the primary method of connecting to the SNA mainframe environment. SNA end users are accustomed to entering abbreviated logon commands, specifying their own logmode, and entering user data from SNA terminals. The USS table itself can be used to customize information such as logmode before it is sent to VTAM. The logmode specified by the user or the USS table overrides the logmode specified by the TELNETDEVICE statement. For ease of migration, Telnet simulates SNA USS processing very closely. This simulation extends to being able to use the same assembled USS tables that are used by VTAM. VTAM-only character substitutions are ignored by Telnet and Telnet-only character substitutions are ignored by VTAM. Blanks are used in their place. To further extend the simulation of SNA terminals, Telnet also supports all of the INTERPRET table function.

Using the Telnet Solicitor logon panel

Telnet sends a Solicitor panel to the end user if one of the following is true:

- No DEFAULTAPPL, LINEMODEAPPL, USSTCP, or LUMAP-DEFAPPL mappings match the client's Client Identifier.
- The requested application is a RESTRICTAPPL.

Below is an example of the Telnet Solicitor panel:

```
Enter Your Userid:
Password:                          New Password:
Application:
```

Initial cursor placement can be specified. Where initial placement should be depends on client macros used and end user preferences. The OLDSOLICITOR parameter is used to implement this choice. The default cursor position is on the 'Application:' field. If OLDSOLICITOR is coded, the cursor is positioned on the 'Enter Your Userid:' field. The OLDSOLICITOR and NOOLDSOLICITOR parameters can be coded at all three parameter block levels for different levels of granularity.

In addition to satisfying RESTRICTAPPL, there are other times when an end user might want to use the user ID/Password fields. For example, the solicitor panel may also be used to change a password by entering the user ID, old password, and new password. The application field does not need to be filled in. If insufficient information was provided by the client (for example, a user ID but no password), then the Telnet Solicitor panel is returned with a message prompting for the required field. A message is also returned if the security program encounters an error when attempting to change the password. Example messages include:

- Password required
- Password is not authorized

Using the Telnet USS and INTERPRET support

The Telnet USS function provides the end user with a USSMSG10 logon panel similar to the logon panel used by native SNA terminals. The Telnet USS function

supports sending USSMSGs to the client, receiving and parsing USSCMDs from the client, and using a translation table defined in the USS table.

Telnet supports both 3270 format and SNA character stream (SCS) format USS tables. The SCS USS table name is optional on the USSTCP mapping statement. If it is provided, the SCS USS table is used for all TN3270E connections.

The CLEAR key is handled differently by the client depending on whether a 3270 or SCS format USS table is used. The CLEAR key has a special function for 3270 format USSMSGs. When the CLEAR key data stream is received by Telnet, pressing the CLEAR key refreshes the screen with USSMSG10 for any USSMSG other than USSMSG10. If the CLEAR key is pressed while at USSMSG10, the screen is cleared. Pressing the CLEAR key a second time refreshes the screen with USSMSG10. For SCS format USSMSGs, the CLEAR key does not provide the same special function. For SCS, the client processes the CLEAR key by clearing the screen and placing the cursor at the upper left corner with no data sent to Telnet.

USS data traffic is also handled differently depending on whether a 3270 or SCS format USS table is used. If a 3270 format USS table is used with a TN3270E connection that has negotiated BIND-IMAGE, a Telnet-generated Bind/Unbind encapsulates the USS traffic. If the connection is TN3270, BIND-IMAGE is not negotiated, or the SCS format USS table is used, Telnet does not encapsulate the USS traffic in a Bind/Unbind. If a 3270 format USS table is used, the Telnet data type for all traffic is 3270-DATA. If an SCS format USS table is used, the Telnet data type for all traffic is SSCP-LU-DATA. Regardless of USS table format, if the SYSREQ function is supported on a TN3270E connection and SYSREQ is received by Telnet, Telnet will accept a LOGOFF command in SSCP-LU-DATA format. If the client sends any command other than LOGOFF, COMMAND UNRECOGNIZED is returned in SSCP-LU-DATA format. If the client sends a second SYSREQ, the Telnet server reverts back to whatever the session state was prior to receiving the first SYSREQ.

USSCMD parsing also includes checking for INTERPRET table entries that might provide more function than USS tables alone can provide. Sample USS tables are in TCP/IP data sets SEZAINST(EZBTPUST) and SEZAINST(EZBTPSCS). A sample INTERPRET table is in TCP/IP data set SEZAINST(EZBTPINT). The 3270 format USS sample has been assembled, linked, and loaded into the product data set. The tables can be used by coding the USSTCP and INTERPTCP mapping statements in BEGINVTAM. For example, the statements below will map the sample tables to the client at IP address 1.1.1.1. See “Mapping Objects to Client Identifiers” on page 467 for mapping details.

```
USSTCP      EZBTPUST,EZBTPSCS  1.1.1.1
INTERPTCP   EZBTPINT  1.1.1.1
```

A new table can be created at any time and link-edited. Customized USS and INTERPRET tables can be created to change messages, commands, and translation tables. For example, messages can be changed to have non-English text or to have different syntax. Commands can be changed to accept different syntax or to have different default values. A VARY TCPIP,,OBEYFILE command will cause Telnet to load the new table with the new profile being processed. Any new connection using the new profile will be assigned the new table. Telnet also supports dynamic updating of same-name USS or INTERPRET tables. The VARY TCPIP,,OBEYFILE command adds the new version of the table to the new profile. New connections use the new copy associated with the new profile while old connections continue to use the old copy associated with the old profile.

USS table customization: Customized USS tables are used by both VTAM and Telnet, with any product-specific character substitutions converted to blanks. For example, @@SSCPNM is blank for Telnet and @@PRT is blank for VTAM. The tables must reside in a data set that is in the system's linklist or is in the STEPLIB statement of the TCP/IP startup procedure. Any changes to a Telnet USS table should be made with supplementary user-defined USS tables. The IBM-supplied USS table should not be changed as it provides a good example of coding most commands and messages. Telnet loads the first table found with the name EZBTPUST and defines it as the default USS table. If this table is not found, there is no default USS table. Whether or not a default USS table should be included depends on the desired message output. When writing a USS Message, Telnet searches the USS table mapped to the client first. If the message does not exist in the mapped table, Telnet searches the default table. If the message does not exist in the default table, Telnet writes USSMSG14. If no default table exists, Telnet generates a USSMSG14. For 3270 format USSMSGs, the end user can get back to the USSMSG10 from any message by pressing the CLEAR key. The default table does not affect the USS commands. The command entered must be in the mapped table or it is not recognized. The sample SCS USS table found in SEZAINST(EZBTPSCS) is not assembled and linked, and it is not loaded into Telnet as a default SCS USS table. The table must be assembled, linked, loaded, and mapped in the Telnet profile to be used.

Creating a USS table: The following macro instructions are used to create the USS table. Telnet USS function supports almost all VTAM session-level USS message and command definitions. Refer to *z/OS Communications Server: IP Configuration Reference* for macro details.

- USSTAB indicates the beginning of the USS table.
- USSCMD defines commands accepted by the Telnet server.
- USSPARM defines each operand or positional parameter that can be specified on the USSCMD macro instruction. It also defines default values for the operand or positional parameter. Multiple USSPARM macro instructions can be associated with a USSCMD macro instruction. For each operand or positional parameter code a USSPARM macro instruction.
- USSMSG defines messages sent from the Telnet server.
- USSEND indicates the end of the USS table.

Following are some of the more common rules to consider when coding a new USS table. Also, refer to the samples found in SEZAINST(EZBTPUST) and SEZAINST(EZBTPSCS) as a guide. The following section discusses general table rules.

- If a DEFAULTAPPL application is mapped at the same Client Identifier level as a USS table, or an LUMAP-DEFAPPL application is mapped, the USS table is used only to return error messages; it is also used optionally after the first session logoff. FIRSTONLY or LOGAPPL options on DEFAULTAPPL will cause Telnet to send a USSMSG10 after the first session logoff. DEFAULTAPPL without the FIRSTONLY or LOGAPPL options will cause Telnet to redrive to the default application after every session logoff.
- Both the 3270 data stream and the SNA character stream (SCS) formats are supported. For more information, refer to *3270 Data Stream Programmer's Reference* and the table samples.
- If a user-defined table is coded as part of another module, code an assembler EXTRN definition statement for the table name in that module so the table will be known externally and can be accessed by other modules.

Below are message related rules.

- 3270 format USSMSGs must contain the 3270 data stream write control characters (WCCs).
- All character substitutions (@@'s) substitute the same number of characters. Any character substitution that is VTAM-specific will be translated to blanks. If the substituted value is smaller, the field is padded to the right with blanks. The parameter LUNAME or SCAN must be coded on the USSMSG macro instruction for Telnet to perform character substitutions. For a complete list of character substitutions, refer to *z/OS Communications Server: IP Configuration Reference* for Telnet and *z/OS Communications Server: SNA Resource Definition Reference* for VTAM. Telnet supports multiple USSPARMs with the DATA keyword. This method can be used to pass multiple data parameters to the host application. For example, two DATA USSPARMs allow the end user to type 'TSO USER1 PROC001' and have both the user ID and the Procname passed to TSO as data.

Below are command related rules.

- LOGON command format

  PL1 - logon applid(tso) logmode(snx32702) data(user1)

  BAL - logon applid=tso,logmode=snx32702,data=user1
- Any application defined in a USSCMD macro instruction must also be specified on either an ALLOWAPPL or a RESTRICTAPPL statement in the Telnet profile.
- If the USS Command rules in *z/OS Communications Server: IP Configuration Reference* cannot be followed, use an interpret table to convert the character-coded command into a formatted SNA request.

Considerations when using mixed-case passwords: Mixed-case passwords can be used by applications if an SAF-compliant security product (such as RACF) has enabled this support. In some cases, the USS LOGON DATA parameter is used to send the password to the application. If a terminal user enters a mixed-case password on the USS LOGON command and it is translated to uppercase by the translation table, the logon will fail if the target application expects the password in mixed case.

The USS LOGON command is displayed on the terminal as it is typed; the password is displayed unless the 3270 format is used and the password is entered into a field with a non-display attribute. For additional security, inform the terminal user to stop entering the password as part of the USS LOGON. Instead, the application should prompt the terminal user for the password in a non-displayed field. If mixed-case passwords are used and the terminal user continues to enter the password as part of the USS LOGON command, the logon will fail when using TRANSLATE=YES (the default) on the USSPARM, because the password has been translated to uppercase.

If you want to continue allowing the terminal user to enter the password on the LOGON command, use one of the following methods to support mixed-case passwords:

- If the current USS translate table is used to set all characters to uppercase, the TRANSLATE=NO operand can be added to the USSPARM macros for the corresponding USSCMD, to prevent the specified DATA USSPARM containing the password from being translated to uppercase. For details, refer to *z/OS Communications Server: IP Configuration Reference*.

- Do not change the USS command and inform the terminal user to enter the DATA portion of the USS LOGON in single quotes. USS will not translate data within single quotes, and the quotes are removed before the data is passed to the application.
- If the terminal user is specifying only APPLID and DATA on the USS LOGON command, you can use an interpret table. The Telnet Interpret function passes all entered data without translation to the application. Specify REMOVE=Y on the LOGCHAR macro to remove the first non blank string from the entered data. Because the terminal user could enter the APPLID in lowercase, you should set up an interpret table that searches for both an uppercase and a lowercase APPLID.

With each of these methods, if the user ID is entered with the password, you must first verify whether the application supports translating the user ID to uppercase. A simple test is to enter the DATA portion of the USS LOGON in single quotes with the user ID specified in lowercase. USS will not translate data within single quotes and the quotes are removed before the data is passed to the application. If the logon fails, the application does not support translating the user ID to uppercase and the terminal user must enter the user ID in uppercase and the password in mixed case for the methods suggested.

INTERPRET table customization: The standard Telnet USS logon support should meet the needs of most installations. However, Telnet does support INTERPRET table function if special circumstances require accepting a sequence of characters outside the normal USS command format. For example, the end user might want to enter logon data that includes blanks. The INTERPRET table defines all entered data, including blanks, as a USSPARM DATA entry. The PL1 USSCMD format treats each blank as a parameter delimiter and cannot properly process a variable number of parameters. The INTERPRET table character sequences are scanned whenever the client is mapped to both a USS table and an INTERPRET table. Both must be mapped because the INTERPRET function is a subset of the USS function. INTERPRET is not a stand-alone function. The sample INTERPRET table found in SEZAINST(EZBTPINT) is not assembled and linked, and it is not loaded into Telnet as a default INTERPRET table. The table must be assembled, linked, loaded, and mapped in the Telnet profile to be used.

Creating an INTERPRET table: Telnet INTERPRET function supports all functions provided by the VTAM INTERPRET definitions. Refer to *z/OS Communications Server: IP Configuration Reference* for macro details. The following macro instructions are used to create an INTERPRET table:

- INTAB indicates the beginning of the INTERPRET table.
- LOGCHAR defines a single logon message and name of an application program.
- ENDINTAB indicates the end of the INTERPRET table.

Below are some of the more common rules to consider when coding a new INTERPRET table. Also, refer to the sample found in SEZAINST(EZBTPINT) as a guide.

- The LOGCHAR APPLID= supports APPLICID, ROUTINE and USERVAR.
- Code the most restrictive, or longest, LOGCHAR SEQNCE values first. Otherwise, unexpected matches can occur. The table is scanned from top to bottom until a match is found whether or not it is the most exact match. For example, assume sequence 'LOGA' is assigned APPL1 and any other 'LOG' sequence is assigned APPL2. If sequence 'LOG' is before 'LOGA', entry 'LOGA' will never be found even when the end user enters 'LOGA' because entry 'LOG'

will be the first match. All sessions will go to APPL2. The problem is corrected by putting 'LOGA' before 'LOG' in the table.

Assemble, link, and load a table: Use the sample JCL in SEZAINST(EZBUSJCL). In the sample, the USS table is in USER1.TABLES(USSTEST). It must be assembled and link-edited into the system's linklist or into a library concatenated as a STEPLIB in the TCP/IP startup procedure. In the sample, the table is link-edited into USER1.LINKLIB(USSTEST). The same procedure can be used for the INTERPRET table. Simply change the name of the input file source and the link-edit target member. The VTAM USS and INTERPRET macros used for the assemble can be found in *hlq*.SISTMAC1.

Timers

Several timers are available in Telnet to control how long connections stay up. The list includes:

- INACTIVE - How long a terminal connection can be idle with no SNA data traffic before the connection is dropped.
- PRTINACTIVE - How long a printer connection can be idle with no SNA data traffic before the connection is dropped.
- KEEPINACTIVE - How long a KEEPOPEN connection can be idle with no SNA session before the connection is dropped. When a KEEPOPEN connection is in session with a SNA application the INACTIVE timer is used instead of the KEEPINACTIVE timer.
- SCANINTERVAL - How often Telnet runs the list of connections looking for potentially lost connections. Because of the methodology, it also determines how long Telnet will wait for a TIMEMARK response before assuming the connection is lost.
- TIMEMARK - How long a connection is active without receiving any data before Telnet sends a TIMEMARK command which acts as an "are you there".
- SSLTIMEOUT - How long Telnet will wait for an SSL handshake initiation from the client before the request is dropped.

To facilitate these timers, Telnet records the time at which data is received from the client, received from VTAM, or sent to VTAM. Data received from the client is used by SCANINTERVAL/TIMEMARK to measure idle time on the connection. Data received from or sent to VTAM is used by the INACTIVE family of timers to measure idle time without SNA data traffic.

SSLTIMEOUT is different than the other timers. Telnet does not run this timer. The time value is passed to the SSL handshake process. If SSL does not get a response from the client within SSLTIMEOUT period of time, the handshake request fails. Telnet will then proceed to the next available connection negotiation method or drop the connection.

The INACTIVE family of timers

INACTIVE, PRTINACTIVE and KEEPINACTIVE all share one timer associated with a port profile to reduce system overhead. The timer with the smallest value defined in TELNETGLOBALS, TELNETPARMS or PARMSGROUP for that port profile is used to define how often the connections are checked.

For example, assume KEEPINACTIVE is defined as 1800, INACTIVE is defined as 3000, and PRTINACTIVE is defined as 5400 in a profile. The Telnet timer will run every 1800 seconds. Therefore, every time the timer pops Telnet will check each KEEPOPEN connection not in session to see if there has been a SNA session

created in the prior 1800 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-K. Telnet will also check each terminal connection to see if there has been any SNA data traffic in the prior 3000 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-S. Telnet also will check each printer connection to see if there has been any SNA data traffic in the prior 5400 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-P.

Setting KEEPOPEN to the smallest time was done as an example. Any of the three timers could be the smallest. Also, since all inactivity checks are done by one timer, a connection check could occur just before the connection would be considered timed out. The connection will remain active until the next check is made.

Using the example above, if a terminal connection has had no SNA activity in the past 2999 seconds when a check is made, the connection is not dropped. Another check is done 1800 seconds later and the connection is dropped, but the connection will have remained active for 4799 seconds instead of the specified 3000 seconds.

SCANINTERVAL and TIMEMARK

SCANINTERVAL and TIMEMARK are used together to determine if a connection has been lost. These parameters can be specified in TELNETGLOBALS, TELNETPARMS, and PARMSGROUP. The smallest SCANINTERVAL value is used to define how often the connections are checked. If TIMEMARK is smaller than SCANINTERVAL, TIMEMARK is set equal to SCANINTERVAL. Whenever data is received from the client, Telnet records the time. Telnet checks all connections at regular intervals defined by the SCANINTERVAL value. Each connection is checked to see if any data has been received from the client in the past TIMEMARK period of time. If not, a TIMEMARK command is sent to the client which acts as an "are you there" and Telnet remembers a TIMEMARK was sent to this client. During the next check at SCANINTERVAL time later, each connection is again checked to see if any data has been received from the client. If not, and a TIMEMARK was sent on the previous check, the connection is dropped with DEBUG SUMMARY message CONN DROP reason TIMEMARK.

For example, assume the values for SCANINTERVAL and TIMEMARK are 1800 and 10800, respectively. That means every 30 minutes all connections are checked to see if any data has been received in the last 3 hours. If not, a TIMEMARK is sent to the client. 30 minutes later Telnet checks the connections again. If the client responded to the TIMEMARK or sent in actual data of some type Telnet leaves the connection active. If nothing has been received Telnet drops the connection.

A SCANINTERVAL check could occur just before the last data received is old enough to trigger sending a TIMEMARK. The connection remains active until the next SCANINTERVAL is made. Then a TIMEMARK is sent, and at the next SCANINTERVAL the connection is dropped.

Using the example above, SCANINTERVAL checks a connection that received data 2 hours, 59 minutes ago. No TIMEMARK is sent. The next SCANINTERVAL runs 30 minutes later. Now the data received time is greater than 3 hours and a TIMEMARK is sent. At the next SCANINTERVAL, the connection is dropped. The connection's last activity was 3 hours 59 minutes ago.

Setting the timers

Caution must be used in setting these timers. Setting the INACTIVE family of timers or SCANINTERVAL timer too low could cause excessive CPU usage. Setting

the TIMEMARK value too low could also cause excessive flooding of the network with TIMEMARK commands or high storage usage. For example, these timers should take into account extended breaks such as lunch. If TIMEMARK is smaller than the lunch break time, the network may be flooded with TIMEMARK commands around the lunch hour. Be aware of the default values and be sure to set appropriate values for the situation.

Telnet diagnostics

In addition to general diagnostic tools such as CTRACE and dumps which are described in *z/OS Communications Server: IP Diagnosis Guide*, there are several Telnet-specific diagnostic tools available.

DEBUG messages

Telnet-specific debug messages can be turned on or off to diagnose Telnet problems. Several types of debug messages are available.

- Summary messages indicate important status changes, such as connection dropped.
- Detail messages indicate a problem was detected, such as LU is not available.
- General messages indicate an important event, such as takeover started.
- Trace messages show data to and from the client, and to and from VTAM.
- Flow messages show entry into modules and exit from modules.

Message generation is controlled with the DEBUG statement. The statement can be specified in TELNETGLOBALS, TELNETPARMS, or PARMSGROUP. A connection must have the DEBUG statement mapped to it for messages to be generated. Whenever a DEBUG message is generated, a CTRACE entry is also generated. The following options are available. DEBUG EXCEPTION is the default.

DEBUG OFF
Turns off all debug message generation.

DEBUG EXCEPTION
If a mapped connection is dropped due to error or inactivity conditions, a summary message is generated.

DEBUG SUMMARY
Summary messages are generated for all mapped connections.

DEBUG DETAIL
Detail, general, and summary messages are generated for all mapped connections.

DEBUG TRACE
Trace messages are generated for two connections. Detail, general, and summary messages are generated for all mapped connections.

DEBUG FLOW
Flow and trace messages are generated for two connections. Detail, general, and summary messages are generated for all mapped connections.

DEBUG PROFILE
Profile debug messages are generated for all profile statements or a subset of profile statements. DEBUG PROFILE can be specified only in TELNETGLOBALS. It has no affect on the other DEBUG statements.

Where the messages are routed is controlled by a second parameter on the DEBUG statement. There are three choices:

CONSOLE
With this option, the messages are assigned routing codes 2 and 8 for the master console and the teleprocessing console. This is the default routing for DEBUG EXCEPTION.

JOBLOG
With this option, the messages are assigned routing code 11. In many cases, the master console is set up to receive messages with routing code 11. To stop the messages from going to the master console, issue VARY CN(01),DROUT=(11), which drops routing code 11 from the master console. This is the default routing for DEBUG SUMMARY, DETAIL, and TRACE.

CTRACE
With this option, no messages are generated. Only the CTRACE entries are generated. This is the default for DEBUG FLOW and DEBUG PROFILE.

The DEBUG SUMMARY statement provides tracking of connection status. A summary message is written when:

- A connection request is accepted by Telnet.
- Connection negotiation is complete.
- A session is established with the host application.
- A session is dropped.
- A connection is dropped.

LU name, Connection ID, and client IP address and port are included in each message. In the example below an end user connects to port 23. The connection is negotiated as a TN3270E connection and a session with TSO is established. The session is dropped because of client disconnect (CLNTDISC) and then the connection is dropped because of client disconnect.

```
EZZ6034I TELNET CONN 00000011 LU **N/A**  ACCEPTED      23
  IP..PORT: 1.12.13.14..456
EZZ6034I TELNET CONN 00000011 LU TCPM1001 NEGOTIATED TN3270E
  IP..PORT: 1.12.13.14..456
EZZ6034I TELNET CONN 00000011 LU TCPM1001 IN SESSION TSO0001
  IP..PORT: 1.12.13.14..456
EZZ6034I TELNET CONN 00000011 LU TCPM1001 SESS DROP  CLNTDISC
  IP..PORT: 1.12.13.14..456
EZZ6034I TELNET CONN 00000011 LU TCPM1001 CONN DROP  CLNTDISC
  IP..PORT: 1.12.13.14..456
```

In addition to tracking major state changes and providing key information, the statements can be used for debug purposes. For example, an end user might be attempting a connection and something is not working. The ACCEPTED, NEGOTIATED, and IN SESSION messages are major connection milestones. Using the information provided and knowing whether or not these messages are displayed can provide many clues about the connection request. The SESS DROP and CONN DROP messages give a variety of reasons about why the drop occurred.

The DEBUG DETAIL statement may be needed if the DEBUG SUMMARY messages do not provide enough information to solve a problem. In addition to the summary messages listed above, DEBUG DETAIL will issue a message at the time of failure which displays the client IP address and port, connection ID, Telnet LU name, detecting module name, unique return code and brief explanation, and additional parameters if relevant. Some messages will be helpful to the system

administrator and others will be helpful to IBM service. Refer to *z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM)* message EZZ6035I for return code details.

In the example below the end user specified an application not in the Telnet profile and then disconnected at the client emulator.

```
EZZ6035I TELNET DEBUG DETAIL CLIENT
  IP..PORT: 1.12.13.14..456
  CONN: 00000011  LU: TCPM1001  MOD: EZBTPGLU
  RCODE: 3012-00  Application name is invalid.
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000

EZZ6035I TELNET DEBUG DETAIL CLIENT
  IP..PORT: 1.12.13.14..456
  CONN: 00000011  LU: TCPM1001  MOD: EZBTTRCV
  RCODE: 1001-02  Client disconnected from the connection.
  PARM1: FFFFFFFF PARM2: 00000461 PARM3: 00000000
```

The DEBUG EXCEPTION statement causes the CONN DROP message to be issued only for error conditions and inactivity reasons. DEBUG EXCEPTION is the default. If more than one connection is dropped for the same reason within 15 seconds, a single message with LU name MULTIPLE will be issued. For example, if MSG07 is not coded in the DEBUG DETAIL example above, the connection will be dropped after the lookup failure. The CONN DROP message will include the return code and indicate that an error caused the connection drop. The following message will be produced whether or not DEBUG was coded because of the error condition.

```
 EZZ6034I TELNET CONN 00000011 LU TCPM1001 CONN DROP  ERR 3012
   IP..PORT: 1.12.13.14..456          EZBTPGLU
```

The DEBUG TRACE statement generates messages containing data to and from the client, to and from VTAM, and to and from an LU name exit for a single connection. The TRACE option allows you to quickly see why a client is not connecting or why a session hangs. Once TRACE is added by using the VARY TCPIP,,OBEYFILE command, the first two clients assigned tracing will be the only clients traced. Another connection cannot be traced until one of these clients currently being traced is dropped.

```
EZZ6034I TELNET CONN 00000080 LU **N/A** ACCEPTED
  IP..PORT: 9.14.6.42..36484
EZZ6035I TELNET DEBUG CLIENT
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: TO  CLNT
  <-C- FFFD28
  PARM1: 00000003 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG CLIENT
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: FRM CLNT
  -C-> FFFB28
  PARM1: 00000003 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG CLIENT
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: TO  CLNT
  <-C- FFFA2808 02FFF0
  PARM1: 00000007 PARM2: 00000000 PARM3: 00000000
```

The DEBUG FLOW statement generates messages at the entry and exit of every Telnet module, in addition to generating the DETAIL and SUMMARY messages. This option should be used carefully, as it could flood the message console quickly if the connection is very active. The FLOW messages can be limited to a smaller, specified set of modules. If a module subset is specified, DEBUG FLOW messages

will only be issued from those modules. All DEBUG SUMMARY messages will be issued. If a large amount of FLOW data is required, it might be best to suppress the DEBUG messages by specifying the CTRACE option on the DEBUG statement. With the CTRACE option, the flow data will be in the CTRACE but not flood the console or joblog. CTRACE is the default option.

The sample below shows the module flow when a WONT_TIMEMARK is received from the client. If only EZBTPTDP and EZBTPOTM are of interest, the DEBUG statement would be written as follows:

```
DEBUG FLOW,EZBTPTDP,EZBTPOTM CONSOLE

EZZ6035I TELNET DEBUG FLOW CLIENT MOD ENTRY
  CONN: 00000024  LU: TCPM1001  MOD: EZBTTRCV
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG TRACE CLIENT 857
  CONN: 00000024  LU: TCPM1001  MOD: FRM CLNT
  -C-> FFFC06
  PARM1: 00000003 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG FLOW CLIENT MOD ENTRY
  CONN: 00000024  LU: TCPM1001  MOD: EZBTPTDP
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG FLOW CLIENT MOD ENTRY
  CONN: 00000024  LU: TCPM1001  MOD: EZBTPOTM
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG FLOW CLIENT MOD EXIT
  CONN: 00000024  LU: TCPM1001  MOD: EZBTPOTM
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG FLOW CLIENT MOD EXIT
  CONN: 00000024  LU: TCPM1001  MOD: EZBTPTDP
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
EZZ6035I TELNET DEBUG FLOW CLIENT MOD EXIT
  CONN: 00000024  LU: TCPM1001  MOD: EZBTTRCV
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000
```

The DEBUG PROFILE statement generates a statement message listing all words associated with the statement and a formatted control block message for each Telnet statement. This option should be used carefully, as it could flood the message console quickly if the connection is very active. The PROFILE messages can be limited to a smaller, specified set of statements. If a statement subset is specified, DEBUG PROFILE messages are only issued from those statements. If a large amount of PROFILE data is required, it might be best to suppress the DEBUG messages by specifying the CTRACE option on the DEBUG statement. With the CTRACE option, the flow data will be in the CTRACE but not flood the console or joblog. CTRACE is the default option.

The sample below shows data sent to the console for profile statements DEFAULTAPPL and LUMAP:

```
DEBUG PROFILE,LUMAP,DEFAULTAPPL CONSOLE

EZZ6035I TELNET DEBUG PROFILE INPUT,LINE:    67 MOD: EZBTMPRF
  PARM1: 00000002 PARM2: 00000000 PARM3: LUMAP
 LUGRP1 IPGRP1

EZZ6035I TELNET DEBUG PROFILE PARSED,LINE: *N/A* MOD: EZBTMPRF
  PARM1: 00000054 PARM2: 00000016 PARM3: LUMAP
       E3C5D3C3 00000000 00000054 00000043 |TELC............|
       00000004 00000016 00000000 D3E4C7D9 |............LUGR|
       D7F14040 40404040 40404040 40404040 |P1              |
       40404040 00000000 00000000 00000000 |    ............|
       00000000 C9D7C7D9 D7F14040 40404040 |....IPGRP1      |
       40404040                            |                |
```

```
EZZ6035I TELNET DEBUG PROFILE INPUT,LINE:    74 MOD: EZBTMPRF
  PARM1: 00000002 PARM2: 00000000 PARM3: DEFAULTAPPL
 APPL2 IPGRP1

EZZ6035I TELNET DEBUG PROFILE PARSED,LINE: *N/A* MOD: EZBTMPRF
  PARM1: 0000003C PARM2: 00000008 PARM3: DEFAULTAPPL
       E3C5D3C3 00000000 0000003C 0000004A |TELC...........¢|
       00000004 00000008 00000000 40404040 |............     |
       40404040 C1D7D7D3 F2404040 C9D7C7D9 |    APPL2   IPGR|
       D7F14040 40404040 40404040          |P1              |
```

The DEBUG OFF statement ensures that all debug messages are suppressed, including the exception CONN DROP messages.

Profile debug warning and error messages can be turned off by coding DEBUG OFF or DEBUG SUMMARY in the TELNETGLOBALS block and putting that block before any other Telnet blocks in the profile.

The DEBUG parameter may cause flooding of the operator's console. Console flooding concerns can be dealt with in several ways.

- Most DEBUG messages are, by default, assigned to routing code 11, the JOBLOG. The DEBUG option JOBLOG can be used for the same effect. However, the master console also receives routing code 11 messages by default. To stop the messages from going to the master console, issue VARY CN(01),DROUT=(11), which drops routing code 11 from the console. Another DEBUG option, CONSOLE, will direct the messages to the master console, routing code 2, and the teleprocessing console, routing code 8. The CTRACE option suppresses messages completely while continuing tracing at the debug level.
- If DEBUG messages are being used primarily for problem diagnosis, the VARY TCPIP,,OBEYFILE command can be used to keep the number of messages low. Bring up Telnet initially without DEBUG coded. When a problem appears, issue a VARY TCPIP,,OBEYFILE command for a Telnet profile that includes the DEBUG statement. Only new connections to the new profile will produce messages. Once data is obtained, issue another VARY TCPIP,,OBEYFILE command for a Telnet profile that omits the DEBUG statement.
- If the Client Identifier of the client having the problem is known, include DEBUG in a PARMSGROUP statement. Using PARMSMAP, map that group to the client. Debug messages for only that client will be issued.

The VARY TCPIP,,TELNET,DEBUG,OFF command can be issued to turn off DEBUG for all connections associated with all profiles, including the current profile. To turn on DEBUG again, issue a VARY TCPIP,,OBEYFILE command with the Debug option specified in the Telnet profile. Summary messages for CONN DROP due to errors or time-outs will also be suppressed. Use DEBUG EXCEPTION to retain these messages.

MSG07

The MSG07 parameter is very helpful when diagnosing problems. It allows Telnet to send a message to the client indicating an error occurred and what the error was. Something simple like a mistyped application name can be corrected by the end user without additional help. Even for more difficult problems, MSG07 provides a good starting point. It is recommended that MSG07 always be coded unless there are reasons not to send error messages to the client.

Abend trap

The VARY TCPIP,,TELNET,ABENDTRAP,*module,rcode,instance* command can be used to set up an abend based on the variables specified. Abendtrap has three variables:

module Is required. It is the module detecting the error. It can be wildcarded with asterisk (*) at the end. If a single * is used, any module reporting the specified return code will cause an abend. The module name ″OFF″ turns off an active trap.

rcode Is optional. It is the return code reported and cannot be wildcarded.

instance Is optional. It is the instance of the return code and cannot be wildcarded.

Below is an example setting the abend trap and then issuing a profile display to verify the trap is set. In the example, when EZBTTRCV reports an error code of 1001, Telnet will issue an abend with reason code ’3133’x. The state of the trap changes from ″ACTIVE″ to ″TRIPPED″. No more abends will be issued. Once tripped, the abendtrap command must be issued again to activate the trap. While the trap is active, the abend trap can be turned off by specifying ″OFF″ as the module name. An active trap cannot be changed directly. The current trap must be tripped or turned off before a new command is accepted.

```
V TCPIP,TCPCS6,T,ABENDTRAP,EZBTTRCV,1001
EZZ0060I PROCESSING COMMAND: VARY TCPIP,TCPCS6,T,ABENDTRAP,EZBTTRCV,1001
EZZ6013I TELNET COMMAND ABENDTRAP EZBTTRCV COMPLETE

D TCPIP,TCPCS6,T,PROF
EZZ6060I TELNET PROFILE DISPLAY
  PERSIS   FUNCTION    DIA  SECURITY  TIMERS  MISC
 (LMTGQAK)(OATSKQSWHT)(DRF)(PCKLECXN)(IKPSTS)(SMLT)
  -------  ----------  ---  --------  ------  ----
  LM*R*P*  **TSBQ*WHT  TJ*  SSH*ESX*  ***STS  SDD*
----- PORT:    23  ACTIVE           PROF: CURR CONNS:      1
------------------------------------------------------------
   ABENDTRAP       : EZBTTRCV  1001  FF  ACTIVE
4 OF 4 RECORDS DISPLAYED
```

SMF

SMF records are written when an end user establishes a session (SMF LOGN or TN3270 Server SNA Session Initiation record) and when the session is ended (SMF LOGF or TN3270 Server SNA Session Termination record). Optional SMF recording is controlled by using the SMFINIT and SMFTERM statements.

Tip: If Telnet is running as its own procedure, the host name will not be in the SMF record if TCPIPJOBNAME has not been specified. The SMF job name will be the job name of the Telnet procedure.

Two different record formats are available: SMF type 118 and 119. The type 119 records were first introduced in z/OS V1R2 Communications Server, and are controlled by use of the TYPE119/NOTYPE119 operands on the SMFINIT and SMFTERM statements. The subtypes cannot be changed for type 119 records and are set to the STD values. The use of the STD operand or the specification of a nonstandard subtype number on the SMFINIT and SMFTERM parameters control the usage of the older type 118 record processing. Data recorded includes the application name, Telnet LU name, client and host IP address and port, time of logon or logoff, and data count in and out. Combined with the SMF utility exit routine, SMF data can be used to track Telnet usage by a number of variables. If statements for both format types are coded then both record types are written. That capability should be used sparingly due to the additional processing costs

involved in generating both types records. Refer to *z/OS Communications Server: IP Configuration Reference* for SMF record layout.

TESTMODE

The TESTMODE parameter in TELNETPARMS allows a profile to be processed by Telnet but then dropped before it becomes an active profile. Using TESTMODE ensures that LU assignments, security levels, and other Telnet parameters are not compromised due to profile syntax errors.

DISPLAYS

Several displays are available which provide summary and detail information. Telnet provides profile-related displays to present the following information. The format of Object and Client Identifier displays has changed to accommodate IPv6 IP addresses. In an IPv4 environment, Telnet continues to use the existing format. If the FORMAT LONG statement in TELNETGLOBALS is specified or an IPv6 address is specified in the profile, the IPv6 display format is used. FORMAT LONG is set by Telnet in an IPv6 environment or if an IPv6 address is specified in the profile in any environment. For display output examples, refer to *z/OS Communications Server: IP System Administrator's Commands*.

General information:

- PROFILE - All parameter information, such as timers and security information.
- WLM - Summary of all workload manager names registered to represent Telnet.

Mapping information:

- Object - Various levels of Object information can be displayed.
  - A list of all Objects sorted by Object type can be generated by omitting the TYPE= and ID= parameters.
  - A detailed display of Objects showing how they are mapped to Client Identifiers can be generated by specifying the TYPE= parameter.
  - A detailed display of Objects showing how they are mapped to Client Identifiers and the details of the Object group can be generated by specifying the ID= parameter.
  - A detailed display showing where an Object is used and how those Objects are mapped to Client Identifiers can be generated by specifying the TYPE=WU and ID= parameters.
- Client Identifier - Various levels of Client Identifier information can be displayed.
  - A list of all Client Identifiers sorted by Client Identifier type can be generated by omitting the TYPE= and ID= parameters.
  - A detailed display of Client Identifiers showing how they are mapped to Objects can be generated by specifying the TYPE= parameter.
  - A detailed display of Client Identifiers showing how they are mapped to Objects and the details the Client Identifier group can be generated by specifying the ID= parameter.
  - A detailed display showing where a Client Identifier is used and how Objects are mapped to those Client Identifiers can be generated by specifying the TYPE=WU and ID= parameters. NULL Client Identifier mappings show up because all clients match the NULL Client Identifier.
- INACTLUS - Summary of all LUs that have been inactivated by the operator or by Telnet as a result of OPEN ACB or multilevel security problems.

Connection information:

- CONNECTION - Summary of all connections, filtered summary of all connections, or detailed data of a single connection.

The above displays give good snapshots of the state of Telnet and its connections. In addition, the WHEREUSED option is very useful as a lookup diagnostic tool. The display will show where any name or IP address is used within the profile. This is particularly helpful when a connection request does not proceed as expected. For example, a client is sent a USSMSG10 screen instead of immediately getting in session with a logon manager. A WHEREUSED display of the client's IP address might show, as expected, the IP address within an IP group that is mapped to the logon manager as a default application. However, another WHEREUSED display of the client's host name might show the host name within an HN group that is mapped to a USS table. Knowing that the selection order of Client Identifiers will cause the USS table to be chosen, the administrator will realize there are conflicting mapping statements for the client and will be able to resolve the problem.

Module information: Telnet module storage can be displayed to verify the level of maintenance. For example, symptoms indicate a problem that has been fixed by an APAR. It is not known for sure if the fixed module is in the current version of Telnet. The module in question can be displayed using the TCP/IP display STOR command. In the example below no APAR is listed, indicating it has not been applied to the system.

```
d tcpip,,stor,mod=ezbtvxrc
  EZZ8456I TCPIP MODULE STORAGE
    EZBTVXRC LOADED AT 0A220328 IN LOAD MODULE EZBTTMST
    +0000 47F0F026 20C5E9C2 E3E5E7D9 C340F0F0 *.00..EZBTVXRC 00
    +0010 4BF0F0F5 40F0F17A F2F97AF2 F240C8E3 *.005 01.29.22 HT
    +0020 C3D7F5F0 C1000DC0 58300224 58403150 *CP50A........ ..
  EZZ8459I DISPLAY TCPIP STOR COMPLETED SUCCESSFULLY
```

Tracing

If a problem is not resolved using the above tools, IBM service will likely need a CTRACE with option Telnet. CTRACE, with only the Telnet option, gives very complete information about the Telnet processes. To debug almost any Telnet problem no other CTRACE option is needed. Generally, the other options simply take up space creating a trace-wrap condition more quickly, especially if Telnet is running as part of the TCP/IP stack. Telnet running as its own procedure continues to use the SYSTCPIP component trace but has fewer trace options available as follows:

- ALL (excludes SERIAL, STORAGE, and TIMER)
- INIT
- MESSAGE
- MINIMUM (includes INIT and OPMSGS)
- NONE
- OFF
- OPMSGS
- SERIAL
- STORAGE
- TELNET
- TIMER
- WORKUNIT

A component trace SYS1.PARMLIB member is supplied for Telnet. The member name is CTIEZBTN. Trace setup for Telnet running as its own procedure is the same as for the TCP/IP stack, except for having fewer trace options available. For a complete description of the trace options and trace setup, refer to *z/OS Communications Server: IP Diagnosis Guide*.

If the problem is data related, use the FULLDATATRACE statement to trace all the data coming into and leaving Telnet rather than tracing only the first 64 bytes of data. FULLDATATRACE will cause a trace-wrap condition more quickly so it should be set only if needed. It should be set in PARMSGROUP instead of TELNETPARMS if a subset of clients can be identified.

For transform problems, the DBCSTRACE statement in TELNETPARMS or PARMSGROUP should be used to produce more trace entries in the SYSPRINT and TNDBCSER data sets.

For Telnet SNMP subagent problems, use the TNSATRACE keyword on the TNSACONFIG statement in PROFILE.TCPIP at startup. This will generate trace points throughout Telnet subagent processing, in addition to tracing data passed between the Telnet subagent, Telnet, and the TCP/IP stack. Subagent tracing can also be enabled after TCP/IP has been started by using the VARY TCPIP,,OBEYFILE command. To enable tracing using the VARY TCPIP,,OBEYFILE command, the subagent must first be disabled and then re-enabled with the TNSATRACE keyword. Trace data is written to the syslog daemon. Subagent tracing can be disabled using the NOTNSATRACE keyword.

WorkLoad Manager for Telnet (WLM)

Telnet can be part of a large sysplex environment where the server is replicated on many machines. One of the goals of such a system is to balance workload across the different machines. WLM is a method used to direct connection requests to various machines within the sysplex. See Chapter 13, “Domain Name System (DNS),” on page 595 for more details about WLM.

Restriction: WLM only supports IPv4 addresses. If Telnet WLM registration occurs on a stack that has only IPv6 addresses, WLM will not resolve the registered name.

Tip: If Telnet is running as its own procedure, TCPIPJOBNAME must be specified for WLM registration to work.

The WLMCLUSTERNAME statement in TELNETPARMS is used by Telnet to specify WLM registration names. For example, assume Telnet resides on three machines.

- Machine 1 supports CICS and TSO.
- Machine 2 supports CICS, TSO, and IMS.
- Machine 3 supports CICS and IMS.

The best approach to implement multiple applications with WLM registration is to create a separate port for each application. For example, CICS can be assigned to port 2023, TSO to port 3023, and IMS to port 4023. Each port number can be associated with a unique WLM clustername. Port 2023 can be set up for CICS and have a WLM name of TNCICS. Port 3023 can be set up for TSO and have a WLM name of TNTSO. Port 4023 can be set up for IMS and have a WLM name of TNIMS. By doing this, the WLM names are associated with ports rather than application names to a single port. This is important when multiple ports already

exist on a Telnet system. If a port has a problem and must be quiesced, the WLM name associated with the port will be deregistered, blocking any new requests from coming to the disabled port.

```
TELNETPARMS       ; Machine 1
  PORT 2023
  WLMCLUSTERNAME TNCICS ENDWLMCLUSTERNAME
ENDTELNETPARMS

TELNETPARMS       ; Machine 1
  PORT 3023
  WLMCLUSTERNAME TNTSO ENDWLMCLUSTERNAME
ENDTELNETPARMS

BEGINVTAM
  PORT 2023
  DEFAULTLUS  TCP0001..TCP0200..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL CICS
ENDVTAM

BEGINVTAM
  PORT 3023
  DEFAULTLUS  TCP0001..TCP0200..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL TSO
ENDVTAM

;*************************************************

TELNETPARMS       ; Machine 2
  PORT 2023
  WLMCLUSTERNAME TNCICS ENDWLMCLUSTERNAME
ENDTELNETPARMS

TELNETPARMS       ; Machine 2
  PORT 3023
  WLMCLUSTERNAME TNTSO ENDWLMCLUSTERNAME
ENDTELNETPARMS

TELNETPARMS       ; Machine 2
  PORT 4023
  WLMCLUSTERNAME TNIMS ENDWLMCLUSTERNAME
ENDTELNETPARMS

BEGINVTAM
  PORT 2023
  DEFAULTLUS  TCP0201..TCP0400..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL CICS
ENDVTAM

BEGINVTAM
  PORT 3023
  DEFAULTLUS  TCP0201..TCP0400..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL TSO
ENDVTAM

BEGINVTAM
  PORT 4023
  DEFAULTLUS  TCP0201..TCP0400..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL IMS
ENDVTAM

;*************************************************

TELNETPARMS       ; Machine 3
  PORT 2023
  WLMCLUSTERNAME TNCICS ENDWLMCLUSTERNAME
ENDTELNETPARMS
```

```
TELNETPARMS       ; Machine 3
  PORT 4023
  WLMCLUSTERNAME TNIMS ENDWLMCLUSTERNAME
ENDTELNETPARMS

BEGINVTAM
  PORT 2023
  DEFAULTLUS  TCP0401..TCP0600..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL CICS
ENDVTAM

BEGINVTAM
  PORT 4023
  DEFAULTLUS  TCP0401..TCP0600..FFFFNNN  ENDDEFAULTLUS
  DEFAULTAPPL IMS
ENDVTAM
```

With this setup, an end user can connect to host TNCICS, TNTSO, or TNIMS and will automatically be directed to the correct machine for load balancing. If a port is disabled by a Quiesce or Stop on one machine, the port will be deregistered from WLM so that no new connection requests will be received. It is important to set up the WLM names as links to port numbers rather than links to applications. By creating a port per application, end users think they are specifying an application but in reality are specifying a port number.

If more than one port is used for connectivity to the same application, then two different WLM names should be used for the two ports. Perhaps TNCICS and TNCICS1 could be used. If the same WLM name is used for both ports, a problem will occur if one of the ports is disabled. Because a second port is still active for the single WLM name, the name is *not* deregistered for the machine. WLM will continue to direct connection requests to the disabled port, where they will be rejected.

Configuring the z/OS UNIX Telnet server (otelnetd)

The z/OS UNIX Telnet server provides access to z/OS UNIX shell applications on the host using the Telnet protocol. The z/OS UNIX Telnet server lets hosts in an IP network log on to the z/OS shell environment directly, without going through TSO. The z/OS UNIX Telnet server supports AIX and UNIX full-screen applications such as the vi editor, so that AIX and UNIX users can use familiar Telnet commands. The z/OS UNIX Telnet server runs in both line mode and raw mode, but does not support TN3270 or TN3270E, as Telnet does.

Installation information

The HFS files used in the z/OS UNIX Telnet server and their locations in the HFS are as follows:

/etc/services

The ports for each application are defined here. For example:

```
otelnet xxxx/tcp
```

where *xxxx* is the port that inetd should listen on for otelnet.

/etc/syslog.conf

The configuration parameters for usage of syslogd are defined in this file. otelnetd writes to syslog facility local1.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/otelnetd
This is a symbolic link to /usr/lpp/tcpip/sbin/otelnetd. /usr/lpp/tcpip/sbin/otelnetd is a sticky-bit file. The OTELNETD member of SEZALOAD contains the executable code for the Telnet server.

/etc/banner
This file contains a login message which will be printed to the client's screen unless the -h option is specified. This banner should be stored here.

/etc/utmpx
This file is updated by the call to fsumoclp. It contains a list of all the users who are logged in with their associated tty.

/dev/ptypXXXX and /dev/ttypXXXX
These special device files represent pseudoterminals (ptys); they are used by the z/OS UNIX Telnet server and other programs.

Note: For information on allocating more of these files for more connections, see *z/OS UNIX System Services Planning*.

/usr/share/lib/terminfo
The descriptions of supported terminals are stored here. For more information, see *z/OS UNIX System Services Planning*.

/usr/lib/nls/msg/C/tnmsgs.cat
The message catalog used by the z/OS UNIX Telnet server is stored here.

If the message catalog does not exist, the software will default to the messages hard-coded within the software. These messages duplicate the English message catalog that is shipped with the product.

/usr/man/C/cat1/otelnetd.1
This file contains the associated manual (man) pages for the z/OS UNIX Telnet server. It provides online help for the user.

Environment variables

Table 26 provides a list of environment variables that can be explicitly set by z/OS UNIX Telnet.

Table 26. Environment variables for z/OS UNIX Telnet

Environment variable	Description
_BPX_SHAREAS	Controls whether a spawned child process is started in the same address space as the login shell.
KRB5_SERVER_KEYTAB	Specifies the location of the key table.
LC_ALL	Determines the values for all local categories.
LC_COLLATE	Determines the local category for character collation.

Table 26. Environment variables for z/OS UNIX Telnet (continued)

Environment variable	Description
LC_CTYPE	Determines the local category for character handling functions, such as tolower(), toupper(), and isalpha(). Also determines the interpretation of sequences of bytes of text data as characters (for example, single as opposed to multibyte characters), the classification of characters (for example, alpha, digit, graph), and the behavior of character classes.
LC_MESSAGES	Determines the local category for processing affirmative and negative responses, and the language and cultural conventions in which messages should be written.
LC_NUMERIC	Determines the local category for numeric formatting information (for example, thousands separator and radix character) in various utilities, as well as the formatted I/O operations in printf() and scanf() and the string conversion functions in strtod().
LC_TIME	Determines the local category for date and time formatting information. It affects the behavior of the time functions in strftime(). Additional semantics of this variable, if any, are implementation defined.
NLSPATH	Contains a sequence of templates that the catopen() function uses when attempting to locate message catalogs. Each template consists of an optional prefix, one or more conversion specifications, a file name, and an optional suffix.
TERMINFO	Specifies the path name for an unsupported terminal that has been added to the terminfo file. Use the TERMINFO variable in /etc/profile or /etc/.login.

Starting, stopping, and administration of z/OS UNIX Telnet

The z/OS UNIX Telnet server is started by inetd for each incoming Telnet connection. When the Telnet session is complete, the z/OS UNIX Telnet server will exit. Each active Telnet session will have a separate instance of the Telnet server which will communicate with the Telnet client.

The z/OS UNIX inetd daemon does not propagate environment variables other than PATH and TZ to its child processes, so the NLSPATH and LANG environment variables cannot be used to point to a different message catalog.

The following standards are supported:
- RFC 854 Telnet Protocol Specification
- RFC 855 Telnet Option Specification
- RFC 856 Telnet Binary Transmission
- RFC 857 Telnet Echo Option
- RFC 858 Telnet Suppress Go Ahead Option
- RFC 859 Telnet Status Option

- RFC 860 Telnet Timing Mark Option
- RFC 861 Telnet Extended Options - List Option
- RFC 885 Telnet End of Record Option
- RFC 1073 Telnet Window Size Option
- RFC 1079 Telnet Terminal Speed Option
- RFC 1091 Telnet Terminal type option
- RFC 1096 Telnet X Display Location Option
- RFC 1123 Requirements for Internet Hosts -- Application and Support
- RFC 1184 Telnet Linemode Option
- RFC 1372 Telnet Remote Flow Control Option
- RFC 1571 Telnet Environment Option Interoperability Issues
- RFC 1572 Telnet Environment Option
- RFC 2941 Telnet Authentication Option
- RFC 2942 Telnet Authentication: Kerberos Version 5
- RFC 2946 Telnet Data Encryption Option
- RFC 2952 Telnet Encryption: DES 64 bit Cipher Feedback
- RFC 2953 Telnet Encryption: DES 64 bit Output Feedback

When a z/OS UNIX Telnet session is started up, otelnetd sends Telnet options to the client side indicating a willingness to do the following:

- WILL ENCRYPT
- DO ENCRYPT
- DO TERMINAL TYPE
- DO TSPEED
- DO XDISPLOC
- DO NEW-ENVIRON
- DO ENVIRON
- WILL SUPPRESS GO AHEAD
- DO ECHO
- DO LINEMODE
- DO NAWS
- WILL STATUS
- DO LFLOW
- DO TIMING-MARK

The z/OS UNIX Telnet server can enable the following options locally.

- WILL BINARY

  This option indicates that the client is willing to send 8 bits of data, rather than the normal 7 bits of network virtual terminal data.

- WILL ECHO

  When the LINEMODE option is enabled, a WILL ECHO or WONT ECHO will be sent to the client to indicate the current state of terminal echoing. When terminal echo is not desired, a WILL ECHO is sent to indicate that Telnet will take care of echoing any data that needs to be echoed to the terminal, and then nothing is echoed. When terminal echo is desired, a WONT ECHO is sent to indicate that Telnet will not be doing any terminal echoing, so the client should do any terminal echoing that is needed.

- WILL LOGOUT

  When a DO LOGOUT is received, a WILL LOGOUT is sent in response and the Telnet session is shut down.

- WILL SGA

  This option indicates that it will not be sending IAC GA, the go ahead command.

- WILL STATUS

Indicates a willingness to send the client, upon request, the current status of all Telnet options.

- WILL TIMING-MARK

  Whenever a DO TIMING-MARK is received, a WILL TIMING-MARK is the response. It is only used in kludge linemode support.
- WILL ENCRYPT

  Indicates a willingness to encrypt the data stream.

The z/OS UNIX Telnet server can enable the following options remotely.

- DO BINARY

  Sent to indicate that Telnet is willing to receive an 8-bit data stream.
- DO ECHO

  If a WILL ECHO is received, a DONT ECHO will be sent in response.
- DO ENVIRON

  Indicates a desire to be able to request environment variable information. (See RFC 1408.)
- DO LFLOW

  Requests that the client handle flow control characters remotely.
- DO LINEMODE

  Supports requests that the client do line-by-line processing.
- DO NAWS

  Requests that the client inform the server when the window size changes.
- DO NEW-ENVIRON

  Indicates a desire to be able to request environment variable information. (See RFC 1572.)
- DO SGA

  Indicates that it does not need to receive IAC GA, the go ahead command.
- DO TERMINAL-TYPE

  Indicates a desire to be able to request the name of the type of terminal that is attached to the client side of the connection.
- DO TERMINAL-SPEED

  Indicates a desire to be able to request information about the speed of the serial line to which the client is attached.
- DO TIMING-MARK

  Only supported if the client responded with WONT LINEMODE. If the client responds with WILL TM, then it is assumed that the client will support kludge linemode. It is not used for any other purposes.
- DO XDISPLOC

  Indicates a desire to be able to request the name of the X Window System display that is associated with the Telnet client.
- DO AUTHENTICATION

  Indicates a willingness to receive authentication information for automatic login.
- DO ENCRYPT

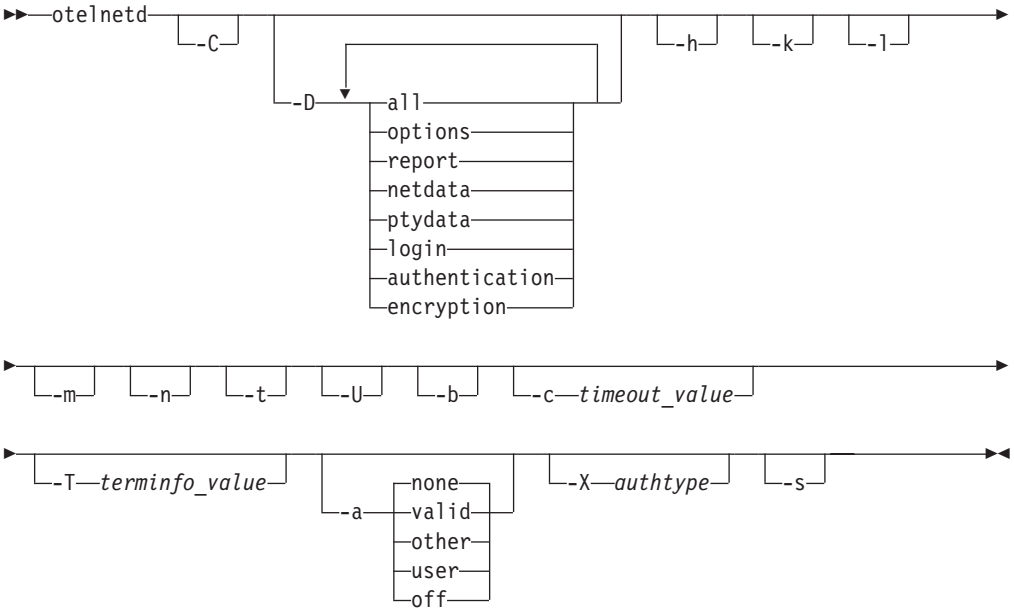
  Indicates a willingness to decrypt the data stream.

otelnetd

Note: The user ID associated with the daemon in /etc/inetd.conf requires superuser authority. Refer to *z/OS UNIX System Services Planning* for a description of the types of authority defined for daemons.

The following syntax is used in the /etc/inetd.conf file to define the arguments used to invoke otelnetd.

Syntax



Parameters

-C

Prints user messages in uppercase. There are several exceptions. Messages issued at startup are not affected by the -C option because the -C option is not processed during the startup. Also, data transmittal messages will not be uppercase. Data transmittal messages are generated from the -D netdata option or the -D ptydata option.

-D The following suboptions apply to -D:

options

Prints information about the negotiation of Telnet options. This information is used for debugging purposes. This suboption allows telnetd to generate debugging information to the connection, which allows the user to view telnetd activity.

report Prints the options information and additional information about processing. This information also includes print information designated for suboption=options. This can be used for debugging purposes. This suboption telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

netdata

Displays the data stream received by telnetd. This information is used

for debugging purposes. It allows telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

ptydata
Displays the data stream written to the pty. This information is used for debugging purposes. It allows telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

all Enables options, report, netdata, ptydata, login, authentication and encryption.

login Records login and logout activity to syslogd facility auth using message EZYTU36I.

authentication
Turns on authentication debugging code.

encryption
Turns on encryption debugging code.

-h Disables the display of the /etc/banner file at the user's terminal.

-k Disables kludge linemode. The server normally attempts to use kludge linemode when the -l option was specified, but the client does not support line mode. Use the -k option when there are remote clients that do not support kludge linemode, but pass the heuristic for kludge line mode support (for example, if they respond with WILL TIMING-MARK in response to a DO TIMING-MARK). This option does not disable kludge line mode when the client requests it. This is accomplished by the client sending DONT SUPPRESS-GO-AHEAD and DONT ECHO.

-l Specifies linemode, which tries to force clients to use linemode. If the LINEMODE option is not supported and the -k option was not specified, it will attempt to use kludge linemode.

Notes:

1. Many clients decline the server's request to operate in linemode.
2. Linemode is not appropriate for full-screen applications like the z/OS UNIX vi editor.

-m Enables the creation of a forked or spawned process to coexist in the same address space. This option can improve performance because the user's login shell runs in the same address space as otelnetd.

-n Disables TCP keep-alives. Normally, telnetd enables the TCP keep-alive mechanism to probe connections that have been idle for some time to determine if the client is still there. In this way, idle connections from machines that have crashed or can no longer be reached can be cleaned up. The cleanup of disabled connections is controlled by the presence of the KEEPALIVEOPTIONS statement in the TCPIP profile.

-t Specifies internal tracing. It also activates the REPORT option, as if the user also specified -D Report.

-U Causes telnetd to drop connections from any IP address that cannot be mapped back into a symbolic name by the gethostbyaddr or getnameinfo routines.

-b Forces the server to DO BINARY in the first pass during negotiations with the client.

-c *timeout_value*
Specifies the number of seconds to wait before terminating the Telnet session for inactive connections. The *timeout_value* is a value between 1 and 86400 seconds.

-T *terminfo_value*
Sets the TERMINFO environment variable to the specified values at startup. This option is needed when terminfo definitions are located in nonstandard directories.

-a This option may be used for specifying what mode should be used for authentication. There are several valid suboptions for authentication mode:

valid
Only allow connections when the remote user can provide valid authentication information to identify the remote user. Thus, for otelnetd, Kerberos authentication will be required. User verification will still occur through the login and password prompt. However, if the login user ID matches the name in the Kerberos principal, then no password will be requested. This is the most secure authentication mode.

other
Only allow connections that supply some authentication information. This option is currently not supported by any of the existing authentication mechanisms, and is thus the same as specifying -a valid.

user
Only allow connections when the remote user can provide valid authentication information to identify the remote user, and is allowed access to the specified account without providing a password. Thus, for otelnetd, Kerberos authentication is required. The NAME received during AUTHENTICATION option negotiation must match the name in the Kerberos principal and be a valid user ID on the host. No user verification will occur through the login or password prompt.

none
This is the default state. Authentication information is not required. User verification will still occur through the login and password prompt. However, if the login user ID matches the name in the Kerberos principal, then no password will be requested.

off
This disables the authentication code. All user verification happens through the login and password prompt. During option negotiation, otelnetd will not send DO AUTHENTICATION and, if necessary, will send DONT AUTHENTICATION.

Note: Authentication is not supported for IPv6 connections. If tcp6 is specified in inetd.conf, -a should not be used as a start option. If tcp6 and -a are both specified, the suboption will be overridden and forced to OFF.

-X *authtype*
This option disables the use of authtype authentication. Currently the only valid value for authtype is KERBEROS_V5. Thus, if otelnetd sends the AUTHENTICATION option SEND command, the authentication-type-pair-list will not contain any KERBEROS_V5 entries and will be empty.

-s Used to set the KRB5_SERVER_KEYTAB environment variable. If this environment variable is set, security runtime uses a local instance of the Kerberos security server to decrypt service tickets instead of obtaining the key

from a key table. To use this capability, the otelnetd application must have at least READ access to the IRR.RUSERMAP facility. For more information, refer to *z/OS Integrated Security Services Network Authentication Service Administration*.

SMF record handling

The SMF records generated are the typical set of records that MVS generates for start of job (login) and end of job (logoff). Additionally, interval records might be issued during the life of the user login. These records are SMF type 30 and type 72 and not the type 118 or type 119 in the current z/OS UNIX Telnet server. The process of issuing these records is external to the specific daemons.

BPX.DAEMON considerations

If the BPX.DAEMON facility class is defined, perform the following additional configuration steps:

1. Provide read access to BPX.DAEMON for the user ID specified in /etc/inetd.conf for otelnetd.
2. Define SEZALOAD to program control.
3. Define the C run-time library, *hlq*.SCEERUN to program control.

Refer to *z/OS UNIX System Services Planning* for more information about the BPX.DAEMON facility class, the security product commands used to perform the required configuration, and the diagnosis procedure for resolving related problems.

Kerberos

otelnetd supports Kerberos Version 5 for authentication on IPv4 connections. Authentication is not supported on IPv6 connections (that is, if tcp6 is specified for otelnetd in inetd.conf). On z/OS, Kerberos is implemented by Security Server. Please refer to*z/OS Integrated Security Services Network Authentication Service Administration* for more information.

The Kerberos principal used by otelnetd will generally be of the form "host/<*hostname*>@*realm*". That is, the first component of the Kerberos principal is "host"; the second component is the fully qualified lowercase hostname of the server; and the realm is the Kerberos realm to which the server belongs.

otelnetd will not accept forwarded credentials from the client.

Successful AUTHENTICATION option negotiation is required for successful ENCRYPT option negotiation. The ENCRYPT option must be negotiated in both directions.

Chapter 11. Transferring files using FTP

The File Transfer Protocol (FTP) allows a user to copy files from one machine to another. The protocol allows for data transfer between the client (the end user) and the server in either direction. In addition to copying files, the client can issue FTP commands to the server to manipulate the underlying file system of the server (for example, to create or delete directories, delete files, rename existing files, and so on.) FTP is the most frequently used TCP/IP application for moving files between computers.

Copying files from one machine to another is one of the most frequently used operations. The data transfer between client and server can be in either direction. The client can send a file to the server machine. It can also request a file from this server.

To access remote files, the user must identify himself or herself to the server. At this point the server is responsible for authenticating the client before it allows the file transfer.

From an FTP user's point of view, the link is connection-oriented. FTP uses TCP as a transport protocol to provide reliable end-to-end connections. Both hosts must run TCP/IP to establish file transfer.

The z/OS model for the FTP server includes a daemon process and a server process. The daemon process starts when you start your cataloged procedure (for example, START FTPD) and it listens for connection requests on a specific port. The port is the well-known port 21 unless otherwise specified. For methods of choosing a different port number, see "Configuring ETC.SERVICES" on page 543 and "Configuring the FTPD cataloged procedure" on page 544. When the daemon accepts an incoming connection, it creates a new process (server's address space) for the FTP server, which handles the connection for the rest of the FTP login session. Each login session has its own server process.

The server process inherits the accepted connection from the daemon process. This connection is called the control connection. The server receives commands from the client and sends replies to the client using the control connection. The control connection port is the same as the daemon's listening port.

The client and server use a different connection for transferring data; this connection is called the data connection. By default, the data port is one less than the control connection port. For example, if the control connection port is 21, the data port is 20. An FTP client can override the default data port by directing the server to run in passive mode. In passive mode, the server uses an ephemeral port for the data port. Passive mode is requested by firewall friendly clients and by clients initiating three-way data transfers.

Note: If you use the environment variable _BPX_JOBNAME when you start FTPD, the server's address space is known as the jobname specified in the _BPX_JOBNAME variable. Having a common naming convention for your installation's FTP address spaces may be needed if your installation uses syslogd isolation or has other workload management requirements.

If you do not use _BPX_JOBNAME, the server's address space assumes the name of the user. For example, if a user logs into an FTP server with the user ID of TCP0001, the FTP server address space servicing the request is also known as TCP0001.

If the FTP daemon accepts a connection that is protected by the TLS security mechanism and you are not using _BPX_JOBNAME, the server's address space name is a name derived from the FTP server jobname. The name is of the form *jobnamex*, where *jobname* is the jobname, and *x* is a numeral from 1 to 9.

Configuring PROFILE.TCPIP for FTP

If you have configured the FTP server to have affinity to a specific stack, or you have configured the FTP server to be a generic server in a single stack environment, the FTP server can be started automatically when the TCP/IP address space is started by specifying the name of the FTP server cataloged procedure in the AUTOLOG statement. If you have configured the FTP server as a generic server in a multiple stack environment, you should not use the AUTOLOG statement to automatically start the server. Instead, use some other automation outside of AUTOLOG to automatically start the server.

In the following example, if your procedure is called FTPD, the following statement allows TCP/IP to issue the MVS start command for procedure FTPD. The job name of FTPD1 will be used on the port statement shown below. If the daemon job name is fewer than eight characters, the FTP daemon forks a process that has the job name of the original daemon appended with the numeral 1.

```
AUTOLOG
  FTPD JOBNAME FTPD1
ENDAUTOLOG
```

To reserve ports 21 and 20 for the FTP server, add the following:

```
PORT
  21 TCP FTPD1             ; FTP server control port
  20 TCP OMVS NOAUTOLOG    ; FTP server data port
```

Specifying FTPD1 on the PORT and AUTOLOG statements directs TCP/IP to restart FTPD if it shound end.

To allow FTP to detect data connection errors when there has been no activity on the data connection for a certain amount of time, set the INTERVAL parameter on the TCPCONFIG statement to a relatively low value. The keepalive packets that the stack sends as specified on the INTERVAL parameter enable the stack to detect errors, such as a reset or terminated peer connection, instead of waiting indefinitely. Be careful when choosing an INTERVAL value on the TCPCONFIG statement because this value will affect all TCP connections at the host for which the interval has been activated, not just FTP connections.

The control connection can also benefit from keepalive packets. Many firewalls require periodic activity on any connection that is made and the control connection can appear idle during a long data transfer. Coding the INTERVAL parameter on the TCPCONFIG statement will, of course, cause keepalive packets to be sent on the control connection as well as the data connection. For FTP control connections only, you can override the keepalive interval you have configured in the stack with the FTPKEEPALIVE statement in FTP.DATA.

Optimally, FTP needs a buffer size of 180K for data connections. Setting TCPMAXRCVBUFRSIZE below 180K is not recommended. The default value for the parameter is 256K.

For more information about the AUTOLOG, PORT, and TCPCONFIG statements, refer to *z/OS Communications Server: IP Configuration Reference*.

If your FTP server accepts connections on a distributed Dynamic VIPA (DVIPA), SYSPLEXPORTS must be specified for the distributed DVIPA if either of the following are true:

- The DVIPA is an IPv6 address.
- Passive mode FTP is used.

For more information about the VIPADYNAMIC statement and specifying the SYSPLEXPORTS option on the VIPADISTRIBUTE parameter, refer to *z/OS Communications Server: IP Configuration Reference*.

Configuring ETC.SERVICES

The ETC.SERVICES file contains the relationship between service names (servers) and port numbers in the z/OS UNIX environment. If necessary, update your ETC.SERVICES file to include the control port that the FTP server is to use. For the search order used to locate the ETC.SERVICES file, see "Configuration files for TCP/IP applications" on page 38. For example, add the following:

```
ftp 21/tcp
```

Notes:

1. In the ETC.SERVICES file, only one port (the one for the control connection) is listed.
2. If the ETC.SERVICES file is changed such that a port other than 21 is specified, that port will become the FTP port for that z/OS host.
3. The port specified for FTP in the ETC.SERVICES file can be overridden by the FTP start parameter, PORT *nnnn*. In either case, the port that is specified should match the port specified for FTP on the PORT statement in PROFILE.TCPIP.

Configuring /etc/syslog.conf

Note: For FTP syslog, you should consider the fact that FTP writes log messages to the system console if syslogd is not running. If you enable FTP server traces without syslogd active, large amounts of data might be written to the system console.

The daemon.*priority* entries in /etc/syslog.conf determine where FTP messages and trace entries are written. The FTP server issues info, warning, and error messages. All trace entries are written with debug priority. To direct trace entries (and all messages) to /tmp/daemon.trace, include the following in /etc/syslog.conf:

```
*.*.daemon.debug    /tmp/daemon.trace
```

Log messages can be isolated within syslogd. For FTP, an installation might want FTP log messages to be written to different files depending on the user ID, or separately for the FTP daemon. If FTP messages are to be isolated for user1, use the first statement below. If FTP messages are to be logged for all the FTP applications, use the second statement below.

```
user1.*.daemon.debug /tmp/daemon.trace
```

```
*.FTPD*.daemon.debug /tmp/daemon.trace
```

In the above statement, it is assumed that _BPX_JOBNAME is set to FTPD.

Configuring the FTPD cataloged procedure

Update the FTP cataloged procedure FTPD by copying the sample in SEZAINST(FTPD) to your system or recognized PROCLIB and modifying it to suit your local configuration. The EXEC PARM=, SYSFTPD DD, and SYSTCPD statements must be updated.

See "Configuring FTP.DATA" on page 551 to configure SYSFTPD DD and "Configuring TCPIP.DATA for FTP" on page 550 to configure SYSTCPD DD.

The system parameters required by the FTP server are passed by the PARM parameter on the EXEC statement of the FTPD cataloged procedure. Add your parameters to `PARMS='` in the PROC statement of the FTPD cataloged procedure, making certain that each parameter is separated by a blank and all parameters are in uppercase

For example: `//FTPD PROC MODULE='FTPD',PARMS='TRACE ANONYMOUS PORT 21'` tells FTP to start up with TRACE active, ANONYMOUS support enabled, and use control PORT 21.

Security considerations for the FTP server

FTP uses the SERVAUTH System Authorization Facility (SAF) class to protect HFS access by FTP users. The use of SERVAUTH is optional. If a SERVAUTH profile is defined, any FTP users who require HFS access must be permitted to it. If the FTP.DATA for the server specifies STARTDIRECTORY HFS and the user does not have HFS access, FTP will make the starting directory the TSO user ID.

The profile name is of the following form:

```
EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS
```

For example, the profile name for FTP daemon FTPD running on system MVSA would be the following:

```
EZB.FTP.MVSA.FTPD1.ACCESS.HFS
```

The profile name can contain wildcards as allowed by the security product. All security product rules (for example wildcards, PROTECTALL, and so on) apply. For example, if all systems will use the same access list and RACF generic profile checking is active for the SERVAUTH class, the following profile name could be used:

```
EZB.FTP.*.FTPD.ACCESS.HFS
```

To restrict access to HFS resources for FTP users, the following RACF setup must be done by a user ID that has authority to issue the specified RACF commands:

- If not active, activate the RACF SERVAUTH class:

  ```
  SETROPTS CLASSACT(SERVAUTH)
  ```

- Define the profile for the FTP HFS ACCESS:

  ```
  RDEFINE  SERVAUTH   EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS
  ```

- Permit the user IDs that require HFS access to the profile:

  ```
  PERMIT EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS CL(SERVAUTH)
         ID(ftpuser)
  ```

- If the SERVAUTH class is not RACLISTed, RACLIST it:
  ```
  SETROPTS RACLIST(SERVAUTH)
  ```
- Refresh the SERVAUTH class before using:
  ```
  SETROPTS RACLIST(SERVAUTH) REFRESH
  ```

Consider the following for security:

- User IDs

  To log into the FTP server, a user ID must have a z/OS UNIX UID or may use the default UNIX UID.

  If the SAF class APPL is activated, a user ID should have READ access to the OMVSAPPL resource profile.
- The FTPD cataloged procedure must be:
  - Defined to the security program.
  - Added to the RACF started class facility or the started procedures table. The user ID associated with the FTP server started class must have a UID of 0. If the FACILITY class is active and the BPX.DAEMON or BPX.POE profiles are defined, the user ID associated with the FTP server must have READ access to them.
  - Given at least READ permission to the FACILITY class resource BPX.STOR.SWAP, if the daemon address space will be configured to run nonswappable.

  See SEZAINST(EZARACF) for more information on SAF resource requirements needed for FTP.
- Access controls
  - Stack access

    If the SERVAUTH class is active and a profile is defined for the EZB.STACKACCESS.mvsname.tcpname resource, the user ID associated with the FTP daemon and each user ID that logs in to FTP must have READ access to the profile.
  - Port access

    The FTP daemon listening port should be reserved for the FTPD job by a PORT statement in the TCPIP PROFILE. If the PORT statement for the FTPD port is protected with the SAF keyword, a SERVAUTH profile for the EZB.PORTACCESS.sysname.tcpname.SAFkeyword resource must be defined and the user ID associated with the FTP daemon must have READ access to it.
  - Network access

    The NETACCESS statement in the TCPIP PROFILE can be used to configure portions of your IP network into named security zones. Each defined security zone must have a SERVAUTH profile for the resource named EZB.NETACCESS.sysname.tcpname.zonename.

    The user ID associated with the FTP daemon must have READ access to the security zone that maps its bind address (0.0.0.0/32 for INADDR_ANY or ::/128 for IN6ADDR_ANY, unless overridden by PORT statement in the TCPIP PROFILE) and to every security zone that maps client IP addresses that FTP is to accept connections from.

    If the client IP address is mapped into a network access security zone, the login user ID must have READ access to the SERVAUTH profile for this resource. This resource can be used as the port of entry at login. This resource will be checked for all reads and writes to the control connection and any

data connections. It will also be checked during connect processing for an active data connection or accept processing for a passive (firewall friendly) data connection.

If the IP address the client connects to is mapped into a network access security zone, the login user ID must have READ access to the SERVAUTH profile for this resource. This resource will be checked during bind processing for active or passive data connections.

– Port of entry

  - Terminal access

    For IPv4 connection partners, the terminal ID passed from FTP to RACF is an 8-byte hexadecimal character string containing an IPv4 address. For example, the IP address 163.97.227.17 is translated to A361E311. RACF interprets this as a terminal login address and rejects it if no profile is defined for this resource name.

    Therefore, if the SETROPTS TERMINAL(NONE) setting is used in RACF, you must define profiles for the IP addresses in class TERMINAL to avoid problems when trying to FTP. You must translate all the IP addresses of any clients connecting to FTP to hexadecimal character strings and add them to the class TERMINAL. Login user IDs must have READ access to the TERMINAL profile covering their client system IP address.

    To allow access by all addresses in the 163.97.227 subnet:

    ```
    RDEFINE TERMINAL A361E3* UACC(READ)
    ```

    If your RACF SETROPTS options are TERMINAL(READ), all terminals are allowed access to your system, and you do not have to add extra resource definitions to your RACF data base.

    For IPv6 connection partners, it is not possible to represent the client's IP address as an eight character terminal ID. FTP uses the Servauth access port of entry support as follows.

  - Servauth access

    The FTP server can use the UNIX System Services _poe() service to identify the control socket as the port of entry. This support is available for both IPv4 and IPv6 connection partners. To activate this support for IPv4 connection partners, the parameter PORTOFENTRY4 SERVAUTH must be specified in the FTP.DATA file. This support is always active for IPv6 connection partners. If the connection partner's IP address is configured within a network access security zone, that SERVAUTH profile name is passed to SAF as the port of entry. For IPv4 connection partners not covered by a network access security zone, the _poe() service continues to pass the terminal ID described under *Terminal access* above. For IPv6 connection partners not covered by a network access security zone, the _poe() service does not pass a port of entry resource name and no port of entry check is performed. The authorizations described above under *Network access* are sufficient for servauth port of entry control.

  - Other resource access

    The port of entry resource passed to SAF at FTP user login is remembered by the system for the duration of the login session. Access to DATASET resources during that session can be further restricted by adding WHEN(TERMINAL=...) or WHEN(SERVAUTH=...) conditions to DATASET resource profiles in SAF.

For more information, see *z/OS UNIX System Services Planning* and *z/OS Security Server RACF Security Administrator's Guide*.

If you are planning to implement a multilevel security environment on your z/OS system, also see Chapter 4, "Preparing for TCP/IP networking in a multilevel secure environment," on page 145.

- Clients may use your server to send random data to other servers.

  Any FTP client in PROXY mode with your FTP server could establish a data connection to any server listening to a port. This could be very disruptive to that server, because the client could then send a very large amount of unexpected data to it. Any malicious FTP client can attack or disrupt the server in a normal server-to-client connection by making the FTP server send a large amount of data to another application server that is listening to a specific port. Since the client itself is not sending the disruptive data, it is difficult to identify the client that is causing the problem. PORTCOMMAND, PORTCOMMANDPORT and PORTCOMMANDIPADDR statements are provided in FTP.DATA to prevent your server from being used in this way.

Table 27. PORTCOMMAND scenarios

When you want your server to...	Code the following statements in the server's FTP.DATA	Comments:
Reject all PORT or EPRT commands	PORTCOMMAND REJECT	Disabling PORT or EPRT commands prevents your server from being used to send random data to other servers. However, your server loses some ability to transfer data in PROXY mode. If a client sends a PORT or EPRT command to your server to set up a proxy transfer, your server will reject the command and the proxy transfer fails. If your client is not firewall friendly, and it does not implement the default port number and IP address for data transfer, that client cannot transfer files to and from your server.
Reject all PORT or EPRT commands that specify well-known ports (port numbers less than 1024)	PORTCOMMANDPORT NOLOWPORTS	When you specify this combination, your server cannot be used to send random data to servers listening on well-known ports. However, a rogue client could use your server to send random data to servers listening on other ports. The server still supports data transfer in PROXY mode.
Reject all PORT or EPRT commands that specify an IP address other than the client's own IP address.	PORTCOMMANDIPADDR NOREDIRECT	When you specify this combination, a client can request data transfer in PROXY mode only between your server and a server on its own IP address. Transfers between client and server are not affected.
Reject all PORT or EPRT commands that specify an IP address other than the client's own IP address or port numbers that are well known.	PORTCOMMANDPORT NOLOWPORTS PORTCOMMANDIPADDR NOREDIRECT	When you specify this combination, a client can request data transfer in PROXY mode only between your server and a server on its own IP address, and the port numbers cannot be well known. The client cannot use PROXY mode to send random data to a server on its own IP address listening to a well-known port.

- Clients can request that your server be the passive server in a three-way transfer.

  When the client sends a PASV or EPSV command to the server, the server opens a listening data socket. This socket is similar to the listening socket associated

with the well-known port, in the sense that any application that knows its IP address and port number can connect to it (not just the client that sent the PASV or EPSV command). The client can exploit this to initiate a three-way data transfer, which is a data transfer between two servers. The client sends PASV to one server followed by PORT to the other. The client sets the PORT command IP address and port number to the information it gets from the PASV reply, and the second server connects to the IP address and port number specified in the PORT command, connecting the two servers. The next data transfer command causes data to move directly between the two servers. The client can also use the EPSV and EPRT commands to set up the three-way data transfer.

Three-way transfers are supported functions in the FTP protocol, but all installations might not want to allow their servers to participate in three-way transfers. To prevent your server from being the passive server (the server that receives the PASV or EPSV command) in a three-way data transfer, code PASSIVEDATACONN NOREDIRECT in the server's FTP.DATA file. This directs the server to verify that the data connection comes from the IP address where the original FTP client resides (the client that sent the PASV or EPSV command). If not, the server closes the data socket and the next data transfer command fails.

To completely disallow use of your FTP server in three-way transfers, code the PASSIVEDATACONN statement as described in the preceding paragraph, and the PORTCOMMANDIPADDR NOREDIRECT statement or PORTCOMMAND REJECT statement described in Table 27 on page 547.

- Using Generic Security Service Application Programming Interface(GSSAPI) to Authenticate Users

  GSS can be used to authenticate FTP clients to FTP servers. A client can attempt to authenticate to the server by sending a command specifying GSS as the authentication type. GSSAPI has Kerberos 5 as just one of many possible security services. For more information on setting up GSS support for the FTP server, refer to "Customizing Transport Layer Security (TLS) and Kerberos security" on page 558.

- Connection Security

  The FTP server and client support TLS. This support enables secure file transfer by providing data privacy, message authentication, and message integrity services for data sent and received using the FTP control and data connections. Optionally, an FTP client certificate that is authenticated during the TLS handshake can be used for end user identification and authentication in addition to user ID and password validation. For more information on setting up TLS support for the FTP server, refer to "Customizing Transport Layer Security (TLS) and Kerberos security" on page 558.

Defining environment variables for the FTP server (optional)

The FTP server optionally uses environment variables to identify the translate table data sets to be used for the control and data connections. These environment variables are used to override a default naming convention as described below. CCXLATE and XLATE statements will be ignored if EXTENSIONS UTF8 is specified in FTP.DATA.

_FTPXLATE_name used for translation

In your FTP.DATA file, you can use the CCXLATE or XLATE statements to specify a name that corresponds to a particular data set that is to be used for the initial translate tables for the control or data connections.

FTP will look for an environment variable defined as *_FTPXLATE_name=fully_qualified_dsn*, where *name* must be one to eight uppercase characters or numbers, and *fully_qualified_dsn* can be a fully qualified MVS data set name or HFS file name.

If the environment variable exists, FTP will use the data set name defined by the environment variable. If no such environment variable is defined, FTP will use the data set name *hlq.name*.TCPXLBIN.

Similarly, from any client you can issue SITE XLATE= to set the translate tables for the data connection for that particular FTP session. The FTP server will look for an environment variable called *_FTPXLATE_name*. If the environment variable does not exist, the server will look for a data set called *hlq.name*.TCPXLBIN.

Note: The CCXLATE and XLATE statements and SITE XLATE command are not case-sensitive, but the name of the optional environment variable is case-sensitive and must be in uppercase or FTP will not recognize it.

TZ and other UNIX environment variables

You can use the ENVAR runtime option in your FTPD start procedure to set environment variables for the FTP server. For information on using the ENVAR runtime option to set environment variables, see *z/OS XL C/C++ Programming Guide*. The following example shows how to specify environment variables in your FTPD started procedure:

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("TZ=EST")/&PARMS')
```

_BPX_JOBNAME

An installation that wants all FTP forked tasks to have similar job names needs to set the _BPX_JOBNAME environment variable. WorkLoad Manager (WLM), accounting, and isolation of syslogd messages are reasons an installation might not want to have each FTP logged-in user to have a job name of its user ID.

The following example sets all FTP forked() tasks to have the job name of FTPD:

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("_BPX_JOBNAME=FTPD"',
//           '"TZ=EST")/&PARMS')
```

_BPXK_SETIBMOPT_TRANSPORT for affinity to a specific stack

As discussed in "Generic server versus server with affinity for a specific transport provider" on page 73, if an installation wants to ensure that FTP has an affinity to a TCP/IP stack, the _BPXK_SETIBMOPT_TRANSPORT keyword should be used.

The example below sets the FTP server to have an affinity to TCPIPOE.

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE"',
//           '"TZ=EST")/&PARMS')
```

Configuring FTP with multiple TCP/IP stacks

Prior to configuring the FTP server with multiple TCP/IP stacks, review "Considerations for multiple instances of TCP/IP" on page 72.

The FTP server can be configured as a server with affinity to a specific transport provider or as a generic server.

To configure the FTP server to have affinity to a specific transport provider, do the following:

- Code the _BPXK_SETIBMOPT_TRANSPORT keyword in the FTP cataloged procedure. The example below sets the FTP server to have an affinity to TCPIPOE.

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE"',
//           '"TZ=EST")/&PARMS')
```

- Reserve ports 21 and 20 for the FTP server in PROFILE.TCPIP as follows:

```
PORT
    21 TCP FTPD1          ; FTP server control port
    20 TCP OMVS NOAUTOLOG ; FTP server data port
```

To configure the FTP server as a generic server, reserve ports 21 and 20 for the FTP server in PROFILE.TCPIP on all transport providers. The FTP server will detect when new transport providers are activated and attempt to bind to port 21. If this port is not reserved for the FTP server, the FTP server will end with the following message:

```
EZYFT13E bind error : EDC5111I Permission denied
```

Configuring TCPIP.DATA for FTP

The following five statements are used by the FTP server:

DATASETPREFIX
Specifies HLQ for dynamic allocation

DOMAINORIGIN
Specifies the domain name to be appended to host name

HOSTNAME
Specifies the TCP host name

LOADDBCSTABLES
Specifies the DBCS tables used by the client and server

MESSAGECASE
Specifies the case that messages should be displayed in

See Chapter 2, "Configuration overview," on page 11 for information about TCPIP.DATA or refer to *z/OS Communications Server: IP Configuration Reference* for information about these statements.

Configuring FTP.DATA

The FTP.DATA data set is optional. The FTP daemon looks for this data set during initialization, following this sequence:

1. A data set specified by the //SYSFTPD DD statement
2. *ftpserve_job_name*.FTP.DATA
3. /etc/ftp.data
4. SYS1.TCPPARMS(FTPDATA)
5. *hlq*.FTP.DATA data set

It is not necessary to include all statements in the FTP.DATA data set. Only include the statements if the default value is not what you want, because the default will be used for any statement not included in the FTP.DATA data set.

To pick up changes made in the FTP.DATA data set, the FTP server must be stopped and restarted. Some FTP server parameters can be changed during an FTP session by the client issuing the SITE subcommand. See *z/OS Communications Server: IP User's Guide and Commands* for more information. The FTP client has an FTP.DATA data set which can also be used to change the defaults for the FTP client local site parameters. See the *z/OS Communications Server: IP User's Guide and Commands* for more information about using the FTP.DATA data set for the FTP client local site parameters.

Optionally configuring user-level server options using FTPS.RC

The default values for the site parameters are coded in the server FTP.DATA. These SITE defaults apply to all login sessions to the server. You can customize settings for a specific user or group of users by creating an FTPS.RC configuration data set containing FTP commands specific to that login session. This file may contain a series of CWD and SITE commands. Refer to the *z/OS Communications Server: IP User's Guide and Commands* for information about these commands.

The FTP server uses the following search order to find the data set or HFS file:

1. tso_prefix.FTPS.RC
2. userid.FTPS.RC
3. $HOME/ftps.rc

Data set attributes

Data set attributes play a significant role in FTP performance. If your environment permits, tune both BLOCKSIZE and LRECL according to the following recommendations:

- Use half a track as the block size.
- For IBM 3380 DASD, use 23424 as the block size with an LRECL of 64 bytes.
- For IBM 3390 DASD or IBM9334, use 27968 as the block size with an LRECL of 64 bytes.
- Use FB as the data set allocation format.
- Use cached DASD controllers.
- If your environment permits, use a preallocated data set for FTP transfers into MVS.

The following configuration data statements apply to FTP server's allocation of data sets.

- AUTOMOUNT

- AUTORECALL
- BLKSIZE
- BUFNO
- CONDDISP
- DATACLASS
- DCBDSN
- DIRECTORY
- LRECL
- MGMTCLASS
- MIGRATEVOL
- PDSTYPE
- PRIMARY
- RECFM
- RETPD
- SECONDARY
- SPACETYPE
- STORCLASS
- UCOUNT
- UMASK
- UNITNAME
- VCOUNT
- VOLUME

Refer to *z/OS Communications Server: IP Configuration Reference* for more detailed information about these keywords.

Some of these allocation variables might provide duplicate information. FTP passes all variables that are specified to z/OS's dynamic allocation function and lets it determine which of the specifications take precedence. The only exceptions to this are the following:

- If the data set organization is physical sequential, directory blocks are not sent.
- If neither primary nor secondary space quantities are specified, the allocation units value is not sent.

For example, the model DCB (DCBDSN) might have a record format (RECFM) that differs from the record format specified by a data class and from the one explicitly specified by the client. The order of precedence for dynamic allocation variables are as follows:

1. Any FTP.DATA statements or SITE parameters explicitly specified or defaulted.
2. Any attributes picked up from the model DCB and not otherwise explicitly specified.
3. Any attributes picked up from the data class and not previously derived from 1 and 2 above.
4. Any system allocation defaults.

Specifying attributes for new MVS data sets

When allocating new data sets, there are two methods you can use to specify the data set attributes. You can customize the data set attributes for your login session using the SITE command, or you can configure the data set attributes for logging

in to your server using statements in FTP.DATA. Or, if your system programmer has used the Storage Management System to group together default attributes into named classes, you can specify those class names on the DATACLASS, STORCLASS, and MGMTCLASS statements.

Dynamic allocation

The FTP server allows a client program to dynamically allocate a new physical sequential data set, a partitioned data set (PDS), or a partitioned data set extended (PDSE), for the purpose of transferring data to be written to that data set. The following optional allocation variables can be used to override and turn off the defaults that affect the allocation of the data set.

Variable	**FTP.DATA statement**
allocation units	SPACETYPE
blocksize	BLKSIZE
data class	DATACLASS
directory blocks	DIRECTORY
logical record length	LRECL
management class	MGMTCLASS
model DCB values	DCBDSN
PDS type	PDSTYPE
primary space	PRIMARY
secondary space	SECONDARY
unit count	UCOUNT
volume count	VCOUNT
record format	RECFM
retention period	RETPD
storage class	STORCLASS
unit	UNITNAME
volume serial number or list	VOLUME

Some of these allocation variables might provide duplicate information. For example, the model DCB might have a record format (RECFM) that differs from the record format specified by a data class and from the one explicitly specified by the client. FTP passes all variables that are specified to dynamic allocation and lets it determine which of the specifications take precedence. The following list describes the exceptions to that policy:

- If neither the primary nor secondary space quantity is specified, the allocation units value is not sent.
- If the data set organization is physical sequential, directory blocks specification is not sent.
- Otherwise, all variables are sent to dynamic allocation where the order of precedence is:
  1. Any FTP.DATA statements or SITE parameters explicitly specified or defaulted
  2. Any attributes picked up from the model DCB and not otherwise explicitly specified
  3. Any attributes picked up from the data class and not previously derived from 1 or 2
  4. Any allocation defaults

Storage Management Subsystem (SMS)

You can specify one or more of the following SMS classes to manage characteristics that are associated with or assigned to data sets.

- Data class is an SMS construct that an installation can define to control data set allocation attributes used by SMS for the creation of data sets. An installation can override all or part of an SMS DATA CLASS definition by using FTP.DATA statements. Note that there is an order of precedence for dynamic allocation. (See "Data set attributes" on page 551 for more information on the precedence.) The fields listed are available attributes that serve as a template for allocation. Each is *optional* and is overridden by any explicit specification of FTP allocation variables or by a model DCB (DCBDSN).

Variable	FTP.DATA statement
directory blocks	DIRECTORY
logical record length	LRECL
primary space	PRIMARY
record format	RECFM
retention period	RETPD
secondary space	SECONDARY
pds type	PDSTYPE

  Note: If either primary or secondary space is explicitly specified, the primary and secondary values from data class are not used.
- Management class (MGMTCLASS) is an SMS construct that determines DFHSM action for data set retention, migration, backup, and release of allocated but unused space. Management class replaces and expands attributes that otherwise would be specified. That is, management class might override any other specification of retention period.
- Storage class (STORCLASS) is a list of storage performance and availability services requests for an SMS-managed data set that SMS attempts to honor when selecting a volume or volumes for the data set. It might conflict with an explicit specification of volume and unit. If storage class is used, volume and unit should not be specified.

Translation of data

Selecting an appropriate translate table for conversion of data from host to network format, and from network to host format, will ensure that data read from or written to the z/OS system are in correct format. The following statements apply to translation of data for the FTP server. Refer to *z/OS Communications Server: IP Configuration Reference* for more information on these statements. The statements are:

- ASATRANS
- CTRLCONN
- DBSUB
- ENCODING
- EXTENSIONS UTF8
- MBDATACONN
- MBSENDEOL
- SBDATACONN
- SBSENDEOL
- SBSUB
- SBSUBCHAR
- UCSHOSTCS
- UCSSUB
- UCSTRUNC

FTP code page conversion

Code page conversion must be performed for:

- FTP commands and replies sent over the control connection
- Data transferred over the data connection

FTP uses the *iconv* function to establish ASCII-to-EBCDIC and EBCDIC-to-ASCII translate tables for the control connection. The default network transfer code page for the control connection is 7-bit ASCII. In addition, FTP maintains support for the use of translate tables by the CONVXLAT utility. After an end user has logged in, a SITE subcommand can be used to change the code page being used on the control connection.

FTP uses the *iconv* function to establish network transfer to file system and file system to network transfer translate tables for the data connection. In addition, FTP maintains support for the use of translate tables by the CONVXLAT utility.

Note: Using iconv conversion to retrieve EBCDIC data that was created with CONVXLAT-generated conversion tables could result in data corruption due to possible conversion table differences.

After an end user has logged in, SITE and LOCSITE subcommands can be issued to change the translation tables being used for single-byte translation.

Code page conversions for the control connection

For the control connection, FTP generally uses ASCII for the network code page, as specified in the FTP RFCs. For the host/ASCII conversion for the control connection, FTP uses either iconv() or the support for single-byte translation tables. However, when EXTENSIONS UTF8 is coded in FTP.DATA, FTP starts the connection in 7-bit ASCII and negotiates a switch to UTF-8 encoding of the control connection, as described in RFC 2640. FTP uses iconv() for the host/UTF-8 conversion.

Priority: The priority for establishing the conversion tables used for the control connection is:

1. FTP start parameter (FTP client only)
2. EXTENSIONS UTF8 coded in FTP.DATA
3. CTRLCONN or CCXLATE keyword in FTP.DATA
4. Search order used to locate a TCPXLBIN data set:
   a. Original jobname.SRVRFTP.TCPXLBIN
   b. *hlq*.SRVRFTP.TCPXLBIN
   c. Original jobname.STANDARD.TCPXLBIN
   d. *hlq*.STANDARD.TCPXLBIN
5. 7–bit ASCII
6. Internal (hard-coded) 7–bit tables

Code page conversions for the data connection

For the transfer of data on the data connection, FTP supports:

- All single-byte conversions available through iconv. For example, iconv supports conversions between IBM-1047 and IBM-850, so conversions between IBM-1047 and IBM-850 are available for data transfer.

- Multibyte conversions for the Chinese standard GB18030 using code page IBM-5488 with code page IBM-1388 or UTF-8, as well as certain double-byte character set (DBCS) code page pairs that are equivalent to supported DBCS languages.
- Both single-byte and double-byte data conversions are supported with the translate tables provided with TCP/IP or generated by the CONVXLAT utility.

Priority for single-byte conversions: The priority that determines how to build single-byte translate tables for converting network transfer data and file system data is the following:

- SYSFTSX DD statement in the startup procedure, where the named data connection is CONVXLAT-generated translate tables. The data set can be an MVS data set or an HFS file.
- SBDATACONN or XLATE keyword in FTP.DATA.
- Search order to locate a TCPXLBIN data set, where the MVS data set contains CONVXLAT-generated translate tables:
  1. Original jobname.SRVRFTP.TCPXLBIN
  2. *hlq*.SRVRFTP.TCPXLBIN
  3. Original jobname.STANDARD.TCPXLBIN
  4. *hlq*.STANDARD.TCPXLBIN
- The same conversions established for the control connection.

Multibyte character sets (MBCS) support: MBCS support in the FTP server is provided for the following code page pairs:

Support for:	TYPE command	File system code page	Network transfer code page
Chinese standard GB18030	Not applicable	IBM-1388 or UTF-8	IBM-5488
BIG5	TYPE B 8	IBM-937	IBM-950 or BIG5
EUCKANJI	TYPE B 2	IBM-930	IBM-eucJP
JIS78KJ (JISROMAN	TYPE B 4 R	IBM-930	IBM-5053
JIS78KJ (ASCII	TYPE B 4 A	IBM-939	IBM-5055
JIS83KJ (JISROMAN	TYPE B 3 R	IBM-930	IBM-5052
JIS83KJ (ASCII	TYPE B 3 A	IBM-939	IBM-5054
KSC5601	TYPE B 6	IBM-933	IBM-949
SCHINESE	TYPE B 9	IBM-935	IBM-1381
SJISKANJI	TYPE B 1	IBM-930 or IBM-939	IBM-932 or IBM-eucJC
TCHINESE	TYPE B 7	IBM-937	IBM-948

Guidelines: The following restrictions and limitations exist::

- ENCODING must be specified as MBCS, either in FTP.DATA or on a SITE command.
- The data type must be ASCII.
- The file structure must be FILE (not RECORD), and the transfer mode must be STREAM (not BLOCK or COMPRESS).
- The FTP file type must be SEQ (not JES or SQL).

- If the file is transferred to or from an MVS data set, the record format of the data set must be V, VB, or U.
- If a file is transferred outbound and is an MVS data set with record format V or VB, the request for RDWs is not allowed.
- Translation of ASA or machine control characters is not allowed.
- MBCS can be used as a migration path for the DBCS languages listed that have an associated TYPE B *x* command.

  Restriction: The DBCS languages can be migrated to the MBCS support only if they do not use the following parameters on the TYPE B *x* command:

  S A SOSI ASCII characters X'1E' and X'1F' in the ASCII data stream

  S E SOSI EBCDIC characters X'0E' and X'0F' in the ASCII data stream

  S S SOSI SPACE characters X'20' and X'20' in the ASCII data stream

  N No SOSI characters in the ASCII data stream and none written to the file system

Master catalog access

FTP uses the IGGCSI00 function to request catalog processing. This accesses both the user and master catalog. Users require read access to the master catalog as well as their own user catalog.

Accounting

The following parameters apply to SMF data:

- SMF
- SMFAPPE
- SMFDEL
- SMFEXIT
- SMFJES
- SMFLOGN
- SMFREN
- SMFRETR
- SMFSQL
- SMFSTOR

Refer to *z/OS Communications Server: IP Configuration Reference* for more information on these statements.

Configure the FTP server for SMF (optional)

The FTP server can write SMF type 118 (X'76') or type 119 (X'77') records to record transactions made by the FTP server. SMF records can be written for the following commands:

- APPE (append)
- DELE (delete)
- RNTO (rename)
- RETR (retrieve)
- STOR (store)
- STOU (store unique)

Information about the previous commands can be recorded for:

- FTP server running in normal data transfer mode (FILETYPE=SEQ)
- FTP server running remote job submission (FILETYPE=JES)
- FTP server running Structured Query Language (SQL) queries (FILETYPE=SQL)
- Any combination of SEQ, JES, and SQL

For commands involving data transfer (APPEND, RETR, STOU or STOR) an SMF record will be written for both successfully and unsuccessfully completed data transfer commands which have begun data transfer. For data transfer commands which have completed unsuccessfully, the byte count of transmission field will contain the number of bytes transferred before the failure, and the recent server reply field will contain the 3-digit error reply code sent to the client. Refer to the *z/OS Communications Server: IP Configuration Reference* to find the particular offsets for the record type being used.

The FTP server can also write SMF records when a login attempt fails.

The capability also exists for a user-written exit routine to get control before the SMF records are written. See "Configuring the optional FTP user exits" on page 572 for more information.

If you want the FTP server to write SMF type 118 (X'76') or type 119 (X'77') SMF records, you must include at least one of the SMF subtype statements (SMF, SMFAPPE, SMFDEL, SMFLOGN, SMFREN, SMFRETR, or SMFSTOR) in the FTP.DATA data set.

If SMF subtype statements are not coded in the FTP.DATA data set, no SMF records are written by the FTP server.

Customizing Transport Layer Security (TLS) and Kerberos security

The following terms apply to TLS and Kerberos.

Integrity protected, data integrity, or data authentication
Indicates an algorithm is applied to the data being transferred, which modifies the data such that the receiving program can verify the data was not modified or changed during the transfer.

Privacy protected
Indicates an algorithm is applied to the data being transferred, which encrypts or scrambles the data such that only the receiving program can use a special key to decrypt or unscramble the data to its original format. The original data cannot be seen or interpreted while the data is in transit.

Raw Indicates data is transmitted without being modified by any encryption or data integrity algorithms.

Encipher or cipher algorithm
Data being transferred is encrypted, integrity protected, or both. This term does not imply which algorithm is used and does not imply the data is encrypted.

Steps for customizing the FTP server for TLS

Before you begin: You should understand the following:

- The FTP server can be enabled to support both TLS and Kerberos. Some of the configuration statement settings apply to both TLS and Kerberos and affect the behavior of both.

- To support TLS, the FTP server always provides server certificate authentication to all the clients to validate that the server is what it says it is. Therefore, a server key ring database is required to contain at least the FTP server's digital certificate and private key. For more information about key ring databases, see Appendix B, "TLS/SSL security," on page 1167.

Perform the following steps to customize the FTP server for TLS:

1. Code the following statement in the server's FTP.DATA configuration file to enable the server for TLS:

   ```
   EXTENSIONS AUTH_TLS
   ```

   ---

2. Decide what level of authentication you will use for TLS sessions:
   - Server authentication only
   - Client authentication level 1
   - Client authentication level 2
   - Client authentication level 3

   For more information about server authentication and client authentication, see "Secure Socket Layer overview" on page 1167.

   ---

3. Create the server key ring database and add the certificates you will need to the server key ring database.

   For information on how to create a key ring database and add certificates to that database, see "Overview" on page 1173.

   Every TLS session handshake includes server authentication, so you must always add a certificate for this server to the server key ring database. If a server certificate is self signed, you must also export that certificate to the key ring databases of those clients that will log in using TLS. If a server certificate is signed by a certificate authority (CA), the CA certificate used to sign the server certificate needs to be in the client key ring databases, rather than the server certificate. For more information about server authentication, see "Server authentication" on page 1168 and "Overview" on page 1173.

   If you are using client authentication and self-signed certificates, you must import the client certificates into the server key ring database. If a client certificate is signed by a CA, the CA certificate used to sign the client certificate needs to be in the server key ring database, rather than the client certificate. For more information, see "Client authentication" on page 1169 and "Overview" on page 1173.

   ---

4. To configure the FTP server with the name of the key ring database, code the following statement in FTP.DATA:

   ```
   KEYRING server-keyring-database
   ```

   For information about the KEYRING statement, refer to *z/OS Communications Server: IP Configuration Reference*.

   ---

5. Decide which cipher algorithms the server should use to encipher data transfers and to encipher control information.

   FTP supports TLS through the system SSL cryptographic services base element of z/OS. System SSL supports multiple cipher algorithms that provide both encryption and data authentication (that is, data integrity). Encryption

scrambles the data so it is transferred confidentially and cannot be interpreted without a special key. Data authentication algorithms ensure the data was not modified during transfer. Some of the supplied cipher algorithms provide only data authentication, and some provide both encryption and authentication. Be aware that the actual cipher algorithm used for the session is determined by a negotiation between the server and client. For example, if you configure an FTP server to use the Triple DES encryption, SHA authentication algorithm, but the client does not support that cipher algorithm, Triple DES encryption, SHA authentication will not be used for sessions between the server and that client.

Select which cipher algorithms you prefer to use by coding a CIPHERSUITE configuration statement in the FTP.DATA file for each cipher algorithm the server can use. For a list of the cipher algorithms you can specify on the CIPHERSUITE statement, refer to *z/OS Communications Server: IP Configuration Reference*. List the CIPHERSUITE statements in FTP.DATA in the order of preference, your most preferred cipher algorithm being first. System SSL will negotiate a cipher algorithm with the server on behalf of the client using the same order of preference as is indicated by the order of CIPHERSUITE statements in FTP.DATA.

Restrictions:

- Only RSA key exchange is supported.
- The following algorithms are subject to export regulations and might not be available to your system:
  - Triple DES encryption, SHA authentication
  - RC4 (128-bit) encryption, SHA authentication
  - RC4 (128-bit) encryption, MD5 authentication
  - AES (128-bit and 256-bit) encryption, SHA authentication

---

6. Decide whether clients logging in to this server should be required to use the TLS protocol. The default is to allow the client to decide whether to use TLS. This setting is customized using the SECURE_FTP configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   To allow the client to decide whether to use TLS, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

   This is the default setting, and indicates the following:

   - If the server is enabled for TLS only, clients must either log in using TLS, or with no security mechanism.
   - If the server is enabled for Kerberos only, clients must either log in using Kerberos, or with no security mechanism.
   - If the server is enabled for both TLS and Kerberos, clients can log in using TLS, Kerberos, or with no security mechanism.

   To require that clients log in using a security mechanism, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP REQUIRED
   ```

   This setting indicates:

   - If the server is enabled for TLS only, clients must log in using TLS.
   - If the server is enabled for Kerberos only, clients must log in using Kerberos.

- If the server is enabled for both TLS and Kerberos, clients must log in using either TLS or Kerberos.

---

7. If you do not want to use client authentication, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_LOGIN NO_CLIENT_AUTH
   ```

   This is the default.

   If you do want to use client authentication, the following levels of client authentication are possible:

   - Level 1 authentication is performed by system SSL. The client passes an X.509 certificate to the server. To pass authentication, the Certificate Authority that signed the client certificate must be considered trusted by the server. To use level 1 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN REQUIRED
     ```

   - Level 2 authentication provides level 1 authentication, and additionally requires that the client certificate be registered with RACF (or another SAF compliant security product) and mapped to a user ID. The client certificate received during the SSL handshake is used to query the security product to verify that the certificate maps to a user ID known to the system prior to connection negotiation. To use level 2 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN VERIFY_USER
     ```

   - Level 3 authentication provides level 1 and 2 authentication. In addition, it provides the capability to restrict access to the server based on the user ID returned from RACF. If the SERVAUTH class of RACF is active and the server's port profile is defined, a connection is accepted only if the requester's user ID associated with the client certificate is defined in the server's port profile. To use level 3 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN VERIFY_USER
     ```

     Also, define the server's port profile in the SERVAUTH class of RACF. For more information on the FTP server's port profile, see "Add user IDs to the SERVAUTH profile access list" on page 1184.

   If you choose to use client authentication, you can also use the client certificate authentication process to eliminate the client login password prompt so that a client supplies only the login user ID to establish the session. The certificate received from the client must be registered in the security product and must be associated with the login user ID. You can use the RACDCERT ADD command to register and associate the certificate. If either the certificate is not registered or is not associated with the user ID, you will be prompted for a password.

   If you do not want to use the client authentication process to eliminate the client password prompt, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_PASSWORD REQUIRED
   ```

   This is the default.

   If you want to use the client authentication process to eliminate the client password prompt, along with your client authentication statement (either

SECURE_LOGIN REQUIRED or SECURE_LOGIN VERIFY_USER), code the following statement in the server's FTP.DATA configuration file:

```
SECURE_PASSWORD OPTIONAL
```

---

8. Decide the level of security for the data connection. You can choose to require enciphered data transfers, or to allow the client to decide the level of security for data transfers. The default is to allow the clients to decide the level of security.

   This setting is customized using the SECURE_DATACONN configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   If you want the server to require that data is transferred raw with no cipher algorithm applied to the data and that clients attempting to use ciphers are rejected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN NEVER
   ```

   If you want the client to decide whether data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN CLEAR
   ```

   This is the default.

   For TLS, the client decides whether data is enciphered or not. If it indicates it should be enciphered, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

   If you want the server to require that data is transferred enciphered and that clients attempting to send raw data are rejected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN PRIVATE
   ```

   For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the data must be transferred using both integrity and privacy protection. Clients attempting to send data that is only integrity protected are rejected.

---

9. For information about configuring your security product for TLS, see Appendix B, "TLS/SSL security," on page 1167.

Steps for customizing the FTP server for Kerberos

Before you begin: You should understand that the FTP server can be enabled to support both TLS and Kerberos. Some of the configuration statement settings apply to both TLS and Kerberos and will affect the behavior of both.

Decide which RACF ID the service principal will be associated with. This will help determine whether or not a keytab file is required. If the service principal is associated to the FTP startup procedure ID, a keytab file will not be required. Decide whether or not a keytab file is required. If a keytab file is not required and will not be used, decide how the FTP startup procedure will be updated to identify the environment variable (ENVAR) KRB5_SERVER_KEYTAB.

Perform the following steps to customize the FTP server for Kerberos:

1. Code the following statement in the server's FTP.DATA configuration file to enable the server for Kerberos:

   ```
   EXTENSIONS AUTH_GSSAPI
   ```

---

2. Decide whether clients should be required to use the Kerberos protocol. The default is to allow the client to decide whether to use Kerberos.

   This setting is customized using the SECURE_FTP configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   To allow the client to decide whether to use Kerberos, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

   This is the default setting, and indicates the following:

   - If the server is enabled for TLS only, clients must either log in using TLS, or with no security mechanism.
   - If the server is enabled for Kerberos only, clients must either log in using Kerberos, or with no security mechanism.
   - If the server is enabled for both TLS and Kerberos, clients can log in using TLS, Kerberos, or with no security mechanism.

   To require that clients log in using Kerberos, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP REQUIRED
   ```

   This setting indicates:

   - If the server is enabled for TLS only, clients must log in using TLS.
   - If the server is enabled for Kerberos only, clients must log in using Kerberos.
   - If the server is enabled for both TLS and Kerberos, clients must log in using either TLS or Kerberos.

---

3. Decide the level of security for the data connection. You can choose to require enciphered data transfers, or to allow the client to decide the level of security for data transfers. The default is to allow the clients to decide the level of security.

   This setting is customized using the SECURE_DATACONN configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   If you want the server to require that data is transferred raw with no cipher algorithm applied to the data and that clients attempting to use ciphers are rejected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN NEVER
   ```

   If you want the client to decide whether data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN CLEAR
   ```

   This is the default.

For TLS, the client decides whether data is enciphered or not. If it indicates it should be enciphered, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

If you want the server to require that data is transferred both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_DATACONN PRIVATE
```

For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols, and clients attempting to send raw data are rejected. For Kerberos, the data must be transferred using both integrity and privacy protection, and clients attempting to send raw data or data that is only integrity protected are rejected.

If you want the server to require that data is transferred integrity protected only or both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_DATACONN SAFE
```

For TLS, specifying this option is identical to specifying SECURE_DATACONN PRIVATE. For Kerberos, specifying this option indicates the data can be transferred integrity protected only, or both integrity and privacy protected. Clients attempting to send raw data are rejected.

---

4. Decide the level of security for the control connection (that is, for FTP commands and replies). You can choose to require enciphered control connection data, or to allow the client to decide the level of security. The default is to allow the clients to decide the level of security.

   This setting is customized using the SECURE_CTRLCONN configuration statement. This setting applies only to Kerberos. For TLS, the control connection is required to be enciphered and this setting has no effect on TLS behavior.

   If you want the client to decide whether control data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN CLEAR
   ```

   This is the default.

   The client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

   If you want the server to require that control data is transferred both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN PRIVATE
   ```

   Clients attempting to send raw data or data that is only integrity protected are rejected.

   If you want the server to require that data is transferred integrity protected only or both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN SAFE
   ```

   Clients attempting to send raw data are rejected.

5. Create the service principal against a RACF ID.
   a. Create a RACF user ID to associate with the FTP service principal.
      ```
      adduser FTP NOPASSWORD DFLTGRP(SYS1) omvs(autouid home('/u/ftp') prog('/bin/sh'))
      ```
   b. After the FTP RACF user ID is created, add the Kerberos principal to it.
      ```
      ALTUSER FTP PASSWORD(ftp) NOEXPIRED KERB(KERBNAME(ftp/<hostname>))
      ```
   c. If a password is not assigned to the FTP ID, issue the following command to remove it.
      ```
      ALTUSER FTP NOPASSWORD
      ```
   d. To ensure the Kerberos segment was added, use the following command to display the ID.
      ```
      LU FTP NORACF KERB
      ```

   Result:

   ```
   USER=FTP

   KERB INFORMATION
   ----------------
   KERBNAME= ftp/<hostname>
   KEY VERSION= 001
   KEY ENCRYPTION TYPE= DES DES3 DESD
   ```

6. Either add the FTP service principal to the keytab file or associate it to the FTP started task ID.
   - To add the FTP service principal to the keytab file, do the following:
     a. The keytab file is located in the /etc/skrb directory. Switch to that directory using the following command:
        ```
        cd /etc/skrb
        ```
     b. Use the following command to see what is currently in the keytab file:
        ```
        keytab list
        ```
        If nothing is currently in the keytab file, the following is returned:
        ```
        Key table: /etc/skrb/krb5.keytab
        ```
     c. Add the FTP service principle using the following command:
        ```
        keytab add ftp/<hostname>
        ```
        You will be prompted for the principals' password. For this example, that password is FTP. The password must be entered in uppercase. This password was assigned with the RACF ALTUSER command when the FTP service principal was created.
     d. Issue the keytab list command again.
        The following should be displayed when the FTP service principal is present:
        ```
        Key table: /etc/skrb/krb5.keytab

        Principal: ftp/<hostname>@<realm>
          Key version: 1
          Key type: 56-bit DES
          Entry timestamp: 2005/02/04-16:21:10

        Principal: ftp/<hostname>@<realm>
          Key version: 1
          Key type: 56-bit DES using key derivation
          Entry timestamp: 2005/02/04-16:21:10

        Principal: ftp/<hostname>@<realm>
        ```

```
Key version: 1
Key type: 168-bit DES using key derivation
Entry timestamp: 2005/02/04-16:21:10
```

- An alternate way to run without a keytab file is to associate the FTP service principal to the ID under which the FTP started task runs. If the ID that the FTP started task runs under is FTPD, issue the following command to create the FTP service principal and have it associated to that ID.

  ```
  ALTUSER FTPD PASSWORD(ftpd) NOEXPIRED KERB(KERBNAME(ftp/<hostname>))
  ```

  Rule: In this setup, you must set the KRB5_SERVER_KEYTAB environment variable. Specify it directly in the FTP startup procedure as follows:

  ```
  //FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
  //      PARM=('POSIX(ON) ALL31(ON)',
  //        'ENVAR("KRB5_SERVER_KEYTAB=1")/&PARMS')
  ```

  Another way of specifying the environment variable directly in the startup procedure is to specify a file where the environment variables are listed.

  ```
  //FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
  //      PARM=('POSIX(ON) ALL31(ON)',
  //        'ENVAR("_CEE_ENVFILE=/etc/ftp.envvars")/&PARMS')
  ```

  Then, within the /etc/ftp.envvars file, add the following:

  ```
  KRB5_SERVER_KEYTAB=1
  ```

You know you are done when a client is able to successfully log in to the FTP server using Kerberos. An example of the login is as follows:

1. Obtain the Kerberos credentials by issuing the following command:

   ```
   kinit joe
   ```

2. You will be prompted for the password. Enter it.
3. Issue the FTP command:

   ```
   ftp <hostname>
   ```

   You should see the following:

   ```
   Using /u/JOE/ftp.data for local site configuration parameters.
   IBM FTP CS V1R6
   FTP: using TCPIP
   Connecting to: <hostname><ip address> port:  <port number>.
   220-FTPD1 IBM FTP CS V1R6 at <hostname>, 21:51:51 on 2005-02-04.
   220 Connection will close if idle for more than 5 minutes.
   >>> AUTH GSSAPI
   334 Using authentication mechanism GSSAPI
   >>> ADAT
   235 ADAT=YGgGCSqGSIb3EgECAgIAb1kwV6ADAgEFoQMCAQ+iSzBJoAMC7moS==
   Authentication negotiation succeeded
   NAME (<hostname>:USER):
   JOE
   >>> USER JOE
   331 Send password please.
   PASSWORD:

   >>> PASS
   230 JOE is logged on.  Working directory is "JOE".
   Command:
   ```

Steps for customizing the FTP client for TLS

Before you begin: You should understand the following:

- The FTP client can be enabled to use either TLS or Kerberos, but not both at the same time.
- To support TLS, the FTP server always provides server certificate authentication to all the clients to validate that the server is what it says it is. Therefore, a client key ring database is required to contain at least the certificate for the CA that signed the server certificate (or the server certificate if the server certificate is self-signed). For more information about key ring databases, see Appendix B, "TLS/SSL security," on page 1167.

Perform the following steps to customize the FTP client for TLS:

1. Code the following statement in the client's FTP.DATA configuration file to enable the client for TLS:

   ```
   SECURE_MECHANISM TLS
   ```

   ---

2. Decide what level of authentication you will use for TLS sessions:
   - Server authentication only
   - Client authentication level 1
   - Client authentication level 2
   - Client authentication level 3

   For more information about server authentication and client authentication, see "Secure Socket Layer overview" on page 1167.

   ---

3. Create the client key ring database and add the certificates you need to the client key ring database.

   For information on how to create a key ring database and add certificates to that database, see "Overview" on page 1173.

   Every TLS session handshake includes server authentication. If a server certificate is self-signed, you must import that certificate to the key ring database of any client that will log in using TLS. If the server certificate is signed by a certificate authority (CA), the CA certificate used to sign the server certificate needs to be in the client key ring database, rather than the server certificate. For more information, see "Server authentication" on page 1168 and "Overview" on page 1173.

   If you are using client authentication, you must add a certificate for the client to the client key ring database.

   If you are using client authentication and self-signed client certificates, you must add a certificate for the client to the server key ring database. If a client certificate is signed by a CA, the CA certificate used to sign the client certificate needs to be in the server key ring database, rather than the client certificate.

   For information about the client certificates you must create, see "Client authentication" on page 1169 and "Overview" on page 1173.

   ---

4. To configure the FTP client with the name of the key ring database, code the following statement in FTP.DATA:

   ```
   KEYRING client-keyring-database
   ```

   For information about the KEYRING statement, refer to *z/OS Communications Server: IP Configuration Reference*.

   ---

5. Decide which cipher algorithms the client should use to encipher data transfers and control information.

   FTP supports TLS through the system SSL cryptographic services base element of z/OS. System SSL supports multiple cipher algorithms that provide both encryption and data authentication (that is, data integrity). Encryption scrambles the data so it is transferred confidentially and cannot be interpreted without a special key. Data authentication algorithms ensure that the data was not modified during transfer. Some of the supplied cipher algorithms provide only data authentication, and some provide both encryption and authentication. Be aware that the actual cipher algorithm used for the session is determined by a negotiation between the server and client. For example, if you configure an FTP client to use the Triple DES encryption, SHA authentication algorithm, but the server does not support that cipher algorithm, Triple DES encryption, SHA authentication will not be used for sessions between the client and that server.

   Select which cipher algorithms you prefer to use by coding a CIPHERSUITE configuration statement in the FTP.DATA file for each cipher algorithm the client can use. For a list of the cipher algorithms you can specify on the CIPHERSUITE statement, refer to *z/OS Communications Server: IP Configuration Reference*.

   Restrictions:

   - Only RSA key exchange is supported.
   - The following algorithms are subject to export regulations and might not be available to your system:
     - Triple DES encryption, SHA authentication
     - RC4 (128-bit) encryption, SHA authentication
     - RC4 (128-bit) encryption, MD5 authentication
     - AES (128-bit and 256-bit) encryption, SHA authentication

---

6. Decide whether the client should be required to use the TLS protocol. If the FTP server does not support TLS, you can choose to allow the client to log in without using the TLS security, or require the client to use a secure session, thus failing the login. The default is to not require the client to use TLS. This setting is customized using the SECURE_FTP configuration statement.

   To have the client log in using the TLS protocol when the server supports TLS, and log in without TLS when the server does not support TLS, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

   This is the default.

   To have the client log in using the TLS protocol, but close the server connection and prevent logging in when the server does not support TLS, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_FTP REQUIRED
   ```

---

7. Decide the level of security for the data connection. You can choose to require enciphered data transfers, or to allow the FTP user to decide the level of security for data transfers. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

Note that the level of security for data connections is determined by both the SECURE_DATACONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

The following subcommands can be issued by the user:

clear Resets the security level so that data is transferred raw.

private Resets the security level so that data is transferred enciphered. The cipher algorithm is negotiated between the server and the client using the TLS protocol negotiation.

If you want the client to transfer data raw with no cipher algorithm applied to the data, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_DATACONN NEVER
```

To indicate the data can be transferred raw or enciphered, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_DATACONN CLEAR
```

This is the default.

By default, data is transferred raw. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level, so that data is transferred enciphered. The user can also issue the `clear` subcommand to reset the data connection security level back, so that data is transferred raw again. For TLS, if the `private` subcommand is issued, the cipher algorithm is negotiated between the server and the client using TLS protocols.

If you want to require that data is transferred enciphered, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_DATACONN PRIVATE
```

For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols.

Steps for customizing the FTP client for Kerberos

Before you begin: You should understand that the FTP client can be enabled to use either TLS or Kerberos, but not both at the same time.

Perform the following steps to customize the FTP client for Kerberos:

1. Code the following statement in the client's FTP.DATA configuration file to enable the client for Kerberos:

   ```
   SECURE_MECHANISM GSSAPI
   ```

   ---

2. Decide whether the client should be required to use the Kerberos protocol. If the FTP server does not support Kerberos, you can choose to allow the client to log in without using Kerberos security, or require the client to use a secure session, thus failing the login. The default is to not require the client to use Kerberos. This setting is customized using the SECURE_FTP configuration statement.

   To have the client log in using the Kerberos protocol, but if the server does not support Kerberos allow the client to complete the login without using it, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

This is the default.

To have the client log in using the Kerberos protocol, but if the server does not support Kerberos have the login fail and not allow the client to log in, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_FTP REQUIRED
```

---

3. Decide the level of security for the data connection. You can choose to require enciphered data transfers, or to allow the FTP user to decide the level of security for data transfers. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

   Note that the level of security for data connections is determined by both the SECURE_DATACONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

   The following subcommands can be issued by the user:

   clear Resets the security level so that data is transferred raw.

   private Resets the security level so that data is transferred enciphered. If the client is using the Kerberos security mechanism, the data is transferred both integrity protected and privacy protected. If the client is using the TLS security mechanism, the cipher algorithm is negotiated between the server and the client using the TLS protocol negotiation.

   safe Resets the security level so that data is transferred integrity protected only.

   If you want the client to transfer data raw with no cipher algorithm applied to the data, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN NEVER
   ```

   To indicate the data can be transferred raw or enciphered, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN CLEAR
   ```

   This is the default.

   By default, data is transferred raw. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level, so that data is transferred both integrity and privacy protected. The user can also issue the `safe` subcommand to change the data connection security level so that data is transferred integrity protected only, or the `clear` subcommand to reset the data connection security level back so that data is transferred raw again.

   If you want to require that data is transferred both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN PRIVATE
   ```

   If you want to require that data is transferred integrity protected only, or both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN SAFE
   ```

   By default, data is transferred integrity protected only. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level so that data is transferred both integrity and privacy

protected. The user can also issue the `safe` subcommand to reset the data connection security level back, so that data is transferred integrity protected only.

---

4. Decide the level of security for the control connection (that is, for FTP commands and replies). You can choose to require enciphered data, or to allow the FTP user to decide the level of security. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

   Note that the level of security for data connections is determined by both the SECURE_CTRLCONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

   The following subcommands can be issued by the user:

   cprotect clear
   Resets the security level so that data is transferred raw.

   cprotect private
   Resets the security level so that data is transferred both integrity protected and privacy protected.

   cprotect safe
   Resets the security level so that data is transferred integrity protected only.

   To indicate the data can be transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN CLEAR
   ```

   This is the default.

   By default, data is transferred raw. However, the user can issue the `cprotect private` subcommand during the FTP session to change the security level so that data is transferred both integrity and privacy protected. The user can also issue the `cprotect safe` subcommand to change the security level so that data is transferred integrity protected only, and the `cprotect clear` subcommand to reset the security level back so that data is transferred raw again.

   If you want to require that data is transferred both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN PRIVATE
   ```

   If you want to require that data is transferred integrity protected only, or both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN SAFE
   ```

   By default, data is transferred integrity protected only. However, the user can issue the `cprotect private` subcommand during the FTP session to change the data connection security level so that data is transferred both integrity and privacy protected. The user can also issue the `cprotect safe` subcommand to reset the data connection security level, so that data is transferred integrity protected only.

Port 990

The use of port 990 to implicitly protect FTP sessions was included in the early drafts of the IETF documents that describe how to use TLS with FTP, but has been removed from the latest drafts. For more information, refer to Info APAR II13516.

Recommendation: Do not start the FTP server on this port. If a server is started on this port, some secure FTP clients might not be able to connect to it.

The FTP server can provide equivalent support on a different port by specifying the following in FTP.DATA:

```
EXTENSIONS AUTH_TLS
SECURE_FTP REQUIRED
SECURE_CTRLCONN PRIVATE
SECURE_DATACONN PRIVATE
```

DB2 and JES

The following statements are used when FTP interfaces with DB2® and JES, respectively. For more information, refer to *z/OS Communications Server: IP Configuration Reference* and the optional steps in this chapter.

- DB2
- DB2PLAN
- SPREAD and SQLCOL
- JESLRECL
- JESPUTGETTO
- JESRECFM
- JESINTERFACELEVEL

Configuring the optional FTP user exits

The following describes exit routines you can code and install. For detailed information regarding these exit routines, refer to the *z/OS Communications Server: IP Configuration Reference*.

The FTPSMFEX user exit

Note the FTP server SMF user exit is called before an SMF type 118 record that contains information about an FTP server session is written to the SYS1.MANx data set. The user exit allows site specific modifications to the record and controls whether the record is written to the SYS1.MANx data set.

Note that the exit is called only for type 118 records. SMF type 119 FTP records must use the system-wide SMF user exits (IEFU83, IEFU84, and IEFU85) to obtain this same functionality. For information on these SMF user exits, refer to *z/OS MVS System Management Facilities (SMF)*.

The FTCHKIP user exit

The FTCHKIP user exit is called when a user attempts to log in to the FTP server or when a user issues the OPEN subcommand to establish a new connection. The following information is passed to the exit:

- Client IP address *
- Client port number *
- Server IP address *

- Server port number *
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier

Note: Fields above marked with an asterisk (*) are valid only for IPv4 addresses, including IPv4-mapped IPv6 addresses.

An installation can use this exit to determine if a particular IP address or port number is allowed to access the FTP site. If the connection is denied by the user exit, the following message is sent to the user:

```
421 User Exit rejects open for connection
```

The FTCHKPWD user exit

The FTCHKPWD user exit is called immediately after the user enters the password or e-mail address during login to the FTP server. The following information is passed to the exit:

- The user ID
- The user password or an asterisk (*) if an e-mail address is entered instead of a password
- A userdata buffer

  If an e-mail address is entered to log in, the userdata buffer contains the e-mail address.
- The number of bad passwords entering during this login attempt
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier

The exit can be used to restrict access to a site based on user ID, password, number of bad passwords, or anything in the socket address information for the client or server. If the login is denied by the user exit, the following reply is sent to the user:

```
530 PASS command failed
```

Note: If ACCESSERRORMSGS TRUE is coded in FTP.DATA, an additional 530 reply with information about why the PASS command failed might precede the reply above.

The FTCHKCMD user exit

The FTCHKCMD user exit is called whenever the user enters an FTP command. The following information is passed to the user exit:

- The user ID
- The FTP command to be issued
- The command's arguments
- The directory type (MVS or HFS)
- The FILETYPE (SEQ, JES, or SQL)
- The current working directory
- A buffer to hold a modified argument string
- A buffer to hold a 500 reply extension to explain why the exit denied the request
- The socket address structure of the client's control connection

- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer

The user exit allows an installation to modify the arguments of an FTP command or to deny a user from issuing the command. For example, if a user issues a DIR * ftp command, the exit can either deny the command or modify it to DIR 'USER1.*'. If the user exit denies the request by this user to issue this command, one or both of the following replies will be sent to the user. The first reply is optional and is sent only if the user exit returns a string in the 500 reply extension buffer.

```
500-UX-buffercontents
500 User Exit denies Userid userid from using Command command
```

The FTCHKJES user exit

FTCHKJES is called if the server is in FILETYPE=JES mode and the client tries to submit a job. The following information is passed to the exit:

- The user ID
- A buffer containing the current JCL statement
- Size of statement in the buffer
- JESLrecl value
- Number of this buffer in current series
- Bytes transferred so far (including this buffer)
- Client identifier (see also session instance identifier)
- JESRecfm value
- FTCHKJES exit-specific workarea (4 bytes)
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer

The exit can allow or refuse the job to be submitted to the JES internal reader based on any information passed to the exit. For example, the exit can look for a USER= parameter on the JOB statement and check it against the client's user ID. If the remote job submission is denied, the exit sends the user the following reply:

```
550 User Exit refuses this job to be submitted by userid
```

The FTPOSTPR user exit

FTPOSTPR is called after execution of the FTP commands RETR, STOR, STOU, APPE, DELE, and RNTO. The following information is passed to the exit:

- The user ID
- Client IP address *
- Client port number *
- The directory type (MVS or HFS)
- The current working directory
- The FILETYPE (SEQ, JES, or SQL)
- Most recent reply code number
- Most recent reply text string
- Current FTP command
- Current CONDDISP setting

- Close reason code
- Name of data set or HFS file retrieved or stored
- Bytes transferred
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer
- The 1-byte description of the confidence-of-successful-completion level assigned to this file transfer

Note: Fields above marked with an asterisk (*) are valid only for IPv4 addresses, including IPv4-mapped IPv6 addresses.

The exit allows for post processing at the termination of data transfer functions within the server.

Customizing the FTP-to-JES interface for JESINTERFACELevel 2 (optional)

If FTP.DATA does not change the JESINTERFACELEVEL to 2, the FTP server uses the JES interface provided in releases prior to CS for OS/390 V2R10. At this level, the FTP user is allowed to submit jobs to JES, retrieve held output matching their logged-in user ID plus one character, and delete held jobs matching their logged-in user ID plus one character.

If JESINTERFACELevel is set to 2, FTP users have the ability to retrieve and delete any job in the system permitted by the System Authorization Facility (SAF) resource class JESSPOOL. For that reason, JESINTERFACELevel=2 should only be specified if the proper JES and SDSF security measures are in place to protect access to JES output. The SAF controls used for JESINTERFACELevel=2 are essentially a subset of those used by SDSF. Therefore, if an installation has customized SAF facilities for SDSF, they are configured for FTP JES level 2.

Before customizing the FTP-to-JES interface, complete JES customization. For example, JESJOBS is an SAF class that controls which users can submit jobs to JES. JESSPOOL is the SAF class that controls which users can access output jobs. Customize these SAF classes before beginning customization of the FTP-to-JES interface.

JESSPOOL defines resource names as <nodeid>.<userid>.<jobname>.<Dsid>.<dsname>. An FTP user can delete an output job if they have UPDATE access to the resource that matches their nodeid, userid, and job name. If the FTP user has READ access to the resource, they can list, retrieve, or GET the job output. For more information on JES security, refer to *z/OS JES2 Initialization and Tuning Guide*. For more information on the SAPI interface, refer to *z/OS MVS Using the Subsystem Interface*.

There are three filters used by the FTP server to control the display of jobs:

- JESSTATUS
- JESOWNER
- JESJOBNAME

SDSF resources are employed for this.

JESSTATUS can be changed by an FTP user with the SITE command to filter jobs in INPUT, ACTIVE, or OUTPUT state. The SDSF resources checked for these states are ISFCMD.DSP.INPUT.jesx, ISFCMD.DSP.ACTIVE.jesx, and ISFCMD.DSP.OUTPUT.jesx, respectively. At login time (USER command), the default value is set to ALL if READ access is allowed to all three classes. Otherwise it attempts to set it to OUTPUT, ACTIVE, and then INPUT if the appropriate READ access is allowed. If no READ access is allowed to any of the classes, JESSTATUS is set to OUTPUT but JESOWNER and JESJOBNAME cannot be changed from the default. In this way, SAF controls can be put in place to limit FTP users to whatever status of jobs an installation requires.

At login time, JESOWNER will have the value of the logged-in user ID. Authority to change JESOWNER is obtained through READ access to RACF profile ISFCMD.FILTER.OWNER. An FTP user who has READ access to ISFCMD.FILTER.OWNER will be allowed to change the JESOWNER parameter with the SITE command.

At login time, JESJOBNAME will have the value of the logged-in user ID plus an asterisk (*). Authority to change JESJOBNAME is obtained through READ access to RACF profile ISFCMD.FILTER.PREFIX. An FTP user who has READ access to ISFCMD.FILTER.PREFIX will be allowed to change the JESJOBNAME parameter with the SITE command.

For example, to allow all users except USER1 to be allowed to change JESOWNER enter the following:

```
SETROPTS CLASSACT(SDSF) REFRESH
RDEFINE SDSF (ISFCMD.FILTER.OWNER) UACC(READ)
PERMIT ISFCMD.FILTER.OWNER ACCESS(NONE) CLASS(SDSF) ID(USER1)
SETROPTS CLASSACT(SDSF) REFRESH
```

For more information on SDSF security, refer to *z/OS SDSF Operation and Customization*.

Configuring the FTP server for anonymous logins (optional)

You can configure the FTP server to accept anonymous logins. A login is anonymous when the remote user specifies USER ANONYMOUS instead of an FTP user ID. To enable anonymous logins, add the ANONYMOUS statement to the server FTP.DATA data set.

You can specify three levels of anonymous support via the ANONYMOUSLEVEL keyword. ANONYMOUSLEVEL 1 is the default, and is equivalent to anonymous login support provided by releases prior to OS/390 V2R10. That is, the ANONYMOUS statement is supported. If no operands are specified on the ANONYMOUS statement, the anonymous user needs no password and has unrestricted access to MVS data sets and the z/OS UNIX file system.

You can specify ANONYMOUSLEVEL 2, but this is not recommended. ANONYMOUSLEVEL 2 is provided for migration purposes only. Consider ANONYMOUSLEVEL 3 if ANONYMOUSLEVEL 1 does not meet your anonymous login security requirements.

If you specify ANONYMOUSLEVEL 3, the anonymous user cannot issue the USER command to leave anonymous mode, nor can another user issue USER anonymous to enter anonymous login mode. If you specify ANONYMOUSLEVEL 3 and

STARTDIRECTORY HFS in FTP.DATA, the anonymous user's z/OS UNIX file system access is restricted to the anonymous user's home directory and home directory subtrees.

The ANONYMOUSLEVEL 3 server recognizes additional keywords that restrict the anonymous user's access to FTP resources. These keywords are ignored when ANONYMOUSLEVEL is less than three:

- ANONYMOUSFILEACCESS allows the system programmer to preclude access to either the z/OS UNIX file system or MVS data sets.
- ANONYMOUSFILETYPEJES, ANONYMOUSFILETYPESQL, and ANONYMOUSFILETYPESEQ control whether the anonymous user can set filetype JES, SQL, or SEQ, respectively.
- ANONYMOUSHFSFILEMODE defines the mode bits used for files written to the z/OS UNIX file system.
- ANONYMOUSHFSDIRMODE defines the mode bits used for directories created in the z/OS UNIX file system.

Finally, when ANONYMOUSLEVEL is set to three, the user's e-mail address is requested in lieu of a password when:

- ANONYMOUS is specified without any parameters.
- ANONYMOUS is specified with user ID/password.
- ANONYMOUS is specified with user ID/SURROGATE.

Control the degree of verification of the e-mail address an anonymous user enters as password by using the EMAILADDRCHECK keyword in FTP.DATA. Refer to *z/OS Communications Server: IP Configuration Reference* for details about the EMAILADDRCHECK keyword. The e-mail address entered is logged to the syslog daemon and is also passed to a user exit routine, FTCHKPWD, for user processing.

The FTP server can be defined to process users without passwords by using the ANONYMOUS SURROGATE support. In order to support this, ANONYMOUSLEVEL must be set to 3 in FTP.DATA on the server and BPX.SRV surrogate must be defined in RACF.

z/OS UNIX uses profiles defined to the RACF SURROGAT class to authorize the server to act as a surrogate of a client. Profiles defined to the SURROGAT class are of the form:

```
BPX.SRV.<userid>
```

in which <userid> is the MVS user ID of the user that the server will support without a password.

The steps below are for a sample userid of the FTP daemon (the userid associated with the FTP started task procedure) called FTPD with the ability to support user ID GUEST without a password. As you add more servers, you will need to follow similar procedures.

1. Activate the SURROGAT class support in RACF:

   ```
   SETROPTS CLASSACT(SURROGAT)
   ```

   This has to be done only once on the system. The SURROGAT class may already have been set up on your system. If a daemon or server you are running will be using the SURROGAT support heavily, consider using the RACLIST command to keep the SURROGAT profiles in storage. The following example shows how to cache the SURROGAT profiles in storage:

```
SETROPTS RACLIST(SURROGAT)
```

2. If the SURROGAT profile is in the RACLIST, any changes to the SURROGAT profiles must be followed by a REFRESH command. To create the SURROGAT class profile for user ID GUEST, issue:

   ```
   RDEFINE SURROGAT BPX.SRV.GUEST UACC(NONE)
   SETROPTS RACLIST(SURROGAT) REFRESH
   ```

   A similar SURROGAT profile is required for each user ID that a server must support without a password.

3. To permit the userid of the FTP daemon (the userid associated with the FTP started task procedure), FTPD, to create a security environment for user ID GUEST, issue the PERMIT command:

   ```
   PERMIT BPX.SRV.GUEST CLASS(SURROGAT) ID(FTPD) ACCESS(READ)
   SETROPTS RACLIST(SURROGAT) REFRESH
   ```

If you choose ANONYMOUSLEVEL greater than one and you choose STARTDIRECTORY HFS, you must create an anonymous directory structure in the z/OS UNIX file system.

Creating an anonymous directory structure in the z/OS UNIX file system

The sample shell script, ftpandir.scp, will create an anonymous directory structure for you, containing required and optional structures. Or, a superuser can create the anonymous directory structure. In this section, the steps a superuser would follow to create an anonymous directory structure are outlined.

For the following steps, assume that the RACF user ID that is used when an anonymous user logs in is called GUEST, that the HOME directory in that user's OMVS segment in RACF is /u/guest, and that FTP.DATA contains a statement similar to this: ANONYMOUS GUEST

1. Create a bin subdirectory in the anonymous root containing the executables *ls* and *sh*. This is a required directory. *ls* can be copied from the standard directory. *sh* is part of the standard MVS search order, so you need only create an empty file with the sticky bit.

   The following example shows how to create ls and sh in the user GUEST's home directory:

   ```
   ===> cd /u/guest
   ===> mkdir bin
   ===> chmod 711 bin
   ===> cd bin

   ===> cp /bin/ls ls
   ===> chmod 711 ls
   ===> touch sh
   ===> chmod 711 sh
   ===> chmod +t sh
   ```

   An ls -al command should give the following results. Owner and group attributes may be different in your system.

   ```
   # ls -al
   total 280
   drwx--x--x   2 USER22   0  8192 Sep 21 17:39 .
   drwx--x--x   7 USER22   0  8192 Nov  1 14:44 ..
   -rwx--x--x   1 USER22   0  126976 Sep 21 17:39 ls
   -rwx--x--t   1 USER22   0  0 Sep 21 17:39 sh
   ```

2. Create a usr/sbin sudirectory of the anonymous root containing the executable file ftpdns. This is a required subdirectory. The file ftpdns can be empty with the sticky bin on.

   The following example is for anonymous user GUEST:

   ```
   ===> cd /u/guest
   ===> mkdir usr
   ===> chmod 711 usr

   ===> cd usr
   ===> mkdir sbin
   ===> chmod 711 sbin
   ===> cd sbin
   ===> touch ftpdns
   ===> chmod 711 ftpdns
   ===> chmod +t ftpdns
   ```

   If you do not configure the subdirectories, bin and usr/sbin, and their contents correctly, the FTP server will not be able to accept anonymous logins and message EZYFT731 will be displayed.

3. Create a dev subdirectory within the anonymous root. This is a required subdirectory. A null file is created in this directory and used during the open of syslog.

   The following example is for anonymous user GUEST:

   ```
   ===> cd /u/guest
   ===> mkdir dev
   ===> chmod 711 usr
   ```

   If you do not have the dev subdirectory, syslog might not open correctly. Messages such as EZA2830I will not be logged out correctly.

4. Set up the public directory structure. This is a required directory.

   This is the directory structure into which you place files that can be downloaded by the anonymous FTP user. It does not have to be named pub; it can be any name you choose. A general convention for anonymous FTP sites is to call it pub:

   ```
    ===> cd /u/guest
   ===> mkdir pub
   ===> cd pub
   ```

   If you want to structure the files you allow to be accessed, you can create multiple subdirectories underneath this directory.

   For simplicity, assume a single level directory, the pub directory. Into this directory you copy the files you want to allow the anonymous user to download:

   ```
   ===> cp /x/y/z/prodinfo1.txt prodinfo1.txt
   ===> cp /x/y/z/prodinfo2.txt prodinfo2.txt
   ===> cd ..
   ```

   Make sure that the permission bits are set correctly by using the following shell command when executed in the /u/guest directory. This will set the permission bits of all files in the pub directory and its subdirectories to 755:

   ```
   ===> chmod -R 755 pub
   ```

   If your system does not require an incoming or extract directory, the system is configured for anonymous FTP. An ls -al command of the pub directory should give the following results:

```
drwxr-xr-x   3 IBMUSER  SYS1  8192 May 13 21:15 .
drwxr-xr-x   6 IBMUSER  SYS1  8192 May 20 14:51 ..
-rwxr-xr-x   1 IBMUSER  SYS1  12 May 11 12:41 prodinfo1.txt
-rwxr-xr-x   1 IBMUSER  SYS1  12 May 11 12:41 prodinfo2.txt
```

5. Set up an incoming directory (optional).

   If you want anonymous users to be able to upload files to your FTP server, you need some additional setup. The objective is to allow an anonymous user to upload a file, but not to allow another anonymous user to download or even be aware of the existence of the file until after an administrative user has verified that the content of the file is acceptable. You do not want your FTP server site to become a store-and-forward site for files of questionable ethical content.

   Positioned at the /u/guest directory, a superuser issues the following shell command:

   ```
   ===> cd /u/guest
   ===> mkdir incoming
   ===> chmod 733 incoming
   ```

   It does not have to be named incoming; it can be any name you choose. A general convention for anonymous FTP sites is to call it incoming.

   The 733 permission bits means that a non-superuser cannot list the content of the incoming directory, but can write a file to it. Because the FTP server enforces a UMASK of 777 when an anonymous user logs in, these files will be written with permission bits 000, which means that they cannot be accessed by the anonymous user or by any other user except a superuser.

   An FTP client user can normally change the UMASK via a SITE UMASK command or the user can change the permission bits of files they own through a SITE CHMOD command.

   If you define ANONYMOUSLEVEL 3, you can use the ANONYMOUSHFSDIRMODE keyword to set the permission bits of any directory created by an anonymous user, and the ANONYMOUSHFSFILEMODE to set the permission bits of any file created by an anonymous user.

   If you do allow anonymous users to store files on your FTP server, you should ensure that the directory into which these files are stored is a separate z/OS UNIX file system that can fill up without impacting other work on your z/OS system. The best way to do that is to allocate the /u/guest/incoming directory in its own zFS, HFS data set, or Network File System. If an anonymous user uploads large amounts of data to the incoming directory, only this separate z/OS UNIX file system will be filled up. Filling this separate z/OS UNIX file system prevents other anonymous users from storing new files on the server, but will not affect other functions on your system. At a minimum, you should make sure that the incoming directory is not in the same z/OS UNIX file system as your /tmp directory.

6. Set up the extract directory (optional).

   If you need to make files available to certain anonymous users, but not to everyone, you can create a directory that cannot be listed, but files in it can be downloaded if the anonymous user knows the name of the file.

   Positioned at the /u/guest directory, a superuser issues the following shell commands:

   ```
   ===> cd /u/guest
   ===> mkdir extract
   ===> chmod 711 extract
   ```

   It does not have to be named extract; it can be any name you choose. A general convention for anonymous FTP sites is to call it extract.

A superuser can then copy files into this directory, ensure they have permissions of 755, inform the intended anonymous user of the file name, and that user can then log on as anonymous and retrieve the file.

An ls -al command at the /u/guest location should give the following result, if you created all four subdirectories:

```
drwxr-xr-x   6 IBMUSER  SYS1 8192 May 20 14:51 .
dr-xr-xr-x   6 IBMUSER  SYS1  0 Jun 10 15:43 ..
drwx--x--x   2 IBMUSER  SYS1  8192 May 11 12:44 bin
drwx--x--x   3 IBMUSER  SYS1  8192 May 11 13:39 extract
drwx-wx-wx   3 IBMUSER  SYS1  8192 May 25 09:35 incoming
drwxr-xr-x   3 IBMUSER  SYS1  8192 May 13 21:15 pub
```

Configure the Welcome Banner Page, Login, and Directory Message (optional)

Starting in V2R10, the FTP server now provides support to allow FTP administrators to provide useful information about the site to FTP users. The following FTP.DATA statements are available:

- BANNER
- LOGINMSG
- ANONYMOUSLOGINMSG
- MVSINFO
- ANONYMOUSMVSINFO
- HFSINFO
- ANONYMOUSHFSINFO

You can use the LOGINMSG statement in FTP.DATA to point to a set of messages displayed when a known user logs in to FTP. Similarly, ANONYMOUSLOGINMSG can point to a set of messages displayed when an anonymous user logs in to FTP.

You can use the MVSINFO statement to point to a set of messages displayed when a known user changes the working directory to a particular MVS data set path. Likewise, use the ANONYMOUSMVSINFO statement to point to a set of messages displayed when an anonymous user changes working directory to a particular MVS data set path.

You can use the HFSINFO statement to point to a set of messages displayed when a client changes the working directory to a particular HFS directory. Likewise, use the ANONYMOUSHFSINFO statement to point to a set of messages displayed when an anonymous user changes working directory to a particular HFS directory.

Using magic cookies to represent information

The content of all the informational messages may include a predefined set of magic cookies, which are substituted by the FTP server before the data is sent to the FTP client. The following magic cookies are supported:

- %T — Local time
- %C — Current working directory
- %E — The FTP server administrators e-mail address
- %R — Remote host name
- %L — Local host name
- %U — Username (logged in user)

If %R is used, a long delay in login processing might occur as the FTP server will issue a DNS query to resolve the remote host IP address. In order to use %E, the ADMINEMAILADDR keyword must be specified in the server FTP.DATA configuration file.

Configuring the FTP server to log session (user ID) activity

You can configure the FTP server to write log messages for trace activity related to individual sessions by coding the FTPLOGGING or ANONYMOUSFTPLOGGING statements in FTP.DATA. If you have configured syslogd for FTP, the log messages appear in syslog.

Table 28 shows the FTP log message written for various activities. You can identify an FTP log message by the ID=*sessionID* parameter in the message. Each login session is assigned a session ID that can be used to identify all log messages related to that session. For more information about EZYF*xxxx* messages, refer to *z/OS Communications Server: IP Messages Volume 3 (EZY)*.

Table 28. EZYFxxxx messages

FTP log message	Activity
EZYFS50I	A client connected to the FTP daemon.
EZYFS51I	A client connection to the FTP daemon failed.
EZYFS52I	An FTP session ended.
EZYFS54I	The server accepted the security mechanism. The connection with the client is now protected.
EZYFS55I	The server rejected the security mechanism. The connection with the client is not protected.
EZYFS56I	A client logged into the server.
EZYFS57I	A client login to the server failed.
EZYFS58I	The server denied the client access to an MVS data set.
EZYFS59I	The server denied the client access to an HFS file.
EZYFS60I, EZYFS61I, EZYFS62I	The server successfully allocated an MVS data set.
EZYFS63I, EZYFS64I, EZYFS65I	The server could not allocate an MVS data set.
EZYFS67I	The server successfully allocated an MVS data set.
EZYFS68I, EZYFS69I	The server could not allocate an HFS file.
EZYFS70I, EZYFS74I, EZYFS75I	The server deallocated an MVS data set.
EZYFS71I, EZYFS72I, EZYFS73I	The server detected an error while deallocating an MVS data set.
EZYFS77I	The server deallocated an HFS file.
EZYFS78I, EZYFS79I	The server detected an error while deallocating an HFS file.
EZYFS80I	The server sent a reply to the client after data transfer.

Table 28. EZYFxxxx messages (continued)

FTP log message	Activity
EZYFS81I	The server finished processing an MVS data set transfer.
EZYFS82I	The server finished processing an HFS data set transfer.
EZYFS83I	The server stored a data set or file into its file system.
EZYFS84I	The server sent a data set or file to the client (or to a server in the case of a proxy transfer).
EZYFS85I	The server finished returning job output to the client.
EZYFS86I	The server assigned a confidence of success level to the completed data transfer.
EZYFS91I	The server submitted a job for the client.
EZYFS92I	The server returned a SQL report.
EZYFS95I	An abend occurred while the server was transferring data.

Configuring to send detailed login failure replies to an FTP client (optional)

The FTP server returns minimal information to the client when the PASS command fails. However, you can configure the FTP server to send additional information by coding ACCESSERRORMSGS TRUE in FTP.DATA. This directs the server to reply to the client with detailed login failure data. The reply might report server errors, such as failing function calls with diagnostic return codes. It might report user errors, such as an expired or incorrect password, an unknown user ID, or a revoked user ID. You should not code ACCESSERRORMSGS TRUE in FTP.DATA if you do not want to share this type of information with users logging in to FTP.

You can capture the same information in syslog by coding FTPLOGGING TRUE and ANONYMOUSFTPLOGGING TRUE in FTP.DATA. You can also turn on the DEBUG option called ACC to log the error messages in the syslog. For more information on coding FTP.DATA statements, see *z/OS Communications Server: IP Configuration Reference*.

Install the SQL query function (optional) and access the DB2 modules

To use FTP to do SQL queries, bind the DBRM called EZAFTPMQ to the plan used by FTP, and grant execution privileges for that plan to PUBLIC. (The name of the plan can be specified by the DB2PLAN keyword in FTP.DATA or defaulted to EZAFTPMQ.) This FTP facility only performs SELECT operations on the DB2 tables. It does not perform UPDATE, INSERT, or DELETE.

Note: If secondary authorization for SQL queries is required, the DSN3SATH sample exit shipped by DB2 must be modified. The exit will return the primary AUTHID for requests originating from the FTP server.

The following sample job is provided in the FTOEBIND member of the SEZAINST data set. It can be used to enable the FTP server and client to do SQL queries.

```
//FTPSETUP JOB FTPSETUP,
//            CLASS=A,
//            NOTIFY=&SYSUID
//*************************************************************
//*
//*   File name:                 tcpip.SEZAINST(FTOEBIND)
//*   SMP/E distribution name:   EZAFTPAB
//*
//*    Licensed Materials - Property of IBM
//*    This product contains "Restricted Materials of IBM"
//*    5647-A01 (C) Copyright IBM Corp.  1997, 2002
//*    All rights reserved.
//*    US Government Users Restricted Rights -
//*    Use, duplication or disclosure restricted by GSA ADP Schedule
//*    Contract with IBM Corp.
//*    See IBM Copyright Instructions.
//*
//*  This JCL binds the EZAFTPMQ DBRM to the specified
//*  DB2 subsystem and allows execution of the
//*  EZAFTPMQ plan by PUBLIC.
//*
//*  The FTP server and client use this plan. (See
//*  Usage note #7)
//*
//*
//*  Usage notes:
//*
//*    1.  You must execute this job from a user ID that has
//*        the authority to bind the EZAFTPMQ plan.
//*
//*    2.  Change the STEPLIB DD statement in the FTPBIND and
//*        FTPGRANT steps to reflect the DB2 DSNLOAD data set.
//*
//*    3.  Change the DB2 sybsystem name in the FTPBIND and
//*        FTPGRANT steps from SYSTEM(xxx) to the
//*        installation defined DB2 subsystem name.
//*
//*    4.  Change the library parameter in the FTPBIND step from
//*        TCPIP.SEZADBRM to the installation defined TCPIP
//*        SEZADBRM library.
//*
//*    5.  Change the plan name in the FTPGRANT step from
//*        DSNTIAYY to reflect the plan associated with the
//*        program DSNTIAD.
//*
//*    6.  Change the library parameter in the FTPGRANT step
//*        from xxxxxx.RUNLIB.LOAD to reflect the library
//*        where the DSNTIAD program resides.
//*
//*    7.  You can bind the DBRM to a plan name other than EZAFTPMQ
//*        by changing the plan specified in the FTPBIND and
//*        FTPGRANT steps.  If you do this, you must use the
//*        DB2PLAN keyword in FTP.DATA to change the plan name
//*        used by the FTP server and/or client to the plan name
//*        specified here.
//*
//*************************************************************
//FTPBIND  EXEC PGM=IKJEFT01,DYNAMNBR=20
//STEPLIB  DD   DSN=xxxxxx.DSNLOAD,DISP=SHR
//SYSTSPRT DD   SYSOUT=*
//SYSPRINT DD   SYSOUT=*
//SYSOUT   DD   SYSOUT=*
//SYSTSIN  DD   *
 DSN  SYSTEM(xxx)
```

```
 BIND ACQUIRE(USE) -
      ACTION(REPLACE) -
      CACHESIZE(1024) -
      CURRENTDATA(NO) -
      EXPLAIN(NO) -
      ISOLATION(CS) -
      LIBRARY('TCPIP.SEZADBRM') -
      MEMBER(EZAFTPMQ) -
      NODEFER(PREPARE) -
      PLAN(EZAFTPMQ) -
      RELEASE(COMMIT) -
      VALIDATE(RUN) -
      RETAIN
 END
//*
//FTPGRANT EXEC PGM=IKJEFT01,DYNAMNBR=20
//STEPLIB  DD   DSN=xxxxxx.DSNLOAD,DISP=SHR
//SYSTSPRT DD   SYSOUT=*
//SYSPRINT DD   SYSOUT=*
//SYSOUT   DD   SYSOUT=*
//SYSTSIN  DD   *
 DSN  SYSTEM(xxx)
 RUN  PROGRAM(DSNTIAD) -
      PLAN(DSNTIAYY) -
      LIBRARY('xxxxxx.RUNLIB.LOAD')
 END
//SYSIN    DD   *
 GRANT EXECUTE ON PLAN EZAFTPMQ TO PUBLIC;
//*
```

Accessing DB2 modules

The FTP server or client loads 3 DB2 modules into storage to perform an SQL query. These modules are:

- DSNALI
- DSNHLI2
- DSNTIAR

The modules are usually found in the DB2 load library with the suffix DSNLOAD. The DB2 administrator or system programmer should add the DSNLOAD library to the LINKLIST to ensure FTP has access to this library.

Another way to ensure access is to add the DSNLOAD library to the FTP STEPLIB. For the FTP server this means the JCL used to start the FTP server has a STEPLIB DD statement referring to the DSNLOAD library or, if the FTP daemon is started from the z/OS shell, the STEPLIB environment variable is set. For the FTP client, this means a TSO CLIST must allocate the DSNLOAD library as the STEPLIB.

If the FTP client is to be run from a batch job to perform SQL queries, the DSNLOAD library must be added to the STEPLIB DD statement for the batch job.

Usage notes:

To allow FTP access to multiple levels of DB2, link to the libraries that contain the lowest level of DB2 to be accessed.

FTP.DATA updates for SQL query function

To obtain FTP.DATA updates for the SQL query function, follow these steps:

1. Set the FTP.DATA DB2 statement to specify the name of the DB2 subsystem.

2. Set DB2PLAN to specify the DB2 plan to be used by the FTP server.
3. Set the SPREAD statement to specify whether SQL output is in spreadsheet format.
4. Set SQLCOL to specify the column headings of the output data.

Verification of FTP

Verify server

If FTP is in the autolog list and the TCP/IP address space is restarted, FTP should start automatically. For other cases, it should be started manually. To do this, go to the MVS console and enter the following command:

```
S FTPD
```

Note: The above command assumes the FTP procedure name is FTPD.

If the FTP server startup is complete, the following message should be seen on the MVS console:

```
EZY2702I Server-FTP: Initialization completed at 17:37:29 on  12/17/99.
```

If the message is not seen, a message explaining why FTP did not start up will appear in SYSLOG. Even if the above message is issued, it would be beneficial to inspect SYSLOG for warning messages issued during FTP initialization. EZY2700I displays the port FTP uses as the control port, the port it listens to for incoming connections from clients. In this example, FTP is listening to standard port 21.

The file syslog uses is defined in /etc/syslog.conf. The statement **daemon.info /tmp/daemon.log** directs SYSLOGD to save all the daemon messages in /tmp/daemon.log. Below is an example of output error messages.

```
EZYFT18I Using catalog '/usr/lib/nls/msg/C/ftpdmsg.cat' for FTP messages.
EZY2697I IBM FTP CS V1R7 17:34:04 on 10/15/04
EZY2640I Using dd:SYSFTPD=USER1.FTP.DATA for local site configuration parameters
EZYFT46E Error in dd:SYSFTPD file: line 4 near column 9.
EZY2636E SMFLOGN value not specified.
EZYFT46E Error in dd:SYSFTPD file: line 5 near column 8.
EZY2636E SMFREN value not specified.
EZYFT47I dd:SYSFTPD file, line 21: Ignoring keyword "EXTENSIONS REST_STREAM".
EZYFT47I dd:SYSFTPD file, line 29: Ignoring keyword "CTRLCONN".
EZYFT21I Using catalog '/usr/lib/nls/msg/C/ftpdrply.cat' for FTP replies.
EZYFT26I Using 7-bit conversion derived from 'ISO8859-1' and 'IBM-1047' for the control connection.
EZYFT33I Unable to open DDNAME 'SYSFTSX' for the data connection: EDC5129I No such file or directory.
EZYFT31I Using //'TPOUSER.STANDARD.TCPXLBIN' for FTP translation tables for the data connection.
EZYFT09I system information for VIC135: z/OS version 1 release 7 (2084)
EZY2700I Using port FTP control (21)
EZY2701I Inactivity time is 0
EZYFT57I FTP registering with WLM as group = ftpgroup host = VIC135
EZY2702I Server-FTP: Initialization completed at 17:35:05 on 10/15/03.
EZYFT41I Server-FTP: process id 16777255, server job name FTPD11
```

Verify client

To verify that the FTP client works correctly, log onto TSO and issue the NETSTAT HOME command, or issue NETSTAT –h from the z/OS UNIX shell. These commands will show the interface addresses that are known to the system. Below is an example of the output from NETSTAT HOME:

```
MVS TCP/IP NETSTAT CS V1R7       TCPIP Name: TCPCS           14:22:17
Home address list:
LinkName:   OSAQDIO6L
  Address:  9.67.115.13
    Flags:  Primary
LinkName:   LSAMEH
  Address:  9.1.1.1
    Flags:
```

```
LinkName:   LOOPBACK
  Address:  127.0.0.1
    Flags:
Address:  fe80::9:6b00:671a:586
    Type:   Link_Local
    Flags:  Autoconfigured
IntfName:   V6SAMEH
  Address:  1::8
    Type:   Global
    Flags:
IntfName:   V6VIRT
Address:  2::55
    Type:   Global
    Flags:
IntfName:   LOOPBACK6
  Address:  3::1
    Type:   Global
    Flags:
  Address:  ::1
    Type:   Loopback
    Flags:
```

To invoke the FTP client, use any address shown on the NETSTAT HOME address list. The first example below shows how you could log in to the FTP server at 9.67.115.13 using a batch job (the output of the batch job is not shown). The second example shows logging in to the FTP server at 9.67.113.37 from the TSO environment.

```
//FTPBATCH JOB FTPUSER,
// USER=USER1,PASSWORD=TCPSUP
//BATCH  EXEC PGM=FTP
//OUTPUT  DD SYSOUT=*
//INPUT DD *
  9.67.115.13
  USER10 tcpusr
  SITE FILE=SEQ
  QUIT
//*
```

```
Using 'USER1.FTP.DATA' for local site configuration parameters.
IBM FTP CS V1R7
FTP: using TCPCS
Connecting to: vic135.tcp.raleigh.ibm.com 9.67.113.37 port: 21.
220-FTPD1 IBM FTP CS V1R7 at vic135, 19:07:34 on 2004-10-08.
220 Connection will close if idle for more than 5 minutes.
>>> FEAT
211- Extensions supported
 AUTH TLS
 PBSZ
 PROT
211 End
NAME (vic135:USER1):

user1
NAME (vic135:USER1):
>>> USER USER1
331 Send password please.
PASSWORD:

>>> PASS
230 USER1 is logged on.  Working directory is "/".
Command:
```

Verify FTP.DATA statements

Many FTP.DATA statements can be verified via the FTP client STAT and LOCSTAT commands. The output from each installation's STAT and LOCSTAT will depend on the client and server copy of FTP.DATA. Below is sample output of one system.

```
stat
 EZA1701I >>> STAT
 211-Server FTP talking to host 127.0.0.1, port 1027
 211-User: USER1  Working directory: USER1.
 211-The control connection has transferred 2006 bytes
 211-There is no current data connection.
 211-The next data connection will be actively opened
 211-to host 127.0.0.1, port 1027,
 211-using Mode Stream, Structure File, type ASCII, byte-size 8
 211-Automatic recall of migrated data sets.
 211-Automatic mount of direct access volumes.
 211-Auto tape mount is allowed.
 211-Inactivity timer is set to 600
 211-VCOUNT is 59
 211-ASA control characters in ASA files opened for text processing
 211-will be transferred as ASA control characters.
 211-Trailing blanks are removed from a fixed format
 211-data set when it is retrieved.
 211-Data set mode.  (Do not treat each qualifier as a directory.)
 211-ISPFSTATS is set to FALSE
 211-Primary allocation 5 tracks.  Secondary allocation 2 tracks.
 211-Partitioned data sets will be created with 15 directory blocks.
 211-FileType SEQ (Sequential - default).
 211-Number of access method buffers is 5
 211-RDWs from variable format data sets are discarded.
 211-Records on input tape are unspecified format
 211-SITE DB2 subsystem name is DB2
 211-Data not wrapped into next record.
 211-Tape write is not allowed to use BSAM I/O
 211-Truncated records will not be treated as an error
 211-JESLRECL is 80
 211-JESRECFM is Fixed
 211-JESINTERFACELEVEL is 2
 211-ENcoding is set to SBCS
 211-SBSUB is set to FALSE
 211-SBSUBCHAR is set to SPACE
 211-SMS is active.
 211-Mgmtclass for new data sets is TCPMGMT
 211-New data sets will be catalogued if a store operation ends abnormally
 211-Single quotes will override the current working directory.
 211-UMASK value is 027
 211-Process id is 12
 211-Checkpoint interval is 0
 211-Authentication type: None
 211-Record format VB, Lrecl: 128, Blocksize: 6144
 211 *** end of status ***


locstat
 EZA1600I Trace: FALSE, Send Port: TRUE
 EZA1601I Send Site with Put command: TRUE
 EZA2676I Connected to:127.0.0.1, Port: FTP control (21), logged in
 EZA1605I Local Port: 1027
 EZA1606I Data type:a, Transfer mode:s, Structure:f
 EZA2098I Automatic recall of migrated data sets.
 EZA2100I Automatic mount of direct access volumes.
 EZA2101I Data set mode. (Do not treat each qualifier as a directory.)
 EZA2844I ISPFSTATS is set to FALSE
 EZA2134I Primary allocation 5 tracks, Secondary allocation 2 tracks.
 EZA2138I Partitioned data sets will be created with 15 directory blocks
 EZA2103I FileType is SEQ (Sequential - the default).
```

```
 EZA2141I Number of access method buffers is 5.
 EZA2948I ENcoding is set to SBCS
 EZA2943I SBSUB is set to FALSE
 EZA2944I SBSUBCHAR is set to SPACE
 EZA2142I Mgmtclass for new data sets is TCPMGMT
 EZA2145I RDW's from VB/VBS files are discarded.
 EZA2518I Records on input tape are unspecified format
 EZA2148I DB2 subsystem name is DB2
 EZA2152I Volid of Migrated Data Sets is MIGRAT
 EZA2154I Trailing blanks in records read from RECFM F datasets are discarded.
 EZA2535I Record format: VB, Lrecl: 128, Blocksize: 6144.
 EZA2801I Data not wrapped into next record.
 EZA2529I Truncated records will not be treated as an error.
 EZA2494I Checkpoint interval is 0
 EZA2511I Checkpoint data set will be opened for GET
 EZA2428I CHKPTPrefix uses Home to determine the HLQ of the FTP.CHECKPOINT file.
 EZA2817I Automatic mount of tape volumes.
 EZA2809I CCONNTIME is 120
 EZA2810I DATACTTIME is 120
 EZA2811I DCONNTIME is 120
 EZA2812I INACTTIME is 120
 EZA2813I MYOPENTIME is 120
 EZA2815I VCOUNT is 59
 EZA2689I Prompting: ON, Globbing: ON
 EZA2719I ASA control characters transferred as ASA control characters
 EZA2720I New data sets catalogued if a store operation terminates abnormally
 EZA2722I Single quotes will override the current working directory
 EZA2724I UMASK value is 027
 EZA2819I Data connections for the client are not firewall friendly.
 EZA2889I Authentication mechanism: None
 EZA2866I Tape write is not allowed to use BSAM I/O
 EZY2640I Using 'SYS1.TCPPARMS(FTPDATA)' for local site configuration parameters.
 EZA1460I Command:
```

Verifying anonymous, banner, and other optional configuration information

Depending on your installation's choices for anonymous level, banner support chosen, exits, and so on, verification of support output will differ. To verify anonymous configuration at a particular installation, log in as anonymous and verify the behavior is as expected. For example, if EMAILADDRCHECK FAIL is specified in FTP.DATA, try to log in as anonymous using an incorrect e-mail address as password. To verify banner support, login and verify the banners are displayed as expected. Below is a sample of FTP.DATA and FTP client output for one such installation.

```
; BANNER STUFF
EMAILADDRCHECK FAIL
BANNER USER1.TEST1
ADMINEMAILADDR FTPADMIN@MYSYSTEM.COM
; ANONYMOUS STUFF
ANONYMOUSLEVEL 3
STARTDIRECTORY HFS
```

```
ftp 9.67.113.63
 IBM FTP CS V1R7  2004 349 01:35 UTC
 FTP: using TCPCS
 Connecting to:  9.67.113.63 port: 21.
 220-FTPD1 IBM FTP CS V1R7 at HOSTA, 19:07:34 on 2004-01-08.
 220-You have just read 'USER1.TEST1'
 220-ADMINEMAILADDRESS is FTPADMIN@MYSYSTEM.COM
 220 Connection will not timeout.
 NAME (9.67.113.63:USER4):
anonymous no-email-pw
 >>> USER anonymous
```

```
 331 Send password please.
 >>> PASS
 530 PASS command failed.
 Command:
```

Verify FTP-JES interface (optional)

As with the other optional configuration information, FTP-JES support can best be verified by logging in and confirming the FTP.DATA parameters chosen. To verify JES support, a simple batch job can be created if the JESINTERFACELEVEL is set to the security requirements of an installation. Below is the batch job and FTP client output for JESINTERFACELEVEL 2.

```
EDIT       USER1.FTP.JCL.TEST                              Columns 00001 00072
Command ===>                                                  Scroll ===> CSR
****** ***************************** Top of Data ****************************
000100 //JOBTEST  JOB MSGCLASS=H,MSGLEVEL=(1,1),CLASS=A,
000200 //         USER=USER1
000300 //STEP1    EXEC PGM=IEBGENER
000400 //OBJTMP1  DD DSN=&PRLOBJ,DISP=(NEW,PASS,DELETE),
000500 //            SPACE=(CYL,(1,1,10)),
000600 //            DCB=(RECFM=FB,LRECL=80,BLKSIZE=800)
000700 //SYSPRINT DD SYSOUT=A
000800 //SYSUT1   DD DSN=SYS1.PROCLIB(JES2),DISP=SHR
000900 //SYSIN    DD DUMMY
001000 //SYSUT2   DD SYSOUT=H
001100 //
001200 //         EXEC PGM=IEFBR14
****** **************************** Bottom of Data ***************************

site file=jes jesjobname=jobtest jesowner=* jesstatus=all
 EZA1701I >>> SITE file=jes jesjobname=jobtest jesowner=* jesstatus=all
 200 Site command was accepted
 EZA1460I Command:
put 'user1.ftp.jcl.test'
 EZA1701I >>> SITE FIXrecfm 80 LRECL=80 RECFM=FB BLKSIZE=32720
 200 Site command was accepted
 EZA1701I >>> PORT 127,0,0,1,4,12
 200 Port request OK.
 EZA1701I >>> STOR 'user1.ftp.jcl.test'
 125 Sending Job to JES internal reader FIXrecfm 80
 250-It is known to JES as JOB00076
 250 Transfer completed successfully.
 EZA1617I 984 bytes transferred in 0.005 seconds.  Transfer rate 196.80 Kbytes/sec.
 EZA1460I Command:
dir j76
 EZA1701I >>> PORT 127,0,0,1,4,13
 200 Port request OK.
 EZA1701I >>> LIST j76
 125 List started OK for JESJOBNAME=JOBTEST, JESSTATUS=ALL and JESOWNER=*
 EZA2284I JOBNAME  JOBID    OWNER    STATUS CLASS
 EZA2284I JOBTEST  JOB00076 USER1    OUTPUT A        RC=0000
 EZA2284I          ID  STEPNAME PROCSTEP C DDNAME   BYTE-COUNT
 EZA2284I          001 JESE              H JESMSGLG      1084
 EZA2284I          002 JESE              H JESJCL        1023
 EZA2284I          003 JESE              H JESYSMSG      1143
 EZA2284I          004 STEP1             H SYSUT2         741
 EZA2284I          005 STEP1             A SYSPRINT       209
 EZA2284I 5 spool files
 250 List completed successfully.
 EZA1460I Command:
```

Chapter 12. Trivial File Transfer Protocol (TFTP)

TFTP is a UDP protocol used to transfer files. TFTP can read or write files from or to a remote server. On the z/OS system, TFTP is a server you can configure with the command line option during TFTP invocation. TFTP can also be started as a procedure.

TFTP is installed in the /usr/lpp/tcpip/sbin/ directory.

Recommendation: The TFTP server uses well-known port 69. The TFTP server has no user authentication. Any client that can connect to port 69 on the server has access to TFTP. If the TFTP server is started without a directory, it allows access to the entire HFS. To restrict access to the HFS, start the TFTP server with a list of directories.

The TFTP server preforks a child process to handle incoming requests when the concurrency limit is exceeded. Consequently, immediately after starting the TFTP server, two TFTP processes exist.

In case of a flood of concurrent TFTP commands, the TFTP server may fork additional processes. When the number of concurrent requests being processed drops below the concurrency limit, the number of TFTP processes is decreased back to two.

Starting TFTP from the command line

To start the TFTP server from the command line, type the tftpd command.

```
tftpd [-l] [-p port] [-t timeout] [-r maxretries] [-c concurrency_limit]
   [-s maxsegsize] [-f file] [-a archive directory [-a ...]]
   [-b IP address] [directory ...]
```

The following are parameters used for the tftpd command:

-l Logs all incoming read and write requests and associated information to the system log. Logged information includes the IP address of the requestor, the file requested, and whether the request was successful.

-p port
Uses the specified port. The TFTP server usually receives requests on well-known port 69. You can specify the port in which requests are to be received.

-t timeout
Sets the packet timeout. The TFTP server usually waits 5 seconds before assuming a transmitted packet has been lost. You can specify a different timeout period in seconds.

-r maxretries
Sets the retry limit. The TFTP server usually limits the number of retransmissions it performs due to lost packet to 5. You can specify a different retry limit.

-c concurrency_limit
Sets the concurrency limit. The TFTP server spawns both threads and processes to handle incoming requests. You can specify the limit for the

number of threads that may be concurrently processing requests under a single process. When the limit is exceeded, a new process is spawned to handle requests. The default is 200 threads.

-s maxsegsize
Sets the maximum block size that can be negotiated by the TFTP block size option. The default is 8192.

-f file Specifies a cache file. You can specify a file containing information on files to be preloaded and cached for transmission. A cache file consists of one or more entries. For clarity, place each entry on a separate line. An entry has the form:

a | b <pathname>

where:

- *a* indicates that the specified file is cached in ASCII form. The file is preconverted to netascii format.
- *b* indicates that the specified file is cached in binary form, with no conversion.

Following are examples of cache file entries,

```
a  /usr/local/textfile
b  local/binaryfile
```

If a relative pathname to the file is specified, the TFTP server searches the specified directories for the file.

The cached version of a file is only used for requests requiring the specified format. For example, the binary cached version of a file is not used in satisfying a request for the file in netascii format. If a file is to be retrieved in both binary and ASCII formats, the user must specify that two copies of the file be cached with one in binary format, and the other in netascii format.

Caching is not dynamic. The cache files are read in when the TFTP server is started and are not updated, even if the file on disk is updated. To update or refresh the cache, the TFTP server must be recycled.

-a archive directory
Specifies an archive directory. The files in this directory and its subdirectories are treated as binary files for downloading. This option is useful on EBCDIC machines that act as file servers for ASCII clients. Multiple -a options can be specified; one directory per -a option. Directories must be specified as absolute path names. You can specify no more than 20 directories.

-b IP address
Uses the specified IP address. The TFTP server usually binds to in6addr_any or inaddr_any. You can specify the IP address on which requests are to be received. TFTP requests that come in on other IP addresses will not be accepted by this instance of TFTPD.

directory
Specifies an absolute path name for a directory. You may specify no more than 20 directories on the tftpd command line.

If the TFTP server is started without a list of directories, all mounted directories are considered active.

If a list of directories is specified, only those directories specified are active. That list is used as a search path for incoming requests specifying a relative path name for a file.

Activating a directory activates all of its subdirectories.

For a file to be readable by the TFTP server, the file must be in an active directory and have world ("other") read access enabled. For a file to be writable by the TFTP server, the file must already exist in an active directory and have world ("other") write access.

Starting TFTPD as a procedure

Before you begin: Obtain a copy of the sample procedure, shipped as SEZAINST(TFTPD), and store it in one of your PROCLIB concatenation data sets.

Perform the following step to start TFTPD as a procedure:

- Invoke the procedure using the system operator start command. Following is a copy of the sample, which shows how to start TFTPD as a procedure:

```
//TFTPSD   PROC
//*
//* Communications Server IP
//* SMP/E distribution name: EZATTFDP
//*
//*   5694-A01 5655-HAL (C) Copyright IBM Corp. 1997, 2004.
//*   Licensed Materials - Property of IBM
//*   This product contains "Restricted Materials of IBM"
//*   All rights reserved.
//*   US Government Users Restricted Rights -
//*   Use, duplication or disclosure restricted by
//*   GSA ADP Schedule Contract with IBM Corp.
//*   See IBM Copyright Instructions.
//*
//* Function: Trivial File Transfer Protocol Server start
//*
//* Please note:
//*
//* -a  Specify an archive directory.  TFTPD treats files in this
//*     directory and its subdirectories as binary files for uploads
//*     and downloads, regardless of how they were requested by the
//*     client.  Use this option on EBCDIC machines that act as file
//*     servers for ASCII clients.
//*
//*     You can specify up to 20 -a options, one directory per -a
//*     option.  You must specify directories as absolute pathnames.
//*
//* -c  the number of threads to use concurrently.  Make the number
//*     a some reasonable number like 20 and not the default, which
//*     is 200.
//*
//* -t  Set the packet timeout. The TFTP server usually waits 5 seconds
//*     before presuming that a transmitted packet has been lost. You can
//*     specify a different timeout period in seconds.
//*
//* -l  Log all incoming read and write requests and associated
//*     information to the system log. Logged information includes the IP
//*     address of the requestor, the file requested, and whether the
//*     request was successful.
//*
```

```
//* -p  Specify the port. The TFTP server usually receives requests on
//*     well-known port 69. You can specify the port in which requests
//*     are to be received.
//*
//* -r  Set the retry limit. The TFTP server usually limits the number
//*     of retransmissions it performs due to lost packet to 5. You
//*     can specify a different retry limit.
//*
//* -s  Set the maximum block size that can be negotiated by the
//*     TFTP block size option. The default is 8192.
//*
//* -f  Secify a cache file. You can specify a file containing
//*     information on files to be preloaded and cached for transmission.
//*     A cache file consists of one or more entries. For clarity, place
//*     each entry on a separate line. See the IP Configuration Guide for
//*     details on this option.
//*
| //* -b  Specify IP address. The TFTP server usually binds to in6addr_any
| //*     or inaddr_any. You can specify the IP address on which requests
| //*     are to be received.
| //*     TFTP requests that come in on other IP addresses will not be
| //*     accepted by this instance of TFTPD.
| //*
//TFTP     EXEC PGM=TFTPD,REGION=0K,TIME=NOLIMIT,
//             PARM='POSIX(ON),ALL31(ON),TRAP(OFF)/
| //             -c 20 -t 300'
//*STEPLIB DD DISP=SHR,DSN=TCP.SEZALOAD,
//*           VOL=SER=,UNIT=
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT   DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
//SYSABEND DD SYSOUT=*
//         PEND
```

You know that TFTPD is starting when the following message appears on the console:

```
EZZ7001I starting
```

Stopping the TFTP server

To terminate the TFTP server, send a SIGTERM signal using the UNIX kill command to the oldest existing TFTP process. This is the process with a parent process ID of 1. Termination of this process will cause all of its children to terminate.

If multiple instances of the TFTP server are running, to determine the pid and its corresponding start options for each instance of the TFTP server, invoke the following UNIX command:

```
ps -o pid,ppid,args
```

Terminate an instance of the TFTP server using the UNIX kill command with the appropriate pid.

Chapter 13. Domain Name System (DNS)

This chapter describes the z/OS UNIX Domain Name System (DNS) name server, which uses the Berkeley Internet Name Domain (BIND) software, the accepted standard of DNS. BIND was developed at the University of California, Berkeley and is currently maintained by the Internet Software Consortium (ISC). The name servers discussed in this chapter are based on BIND 4.9.3 and BIND 9.

The Domain Name System is a client/server model in which programs called *name servers* contain information about host systems and IP addresses. Name servers provide this information to clients called *resolvers*.

z/OS V1R2 Communications Server BIND 9.1 was the first implementation of BIND 9 on the z/OS platform, which was a complete rewrite of the name server and associated utilities. This allowed IPv6-type records in zone data. It also introduced better transaction security among servers and clients, as well as zone data authentication capability. It introduced the rndc utility to replace and complement UNIX signals for name server local and remote control.

The BIND 9.2 name server was introduced in z/OS V1R4. It makes DNS server-to-server and client-to-server IPv6 connections possible, adding new name server configuration options for IPv6 connections and tuning. BIND 9.2 also provides a new rndc utility with a larger set of commands than was available with BIND 9.1. BIND 9.2 rndc is not compatible with BIND 9.1 rndc.

The configuration file for the name server has changed in both name and syntax between BIND 4.9.3 and BIND 9. The dnsmigrate tool aids in converting a BIND 4.9.3 named.boot file into a BIND 9 named.conf file. This utility is available as of z/OS V1R2.

BIND 9 nsupdate utility enables client hosts and many DHCP servers to dynamically and securely register their name and address mappings. The z/OS BIND 9 name server is generally compatible with network DHCP servers. The z/OS DHCP server is compatible with the z/OS BIND 4.9.3 name server, but is incompatible with the BIND 9 name server.

This chapter also contains information about connection optimization, which uses DNS for distributing connections among hosts or server applications within a sysplex domain.

This chapter is not intended to be a comprehensive description of DNS or of BIND. For more complete descriptions, refer to the latest edition of *DNS and BIND* by Paul Albitz and Cricket Liu (O'Reilly & Associates, Inc.).

DNS and BIND overview

TCP/IP applications map fully qualified domain names to 32-bit IPv4 IP addresses or 128-bit IPv6 addresses to identify network nodes. The z/OS BIND 9 name server supports resource records for IPv6 address mapping. It also accepts IPv6 connections, depending on the z/OS TCP/IP stack setup and profile, and on the name server configuration. While TCP/IP applications refer to host computers by their IP addresses, it is easier to use host names. To enable the use of host names in a network, the Domain Name System (DNS) translates host names to IP

addresses. Mapping must be consistent across the network to ensure interoperability. DNS provides the host name-to-IP address mapping through network server hosts called *domain name servers*. For detailed information about name servers, see "Domain name servers" on page 597. DNS can also provide other information about server hosts and networks such as the TCP/IP services available at a server host and the location of domain name servers in a network.

DNS organizes the hosts in a network into domains. A *domain* is a group of hosts that share the same name space in the domain hierarchy and are usually controlled within the same organization. Domains are arranged in a hierarchy. A special domain known as the *root domain* exists at the top of the hierarchy. The root domain servers store information about server hosts in the root domain and the name servers in the delegated, *top-level* domains, such as *com* (commercial), *edu* (education), and *mil* (military). The name servers in the top-level domain, in turn, store the names of name servers for their delegated domains, and so on.

The complete name of a host, also known as the *fully qualified domain name* (FQDN), is a series of labels separated by dots or periods. Each label represents an increasingly higher domain level within a network. The complete name of a host connected to one of the larger networks generally has more than one subdomain, as shown in the following examples:

```
host1.subdomain2a.subdomain2.rootdomain
user4720.eng.mit.edu
```

A domain name server requires the FQDN. The client resolver combines the host name with the domain name to create the FQDN before sending the name resolution request to the domain name server.

DNS also provides IP address-to-host name mapping. The DNS defines a special domain called *in-addr.arpa* to translate IPv4 addresses to host names, and the *ip6.int* and *ip6.arpa* domains for IPv6 address-to-host name translation. This kind of mapping is useful for producing output (host names) that is easy to read. An in-addr.arpa name is composed of the reverse octet order of an IP address concatenated with the *in-addr.arpa* string. For example, a host named `Host1` has `9.67.43.100` as an IP address. The in-addr.arpa domain translates the `Host1` IP address `9.67.43.100` to `100.43.67.9.in-addr.arpa`.

For IPv6 reverse lookups, BIND 9 supports the *bitstring* and *nibble* formats.

A system administrator can name the host systems and domains in the local, private network with any name you want, but to link with name servers in a public network like the Internet, you need to determine which domain you want to be in (which parent domain) and then contact the registrar in that domain to register the names and IP addresses of your name servers. This ensures that queries from outside the domain being defined can be answered by this name server if need be.

Note: Contact the InterNetwork Information Center (InterNIC) for more information about Internet registration. You can contact InterNIC by pointing your Web browser at http://www.internic.net.

Domain names

The DNS uses a hierarchical naming convention for naming hosts. Each host name is composed of domain labels separated by periods. Local network administrators have the authority to name local domains within an intranet. Each label represents

an increasingly higher domain level within an intranet. The fully qualified domain name of a host connected to one of the larger intranets generally has one or more subdomains:

- *host.subdomain.subdomain.rootdomain*
- *host.subdomain.rootdomain*

Domain names often reflect the hierarchy level used by network administrators to assign domain names. For example, the domain name `eng.mit.edu.` is the fully qualified domain name, where `eng` is the host, `mit` is the subdomain, and `edu` is the highest level domain (root domain).

Figure 63 is an example of the DNS used in the hierarchy naming structure across an intranet.

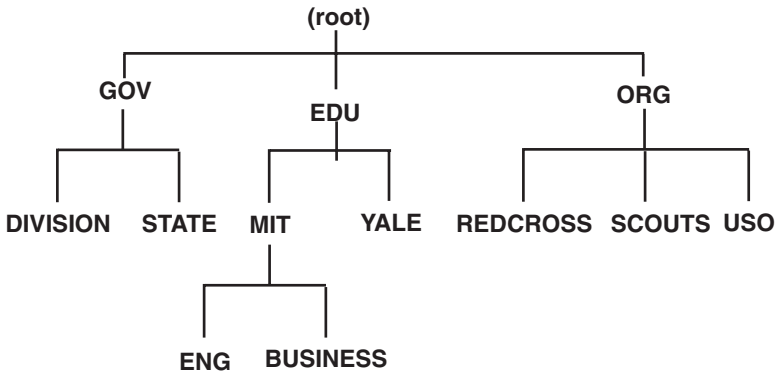


Figure 63. Hierarchical naming tree

You can refer to hosts in your domain by host name only; however, a name server requires a fully qualified domain name. The local resolver appends the domain name before sending the query to the Domain Name Server for address resolution.

Domain name servers

Domain name servers are designated network nodes that maintain a database of information about all nodes in some part of the domain name space, called a *zone*. A name server is said to be *authoritative* for its zone. A zone consists of the resources within a single domain (for example, commercial or `.com`) or subdomain (for example, `raleigh.ibm.com`). Typically, a zone is administered by a single organization or individual. The complete database is not kept by any one name server on a network. A name server is authoritative only within its zone of authority.

All host systems in a given zone share the same higher level domain name (for example, `host1.raleigh.ibm.com`, `host2.raleigh.ibm.com`, `host3.raleigh.ibm.com`, and so on). As system administrator, you create a zone of authority by listing all the host systems in your zone in the database file of the name server that is authoritative for the zone.

If a domain name server receives a query about a host for which it has information in its database or in its cache, it performs the name resolution and returns all the address records associated with the host to the client. Some hosts (for example, routers or gateways between two or more networks) might have more than one IP address.

Alternatively, the name server can query other name servers for information. This process is called *iterative resolution*. The local name server successively queries other name servers, each of which responds by referring the local name server to a remote name server that is closer to the name server authoritative for the target domain. Finally, the local name server queries the authoritative name server and gets an answer. If the information about a requested host name does not exist or if a name server does not know where to go for the information, it sends a negative response back to the client.

There are multiple name server modes in the DNS:

- Authoritative
  - Master (primary)
  - Slave (secondary)
- Caching-only servers
- Forwarders
- Stealth

A single server can perform multiple functions. For example, it can be a primary server and a slave server for different zones. The purpose of having these different kinds of servers is to provide redundancy (in case of system failure), to distribute the workload among multiple servers, to speed up the name-resolution process, and to provide flexibility in network design. In addition to being an authoritative or caching-only server, a name server can be defined to only contact a specific set of name servers if queries cannot be resolved locally (through the use of forwarders).

The following sections discuss authoritative servers, caching-only servers, and forwarding.

Authoritative servers

An authoritative server is the authority for its zone. It queries and is queried by other name servers in the DNS. The data it receives in response from other name servers is cached. Authoritative servers are not authoritative for cached data.

There are two types of authoritative servers: master (primary) and slave (secondary). Each zone must have only one master name server, and it should have at least one slave name server for backup to minimize dependency on a particular node. Calling a *particular* name server a master or slave is misleading. Any given name server can take on either or both roles, as defined by the boot or conf file.

The zone data updates and maintenance are reflected in the master name server. The slave name servers update their databases by contacting the master name server at regular intervals or possibly (BIND 9) after being notified of an update by the master name server. Both master and slave name servers are authoritative for a zone.

The zones of authority are arranged in a hierarchy based on the domain origin components. A special zone known as the *root* exists at the top of the domain name hierarchy in a network. The root zone contains a list of all the root servers. For example (see Figure 63 on page 597), in the Internet, the root name servers store information about nodes in the root domain, and information about the delegated domains, such as `com` (commercial), `edu` (education), and `mil` (military). The root name servers store the names of name servers for each of these domains, which in turn store the names of name servers for their delegated subdomains.

TCP/IP applications contact a name server whenever it is necessary to translate a domain name into an IP address, or when information is required about a domain. The name server performs the translation if it has the necessary information. If it does not have the necessary information, the name server can contact other name servers, which in turn can contact other name servers. This process is called a *recursive query*. Alternatively, a name server can simply return the address of another name server that might hold the requested information. This is called a *referral response* to a query. Name server implementations must support referrals, but are not required to perform recursive queries. See "Resolvers" on page 600 for more information about query responses.

Master name servers: A master name server maintains all the data for its zone. Static resources are kept in database files called *domain data files*. For information on creating domain data files, see "Step 5. Create the domain data files (master name server only)" on page 612. Master name servers can also receive zone updates dynamically. For information on dynamic DNS, see "Dynamic IP" on page 674. For information on dynamic generation of resources, see "Connection optimization in a sysplex domain" on page 660.

Slave name servers: A slave name server acts as an alternate to the master server if the master name server becomes unavailable or overloaded. The slave name server receives zone data directly from the master name server in a process called *zone transfer*. Zone transfers, which only occur when data has changed, are based on the refresh interval in the Start of Authority (SOA) resource record or, for BIND 9 name servers only, on using the DNS Notify function. For a description of the SOA resource record, see *z/OS Communications Server: IP Configuration Reference*. A slave server, like a master server, is authoritative for a zone.

Caching-only servers

All name servers cache (store) the data they receive in response to a query. A caching-only server, however, is not authoritative for any domain. Responses derived from cached information are flagged in the response. When a caching-only server receives a query, it checks its cache for the requested information. If it does not have the information, it queries a local name server or a root name server, passes the information to the client, and caches the answer for future queries. The names and addresses of the root name servers are acquired from the servers listed in the hints file, the name and file path of which are specified in the name server's configuration file.

You can use caching servers to create a large cache of responses to frequently requested queries and reduce the number of queries made to master servers. The caching server stores data for a period of time determined by the time-to-live (ttl) value, and the cached information is lost if the name server is restarted.

Forwarders

Normally, name servers answer queries from cached data or, if that does not succeed, they attempt to contact other name servers identified in their data files as authoritative for certain domains. However, name servers can also be configured to contact special servers called *forwarders* before contacting the name servers listed in their data files. If a forwarder cannot process the query and if the local name server is not a forward-only name server, the local name server contacts the name servers in its data files. A forward-only name server relies completely on its forwarders. It does not try to contact other servers to find out information if the forwarders do not give it an answer.

The forwarding function is useful for reducing the number of queries to servers on the Internet and for creating a large cache of information on forwarders. It is also a useful function for providing Internet access for local servers that, for one reason or another, do not have access themselves.

Stealth server

A *stealth server* is a server that answers authoritatively for a zone, but is not listed in that zone's NS records. Stealth servers can be used as a way to centralize distribution of a zone, without having to edit the zone on a remote nameserver. When the master file for a zone resides on a stealth server in this way, it is often referred to as a *hidden primary* configuration. Stealth servers can also be a way to keep a local copy of a zone for rapid access to the zone's records, even if all official nameservers for the zone are inaccessible.

Resolvers

Programs that query a name server are called *resolvers*. Because many TCP/IP applications need to query the name server, a set of routines is usually provided for application programmers to perform queries. On z/OS, these routines are available in the resolver provided by z/OS Communications Server.

z/OS Communications Server provides programs for interactively querying a name server:

- NSLOOKUP (TSO)
- onslookup/nslookup (z/OS UNIX)
- DIG (TSO)
- dig (z/OS UNIX)
- host

Note: The nsupdate program also makes queries to name servers as part of its operations.

For information on these programs, see *z/OS Communications Server: IP System Administrator's Commands*.

The BIND 4 onslookup and TSO DIG commands use the resolver provided by z/OS Communications Server for all their resolver facilities. The BIND 9 onslookup and dig commands use the resolver initialization facilities of the resolver provided by z/OS Communications Server, but use their own resolver for any additional resolver facilities needed.

Resolver directives for nslookup

The onslookup program uses the following resolver directives (TCPIP.DATA statements):

- domain/domainorigin
- search
- nameserver/nsinteraddr
- sortlist
- options debug/options ndots

Resolver directives for dig

The dig program uses the following resolver directives (TCPIP.DATA statements):

- domain/domainorigin
- search

- nameserver/nsinteraddr
- options ndots

Query Packets

Resolvers operate by sending query packets to a name server, either over the network or to the local name server.

A query packet contains the following fields:

- Domain name
- Query type
- A query class

For information on valid query class (network class) and query type (data type) values, see *z/OS Communications Server: IP Configuration Reference*. The name server attempts to match the three fields of the query packet to its database. For flexibility, the following wildcard query types are defined:

Type	Description
ANY	Indicates any record type for the domain name.
AXFR	Indicates the query type used by secondary name servers to transfer all records in the zone. (The query class is set to IN when using the AXFR query type.)
MAILB	Indicates any mailbox records for the domain name.

The name server can return the following query responses:

Response	Description
Authoritative	Is returned from a primary or secondary name server. The name server contains all the domain data used to define the zone for the specified query.
BADVERS	The name server received a request which contained a bad EDNS version.
Nonauthoritative	Is returned from a cache kept by a name server. The cache does not contain the domain data used to define the zone for the specified query.
Format Error	The name server found an error in the query packet sent by the resolver.
Name Error	No resource records of any type (including wildcards) exist for the domain name specified.
NXDOMAIN (negative)	No records of the requested type were found for the domain name specified.
Not-implemented	The name server does not support the type of query requested.
NOTAUTH	The name server is not authoritative for the zone.
NOTZONE	A dynamic update failed because the name to be updated is not contained within the given zone.
NXRRSET	A dynamic update failed because the prerequisites were not satisfied. The Resource Record set existed when the prerequiste stated it should not.

Response	Description
Referral	Contains the addresses of other name servers that might be able to answer the query. A referral response is returned when a recursive query is not supported, not requested, or cannot be answered because of network connectivity.
Refused	The name server refuses to perform the specified operation. For example, some root name servers limit zone transfers to a set number of IP addresses.
YXDOMAIN	DNAME mapping failed because the new name was too long.
YXRRSET	A dynamic update failed because the prerequisites were not satisfied. The Resource Record set did not exist when the prerequiste stated it should.

Resource Records

Data from a name server is stored and distributed in a format known as a resource record. Resource record fields are described in detail in *z/OS Communications Server: IP Configuration Reference*. Each response from a name server can contain several resource records, which can contain a variety of information. The format of a response is defined in RFC 1035. It includes the following sections:

- A question section, echoing the query for which the response is returned.
- An answer section, containing resource records matching the query.
- An additional section, containing resource records that do not match the query, but might provide useful information for the client. For example, the response to a query for the host name of a name server for a specific zone includes the IP address of that name server in the additional section.
- An authority section, containing information specific to the type of response made to the query. If a referral is returned, this section contains the domain names of name servers that could provide an authoritative answer. If a negative response is returned indicating the name does not exist, this section contains a Start Of Authority (SOA) record defining the zone of authority of the responding name server.

Recommended reading

DNS and BIND, 4th Edition by Paul Albitz and Cricket Liu (O'Reilly & Associates, Inc.) gives a comprehensive description of DNS and BIND, and specifically contains information on BIND 9.1.0. The BIND 9 name server was upgraded in V1R4 and is now based upon BIND 9.2.0.

For additional information on DNS in a sysplex, refer to the following Redbooks:

- *z/OS eNetwork Communications Server V2R7 TCP/IP Implementation Guide Volume 2: UNIX Applications, SG24-5228*
- *TCP/IP in a Sysplex, SG24-5235-01*

For information on Dynamic IP, see "Dynamic IP" on page 674.

If you wish to participate, a BIND users mailing list can be subscribed to at `http://www.isc.org/products/BIND/`

DNS protocols are described in various Request for Comments (RFC) papers and Internet drafts. RFCs outline existing protocols, suggest new protocols, and

establish standards for the Internet protocol suite. Internet drafts are proposals, techniques, and mechanisms that document Internet Engineering Task Force (IETF) work-in-progress.

For information about obtaining RFCs, refer to `http://ftp.ietf.org/rfc.html`.

A list of RFCs related to DNS are located in "DNS-related RFCs" on page 714.

Performance issues

The BIND 9 name server supports multithreading and DNSSEC which creates extra overhead. Multithreading might improve performance for large zones but can be a disadvantage for small zones. The BIND 9 name server might perform slower than the BIND 4.9.3 name server for small zones on simple query and response operations, or might be as fast as a BIND 4.9.3 name server but with higher CPU consumption, depending on whether your system is already CPU constrained.

Because of the multithreading, BIND 9 name servers are able to answer queries during zone transfers. BIND 4.9.3 name servers are unable to answer queries for a period of time during zone transfers; the larger the zone, the more noticeable this can become. BIND 9 name servers are also capable of Incremental Zone Transfers, while BIND 4.9.3 name servers are not. Incremental Zone Transfer allows only the changed information in a zone to be sent to slave name servers instead of the entire zone. If your name servers employ dynamic update for frequent zone changes, the Incremental Zone Transfer feature of BIND 9 might offer some performance advantages while reducing network traffic.

The use of DNSSEC (authenticating DNS data with digital signatures) will have a performance cost. The authentication process requires more CPU, and signing a zone greatly increases the zone's size. DNS message sizes will also increase between client and server, and between DNS servers. If the message size becomes too large for UDP, the message will be sent by TCP, which is more resource intensive.

Since the BIND 9 name server is multithreaded, it can take advantage of any additional processors you add to the system. The BIND 9 name server will detect the number of logical CPUs configured for the system (if not running partitioned) or LPAR (if running partitioned), and create additional worker threads accordingly. For simply configured name servers that are small, are not using DNSSEC, or are not kept busy, the overhead in managing the extra threads created on a multiprocessor image can actually be disadvantageous. If you feel this might be the case, you can override the number of worker threads created by using the *-n* option when starting the name server. The number of logical CPUs detected (and therefore, the number of worker threads created by default) is logged when the name server is started.

Compatibility considerations

Zone transfers

It is not recommended to have a BIND 4.9.3– DNS act as a slave name server to a BIND 9 master. However, if required, it can be done under certain conditions. If the BIND 9 master contains any resource records (RRs) that BIND 4.9.3–DNS does not understand, the zone will fail to load. A BIND 9 nameserver with BIND 4 slaves should specify the 'transfer-format one-answer' option in its named.conf. Otherwise, any transfer will also fail.

BIND 4.9.3 does not understand NOTIFY, therefore, if BIND 4.9.3 is running as a slave name server to a BIND 9 master, the DNS Change Notification protocol will not work. The standard method of zone transfers applies, where the slave periodically polls the master for an updated SOA serial number.

Queries

BIND 4.9.3 can participate in DNS queries when BIND 9 name servers are in the DNS tree structure, even when the queries are for RR types that BIND 4.9.3 does not understand. For example, if a resolver is pointed to a BIND 4.9.3–DNS and is asked for an AAAA record (an RR type that BIND 4.9.3 name server does not understand), BIND 4.9.3–DNS will recursively query other (possibly BIND 9) name servers and return the answer to the client. The BIND 4.9.3–DNS will not cache the response (in the case of RR types it does not understand), which may mean a little more network traffic. This caching issue may or may not be significant depending on your particular network traffic patterns.

Dynamic update

BIND 4.9.3 version of Dynamic Update is incompatible with the BIND 9 version. For dynamic update on BIND 4.9.3-DNS, use nsupdate with -V v4 start option. For dynamic update on BIND 9-DNS, use nsupdate with -V v9 start option. Additionally, only the DHCP server on z/OS and OS/2 can successfully dynamically update the BIND 4.9.3–DNS.

DNSSEC

BIND 4.9.3 does not support DNSSEC.

TSIG

BIND 4.9.3 does not support TSIG security which may be used on queries, update and zone transfers on BIND 9 name servers.

DNS/WLM (Sysplex connection balancing)

Cannot be done by a name server running in BIND 9 mode.

IPv6 support

Not supported by a BIND 4.9.3 name server.

Stack affinity

The BIND 9 name server is a generic server, unlike the BIND 4.9.3 name server which has stack affinity. If stack affinity is desired for the BIND 9 name server, use the _BPXK_SETIBMOPT_TRANSPORT environment variable.

NOTIFY

This function notifies the slave of a change in the master. It is not supported by a BIND 4.9.3 name server.

Running the name server in BIND 9 and BIND 4.9.3 mode simultaneously

The following describes the tasks involved in setting up the name server to run in BIND 9 and BIND 4.9.3 modes simultaneously:

1. Bind each name server to its own set of IP interfaces.

a. Bind the BIND 9 name server to the set of IPv4 interfaces you wish to be serviced by the BIND 9 name server using the 'listen-on{}' option in the named.conf file. The BIND 4.9.3 name server is unable to listen on IPv6 interfaces while the BIND 9 name server can. Therefore, IPv6 interfaces cannot be shared among the BIND 4.9.3 and BIND 9 name servers. To further specify interfaces for queries, transfers and notifies from a BIND 9 name server, the following BIND 9 configuration file options are also available:
   - query-source{}
   - transfer-source{}
   - notify-source{}

b. The BIND 4.9.3 name server cannot specify its own IPv4 interfaces in its configuration file but it will listen, query and transfer on the IPv4 interfaces not taken by the BIND 9 name server.

Both servers can share port 53 if they connect to clients or other servers on different IP addresses.

2. Specify port ownership.

   Assign unique job names to the BIND 4.9.3 and BIND 9 name servers. Reserve port 53 for TCP for both job names. Port 53 TCP port reservation jobname must have a suffix of '2' for BIND 4.9.3 and a suffix of '1' for BIND 9. UDP port reservation for multiple job names is not allowed.

3. Use the _BPX_JOBNAME environment variable if starting the name server from the z/OS UNIX shell to distinguish the job names of the two name server daemons.

   Notes:

   a. MVS jobname cannot be *named* for both BIND 4 and BIND 9 servers (for example, call them namedv4 and namedv9).
   b. PORT 53 UDP can only be reserved for one jobname because of a TCP/IP profile restriction.
   c. Jobname/step is unpredictable if the name server is directly started from the z/OS UNIX shell.
   d. Whether started from an MVS procedure or the z/OS UNIX shell, the port can be generically reserved to UNIX applications: PORT 53 TCP (also UDP) OMVS.

4. Store the name server process IDs (PIDs) in unique files.

   Configure the BIND 9 name server to store the process ID (PID) in a file other than the one used for the BIND 4.9.3 name server (/etc/named.pid). This is done with the 'pid-file' named.conf file option.

5. Configure client resolvers.

   Configure clients' resolver configuration data set or HFS file so that it points to the interface or interfaces of the desired name server. This is typically specified by the NSINTERADDR statement, or an equivalent statement. For more information on how to configure the resolver, refer to "Understanding resolvers" on page 20.

Setting up and running the name server

This section describes the tasks involved in configuring the name server and verifying that the name server is working correctly.

Name server configuration files are arranged in a Hierarchical File System (HFS). Before configuring DNS, the TSO user ID from which the name server is started must have the proper authority to access the name server configuration and zone files. For a complete description of file permissions within the HFS, refer to *z/OS UNIX System Services Planning*.

Configuring a master (primary) name server

The name resolution process is an example of a client/server relationship in which clients, through their resolvers, request a service (name resolution) from name servers. For a general overview of name servers, see "Domain name servers" on page 597.

The following summary lists the steps for configuring a master server or a caching-only server:

1. Create a configuration file for your environment. Select a) or b):
   a. Create the boot file for BIND 4.9.3–DNS .
   b. Create the configuration file for BIND 9–DNS.
2. For BIND 4.9.3 DNS only — Specify stack affinity (multiple stack environment).
3. Specify port ownership.
4. Update the name server start procedure.
5. Create the domain data files (master name server only).
6. Create the hints (root server) file.
7. Create the loopback file.
8. For BIND 9 only - configure logging.
9. Ensure that the syslog daemon is running on your system.
10. Specify whether the name server is to run swappable or nonswappable.
11. Start the name server.
12. Verify that the name server started correctly.
13. Verify that the name server can accept queries.

The difference between configuring a master (primary) name server and slave (secondary) and caching-only servers is the creation of domain data files (the database files containing host-to-address and address-to-host mappings). The domain data files are maintained on the master name server, and the slave name server transfers this data to its own database. Examples of slave, caching-only, and forward-only configurations are in "Configuring a slave name server" on page 629, "Configuring a cache-only name server" on page 632, and "Adding forwarding to your name server" on page 635.

Note: Continue with selecting Step 1a. *or* 1b.

Step 1a. Create the boot file for BIND 4.9.3–DNS

The boot file is the main configuration file for a domain name server. The named daemon reads the boot file for information about how to set up the local name server. The records in the boot file identify the type of name server, the zones over which it has authority, the location of data for setting up its name resolution database, and other configuration options. The default name of the boot file is `/etc/named.boot`. You can specify an alternate boot file using the `-b` named start option. For information about named options, refer to *z/OS Communications Server: IP Configuration Reference*.

Note: The named daemon reads the boot file only when the named daemon starts or when it receives a SIGHUP signal. For a description of named signals, refer to *z/OS Communications Server: IP Configuration Reference*.

Each type of name server has a special boot file configuration. You create a boot file using directives. Refer to *z/OS Communications Server: IP Configuration Reference* for more information.

Boot files created locally for use by the name server are assumed to be in code page IBM-1047. For systems using other code pages, use the `iconv` command to translate from the local code page to code page IBM-1047. See *z/OS UNIX System Services Command Reference* for more information.

A boot file for a master name server (a name server that maintains all the data for its zone in database files) will need, at a minimum, to specify the zones for which the name server will be authoritative, their locations in the HFS, and the location of the hints file (the location of root name servers). A loopback file is also recommended.

This example illustrates a name server acting as master for the forward zone `mycorp.com` and the reverse zone `34.37.9.in-addr.arpa`.

The sample BIND 4.9.3–DNS boot file shipped in /usr/lpp/tcpip/samples/named.boot is shown below. Refer to the program directory for its location.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
; (C) COPYRIGHT International Business Machines Corp. 1985, 1993
; All Rights Reserved
; US Government Users Restricted Rights - Use, duplication or
; disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
;
; Licensed Materials - Property of IBM
;
;
;       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
;
; INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
; EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
; WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
; LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
; PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
; OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
; IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
; YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
; CORRECTION.
;
;  Note:  This file must be copied and renamed to /etc/named.boot and all
;         zone files referenced below must be copied to /etc/dnsdata/ for
;         this file to function as intended.
;
;  /etc/named.boot
;
;      boot file for name server
;
;
;type    domain                  source file or host
;
directory /etc/dnsdata     {1}
primary  mycorp.com   {2}        db.mycorp.v4        {3}
```

```
primary  34.37.9.in-addr.arpa    db.34.37.9.v4   {4} {5}
primary  0.0.127.in-addr.arpa    db.loopback.v4      {7}
cache  .                       db.cache              {6}
options query-log     {8}
```

This boot file specifies:

1. The location of the files (/etc/dnsdata).
2. The name server will be the primary name server for the `mycorp.com` zone.
3. The data for `mycorp.com` is contained in `db.mycorp.v4`.
4. The name server will be the primary name server for the reverse mapping zone, `34.37.9.in-addr.arpa`.
5. The data for addresses `9.37.34.x` contained in the zone will be specified in `db.34.37.9.v4`.
6. The list of root name servers is in `db.cache`.
7. The name server defines the loopback address in the `0.0.127.in-addr.arpa` zone and the data is contained in the file, `db.loopback.v4`.
8. All queries coming in to this name server will be logged in the syslog daemon output file.

Step 1b. Create the configuration file for BIND 9–DNS.

The sample BIND 9-DNS configuration file shipped in /usr/lpp/tcpip/samples/named.conf is shown below. Refer to the program directory for its location. All zone data files referenced within the sample configuration file, with the exception of the hints file, can be found in the samples directory. To obtain the hints file, follow the instructions in "Step 6. Create the hints (root server) file" on page 618.

```
#           LICENSED MATERIALS - PROPERTY OF IBM
#             "RESTRICTED MATERIALS OF IBM"
#          5694-A01 (C) COPYRIGHT IBM CORP. 2001
#
# (C) COPYRIGHT International Business Machines Corp. 1985, 1993
# All Rights Reserved
# US Government Users Restricted Rights - Use, duplication or
# disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
#
# Licensed Materials - Property of IBM
#
#
#       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
#
# INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
# EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
# WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
# LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
# PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
# OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
# IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
# YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
# CORRECTION.
#
# Note:  This file must be copied and renamed to /etc/named.conf and all
#        all zone files referenced below must be copied to /etc/dnsdata/ for
#        this file to function as intended.  In addition, the default location
#        for the process id file is in /var/run/pid.file; if that directory
#        does not exist a different one can be configured with the option:
#
#        pid-file "path/file-name";
#
```

```
#  /etc/named.conf
#
#        conf file for name server
#

options {
   directory "/etc/dnsdata";
};

logging {
   category "queries" {
      default_syslog;
   };
};

zone "mycorp.com" in {
   type master;
   file "db.mycorp.v9";
};

zone "34.37.9.in-addr.arpa" in {
   type master;
   file "db.34.37.9.v9";
};

zone "0.0.127.in-addr.arpa" in {
   type master;
   file "db.loopback.v9";
};

zone "." in {
   type hint;
   file "db.cache";
};
```

Step 2. For BIND 4.9.3–DNS only: specify stack affinity (Multiple stack environment)

In a multiple stack environment, the name server is like any application. It binds to the stack specified by TCPIPJOBNAME. To run multiple name servers, each must use a different TCPIPJOBNAME. See "Considerations for multiple instances of TCP/IP" on page 72 for more information about specifying TCPIPJOBNAME.

Note: When changing TCPIPJOBNAME, any client needing to be connected to the new stack name must be restarted (for example, any application, including the name server, that is currently running, that is bound to the new stack name specified by TCPIPJOBNAME).

Step 3. Specify port ownership

The name server uses a single port (53) for TCP and UDP sessions. Both servers can share port 53 if they connect to clients or other servers on different IP addresses:

- A BIND 9 server can specify IP addresses to listen on and from which to send queries, notifies and zone transfers in its configuration file.
- A BIND 4 server cannot specify IP addresses in its configuration file but will use any IP address not already selected by a BIND 9 server.

To specify port ownership when using the named start procedure for BIND 4.9.3, if NAMED is started using BPXBATCH, add the following statements to the PROFILE.TCPIP data set:

```
PORT
  53 TCP NAMED2
  53 UDP NAMED2
```

To specify port ownership when using the named start procedure for BIND 9, add the following statements to the PROFILE.TCPIP data set:

```
PORT
  53 TCP NAMED1
  53 UDP NAMED1
```

Notes:

1. MVS jobname cannot be *named* for both BIND 4 and BIND 9 servers (for example, call them namedv4 and namedv9).
2. The jobname on the TCP and UDP port reservation statements requires a suffix of *2* for BIND 4.9.3, when NAMED is started using BPXBATCH.
3. The jobname on the TCP and UDP port reservation statements requires a suffix of *1* for BIND 9.
4. PORT 53 UDP can only be reserved for one jobname because of a TCP/IP profile restriction.
5. Jobname/step is unpredictable if the name server is directly started from the z/OS UNIX shell.
6. Whether started from an MVS procedure or the z/OS UNIX shell, the port can be generically reserved to UNIX applications: PORT 53 TCP (also UDP) OMVS.

For more information on the PORT statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Note: In order to pick up changes in the PROFILE.TCPIP data set, stop and restart TCP/IP. As an alternative to stopping the stack, use the VARY TCPIP,,OBEYFILE command to reserve the ports while the stack is up.

Step 4. Update the name server start procedure (Optional)

When choosing to start the name server from MVS, create a start procedure. This is not necessary if the name server is started from the z/OS UNIX shell. Move the sample start procedure, `SEZAINST(NAMED)`, to a recognized PROCLIB. Specify name server parameters and change the data set names as required to suit local configuration. The boot file path (for BIND 4.9.3–DNS) or the conf file path (for BIND 9–DNS) can also be changed as shown below in the sample start procedures. If you want to have NAMED messages written out to SYSLOGD (HFS file) instead of the system console (syslog), then you must start NAMED via BPXBATCH as shown below:

BIND 4.9.3–DNS:

```
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: EZANSPRO
//*
//*        Licensed Materials - Property of IBM
//*        This product contains "Restricted Materials of IBM"
//*        5647-A01 (C) Copyright IBM Corp. 1997, 2000.
//*        All rights reserved.
//*        US Government Users Restricted Rights -
//*        Use, duplication or disclosure restricted by
//*        GSA ADP Schedule Contract with IBM Corp.
//*        See IBM Copyright Instructions.
//*
//*        NAMED can be started with a variety of parameters.
```

```
//*        In this example, the "-b" parameter describes which
//*        boot file NAMED should be started with.
//*
//NAMED    PROC B='/etc/named.boot'
//NAMED    EXEC PGM=BPXBATCH,REGION=0K,TIME=NOLIMIT,
//         PARM='PGM /usr/lpp/tcpip/sbin/named -b &B '
//*
//*        NAMED can use certain environmental variables, such
//*        as NLSPATH (to determine the location of the message
//*        catalog), and RESOLVER_CONFIG (to determine the location
//*        of the file that contains the parameter TCPIPjobname).
//*        These variables can be specified in a file defined
//*        by STDENV.
//*        An example of the contents of this file follows:
//*
//*             RESOLVER_CONFIG=//'SYS1.TCPPARMS(TCPDATA2)'
//*                          or
//*             RESOLVER_CONFIG=/etc/resolv.conf.tcp2
//*                          or
//*             _BPXK_SETIBMOPT_TRANSPORT-TCPCS
//*                          or
//*             DNS_VERSION=v4
//*
//*        Define STDENV with the name of the file that contains
//*        the environmental variables to be used for this
//*        invocation of NAMED.
//*
//*STDENV   DD PATH='/etc/named.env',
//*            PATHOPTS=(ORDONLY)
//*STDENV  DD DSN=SAMPLE.NAMED(ENV&SYSCLONE),DISP=SHR
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
```

BIND 9–DNS:

```
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: EZANSPR9
//*
//*        Licensed Materials - Property of IBM
//*        This product contains "Restricted Materials of IBM"
//*        5694-A01 (C) Copyright IBM Corp. 2001.
//*        All rights reserved.
//*        US Government Users Restricted Rights -
//*        Use, duplication or disclosure restricted by
//*        GSA ADP Schedule Contract with IBM Corp.
//*        See IBM Copyright Instructions.
//*
//*        NAMED can be started with a variety of parameters.
//*        In this example, the "-c" parameter describes which
//*        configuration file NAMED should be started with.
//*
//NAMED    PROC C='/etc/named.conf'
//NAMED    EXEC PGM=BPXBATCH,REGION=0K,TIME=NOLIMIT,
//         PARM='PGM /usr/lpp/tcpip/sbin/named -c &C '
//*
//*        NAMED can use certain environmental variables, such
//*        as NLSPATH (to determine the location of the message
```

```
//*        catalog), and RESOLVER_CONFIG (to determine the location
//*        of the file that contains the parameter TCPIPjobname).
//*        These variables can be specified in a file defined
//*        by STDENV.
//*        An example of the contents of this file follows:
//*
//*             RESOLVER_CONFIG=//'SYS1.TCPPARMS(TCPDATA2)'
//*                          or
//*             RESOLVER_CONFIG=/etc/resolv.conf.tcp2
//*                          or
//*             DNS_VERSION=v9
//*
//*        Define STDENV with the name of the file that contains
//*        the environmental variables to be used for this
//*        invocation of NAMED.
//*
//*STDENV   DD PATH='/etc/named.env',
//*            PATHOPTS=(ORDONLY)
//*STDENV  DD DSN=SAMPLE.NAMED(ENV&SYSCLONE),DISP=SHR
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
```

Step 5. Create the domain data files (master name server only)

The domain data files contain information about a domain, such as the IP addresses and names of the hosts in the domain for which the master name server is authoritative. The *forward* domain data file contains entries that provide forward mapping (host names-to-IP addresses for each host system in the zone) as well as additional information about system resources. The *reverse* domain data file contains entries that provide reverse mapping (IP addresses-to-host names). A separate reverse domain data file for each network (or subnet) in a domain can be created. Definition of WLM and dynamic zones is covered in "Advanced BIND 4.9.3–Name server topics" on page 660. DNS/WLM is only supported by the BIND 4.9.3 name server.

Note: The TSO user ID from which the name server is started must have the proper authority to access the name server configuration and zone files. For a complete description of file permissions within the HFS, see the *z/OS UNIX System Services Planning* (SC28–1890–02).

Naming of domain data files is flexible. For convenience in maintaining the database files, it is common to give them names such as *db.extension*, where *extension* identifies the domain of the data contained within. This document uses this convention. It also uses the .v4 and .v9 suffixes to indicate the appropriate nameserver version and the suffix *.bak* to specify a slave backup file.

Use the following to create domain data files:

- Control entries
- Resource records
- Special characters

Note: Refer to *z/OS Communications Server: IP Configuration Reference* for more information about these files.

BIND 4.9.3–DNS: The sample forward domain file, /usr/lpp/tcpip/samples/db.mycorp.v4, is listed below. Continuing the example that started in the boot file, the file would be /etc/dnsdata/db.mycorp.v4.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;
; (C) COPYRIGHT International Business Machines Corp. 1985, 1993
; All Rights Reserved
; US Government Users Restricted Rights - Use, duplication or
; disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
;
; Licensed Materials - Property of IBM
;
;
;       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
;
; INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
; EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
; WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
; LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
; PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
; OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
; IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
; YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
; CORRECTION.
;
;
;  /etc/dnsdata/db.mycorp.v4
;        name server zone data
;
$ORIGIN com.
mycorp         IN      SOA   ns1.mycorp admin.mycorp (                        {1}
             1               ; Serial (incremented when database is changed
             10800           ; Refresh (slave will check every 3 hours
             3600            ; Retry  (retry every hour after refresh failure
             604800          ; Expire (slave gives up retry after one week
             86400 )         ; Time to Live (data cached in other servers 1 day
;
$ORIGIN mycorp.com.                                                           {2}
; define domain nameservers
                   IN      NS      ns1                                        {3}
                   IN      NS      ns2

; example delagation of a subdomain
;intranet           IN      NS      ns1.intranet
;intranet           IN      NS      ns2.intranet
;ns1.intranet       IN   A    9.37.35.10
;ns2.intranet       IN   A    9.37.35.11

_http._tcp              SRV  0  0 80   www.mycorp.com.                       {4}
                        SRV  10 0 8000 www2.mycorp.com.                     {4}
_http._tcp.w3           SRV  0  0 80   www.mycorp.com.                       {5}
                        SRV  10 0 8000 www2.mycorp.com.                     {5}

localhost          IN   A    127.0.0.1
ns1                IN   A    9.37.34.10
ns2                IN   A    9.37.34.11

gateway            IN   A    9.37.34.30                                     {6}
                   IN   A    9.37.35.30                                     {6}

host1              IN   A    9.37.34.1
host2              IN   A    9.37.34.2
host3            IN   A    9.37.34.3
```

```
host4              IN   A    9.37.34.4

www2               IN   A    9.37.34.5
www                IN   A    9.37.34.6
www                IN   A    9.37.34.7

mail               IN   CNAME ns1                                        {7}
ftp                IN   CNAME ns2

{1}
    The SOA (Start of Authority) record specifies the name server ns1 as the
authoritative name server for the domain mycorp.com. The mail address of the
person responsible for domain data is admin@mycorp.com.
The numbers enclosed in parentheses are parameters used to set different values
 for the zone.

{2}
   The control entry $ORIGIN appends the string mycorp.com. to all the following
 host names that do not end with a dot ('.').

{3}
   The NS (Name Server) records specify the name servers in the zone. Note that
NS records do not distinguish between primary and secondary name servers.

{4}
  The SRV records specify the location for the 'http' service using the 'tcp'
protocol.  The first record has a priority of 0, a weight of 0, uses port 80 and
the service is provided at host, www.mycorp.com.  The second record has a priority
of 10 which is lower, a different port and target.  A web client capable of using
SRV records requesting http://mycorp.com/ would be directed to www.mycorp.com and
www2.mycorp.com.  The client would be responsible for determining which site to
connect to first based first on priority and then on weight.

{5}
  The SRV record also specifies the location for the 'http' service using the
'tcp' protocol.  A web client capable of using SRV records requesting
http://w3.mycorp.com/ would also be directed www.mycorp.com and www2.mycorp.com.

{6}
    These A (Address) records map the host name (gateway.mycorp.com) to the IP
 addresses of the two networks to which it is connected.

{7}
    The CNAME record specifies that the name mail is an alias for the host name
ns1.mycorp.com.
```

BIND 9–DNS: The sample forward domain file, /usr/lpp/tcpip/samples/db.mycorp.v9, is listed below. Continuing the example that started in the boot file, the file would be /etc/dnsdata/db.mycorp.v9.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;
; (C) COPYRIGHT International Business Machines Corp. 1985, 1993
; All Rights Reserved
; US Government Users Restricted Rights - Use, duplication or
; disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
;
; Licensed Materials - Property of IBM
;
;
;       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
;
; INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
; EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
; WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
```

```
; LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
; PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
; OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
; IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
; YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
; CORRECTION.
;
;
;  /etc/dnsdata/db.mycorp.v9
;        name server zone data
;
; Default TTL value
$TTL 86400                                                                  {1}
$ORIGIN com.
mycorp         IN      SOA   ns1.mycorp admin.mycorp (                      {2}
             1               ; Serial (incremented when database is changed)
             10800           ; Refresh (slave will check every 3 hours)
             3600            ; Retry  (retry every hour after refresh failure)
             604800          ; Expire (slave gives up retry after 1 week)
             86400 )         ; Negative caching (NXDOMAIN/NXRRSET responses, 1 day){3}
;
$ORIGIN mycorp.com.                                                         {4}
; define domain nameservers
                   IN      NS      ns1                                      {5}
                   IN      NS      ns2

; example delagation of a subdomain
;intranet           IN      NS      ns1.intranet
;intranet           IN      NS      ns2.intranet
;ns1.intranet       IN   A    9.37.35.10
;ns2.intranet       IN   A    9.37.35.11

_http._tcp              SRV  0  0 80   www.mycorp.com.                     {6}
                        SRV  10 0 8000 www2.mycorp.com.                   {6}
_http._tcp.w3           SRV  0  0 80   www.mycorp.com.                     {7}
                        SRV  10 0 8000 www2.mycorp.com.                   {7}

localhost          IN   A    127.0.0.1
ns1                IN   A    9.37.34.10
ns2                IN   A    9.37.34.11

gateway            IN   A    9.37.34.30                                   {8}
                   IN   A    9.37.35.30

host1              IN   A    9.37.34.1
host2              IN   A    9.37.34.2
host3              IN   A    9.37.34.3
host4              IN   A    9.37.34.4

www2               IN   A    9.37.34.5
www                IN   A    9.37.34.6
www                IN   A    9.37.34.7

;IPv6 addresses
www                IN   AAAA 3ffe:8050:201:1860:42::1                      {9}
www                IN   A6   0 3ffe:8050:201:1860:42::1                    {10}

mail               IN   CNAME ns1                                          {11}
ftp                IN   CNAME ns2
```

{1}
The rules for time-to-live values have been complicated somewhat in BIND v9. If named finds a $TTL directive it follows TTL semantics defined in RFC 2308, which states that records not explicitly setting a TTL inherit the TTL from the $TTL value. If there is no $TTL set, it follows TTL semantics from RFCs 1034 and 1035, which state that records with no explicit TTL inherit one from the previous record. This implies that to follow RFC 1034/1035 semantics, the SOA RR must set its TTL value. For simplicity, it is recommended that you always specify a $TTL value. This line sets the default TTL for all records to 86400 seconds (one day).

{2}
The SOA (Start of Authority) record specifies the name server ns1 as the

```
authoritative name server for the domain mycorp.com. The mail address of the
person responsible for domain data is admin@mycorp.com. The numbers enclosed
in parentheses are parameters used to set different values for the zone.

{3}
     The meaning of the last SOA value has changed from BIND v4 to v9.  It
now represents length of time other servers should cache negative responses
from this zone.  This line sets that value to 86400 seconds (1 day).

{4}
    The control entry $ORIGIN appends the string mycorp.com. to all the
following host names that do not end with a dot ('.').

{5}
    The NS (Name Server) records specify the name servers in the zone. Note
that NS records do not distinguish between primary and secondary name servers.

{6}
  The SRV records specify the location for the 'http' service using the
'tcp' protocol.  The first record has a priority of 0, a weight of 0, uses
port 80 and the service is provided at host, www.mycorp.com.  The second record
has a priority of 10 which is lower, a different port and target.  A web client
capable of using SRV records requesting http://mycorp.com/ would be directed to
www.mycorp.com and www2.mycorp.com.  The client would be responsible for
determining which site to connect to first based first on priority
and then on weight.

{7}
 The SRV record also specifies the location for the 'http' service using the
'tcp' protocol.  A web client capable of using SRV records requesting
http://w3.mycorp.com/ would also be directed www.mycorp.com and www2.mycorp.com.

{8}
    These A (Address) records map the host name (gateway.mycorp.com) to the IP
addresses of the two networks to which it is connected.

{9}
    The AAAA IPv6 record type sets the IPv6 address of www.mycorp.com to
3ffe:8050:201:1860:42::1.

{10}
    The A6 IPv6 record type is an experimental way of specifying IPv6 addresses.
This record sets the address of www.mycorp.com to 3ffe:8050:201:1860:42::1. The '0'
indicates that the address value is fully qualified (starts at bit 0).  See the
most recent edition of DNS and BIND by Cricket Liu and Paul Albitz (O'Reilly and
Associates, Inc.) for more information on A6 records.

{11}
    The CNAME record specifies that the name mail is an alias for the host name
ns1.mycorp.com.
```

BIND 4.9.3–DNS: The sample reverse domain file /usr/lpp/tcpip/samples/db.34.37.9.v4 is listed below. Continuing the example that started in the boot file, the file would be /etc/dnsdata/db.34.37.9.v4.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;   /etc/dnsdata/db.34.37.9.v4
;
$ORIGIN 37.9.in-addr.arpa.

34 IN SOA ns1.mycorp.com. admin.mycorp.com. (
              1  10800  3600 604800 86400 )


34              IN      NS     ns1.mycorp.com.
34              IN      NS     ns2.mycorp.com.
$ORIGIN 34.37.9.in-addr.arpa.
```

```
10            IN      PTR    ns1.mycorp.com.
11            IN      PTR    ns2.mycorp.com.

1             IN      PTR    host1.mycorp.com.
2             IN      PTR    host2.mycorp.com.
3             IN      PTR    host3.mycorp.com.
4             IN      PTR    host4.mycorp.com.
5             IN      PTR    www2.mycorp.com.
6             IN      PTR    www.mycorp.com.
7             IN      PTR    www.mycorp.com.

20            IN      PTR    printserver.mycorp.com.
```

Note: Data files created locally for use by the name server are assumed to be in code page IBM-1047. For systems using other code pages, use the `iconv` command to translate from the local code page to code page IBM-1047. Refer to *z/OS UNIX System Services Command Reference* for more detailed information about this command. Files read through a network connection (for example, secondary data files) are converted to IBM-1047 by the name server before they are written to the local file system.

FTP can also be used to convert the files to code page IBM-1047.

BIND 9–DNS: The sample reverse domain file /usr/lpp/tcpip/samples/db.34.37.9.v9 is listed below. Continuing the example that started in the boot file, the file would be /etc/dnsdata/db.34.37.9.v9.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;   /etc/dnsdata/db.34.37.9.v9
;
; Default TTL value
$TTL 86400
$ORIGIN 37.9.in-addr.arpa.

34 IN SOA ns1.mycorp.com. admin.mycorp.com. (
              1  10800  3600 604800 86400 )


34              IN      NS     ns1.mycorp.com.
34              IN      NS     ns2.mycorp.com.
$ORIGIN 34.37.9.in-addr.arpa.
10            IN      PTR    ns1.mycorp.com.
11            IN      PTR    ns2.mycorp.com.

; Build similar reverse lookup records
$GENERATE 1-4 $ PTR host$.mycorp.com.                                    {1}
; The following records are generated by the above $GENERATE directive.
;1             IN      PTR    host1.mycorp.com.
;2             IN      PTR    host2.mycorp.com.
;3             IN      PTR    host3.mycorp.com.
;4             IN      PTR    host4.mycorp.com.
5             IN      PTR    www2.mycorp.com.
6             IN      PTR    www.mycorp.com.
7             IN      PTR    www.mycorp.com.

20            IN      PTR    printserver.mycorp.com.

{1}
  $GENERATE is a v9-specific directive that is useful for creating a series of
records that differ only by an iterator.  This line prompts the name server to
create the records listed below upon zone load.  For more information on
$GENERATE, refer to the z/OS Communications Server:  IP Configuration Reference.
```

Note: Data files created locally for use by the name server are assumed to be in code page IBM-1047. For systems using other code pages, use the `iconv` command to translate from the local code page to code page IBM-1047. Refer to *z/OS UNIX System Services Command Reference* for more detailed information about this command. Files read through a network connection (for example, secondary data files) are converted to IBM-1047 by the name server before they are written to the local file system.

FTP can also be used to convert the files to code page IBM-1047.

Step 6. Create the hints (root server) file

The hints file contains the names and IP addresses of the authoritative root domain name servers. The root name servers contain the names of name servers in the top-level domains such as com, edu, and mil. The name server uses root server information when deciding which name server to contact when it receives a query for a host outside its zone of authority and it does not have the data in its cache.

Note: The hints file does not contain cached data nor does the name server provide other hosts with the information contained in the hints file. A forward-only server is the only type of name server that does not require a hints file.

To obtain a hints file, point your Web browser at `ftp://ftp.rs.internic.net` and retrieve the file `named.root` from the `domain` subdirectory. Update your hints file on a regular basis.

The `cache` directive in a BIND 4.9.3 boot file specifies the path and name of the hints file.

The hints file in a BIND 9 config file is specified with a zone{} statement of type 'hints'.

An example of a hints file originally copied from `ftp://ftp.rs.internic.net/domain/named.root` is listed below. Continuing the example that began in the boot and conf files, the file would be /etc/dnsdata/db.cache.

```
;   This file holds the information on root name servers needed to
;   initialize cache of Internet domain name servers
;   (e.g., reference this file in the "cache   .   <file>"
;   or zone "." { type hint; file "db.cache"; };
;   in a v4 or v9 config file, respectively.
;   configuration file of BIND domain name servers).
;
;   This file is made available by InterNIC registration services
;   under anonymous FTP as
;   file                    /domain/named.root
;   on server               FTP.RS.INTERNIC.NET
;   -OR- under Gopher** at RS.INTERNIC.NET
;       under menu          InterNIC Registration Services (NSI)
;         submenu           InterNIC Registration Archives
;       file                named.root
;
;   last update:       May 19, 1997
;   related version of root zone:      1997051700
;
;   formerly NS.INTERNIC.NET
;
.                       3600000   IN  NS   A.ROOT-SERVERS.NET.
A.ROOT-SERVERS.NET.     3600000       A    198.41.0.4
;
```

```
;   formerly NS1.ISI.EDU
;
.                       3600000       NS   B.ROOT-SERVERS.NET.
B.ROOT-SERVERS.NET.   3600000       A    128.9.0.107
;
;   formerly C.PSI.NET
;
.                       3600000       NS   C.ROOT-SERVERS.NET.
C.ROOT-SERVERS.NET.   3600000       A    192.33.4.12
;
;   formerly TERP.UMD.EDU
;
.                       3600000       NS   D.ROOT-SERVERS.NET.
D.ROOT-SERVERS.NET.   3600000       A    128.8.10.90
;
;   formerly NS.NASA.GOV
;
.                       3600000       NS   E.ROOT-SERVERS.NET.
E.ROOT-SERVERS.NET.   3600000       A    192.203.230.10
;
;  formerly. NS.ISC.ORG
;
.                       3600000       NS   F.ROOT-SERVERS.NET.
F.ROOT-SERVERS.NET.   3600000       A    192.5.5.241
;
;   formerly NS.NIC.DDN.MIL
;
.                       3600000       NS   G.ROOT-SERVERS.NET.
G.ROOT-SERVERS.NET.   3600000       A    192.112.36.4
;
;   formerly AOS.ARL.ARMY.MIL
;
.                       3600000       NS   H.ROOT-SERVERS.NET.
H.ROOT-SERVERS.NET.   3600000       A    128.63.2.53
;
;   formerly NIC.NORDU.NET
;
.                       3600000       NS   I.ROOT-SERVERS.NET.
I.ROOT-SERVERS.NET.   3600000       A    192.36.148.17
;
;   temporarily housed at NSI (InterNIC)
;
.                       3600000       NS   J.ROOT-SERVERS.NET.
J.ROOT-SERVERS.NET.   3600000       A    198.41.0.10
;
;   housed in LINX, operated by RIPE NCC
;
.                       3600000       NS   K.ROOT-SERVERS.NET.
K.ROOT-SERVERS.NET.   3600000       A    193.0.14.129
;
;   temporarily housed at ISI (IANA)
;
.                       3600000       NS   L.ROOT-SERVERS.NET.
L.ROOT-SERVERS.NET.   3600000       A    198.32.64.12
;
;   temporarily housed at ISI (IANA)
;
.                       3600000       NS   M.ROOT-SERVERS.NET.
M.ROOT-SERVERS.NET.   3600000       A    198.32.65.12
; End of File
```

Step 7. Create the loopback file

The loopback file contains the loopback address. This is the address that a host uses to route queries to itself. The preferred loopback address is `127.0.0.1`, although you can configure additional loopback interfaces in the PROFILE.TCPIP.

For BIND 4.9.3, DNS will bind to 127.0.0.1 in addition to the first loopback address configured in PROFILE.TCPIP. BIND 9 mode requires the availability to bind onto loopback address 127.0.0.1.

BIND 9 and BIND 4 server modes coexistence on the same host requires different loopback addresses. If BIND 9 and BIND 4.9.3 are to be started on the same host, once BIND 9 mode is started, BIND 4 mode server will have to use a different loopback address. BIND 4 server master forward and reverse zone files should define A and PTR resource records, respectively, for the appropriate loopback addresses.

This guide uses the extension `.loopback` to specify the loopback file.

Note: In addition to creating the loopback file, add an address resource record called *localhost* to the forward domain data file. This record supports proper two-way resolution.

Use the following elements to create the loopback file:

- Control entries
- Resource records
- Special characters

Note: Refer to *z/OS Communications Server: IP Configuration Reference* for more information about these files.

BIND 4.9.3–DNS: The sample loopback file for BIND 4.9.3 shipped in /usr/lpp/tcpip/samples/db.loopback.v4 is listed below. Continuing the example that started in the boot and .conf files, the file would be /etc/dnsdata/db.loopback.v4.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;   /etc/dnsdata/db.loopback.v4
;
0.0.127.in-addr.arpa. IN SOA  ns1.mycorp.com. admin.mycorp.com. (
     1
     10800
     3600
     604800
     86400   )

0.0.127.in-addr.arpa.   IN   NS  ns1.mycorp.com.
0.0.127.in-addr.arpa.   IN   NS  ns2.mycorp.com.
1.0.0.127.in-addr.arpa. IN   PTR localhost.
```

For DNS/WLM and BIND 4.9.3: To get the sysplex domain name, add the following PTR record to the loopback file in the loopback zone:

128.0.0.127.in-addr.arpa. IN PTR Sysplex_Domain_Name.

Sysplex_Domain_Name is the domain name of the sysplex (specified as cluster zone in the master boot file). Do not forget to put a period (.) after the Sysplex_Domain_Name. Note that all of the sysplex name servers must be updated with this change.

BIND 9–DNS: The sample loopback file for BIND 9 shipped in /usr/lpp/tcpip/samples/db.loopback.v9 is listed below. Continuing the example that started in the boot file, the file would be /etc/dnsdata/db.loopback.v9.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
;   /etc/dnsdata/db.loopback.v9
;
; Default TTL value
$TTL 86400
0.0.127.in-addr.arpa. IN SOA  ns1.mycorp.com. admin.mycorp.com. (
     1
     10800
     3600
     604800
     86400   )

0.0.127.in-addr.arpa.   IN   NS  ns1.mycorp.com.
0.0.127.in-addr.arpa.   IN   NS  ns2.mycorp.com.
1.0.0.127.in-addr.arpa. IN   PTR localhost.
```

A separate loopback file is required for use with the IPv6 loopback address (::1). The following shows an example named.conf configuration and the associated zone file.

```
IPv6 loopback master zone simple definition in named.conf: zone
"1.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.ip6.arpa" { type master; file "loopback.v6"; };

loopback.v6 zone file (with implied domain origin from above master zone definition)

$TTL 86400
@   IN SOA ns1.mycorp.com. admin.mycorp.com. (
     1
     10800
     3600
     604800
     7200 )
        NS ns1.mycorp.com.
        PTR localhost.
```

Step 8. For BIND 9 only — configure logging

A wide variety of logging options for the nameserver can be configured via the *logging* statement. Its *channel* phrase associates output methods, format options and severity levels with a name that can then be used with the *category* phrase to select how various classes of messages are logged.

Only one *logging* statement is used to define as many channels and categories as are wanted. If there is no logging statement, the logging configuration will be:

```
logging {
     category "default" { "default_syslog"; "default_debug"; };
  };
```

In BIND 9, the logging configuration is only established when the entire configuration file has been parsed. When the server is starting up, all logging messages regarding syntax errors in the configuration file go to the default channels. Therefore, if started from a procedure, the logging messages will be written to syslogd. If started from z/OS UNIX, the logging messages may be written to 'named.run' if started with the -d option, in addition to syslog.

All log output goes to one or more channels; you can make as many of them as you want. Every channel definition must include a clause that says whether messages selected for the channel go to a file, to a particular syslog facility, or are

discarded. It can optionally also limit the message severity level that will be accepted by the channel (the default is *info*), and whether to include a named-generated time stamp, the category name, the severity level, and the thread ID (the default is to include all).

Messages written to logging files can be buffered according to the value on the *max-buffered-messages* options statement. The default value is the maximum allowed value of 35. Buffering messages to logging files will provide some amount of a performance advantage. However, it might be misleading when viewing a logging file while the name server is running, since the most recent logging information might not have been written yet to the file.

The word *null* specified as the destination option for the channel will cause all messages sent to it to be discarded; in that case, other options for the channel are meaningless.

The *file* names the HFS path name for the log file and can include limitations, both on how large the file is allowed to become and how many versions of the file will be saved each time the file is opened.

The *size* option for files is simply a hard ceiling on log growth. If the file ever exceeds the size and the version is zero, then named will not write anything more to it until the file is reopened, which will be done after the file is renamed or erased. If version option value is 1 or more, the current file, when full, is renamed with a suffix and another file is opened with the original name. In the latter case, logging will continue in a round robin fashion using the current file name and the suffixed file names. The default behavior is not to limit the size of the file. Note that if debug is enabled, the logs can grow large very quickly and you may run the risk of filling up your HFS. Therefore, it is recommended that you limit the size of the file when debugging is enabled.

If you use the *version* log file option, then named will retain that many backup versions of the file by renaming them when opening. For example, if you choose to keep 3 old versions of the file *lamers.log* then just before it is opened *lamers.log.1* is renamed to *lamers.log.2* , *lamers.log.0* is renamed to *lamers.log.1* , and *lamers.log* is renamed to *lamers.log.0* . No rolled versions are kept by default; any existing log file is simply appended. The *unlimited* keyword is synonymous with 99 in current BIND releases.

Example usage of the size and versions options:

```
channel "an_example_channel" {
        file "example.log" versions 3 size 20m;
        print-time yes;
        print-category yes;
        print-threadid yes;
    };
```

The argument for the *syslog* clause is a syslog facility. Refer to the *z/OS Communications Server: IP Configuration Reference* for more detailed parameter information.

The severity clause works like syslog's priorities, except that they can also be used if you are writing straight to a file rather than using syslog . Messages which are not at least of the severity level given will not be selected for the channel; messages of higher severity levels will be accepted. Severity level decreases from critical down to info, and further decreases from debug 1 down to debug 99.

If you are using syslog , then the syslog.conf priorities will also determine what eventually passes through. For example, defining a channel facility and severity as *daemon* and *debug* but only logging daemon.warning via syslog.conf will cause messages of severity *info* and *notice* to be dropped. If the situation were reversed, with named writing messages of only warning or higher, then syslogd would print all messages it received from the channel.

The server can supply extensive debugging information when it is in debugging mode. If the server's global debug level is greater than zero, then debugging mode will be active. The global debug level is set by starting the named server with the -d flag followed by a positive integer. All debugging messages in the server have a debug level, and higher debug levels give more detailed output. The maximum debug level is 99. Channels that specify a specific debug severity, for example:

```
channel "specific_debug_level" {
      file "foo";
      severity debug 3;
  };
```

will get debugging output of level 3 or less any time the server is in debugging mode, regardless of the global debugging level. Channels with *dynamic* severity use the server's global level to determine what messages to print.

The *print-* options can be used in any combination.

If the following is turned on . . .	Then the following is logged . . .
print-time	date and time
print-category	category of the message
print-severity	severity level of the message
print-threadid	the thread ID that is issuing the message

Note: *print-time* can be specified for a syslog channel, but is usually pointless since syslog also prints the date and time.

The *print-* options are always printed in the following order: time, category, severity, and thread ID. Here is an example:

```
Apr 24 09:28:05.848 queries: info: 0a923850: EZZ8828I client 127.0.0.1#20021:
                    query: host1.mycorp.com IN A
```

There are four predefined channels that are used for named's default logging as follows:

```
channel "default_syslog" {
        syslog daemon;                          // send to syslog's daemon
                                                // facility
        severity info;                          // only send priority info
                                                // and higher
    };
    channel "default_debug" {
        file "named.run";                       // write to named.run in
                                                // the working directory
                                                // Note: stderr is used instead
                                                // of "named.run"
                                                // if the server is started
                                                // with the '-f' option.
        severity dynamic                        // log at the server's
                                                // current debug level
    };
    channel "default_stderr" {                  // writes to stderr
        file "<stderr>";                        // this is illustrative only;
```

```
                                                // there's currently no way of
                                                // specifying an internal file
                                                // descriptor in the
                                                // configuration language.
        severity info;                          // only send priority info
                                                // and higher
    };
    channel "null" {
       null;                                    // toss anything sent to
                                                // this channel
};
```

The *default_debug* channel normally writes to a file named.run in the server's working directory. For security reasons, when the -u command line option is used, the named.run file is created only after named has changed to the new UID, and any debug output generated while named is starting up and still running as root is discarded.

Once a channel is defined, it cannot be redefined. Thus you cannot alter the built-in channels directly, but you can modify the default logging by pointing categories at channels you have defined.

There are many categories, so you can send the logs you want to see wherever you want, without seeing logs you do not want. If you don't specify a list of channels for a category, then log messages in that category will be sent to the *default* category instead. If you don't specify a default category, the following "default default" is used:

```
category "default" { "default_syslog"; "default_debug"; };
```

As an example, say you want to log security events to a file, but you also want to keep the default logging behavior. You would specify the following:

```
channel "my_security_channel" {
    file "my_security_file";
    severity info;
};
category "security" {
    "my_security_channel";
    "default_syslog";
    "default_debug";
};
```

To discard all messages in a category, specify the null channel:

```
category "xfer-out" { "null"; };
category "notify" { "null"; };
```

In the following example:

- Four channels have been defined to make it possible to browse and keep logs for some categories separately. In theory, each category can log to one or more different channels, but keeping the number of channels to a minimum is recommended.
- Most existing categories have been specifically associated with one or more channels to demonstrate logging flexibility. However, categories directed to the main log only may be omitted and instead, covered by the default category. One exception is the *queries* category, which requires a specific channel association to enable queries logging.
- Every channel log entry will be prefixed with time stamp, category, severity, and thread ID.

Note: The default for all print- options is *yes*.

- Every channel has been customized for maximum file size and for keeping 2 archived files in addition to the active log file. Make sure the disk has enough space for the total maximum size of active and archived log files, and extra space for any other growing logs or files.
- Every channel but transfer_log will log messages up to debug level 99 which is the suggested detailed level to gather problem documentation but can fill up logging files quickly. Lower debug levels (e.g. 11, info, error) may be used for normal operation.
- - transfer_log is shown at debug level 7, where minimal zone transfer starting and stopping activity is recorded. Increasing level to 8 and above will considerably increase logging activity, mainly for large zones with one-answer transfer format.
- ″severity dynamic;″ can also be set for any channel, in which case debug level is determined by named -d start option value.
- The default_debug channel can be specified instead of one or more user defined channels, in which case logging goes to ″named.run″ file in the named working directory. No maximum file size will stop logging to named.run.

```
logging {
 channel main_log {
      file "/tmp/named_main.log" versions 2 size 20M;
      print-time yes;
      print-category yes;
      print-severity yes;
      print-threadid yes;
      severity debug 99;
                 };
 channel security_log {
      file "/tmp/named_security.log" versions 2 size 1M;
      severity info;
      severity debug 99;
                    };
 channel query_log {
      file "/tmp/named_query.log" versions 2 size 10M;
      severity info;
      severity debug 99;
                 };
 channel transfer_log {
      file "/tmp/named_transfer.log" versions 2 size 10M;
      severity debug 7;
                     };


     category client   { main_log; };
     category config   { main_log; };
     category database { main_log; };
     category dispatch { main_log; };
     category dnssec { security_log; main_log; };
     category general  { main_log; };
     category network  { main_log; };
     category notify   { main_log; };
     category resolver { main_log; };
     category security { security_log; main_log; };
     category update   { main_log; };
     category queries  { query_log;};
     category lame-servers { query_log; main_log; };
     category xfer-in  { "transfer_log"; };
     category xfer-out { "transfer_log"; };
     category default  { main_log; };
     category unmatched { main_log; };
        };
```

More detail about the available categories and brief descriptions of the types of log information they contain can be found in the *z/OS Communications Server: IP Configuration Reference*.

Step 9. Ensure that the syslog daemon is running on your system

The name server uses the syslog daemon to log messages. To verify that the name server starts correctly or to diagnose problems, the syslog daemon should be running.

Guideline: For BIND 9 only, unless syslogd is running, no messages will be produced by NAMED during name server initialization. This includes event logging and any syntax errors that might be detected in the configuration file. Not performing this step complicates problem determination, especially for failures at startup.

If your syslog daemon is not configured, see "Creating the syslog file" on page 637 for information regarding the syslog daemon.

Step 10. Specify whether the name server is to run swappable or nonswappable

You might want to run the name server in a swappable state, as it has in the past. This is an optional step. Keep in mind that when an application makes an address space nonswappable, it might convert additional real storage in the system to preferred storage. Because preferred storage cannot be configured offline, allowing the name server to run in a nonswappable state can reduce the installation's ability to reconfigure storage in the future.

If you want to run the name server as swappable, you must have the "BPX.STOR.SWAP" FACILITY class profile defined to RACF with no universal access. To do this, enter the following commands from a RACF user ID.

```
RDEFINE FACILITY BPX.STOR.SWAP UACC(NONE)
SETROPTS RACLIST(FACILITY) REFRESH
```

For BIND 4.9.3 only, the name server will also run as swappable if the "BPX.STOR.SWAP" FACILITY is not defined and the name server is started from a user ID with a UID not equal to 0.

If you want the name server to run in a nonswappable state, do one of the following:

- Do not define the BPX.STOR.SWAP facility to RACF and start the name server from a user ID with a UID=0. For BIND 9, the name server must be started from a user ID with UID=0, but for BIND 4.9.3, you have a choice in this matter.
- Define the facility to RACF and allow the appropriate users at least READ access to the facility.

The latter method can be accomplished with the following set of commands:

```
RDEFINE FACILITY BPX.STOR.SWAP UACC(NONE)
PERMIT BPX.STOR.SWAP CLASS(FACILITY) ID(userid) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

Step 11. Start the name server

Your name server is ready to start. Start the name server using the following methods:

- An authorized TSO ID can start a name server from the MVS operator's console by starting the named start procedure. If the boot or config file path is not `/etc`,

specify the correct path in the start procedure. Refer to *z/OS Communications Server: IP Configuration Reference* for start options. A sample start procedure is provided with the product and is found in SEZAINST(NAMED4) for BIND 4.9.3 and in SEZAINST(NAMED9) for BIND 9.

- A user ID with superuser authority can start the name server from the shell by starting z/OS UNIX, then issuing the `named` command and, optionally, any parameters.
- It is also possible to start the server automatically when z/OS UNIX is started by specifying the path and file name of the z/OS UNIX initialization shell script in the `/etc/init.options` file using the `-sc` option:

  ```
  -sc /etc/rc       shell script = /etc/rc
  ```

  The file `/etc/rc` is the default z/OS UNIX initialization shell script that is executed when z/OS UNIX is started. Information such as the following can be entered in `/etc/rc`:

  ```
  FOR BIND 4.9.3-DNS

  # Start name server
  /usr/lpp/tcpip/sbin/named -b /named/production/named.boot &

  FOR  BIND 9-DNS

  # Start name server
  /usr/lpp/tcpip/sbin/named -c /named/production/named.conf &
  ```

  Port 53 may be reserved for the name server in the PROFILE.TCPIP data set. For directions on specifying port ownership, see "Step 3. Specify port ownership" on page 609. Only when in BIND 4 mode, when the stack to which named binds is started, named completes initialization. The BIND 9 name server may only be started after TCP/IP is up. In rare circumstances, the BIND 9 name server may complete initialization before all of the stack's interfaces have been brought up. In this case, the name server will not be listening on all desired interfaces. Eventually, the name server will scan the stack interfaces again, and begin listening on all desired interfaces. The default time period for this rescan is one minute. You can have the name server rescan the interfaces at the interval you desire by specifying the "interface-interval" option in the named.conf file.

- If you are starting the BIND 4.9.3 name server, or you are starting the BIND 9 name server in a single stack environment, use the AUTOLOG statement to start the name server automatically during initialization with z/OS UNIX running. Insert the name of the named start procedure in the AUTOLOG statement of the PROFILE.TCPIP data set.

  ```
  AUTOLOG
    NAMED
  ENDAUTOLOG
  ```

  The JOBNAME keyword should not be added to the AUTOLOG statement for named.

  If you are starting the BIND 9 name server in a multiple stack environment, use some other automation outside of AUTOLOG to automatically start the name server, since it is a generic server. For more information on the AUTOLOG statement, refer to *z/OS Communications Server: IP Configuration Reference*.

  Note: Named cannot be started from `INETD`.

Step 12. Verify that the name server started correctly

BIND 4.9.3–DNS: After starting the name server, ensure that no errors occurred when it was started. Look at the syslog daemon output data set for name server messages. If startup is successful, messages similar to the following are displayed:

```
named[22]: EZZ6698I name server starting. @(#) ddns/ns/ns_main.c,
           dns_ns, dns_r1.1 1.62 9/23/97 10:57:21
named[22]: EZZ6701I named established affinity with 'TCPCS'
named[22]: EZZ6540I Static primary zone 'raleigh.ibm.com' loaded (serial 1)
named[22]: EZZ6540I Static primary zone '34.37.9.inaddr.arpa' loaded (serial 1)
named[22]: EZZ6540I Static primary zone '0.0.127.inaddr.arpa' loaded (serial 1)
named[22]: EZZ6540I Static cache zone '' loaded (serial 0)
named[23]: EZZ6475I named: ready to answer queries.
```

To correct errors, either stop and restart the name server to pick up the changes, or reload the name server with the -SIGHUP signal.

To stop the name server from the z/OS UNIX shell, issue:

```
kill -TERM $(cat /etc/named.pid)
```

To stop the name server from the MVS console, issue the following:

```
p named3
(Use the name of the procedure that is currently active. This is
usually the proc name that was used to start the name server, followed
by a '1'  for  BIND 9 server, by a '3' for BIND 4 server, due to extra forking
steps on startup)
```

To reload the BIND 4.9.3 name server with a signal, issue the following command from the z/OS UNIX shell:

```
kill -HUP $(cat /etc/named.pid)
```

BIND 9–DNS: After starting the name server, ensure that no errors occurred when it was started. Look at the syslog daemon output data set for name server messages. If startup is successful, messages similar to the following are displayed:

```
Mar 26 ... mvsw named[...29]: EZZ9172I VM mode detected. Using 1 CPU(s) for -n option
Mar 26 ... mvsw named[...56]: EZZ9547I starting named, BIND 9.2.0 -c /etc/namedgm.conf
Mar 26 ... mvsw named[...56]: EZZ9095I STARTING NAMED, BIND 9.2.0
Mar 26 ... mvsw named[...56]: EZZ9217I Running non-swappable
Mar 26 ... mvsw named[...56]: EZZ9540I using 1 CPU
Mar 26 ... mvsw named[...56]: EZZ9126I loading configuration from '/etc/namedgm.conf'
Mar 26 ... mvsw named[...56]: EZZ8842I the default for the 'auth-nxdomain' option is now 'no'
Mar 26 ... mvsw named[...56]: EZZ9052I no IPv6 interfaces found
Mar 26 ... mvsw named[...56]: EZZ9046I listening on IPv4 interface VLINK1, 9.67.116.122#53
Mar 26 ... mvsw named[...56]: EZZ9046I listening on IPv4 interface TR1, 9.67.113.75#53
Mar 26 ... mvsw named[...56]: EZZ9046I listening on IPv4 interface loopback127, 127.0.0.1#53
Mar 26 ... mvsw named[...56]: EZZ9111I command channel listening on 9.67.113.75#953
Mar 26 ... mvsw named[...56]: EZZ9130I NAMED, BIND 9.2.0 IS RUNNING
```

To stop the name server from the z/OS UNIX shell, issue:

```
kill -TERM $(cat /etc/named.pid)
```

To stop the name server from the MVS console, issue the following:

```
p named1
(Use the name of the procedure that is currently active. This is
usually the proc name that was used to start the name server, followed
by a '1'  for  BIND 9 server, by a '3' for BIND 4 server, due to extra
forking steps on startup)
```

To reload the BIND 9 name server with a signal, issue the following command from the z/OS UNIX shell:

```
kill -HUP $(cat bind9_pid_file)
```

where *bind9_pid_file* is derived from pid-file option in named v9 configuration file.

rndc can also be used to reload or stop the server. Refer to "Remote Name Daemon Control (rndc)" on page 638 for more details on the rndc command.

Step 13. Verify the name server can accept queries

When the name server is up with no logged errors, ensure that it can accept queries. Ensure that the name server can accept queries locally from both the MVS and z/OS UNIX environments. In order to correctly set up these environments, see "Understanding search orders of configuration information" on page 28 for instructions.

After the resolver configuration is correct, test with the nslookup or dig command. An example using nslookup follows.

Issue the following command from both the z/OS UNIX shell and the TSO ready prompt. In the following example, the name 'host1.mycorp.com.' is used for the search.

Note: Choose any name in the domain you have defined.

```
nslookup host1.mycorp.com
```

Using the sample files in this example, the following should be the result when the command is issued:

```
$ nslookup host1.mycorp.com
Defaulting to nslookup version 4
Starting nslookup version 4
Server:  localhost
Address:  127.0.0.1

Name:    host1.mycorp.com
Address:  9.37.34.1
```

Configuring a slave name server

After setting up a working master name server, one or more slave name servers can be set up. This process is very similar as configuring a master name server. The differences are in the boot file for BIND 4.9.3 and the conf file for BIND 9 and the absence of the domain data files.

For example, see "Configuring a master (primary) name server" on page 606 to configure a slave server for the forward and reverse mapping zones. The steps are identical to the steps for configuring a master name server, except for step 1. For step 1, more information is included below in the subsections following the steps.

1. Create a configuration file for your environment:
   a. Create the boot file for BIND 4.9.3–DNS. See "Step 1a. Create the boot file for BIND 4.9.3–DNS" on page 630.
   b. Create the configuration file for BIND 9–DNS. See "Step 1b. Create the configuration file for BIND 9–DNS." on page 631.
2. See "Step 2. For BIND 4.9.3–DNS only: specify stack affinity (Multiple stack environment)" on page 609.
3. See "Step 3. Specify port ownership" on page 609.
4. See "Step 4. Update the name server start procedure (Optional)" on page 610.
5. See "Step 5. Create the domain data files (master name server only)" on page 612.

6. See “Step 6. Create the hints (root server) file” on page 618.
7. See “Step 7. Create the loopback file” on page 619.
8. See “Step 8. For BIND 9 only — configure logging” on page 621.
9. See “Step 9. Ensure that the syslog daemon is running on your system” on page 626.
10. See “Step 10. Specify whether the name server is to run swappable or nonswappable” on page 626.
11. See “Step 11. Start the name server” on page 626.
12. See “Step 12. Verify that the name server started correctly” on page 628.
13. See “Step 13. Verify the name server can accept queries” on page 629.

The difference between configuring a master and slave name server is the creation of domain data files (the database files containing host-to-address and address-to-host mappings). The domain data files are maintained on the master name server, and the slave name server transfers this data to its own database.

Instructions for creating the boot file for a slave name server are below. All remaining steps are identical to those in “Configuring a master (primary) name server” on page 606.

Step 1a. Create the boot file for BIND 4.9.3–DNS

The easiest way to create a boot file for a slave name server is to start from the boot file for the master name server. The sample boot file, /usr/lpp/tcpip/samples/slave.boot (based on named.boot), reflects the setup for a slave name server and is shown below.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2001
;
; (C) COPYRIGHT International Business Machines Corp. 1985, 1993
; All Rights Reserved
; US Government Users Restricted Rights - Use, duplication or
; disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
;
; Licensed Materials - Property of IBM
;
;
;       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
;
; INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
; EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
; WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
; LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
; PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
; OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
; IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
; YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
; CORRECTION.
;
; Note:  This file must be copied and renamed to /etc/named.boot and all
;        all zone files referenced below must be copied to /etc/dnsdata/ for
;        this file to function as intended.  Also note this example is
;        built on the assumption that there is a master server for the zones
;        at 9.37.34.10.
;
;  /etc/named.boot
;
;        boot file for name server
```

```
;
;
;type    domain                  source file or host
;
directory   /etc/dnsdata
secondary   mycorp.com          9.37.34.10     db.mycorp.bak
secondary   34.37.9.in-addr.arpa 9.37.34.10     db.34.37.9.bak
primary     0.0.127.in-addr.arpa db.loopback.v4
cache       .                   db.cache
options query-log
```

This boot file specifies:

1. The location of the files (/etc/dnsdata).
2. The name server will be the slave name server for the `mycorp.com` zone.
3. The primary name server is located at IP address `9.37.34.10`.
4. The data for `mycorp.com` will be stored in /etc/dnsdata/db.mycorp.bak after it is retrieved from the primary name server by doing a zone transfer.
5. The name server will be the slave name server for the reverse mapping zone, `34.37.9.in-addr.arpa`.
6. The primary name server is located at IP address `9.37.34.10`.
7. The data for addresses `9.37.34.x` contained in the zone will be stored in /etc/dnsdata/db.34.37.9.bak after it is retrieved from the primary name server by doing a zone transfer.
8. The name server will continue to function in the primary role for the loopback address (`127.0.0.1`).
9. The list of root name servers is in /etc/dnsdata/db.cache.
10. All queries coming in to this name server will be logged in the syslog daemon output file.

Step 1b. Create the configuration file for BIND 9–DNS.

This example illustrates the equivalent configuration for a v9 name server where the sample configuration file, /usr/lpp/tcpip/samples/slave.conf (based on named.conf), reflects the setup for a slave name server. All zone data files referenced within the sample configuration file, with the exception of the hints file, can be found in the samples directory. To obtain the hints file, follow the instructions in “Step 6. Create the hints (root server) file” on page 618.

```
#           LICENSED MATERIALS - PROPERTY OF IBM
#             "RESTRICTED MATERIALS OF IBM"
#          5694-A01 (C) COPYRIGHT IBM CORP. 2000
#
# (C) COPYRIGHT International Business Machines Corp. 1985, 1993
# All Rights Reserved
# US Government Users Restricted Rights - Use, duplication or
# disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
#
# Licensed Materials - Property of IBM
#
#
#       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
#
# INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
# EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
# WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
# LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
# PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
# OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
```

```
# IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
# YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
# CORRECTION.
#
# Note:  This file must be copied and renamed to /etc/named.conf and all
#        all zone files referenced below must be copied to /etc/dnsdata/ for
#        this file to function as intended.  Also note this example is
#        built on the assumption that there is a master server for the zones
#        at 9.37.34.10.  In addition, the default location for the process
#        id file is in /var/run/pid.file; if that directory does not exist
#        a different one can be configured with the option:
#
#        pid-file "path/file-name";
#
#  /etc/named.conf
#
#        conf file for name server
#
options {
  directory "/etc/dnsdata";
};

logging {
  category "queries" {
    default_syslog;
  };
};

zone "mycorp.com" in {
  type slave;
  file "db.mycorp.bak";
  masters { 9.37.34.10; };
};

zone "34.37.9.in-addr.arpa" in {
  type slave;
  file "db.34.37.9.bak";
  masters { 9.37.34.10; };
};

zone "0.0.127.in-addr.arpa" in {
  type master;
  file "db.loopback.v9";
};

zone "." in {
  type hint;
  file "db.cache";
};
```

Configuring a cache-only name server

If the name server does not need to be authoritative for any data, choose a special type of name server, called a caching-only name server. A caching-only name server can improve performance by reducing the number of network flows required for names or addresses that are frequently requested.

Use the following steps to configure a basic name server. More information for step 1 is included in the subsections following the steps. For the subsequent steps, see the appropriate sections elsewhere in the book.

1. Create a configuration file for your environment:
   a. Create the boot file for BIND 4.9.3–DNS. See “Step 1a. Create the boot file for BIND 4.9.3–based DNS..”
   b. Create the configuration file for BIND 9–DNS. See “Step 1b. Create the configuration file for BIND 9–DNS.” on page 634.
2. See “Step 2. For BIND 4.9.3–DNS only: specify stack affinity (Multiple stack environment)” on page 609.
3. See “Step 3. Specify port ownership” on page 609.
4. See “Step 4. Update the name server start procedure (Optional)” on page 610.
5. See “Step 6. Create the hints (root server) file” on page 618.

   Following is an example of a hints (root server) file for a cache-only server:

   ```
   ;
   ;  Cache-only "hints" file
   ;
   .                       3600000 IN NS   hostname
   hostname.              3600000    A    ipaddress
   ```

   where *hostname* is the fully qualified host name of an authoritative name server for the root (’.’) domain and *ipaddress* is the IP address for the specified hostname. How this file is configured depends on whether you are behind a firewall or not. If behind a firewall, hostname should be the name of an internal root name server if internal roots are being used. If you are behind a firewall and not using internal roots, then requests are probably being forwarded to a name server on a bastion host, which can resolve internal and internet names. In the latter case, what is in the hints file is unimportant since it will not be used, and if the name server does attempt to use it, the firewall would block it from contacting the internet root name servers. If you are not behind a firewall, follow the example in “Step 6. Create the hints (root server) file” on page 618 and instructions on getting a recent copy of the internet root name servers.
6. See “Step 7. Create the loopback file” on page 619.
7. See “Step 8. For BIND 9 only — configure logging” on page 621.
8. See “Step 9. Ensure that the syslog daemon is running on your system” on page 626.
9. See “Step 10. Specify whether the name server is to run swappable or nonswappable” on page 626.
10. See “Step 11. Start the name server” on page 626.
11. See “Step 12. Verify that the name server started correctly” on page 628.
12. See “Step 13. Verify the name server can accept queries” on page 629.

The difference between configuring a master name server and configuring a caching-only server is the creation of domain data files (the database files containing host-to-address and address-to-host mappings). A caching-only name server will only contain the loopback zone file and the hints file.

Step 1a. Create the boot file for BIND 4.9.3–based DNS.

The easiest way to create a boot file for a caching-only name server is to use the boot file for the master name server. The sample boot file /usr/lpp/tcpip/samples/caching.boot (based on named.boot), reflects the setup for a caching-only name server and is shown below.

```
;           LICENSED MATERIALS - PROPERTY OF IBM
;             "RESTRICTED MATERIALS OF IBM"
;          5694-A01 (C) COPYRIGHT IBM CORP. 2000
```

```
;
; (C) COPYRIGHT International Business Machines Corp. 1985, 1993
; All Rights Reserved
; US Government Users Restricted Rights - Use, duplication or
; disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
;
; Licensed Materials - Property of IBM
;
;
;       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
;
; INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
; EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
; WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
; LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
; PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
; OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
; IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
; YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
; CORRECTION.
;
; Note:  This file must be copied and renamed to /etc/named.boot and all
;        all zone files referenced below must be copied to /etc/dnsdata/ for
;        this file to function as intended.
;
;  /etc/named.boot
;
;        boot file for name server
;
;
;type    domain                  source file or host
;
directory /etc/dnsdata                              {1}
primary  0.0.127.in-addr.arpa  db.loopback.v4       {2}
cache  .                       db.cache             {3}
options query-log                                   {4}
```

This boot file specifies:

1. The location of the files (/etc/dnsdata).
2. The name server will continue to function in the primary role for the loopback address (127.0.0.1).
3. The list of root name servers is in /etc/dnsdata/db.cache.
4. All queries coming in to this name server will be logged in the syslog daemon output file.

Step 1b. Create the configuration file for BIND 9–DNS.

The following sample configuration sets up an equivalent configuration for a BIND 9 nameserver where the sample configuration file, /usr/lpp/tcpip/samples/caching.conf (based on named.conf), reflects the setup for a caching-only name server. All zone data files referenced within the sample configuration file, with the exception of the hints file, can be found in the samples directory. To obtain the hints file, follow the instructions in “Step 6. Create the hints (root server) file” on page 618.

```
#           LICENSED MATERIALS - PROPERTY OF IBM
#             "RESTRICTED MATERIALS OF IBM"
#          5694-A01 (C) COPYRIGHT IBM CORP. 2001
#
# (C) COPYRIGHT International Business Machines Corp. 1985, 1993
# All Rights Reserved
# US Government Users Restricted Rights - Use, duplication or
# disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
#
```

```
# Licensed Materials - Property of IBM
#
#
#       NOTICE TO USERS OF THE SOURCE CODE EXAMPLES
#
# INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THE SOURCE CODE
# EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS, "AS IS" WITHOUT
# WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT
# LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A
# PARTICULAR PURPOSE.  THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE
# OF THE SOURCE CODE EXAMPLES, BOTH INDIVIDUALLY AND AS ONE OR MORE GROUPS,
# IS WITH YOU.  SHOULD ANY PART OF THE SOURCE CODE EXAMPLES PROVE DEFECTIVE
# YOU ASSUME THE ENTIRE COST OF ALL NECESSARY SERVICING, REPAIR OR
# CORRECTION.
#
#
# Note:  This file must be copied and renamed to /etc/named.conf and all
#        all zone files referenced below must be copied to /etc/dnsdata/ for
#        this file to function as intended.  In addition, the default location
#        for the process id file is in /var/run/pid.file; if that directory
#        does not exist a different one can be configured with the option:
#
#        pid-file "path/file-name";
#
#  /etc/named.conf
#
#        conf file for name server
#
options {
   directory "/etc/dnsdata";
};

logging {
   category "queries" {
     default_syslog;
  };
};

zone "0.0.127.in-addr.arpa" in {
  type master;
  file "db.loopback.v9";
};

zone "." in {
  type hint;
  file "db.cache";
};
```

Configuring a stealth name server

A stealth server is a server that answers authoritatively for a zone, but is not listed in that zone's NS records. Configure a master or slave stealth server for a zone like you would configure a visible master or slave server for the zone, except do not create an NS record for the stealth server in the master zone. The named server configuration option *also-notify* may be used to notify stealth slave servers of a zone update. In order to configure a stealth server, follow the steps in "Configuring a slave name server" on page 629.

Adding forwarding to your name server

In order to use forwarding in any BIND 4.9.3 or BIND 9 name server, update the boot file for BIND 4.9.3 or the conf file for BIND 9.

Add the following statement to the BIND 4.9.3 boot file:

```
forwarders 9.4.2.1
```

Add the following statement to the BIND 9 conf file in the options section:

```
options {
     ..............
     forwarders {9.4.2.1;};
     ..............
};
```

where 9.4.2.1 is the IP address of the machine where queries should be forwarded. This sends unresolved queries to 9.4.2.1 before trying to resolve the query using root name servers (specified in the hints file) or other cached name servers authoritative for or 'closer' to the authoritative name server.

For a name server to *only* use forwarders and not use the root servers, in addition to the forwarders directive, also add the following directive to its BIND 4.9.3 file:

```
options forward-only
```

OR to its BIND 9 conf file:

```
options {
     ..............
     forward only;
};
```

A name server with this option can still answer queries from its cached data. The cache is checked first and if the cache does not contain the answer, the query is sent to the name servers in the forwarders list.

Configuring host resolvers: Name server considerations

If the name server will run on the host being configured, create a loopback file. Specify the loopback address in the *first* name server directive of the resolver configuration file so local clients can access the name server. Refer to "Step 7. Create the loopback file" on page 619 for loopback address considerations.

The BIND 9 name server and BIND 9 DNS utilities (e.g. v9 nslookup and z/OS UNIX dig) use a private resolver that is different from the resolver used by other z/OS UNIX socket programs. The name server has the following functional differences:

- z/OS UNIX nslookup does not use site tables (for example, /etc/hosts) for host name resolution.
- Only the built-in translation table is used for BIND 4.9.3, BIND 9 and all DNS utilities (for example, v9 and v4 nslookup, v9 dig, etc.).

For a complete discussion of resolver configuration files, see *z/OS Communications Server: IP Configuration Reference*.

Configuring host resolvers: onslookup considerations

Programs that query a name server are called resolvers. Because many TCP/IP applications need to query the name server, a set of routines is usually provided for application programmers to perform queries. Under MVS, these routines are available in the TCP/IP application programming interface (API) for each supported language, Language Environment for z/OS UNIX C/C++ Sockets API or z/OS UNIX assembler callable services API.

The v9 nslookup command uses a private resolver that is different from the z/OS UNIX resolver used by other z/OS UNIX socket programs. The v4 nslookup command uses the same resolver as other z/OS UNIX socket programs. The onslookup command has the following functional differences:

- The HFS file, /etc/hosts, is required for host table lookup if name services do not exist. Following is a sample /etc/hosts file:

```
#
# z/OS UNIX Resolver /etc/hosts file on mvss18oe.
#
# The format of this file is:
#
# Internet Address      Hostname   Aliases     # Comments
#
# Items are separated by any number of blanks and/or tabs.  A '#'
# indicates the beginning of a comment; characters up to the end of the
# line are not interpreted by routines which search this file.  Blank
# lines are allowed in this file.

9.24.104.126    mvs18oe mvsoe # z/OS UNIX host
192.168.210.1   mvs18an       # AnyNet MVS host
192.168.210.8   mypcaa        # AnyNet gw host
9.24.104.79     mypc          # A workstation
```

  Note: The presence of /etc/hosts will prevent the z/OS UNIX resolver from accessing *prefix*.HOSTS.SITEINFO and *prefix*.HOSTS.ADDRINFO data sets. The use of /etc/hosts is not recommended unless it is used for purposes other than onslookup.

- Only the built-in translation table is used for both versions of BIND (V4 and V9).

If the z/OS UNIX name server will run on the host being configured, you need to configure the *first* name server (or NsInterAddr) directive in the resolver configuration file as the loopback address (127.0.0.1) or any address in your home list.

Creating the syslog file

If your syslog daemon is not configured, see "Configuring the syslog daemon (syslogd)" on page 175 for information regarding the syslog daemon.

Syslog daemon (syslogd) is a server process that is typically started as one of the first processes in a z/OS UNIX environment. Servers and stack components use syslogd for logging purposes and can also send trace information to syslogd. The named daemon logs messages to the syslog daemon. For BIND 9, specify the *syslog* option in the *channel* phrase of the *logging* statement in the named.conf file in order to use this function. Also for BIND 9, you can direct a category to the *default_syslog* channel. For information about the syslog daemon, see "Configuring the syslog daemon (syslogd)" on page 175.

If you will be using syslogd with BIND 9, refer to "Step 8. For BIND 9 only — configure logging" on page 621 for detailed information.

The name and location of your syslog file is specified in /etc/syslog.conf.

BIND 9 security considerations

Remote Name Daemon Control (rndc)

rndc is a tool that allows the system administrator some degree of control over the name server. The functions available are:

- Reload configuration file and zones.
- Reload the given zone.
- Schedule zone maintenance for the given zone.
- Reload the configuration file and load new zones, but do not reload existing zone files even if they have changed. This is faster than a full reload, when there is a large number of zones, because it avoids the need to examine the modification times of the zone files.
- Write server statistics to the statistics file.
- Toggle query logging.
- Dump the current contents of the cache.
- Stop the server, making sure any recent changes made through dynamic update or IXFR are first saved to the master files of the updated zones.
- Stop the server immediately. Recent changes made through dynamic update or IXFR are not saved to the master files, but will be rolled forward from the journal files when the server is restarted.
- Increment the server's debugging level by one.
- Set the server's debugging level to an explicit value.
- Set the server's debugging level to 0.
- Flush the server's cache.
- Display status of the server.

For more detail, refer to *z/OS Communications Server: IP System Administrator's Commands*, the rndc man page, the rndc.conf man page, and the rndc-confgen man page.

A configuration file is required, since all communication with the server is authenticated with digital signatures that rely on a shared secret, and there is no way to provide that secret other than with a configuration file. The default location for the rndc configuration file is /etc/rndc.conf, but an alternate location can be specified with the -c option. If the configuration file is not found, rndc will also look in /etc/rndc.key. The rndc.key file is generated by running rndc-confgen -a.

The format of the configuration file is similar to that of named.conf, but limited to only four statements:

- options
- key
- server
- include

These statements are what associate the secret keys to the servers with which they are meant to be shared. The order of statements is not significant.

The options statement has three clauses:

- default-server
- default-key
- default-port

The default-server clause takes a host name or address argument and represents the server that will be contacted if no -s option is provided on the command line. The default-key clause takes the key name as its argument, as defined by a key statement. The default-port clause specifies the port to which rndc should connect if no port is given on the command line or in a server statement.

The key statement names a key with its string argument. The string is required by the server to be a valid domain name, though it need not actually be hierarchical; thus, a string like ″rndc_key″ is a valid name. The key statement has two clauses:

- algorithm
- secret

While the configuration parser will accept any string as the argument to algorithm, currently only the string ″hmac-md5″ has any meaning. The secret is a base-64 encoded string.

Since the tool may be used remotely, rndc and the name server must communicate using digital transaction signatures (TSIG). Therefore, rndc and the name server must be configured with a *shared-secret*. There are two ways to configure a shared-secret key. One way is to use the /etc/rndc.key file generated by the rndc-confgen -a command. This file is shared between the name server and the rndc utility. The other way is to generate a shared-secret TSIG key with the HMAC-MD5 algorithm using the dnssec-keygen utility. The key must be configured in the name server under the *controls* section, and in the rndc.conf file on the key clause.

The server statement uses the key clause to associate the server with a key. The argument to the server statement is a host name or address (addresses must appear in double quotation marks). The argument to the key clause is the name of the key as defined by the key statement. The port clause can be used to specify the port to which rndc should connect on the given server.

The include statement can be used to insert the contents of another file within the rndc configuration file (for example, to include the contents of a file that contains sensitive key information).

A sample minimal configuration file is as follows:

```
key rndc_key {
     algorithm "hmac-md5";
     secret "c3Ryb25nIGVub3VnaCBmb3IgYSBtYW4gYnV0IG1hZGUgZm9yIGEgd29tYW4K";
};
options {
     default-server localhost;
     default-key    rndc_key;
};
```

This file, if installed as /etc/rndc.conf, would allow the rndc reload command to connect to 127.0.0.1 port 953 (the default port) and cause the nameserver to reload, if a nameserver on the local machine were running with the following controls statements and it had an identical key statement for rndc_key:

```
controls {
        inet 127.0.0.1 allow { localhost; } keys { rndc_key; };
};
```

Running the rndc-confgen program will conveniently create a rndc.conf file for you, and also display the corresponding controls statement that you need to add to named.conf. Alternatively, you can run rndc-confgen -a to set up a rndc.key file and not modify named.conf at all.

Access Control Lists

Access Control Lists (ACLs), are address match lists that you can set up and nickname for future use in allow-query , allow-recursion , blackhole , allow-transfer , etc. Using ACLs allows you to have finer control over who can access your nameserver, without cluttering up your config files with huge lists of IP addresses. It is a good idea to use ACLs, and to control access to your server. Limiting access to your server by outside parties can help prevent spoofing and DoS (Denial of Service) attacks against your server. Here is an example of how to properly apply ACLs:

```
// Set up an ACL named "bogusnets" that will block RFC1918 space,
// which is commonly used in spoofing attacks.

acl bogusnets { 0.0.0.0/8; 1.0.0.0/8; 2.0.0.0/8; 192.0.2.0/24; 224.0.0.0/3;
10.0.0.0/8; 172.16.0.0/12; 192.168.0.0/16; };

// Set up an ACL called our-nets. Replace this with the real IP numbers.

acl our-nets { x.x.x.x/24; x.x.x.x/21; };

options {
  ...
  ...
  allow-query { our-nets; };
  allow-recursion { our-nets; };
  ...
  blackhole { bogusnets; };
  ...
};
zone "example.com" {
  type master;
  file "m/example.com";
  allow-query { any; };
};
```

This allows non-recursive queries for example.com from inside and outside nets, except from bogusnets, and allows recursive queries from our-nets to outside nets unless *recursion no;* is also specified in the configuration file.

For more information on how to use ACLs to protect your server, see AUSCERT advisory AL-1999.004 at the Australian Computer Emergency Response Team Web site.

chroot and setuid

It is possible to run BIND in a chrooted environment (**chroot()**) by specifying the *-t* option. This can help improve system security by placing BIND in a sandbox, which will limit the damage done if a server is compromised.

Another useful feature is the ability to run the daemon as a nonprivileged user (-u user). We suggest running as a nonprivileged user when using the **chroot** feature.

Here is an example command line to load BIND in a **chroot()** sandbox, **/var/named**, and to run **named setuid** to user 202:

```
<>/usr/local/bin/named -u 202 -t /var/named
```

The chroot environment: In order for a **chroot**() environment to work properly in a particular directory (for example, /var/named), you will need to set up an environment that includes everything BIND needs to run. From BIND's point of view, /var/named is the root of the filesystem. You will need /dev/null, and any library directories and files that BIND needs to run on your system.

Using the setuid function: Prior to running the **named** daemon, use the **touch** utility (to change file access and modification times) or the **chown** utility (to set the user id and/or group id) on files to which you want BIND to write.

Dynamic update security

Access to the dynamic update facility should be strictly limited. For these reasons, we strongly recommend that updates be cryptographically authenticated by means of transaction signatures (TSIG). That is, the **allow-update** option should list only TSIG key names, not IP addresses or network prefixes. Alternatively, the **update-policy** option can be used.

Some sites choose to keep all dynamically updated DNS data in a subdomain and delegate that subdomain to a separate zone. This way, the top-level zone containing critical data such as the IP addresses of public web and mail servers need not allow dynamic update at all.

General VIPA considerations

If any VIPA addresses are used for zone delegation in the name server zone data files, ensure that IPCONFIG SOURCEVIPA is coded in PROFILE.TCPIP on the host to which a delegation has been made and that the SOURCEVIPA address that is used to respond to queries matches the VIPA used in the zone delegation. If this is not done, recursive queries could fail (nslookup will fail with ″No response from server″, dig will respond with, ″no servers could be reached″). The UDP query response packets will be discarded when they are received because the source IP address of the response will not match the destination IP address of the request. These addresses are required to match by the name server as a minimal security check.

Special considerations when using Dynamic VIPA

If you run a name server on a host that is using Dynamic VIPA (DVIPA), you may be required to do some additional configuration. Name servers running on a host using DVIPA need to BIND the UDP port that the name server listens on (usually 53) to the DVIPA, if you wish DNS to make use of the DVIPA. This can be done by using the BIND option on the UDP PORT statement in PROFILE.TCPIP. If you do BIND the DNS UDP port to the DVIPA, then all references to that name server must use the DVIPA whether those references are from other name servers or from resolvers.

References to a name server could occur in a number of places, and should be changed to use the DVIPA if BINDing a DNS UDP port to the DVIPA. This list is not exhaustive, but is intended to aid you for some of the most common cases. In general, you may need to change any place that references a name server by its IP address when using DVIPAs.

Task	Location
Delegating a DNS subdomain to a name server running on a host using DVIPA	The 'A' record in the glue records for the delegated (child) name server in the domain data file of the delegating (parent) name server

Task	Location
Designating a slave (secondary) name server when master (primary) name server is running on a host using DVIPA	The 'secondary' statement in the boot file of the secondary (slave) name server for BIND 4.9.3 or the 'masters' option of the 'zone' statement for BIND 9
Configuring resolvers	'NSINTERADDR' or 'nameserver' directive of the resolver configuration file
Using a name server as the target of other forwarding name servers when the target name server resides on a host using DVIPA	'forwarders' directive in the boot file of the forwarding name servers for BIND 4.9.3 or the 'forwarders{}' option of the 'options{}' statement for BIND 9
Using a name server as an intranet root name server when the root name server is running on a host using DVIPA	'A' record of the intranet root name server in the hints file on all name servers within the intranet

Dynamic primary DNS movement using Dynamic VIPA

To use DNS along with DVIPA takeover functionality to provide a high availability environment, perform the following steps:

1. Define a DVIPA. If using a BIND 4.9.3 name server, also configure a VIPADISTribute statement, specifying a port on the PORT parameter that applications will not listen on (to prevent undesired distribution of connections through Sysplex Distributor). DESTIP should be the Dynamic XCF addresses of all the members that will be backup servers. For example:

   ```
   VIPADYNAMIC
   VIPADEFINE     255.255.255.192 10.134.61.190
   ;The following VIPADISTRIBUTE line applies to the BIND 4.9.3 name server only
   VIPADISTRIBUTE 10.134.61.190 PORT 8081 DESTIP ALL
   ENDVIPADYNAMIC
   ```

2. Define VIPABACKUPs on all the members that will be backup servers. For example:

   ```
   VIPADYNAMIC
   VIPABACKUP  50  10.134.61.190
   ENDVIPADYNAMIC
   ```

3. On both the Distributor and all the targets, define the following statements.

   Note: The following example applies to the BIND 4.9.3 name server only.

   ```
   PORT
   .
   .
   .
   53 TCP NAMED2 BIND 10.134.61.190
   53 UDP NAMED2 BIND 10.134.61.190
   ```

4. Update /etc/resolv.conf and TCPDATA on all SYSPLEX members to point to that address. For example:

   ```
   .
   .
   .
   NSINTERADDR  10.134.61.190
   .
   .
   .
   ```

5. Update all of the glue records (NS records and their corresponding A records), in the sysplex name servers and the sysplex parent's name servers that point to the sysplex name server, to use the dynamic VIPA.

After doing the above configuration, stack termination will cause the DVIPA to be taken over, making DNS on the new DVIPA owning stack reachable after route convergence completes (OMPROUTE recommended).

Querying name servers

This section describes how to use the nslookup command to query the name server. onslookup is an alias of nslookup in the z/OS UNIX environment.

Notes:

1. The z/OS UNIX nslookup command runs only from the z/OS shell. The nslookup command can query the name server from TSO or the z/OS shell. However, only the legacy TSO version of NSLOOKUP is available from TSO. Refer to *z/OS Communications Server: IP System Administrator's Commands* for detailed information.
2. The host and dig commands are another way to query name servers from the z/OS shell. See the *z/OS Communications Server: IP User's Guide and Commands* for a list of host commands.

nslookup command

The z/OS UNIX nslookup and TSO NSLOOKUP commands can be used to query the name server to perform the following tasks:

- Identifying the location of name servers
- Examining the contents of a name server database
- Establishing the accessibility of name servers

Note: sortlist is not supported by nslookup

The z/OS UNIX nslookup and TSO NSLOOKUP commands have two modes of operation: interactive mode and command mode. nslookup also has two versions in z/OS UNIX: v4 and v9, where v4 gives the legacy z/OS UNIX (o)nslookup function, and v9 gives the BIND 9 version of nslookup. In either mode, the address of the default name server comes from the resolver configuration data. In the sample data below, the default domain is raleigh.ibm.com, and the default name server is at 9.37.34.149. If that name server fails to respond, the one at 9.37.34.7 is used.

```
domain   raleigh.ibm.com
nameserver   9.37.34.149
nameserver   9.37.34.7
```

Entering the interactive mode

Interactive mode can be used to repetitively query one or more name servers for information about various hosts and domains, to display that information on the console, and, in some cases, to write response data to a file.

You can enter the interactive mode under the following conditions only:

- No arguments are supplied on command invocation or the -v option is specified; the default name server is used.
- The first argument is a hyphen, and the second argument is the host name or Internet address of a name server.

For a complete description of the z/OS UNIX and TSO nslookup interactive modes, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Entering the command line mode

The command line mode displays or stores the output from the query supplied as part of the invocation string and then exits.

To enter the command line mode, provide a complete query with the z/OS UNIX nslookup command invocation string.

For a complete description of the z/OS UNIX and TSO nslookup command line modes, refer to *z/OS Communications Server: IP System Administrator's Commands*.

nslookup configuration

nslookup for BIND 4.9.3: The configuration options of z/OS UNIX nslookup determine the operation and results of the name server queries. The values for z/OS UNIX nslookup options can be specified in more than one location, as shown in Table 29. Values specified as z/OS UNIX nslookup command options have priority over values specified in the `.onslookuprc` file (for nslookup version 4), which have priority over the value specified by the environment variable, and so on. For example, the value specified by the `querytype` option in the z/OS UNIX nslookup command has priority over the value specified by the `querytype` option in the `.onslookuprc` file.

The letters beside some settings indicate that the *terms* are functionally equivalent. For example, the term `domain` (the letter "A") is functionally equivalent to the terms `DomainOrigin` and `LOCALDOMAIN`. If two functionally equivalent settings are listed in the same file, the one listed last has priority. For example, if domain and DomainOrigin are both listed in the Resolver configuration data set, and domain is listed first, the value specified by DomainOrigin has priority.

The numeric column headings in Table 29 correspond to the following:

1 z/OS UNIX nslookup command options. For more details about the z/OS UNIX nslookup command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

2 .onslookuprc file in the home directory. For more details, refer to *z/OS Communications Server: IP System Administrator's Commands*.

3 Environment variable (set by the z/OS UNIX command "export LOCALDOMAIN=domain_origin").

4 `Resolver configuration data set.` Refer to *z/OS Communications Server: IP Configuration Reference* for more details.

For example, column 1 lists z/OS UNIX nslookup options and column 2 lists options you can set in the `.onslookuprc` file.

Table 29. Settings that affect nslookup operation

Settings	1	2	3	4
All	x	x		
Class	x	x		
no[d2]	x	x		
[no]debug	x	x		
[no]defname	x	x		
domain (A)	x	x		x
[no]ignoretc	x	x		
port (B)	x	x		
querytype	x	x		
[no]recurse	x	x		

Table 29. Settings that affect nslookup operation (continued)

Settings	1	2	3	4
retry (C)	x	x		
root	x	x		
[no]search	x	x		
srchlist (D)	x	x		
timeout (E)	x	x		
[no]vc (F)	x	x		
search (D)				x
nameserver (G)				x
sortlist				x
options debug				x
options ndots				x
DomainOrigin (A)				x
NsInterAdd (G)				x
NsPortAddr (B)				x
ResolveVia (F)				x
ResolverTimeout (E)				x
ResolverUdpRetries (C)				x
LOCALDOMAIN (A)			x	

nslookup for BIND 9: There are fewer ways to configure BIND 9 (v9) nslookup in comparison to BIND 4.9.3 (v4) nslookup. There are only two places to specify v9 nslookup options: as command options, or from the resolver configuration data set. Only a few of the options may be set in the resolver configuration data set. The command options always have precedence over any option configured in the resolver configuration data set.

Only the following options can be specified in the resolver configuration data set for v9 nslookup:

- nameserver/nsinteraddr
- options ndots: *n*
- search
- domain/domainorigin

Programs that query a name server are called resolvers. Refer to "Understanding resolvers" on page 20 for more detailed information. Because many TCP/IP applications need to query the name server, a set of routines is usually provided for application programmers to perform queries. Under MVS, these routines are available in the TCP/IP application programming interface (API) for each supported language, Language Environment for z/OS UNIX C/C++ Sockets API or z/OS UNIX assembler callable services API.

The z/OS UNIX v9 nslookup command uses a private resolver that is different from the resolver used by other z/OS UNIX socket programs. The z/OS UNIX nslookup command has the following functional differences:

- z/OS UNIX nslookup does not use SiteTables (for example, /etc/hosts) for host name resolution.
- Only the built-in translation table is used for v4 and v9 nslookup.
- z/OS UNIX v4 nslookup uses the LOCALDOMAIN environment variable, whereas no other resolvers use this.

For a complete discussion of resolver configuration files, see Chapter 2, "Configuration overview," on page 11.

If the name server will run on the host being configured, you need to configure the *first* name server (or NsInterAddr) directive in the resolver configuration file as the loopback address (127.0.0.1 or any address in your home list). If any VIPA addresses are used with the NsInterAddr statement, ensure that IPCONFIG SOURCEVIPA is coded in PROFILE.TCPIP. If it is not, UDP packets returned from the VIPA address will have the physical interface address as the destination address instead of the VIPA address that it sent. The UDP packet will be discarded when it is received because the addresses do not match.

Diagnosing problems

This section describes the following methods for diagnosing problems:

- Checking messages on the operators console
- Checking the syslog messages
- Using name server signals
- Using rndc to diagnose BIND 9 problems
- Checking name server logging files to diagnose BIND 9
- Using nslookup program
- Using the dig command

These methods are discussed below. In addition to these methods, diagnosing problems for a dynamic zone can be done with nsupdate. This allows debugging of problems between DHCP and the dynamic zone in a more controlled manner.

For DNS configuration firewall considerations, refer to "Split DNS" on page 650, and the latest edition of *DNS and BIND* by Paul Albitz and Cricket Liu (O'Reilly & Associates, Inc.).

Checking messages sent to the operators console

Messages displayed on the operators console indicate the status of your DNS. Messages fall into the following categories:

- Name server initialization
- Name server initialization failure
- Name server initialization complete
- Name server termination
- Assertion failures (unexpected errors) (v9 only)

Regularly check console messages to identify problems.

Checking the syslog messages

Error messages may also be displayed in the syslog output file, which is pointed to by the syslog configuration file. (`/etc/syslog.conf` is the default configuration file.) For BIND 9, refer to "Step 8. For BIND 9 only — configure logging" on page 621. For the BIND 9 name server, initial startup messages go to syslog, later

messages will be directed to other defined or default logs according to logging statements found or implied in the configuration file. For descriptions of the syslog file and the syslog daemon, see "Configuring the syslog daemon (syslogd)" on page 175.

Using name server signals to diagnose BIND 4.9.3 DNS problems

You can use name server signals to send messages to a BIND 4.9.3 or a BIND 9 DNS. Note that some of the signals may have different consequences if a signal is sent to a BIND 4.9.3 name server instead of a BIND 9 name server. These signals control various functions that can be used to diagnose problems.

Diagnostic functions include the following:

- Enabling and disabling debug message logging
- Dumping the contents of the name server database
- Getting short status
- Logging queries

For an explanation of the format of the dumped database or the name server statistics file that can be generated with these signals, refer to publications such as *DNS and Bind* by Albitz and Liu.

The issuing of a signal is done by using the z/OS UNIX kill command and also involves specifying the process ID. The BIND 4.9.3 name server always stores its process ID in the file, /etc/named.pid.

Using name server signals to diagnose BIND 9 DNS problems

You can use name server signals to send messages to a BIND 4.9.3 or a BIND 9 DNS name server. Note that some of the signals may have different consequences if a signal is sent to a BIND 4.9.3 name server instead of a BIND 9 name server. These signals control various functions that can be used to diagnose problems.

The BIND 9 name server relies on a start option, rndc, or the configuration file to define and alter the debug level. You can change the logging options in named.conf to gather more information, and then issue the SIGHUP signal or 'rndc reload' to have the new logging options take effect. However, the preferred method is by using 'rndc trace level'.

For an explanation of the format of the dumped database or the name server statistics file that can be generated with these signals, refer to publications like *DNS and Bind* by Albitz and Liu.

Signals are issued with the z/OS UNIX kill command using the name server process ID as a parameter. The file where the BIND 9 process ID is stored is determined by the user. The file name is specified by the 'pid-file' option in the named.conf file. Refer to *z/OS Communications Server: IP Configuration Reference* for further details.

Using rndc to diagnose BIND 9 problems

The rndc utility can be used to provide a variety of functions that can be helpful in debugging name server problems. For example, the name server's cache can be viewed using the dumpdb parameter and debug trace can be turned on or off using the trace parameter. If you suspect your cache is corrupted, you can flush

the name server's cache with the flush parameter. For more information, see *z/OS Communications Server: IP System Administrator's Commands*.

Checking name server logging files to diagnose BIND 9

Error, debug and informational messages can be written to the name server's logging files. Refer to "Step 8. For BIND 9 only — configure logging" on page 621 for more information.

Using nslookup to diagnose problems

The z/OS UNIX nslookup program lets you query other name servers with the same query packet another name server would use. This is helpful in diagnosing lookup problems in TCP/IP.

It is recommended that you use z/OS UNIX or TSO nslookup with each NSINTERADDR used in TCPIP.DATA to ensure you receive the expected results. Some name server clients, on other platforms, may require the address you specify for the name server to match the source IP address in the response from the name server. For example, if a static VIPA address is specified as the address of the name server, and IPCONFIG SOURCEVIPA is not specified in PROFILE.TCPIP, then nslookup on some platforms will discard the returned packet because it will have the destination address of the physical interface instead of the VIPA interface. If you wish to specify a Dynamic VIPA (DVIPA) as the address of the name server, then the name server must BIND the UDP port to the DVIPA. Refer to *z/OS Communications Server: IP Configuration Reference* for information on how to specify the BIND parameter on the PORT statement in PROFILE.TCPIP. Zone delegation using VIPA addresses also has special considerations and can cause name server operation to fail in some situations if not configured properly. For more information, see "General VIPA considerations" on page 641.

Debugging is available at different levels for BIND 9 and 4.9.3. The commands are similar but the output differs between BIND 9 and BIND 4.9.3. Refer to the *z/OS Communications Server: IP System Administrator's Commands* for more detailed information on the z/OS UNIX nslookup command.

Using dig to diagnose problems

The dig command is an alternate and recommended choice for resource record lookup. The dig response format is similar to the resource record (RR) definitions in master zone files. The dig command will not attempt a reverse lookup on the address provided for the server, which sometimes makes nslookup fail initialization if reverse lookup fails. The dig command offers flexible options, including toggling flags for checking response authority or authentication. It is the only v9 lookup utility that can list the complete contents of a zone. This can be accomplished with the -t AXFR option. For more detailed information on the z/OS UNIX dig command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Advanced BIND 9 name server topics

Multiple TCP/IP stack (common INET) considerations

The BIND 9 name server is a *generic* server which does not have stack affinity. This is in contrast to the BIND 4.9.3 name server which does have stack affinity. This has certain implications.

- If you wish to run multiple BIND 9 name servers, you must divide the interfaces between the name servers with the **listen-on** named.conf file option. For example, you may want one stack serviced by one BIND 9 name server, and a second stack serviced by a second BIND 9 name server. Each name server would contain only the IP addresses of the assigned stack in its **listen-on** option.
- Any time a stack is brought up or a stack is brought down, the name server will essentially restart. On the MVS console, you will see the NAMED Exiting, NAMED Starting, and NAMED Running messages (messages EZZ9096I, EZZ9095I, and EZZ9130I).
- Be aware, that once you start a TCP/IP stack, all of the adapters may not be active immediately, and therefore will not be usable by the name server immediately. When the name server is started manually or restarted automatically by stack bring up or bring down, it immediately queries the available TCP/IP stacks for active adapters. Often times, it will take some time for all of the adapters to become active (this is independent of the name server). The name server will re-query the stacks every minute, by default, for any changes in the active/inactive status of adapters and then make use of them once they are active. The one minute interval can be lengthened by the **interface-interval** named.conf file option if desired, but this is not recommended.
- By default, the name server will unpredictably choose one adapter from any of the active stacks to use when it must communicate with other name servers. If some adapters do not have the capability to route into the network, you might see undesired results on name server queries. This unpredictable behavior can be eliminated by making use of the **query-source** option in the named.conf file. The **query-source** option should specify an adapter address that will always have network routing capability. The **query-source** option then places a dependency on the stack that owns that address to be active. If the owning TCP/IP stack of the **query-source** option address is taken down, the name server will end, since it will no longer have a way to communicate with the network, and thus, other name servers.
- BIND 9 is restricted to listening on all IPv6 interfaces or none of them. Therefore, if you want to run multiple BIND 9 name servers, you need to choose one of them to answer all IPv6 queries. Use the listen-on-v6 named.conf option with the value of *any;* to get BIND 9 to listen on your IPv6 interfaces.

Dynamic update

Dynamic update is the term used for the ability under certain specified conditions to add, modify or delete records or RRsets in the master zone files. Dynamic update is fully described in RFC 2136.

Dynamic update is enabled on a zone-by-zone basis, by including an **allow-update** or **update-policy** clause in the **zone** statement. Preferably, use TSIG security between the nsupdate utility and the targeted name server. BIND 9 name server configuration processing messages will remind you when nsupdate authorization is only based on client IP address.

Updating of secure zones (zones using DNSSEC) is modelled after the simple-secure-update proposal, a work in progress in the DNS Extensions working group of the IETF. (See http://www.ietf.org/html.charters/dnsext-charter.html for information about the DNS Extensions working group.) SIG and NXT records affected by updates are automatically regenerated by the server using an online zone key. Update authorization is based on transaction signatures and an explicit server policy. On z/OS, dynamic DNSSEC zones should use the RSA encryption

algorithm when creating the zone key, or they can use the DSA algorithm if the 'random-device' option is also specified in the named.conf file. Non-dynamic DNSSEC zones can use any other supported encryption algorithm.

The zone files of dynamic zones must not be edited by hand. If the zone file of a dynamic zone is edited by hand, corrupt .jnl files can result and all changes not written to the zone file may be lost. The zone file on disk at any given time may not contain the latest changes performed by dynamic update. The zone file is written to disk only periodically, and changes that have occurred since the zone file was last written to disk are stored only in the zone's journal (.jnl) file. Depending on signal or rndc stop options, BIND 9 name server may or may not update the zone file. Therefore, editing the zone file manually is unsafe even when the server has been shut down.

Incremental zone transfers (IXFR)

The incremental zone transfer (IXFR) protocol is a way for slave servers to transfer only changed data, instead of having to transfer the entire zone. The IXFR protocol is documented in RFC 1995.

When acting as a master server, BIND 9 supports IXFR for those zones where the necessary change history information is available. These include master zones maintained by dynamic update and slave zones (to transfer to other slave servers) whose data was obtained by IXFR, but not manually maintained master zones or slave zones obtained by performing a full zone transfer (AXFR).

When acting as a slave server, after the initial full zone transfer, BIND 9 will request IXFR by default when notified of a change by the zone master server. IXFR updates are applied to the slave zone database and also kept in a journal file (*.jnl) associated with any existing backup file for the slave zone.

Master and slave servers can each have IXFR globally or partially disabled through the use of the *provide-ixfr* and *request-ixfr* options under the general options or the *server* statements of the BIND 9 configuration file.

Split DNS

Setting up different views, or visibility, of DNS space to internal and external resolvers is usually referred to as a Split DNS setup. There are several reasons an organization would want to set up its DNS this way.

One common reason for setting up a DNS system this way is to hide *internal* DNS information from *external* clients on the Internet. There is some debate as to whether or not this is actually useful. Internal DNS information leaks out in many ways (via e-mail headers, for example) and most savvy attackers can find the information they need using other means.

Another common reason for setting up a Split DNS system is to allow internal networks that are behind filters or in RFC 1918 space (reserved IP space, as documented in RFC 1918) to resolve DNS on the Internet. Split DNS can also be used to allow mail from outside back in to the internal network.

Here is an example of a split DNS setup:

A company named Example, Inc. (example.com) has several corporate sites that have an internal network with reserved Internet Protocol (IP) space and an external demilitarized zone (DMZ), or *outside* section of a network, that is available to the public.

Example, Inc. wants its internal clients to be able to resolve external hostnames and to exchange mail with people on the outside. The company also wants its internal resolvers to have access to certain internal-only zones that are not available at all outside of the internal network.

In order to accomplish this, the company will set up two sets of nameservers. One set will be on the inside network (in the reserved IP space) and the other set will be on bastion hosts, which are *proxy* hosts that can talk to both sides of its network, in the DMZ.

The internal servers will be configured to forward all queries, except queries for site1.internal, site2.internal, site1.example.com, and site2.example.com, to the servers in the DMZ. These internal servers will have complete sets of information for site1.example.com, site2.example.com, site1.internal, and site2.internal.

To protect the site1.internal and site2.internal domains, the internal nameservers must be configured to disallow all queries to these domains from any external hosts, including the bastion hosts.

The external servers, which are on the bastion hosts, will be configured to serve the *public* version of the site1 and site2.example.com zones. This could include things such as the host records for public servers (www.example.com and ftp.example.com), and mail exchange (MX) records (a.mx.example.com and b.mx.example.com).

In addition, the public site1 and site2.example.com zones should have special MX records that contain wildcard (*) records pointing to the bastion hosts. This is needed because external mail servers do not have any other way of looking up how to deliver mail to those internal hosts. With the wildcard records, the mail will be delivered to the bastion host, which can then forward it on to internal hosts.

Here's an example of a wildcard MX record:

```
*   IN MX 10 external1.example.com.
```

Now that they accept mail on behalf of anything in the internal network, the bastion hosts will need to know how to deliver mail to internal hosts. In order for this to work properly, the resolvers on the bastion hosts will need to be configured to point to the internal nameservers for DNS resolution. Queries for internal hostnames will be answered by the internal servers, and queries for external hostnames will be forwarded back out to the DNS servers on the bastion hosts. In order for all this to work properly, internal clients will need to be configured to query only the internal nameservers for DNS queries. This could also be enforced via selective filtering on the network.

If everything has been set properly, Example, Inc.'s internal clients will now be able to:

- Look up any hostnames in the site1 and site2.example.com zones.
- Look up any hostnames in the site1.internal and site2.internal domains.
- Look up any hostnames on the Internet.

- Exchange mail with internal and external people.

Hosts on the Internet will be able to:

- Look up any hostnames in the site1 and site2.example.com zones.
- Exchange mail with anyone in the site1 and site2.example.com zones.

Here is an example configuration for the setup we just described above. Note that this is only configuration information; for information on how to configure your zone files, see "Step 5. Create the domain data files (master name server only)" on page 612.

Internal DNS server config:

```
acl internals { 172.16.72.0/24; 192.168.1.0/24; };

acl externals { bastion-ips-go-here; };

options {
    ...
    ...
    forward only;
    forwarders {     // forward to external servers
     bastion-ips-go-here;
    };
    allow-transfer { none; };   // sample allow-transfer (no one)
    allow-query { internals; externals; }; // restrict query access
    allow-recursion { internals; };  // restrict recursion
    ...
    ...
};

zone "site1.example.com" {
  type master;
  file "m/site1.example.com";
  forwarders { };     // do normal iterative
      // resolution (do not forward)
  allow-query { internals; externals; };
  allow-transfer { internals; };
};

zone "site2.example.com" {
  type slave;
  file "s/site2.example.com";
  masters { 172.16.72.3; };
  forwarders { };
  allow-query { internals; externals; };
  allow-transfer { internals; };
};

zone "site1.internal" {
  type master;
  file "m/site1.internal";
  forwarders { };
  allow-query { internals; };
  allow-transfer { internals; }
};

zone "site2.internal" {
  type slave;
  file "s/site2.internal";
  masters { 172.16.72.3; };
```

```
  forwarders { };
  allow-query { internals };
  allow-transfer { internals; }
};
```

External (bastion host) DNS server config:

```
acl internals { 172.16.72.0/24; 192.168.1.0/24; };

acl externals { bastion-ips-go-here; };

options {
  ...
  ...
  allow-transfer { none; };   // sample allow-transfer (no one)
  allow-query { internals; externals; }; // restrict query access
  allow-recursion { internals; externals; }; // restrict recursion
  ...
  ...
};

zone "site1.example.com" {
  type master;
  file "m/site1.foo.com";
  allow-query { any; };
  allow-transfer { internals; externals; };
};

zone "site2.example.com" {
  type slave;
  file "s/site2.foo.com";
  masters { another_bastion_host_maybe; };
  allow-query { any; };
  allow-transfer { internals; externals; }
};
```

In the resolv.conf (or equivalent) on the bastion host(s):

```
search ...
nameserver 172.16.72.2
nameserver 172.16.72.3
nameserver 172.16.72.4
```

Implementing split DNS with views

The view statement is a powerful new feature of BIND 9 that lets a name server answer a DNS query differently depending upon who is asking. It is particularly useful for implementing split DNS setups without having to run multiple servers.

Each view statement defines a view of the DNS namespace that will be seen by a subset of clients. A client matches a view if its source IP address matches the address_match_list of the view's match-clients clause and its destination IP address matches the address_match_list of the view's match-destinations clause. If not specified, both match-clients and match-destinations default to matching all addresses. A view can also be specified as match-recursive-only, which means that only recursive requests from matching clients will match that view. The order of the view statements is significant; A client request will be resolved in the context of the first view that it matches.

Zones defined within a view statement will only be accessible to clients that match the view. By defining a zone of the same name in multiple views, different zone data can be given to different clients (for example, internal and external clients in a split DNS setup.

Many of the options given in the options statement can also be used within a view statement, and then apply only when resolving queries with that view. When no view-specific value is given, the value in the options statement is used as a default. Also, zone options can have default values specified in the view statement. These view-specific defaults take precedence over those in the options statement.

Views are class specific. If no class is given, class IN is assumed. Note that all non-IN views must contain a hint zone, since only the IN class has compiled-in default hints.

If there are no view statements in the configuration file, a default view that matches any client is automatically created in class IN, and any zone statements specified on the top level of the configuration file are considered to be part of this default view. If any explicit view statements are present, all zone statements must occur inside view statements.

Following is an example of a typical split DNS setup implemented using view statements:

```
view "internal" {
               // This should match our internal networks.
      match-clients { 10.0.0.0/8; };
               // Provide recursive service to internal clients only.
      recursion yes;
               // Provide a complete view of the example.com zone
               // including addresses of internal hosts.
      zone "example.com" {
            type master;
            file "example-internal.db";
      };
};
view "external" {
      match-clients { any; };
               // Refuse recursive service to external clients.
      recursion no;
               // Provide a restricted view of the example.com zone
               // containing only publicly accessible hosts.
      zone "example.com" {
           type master;
           file "example-external.db";
      };
};
```

TSIG

This is a short guide to setting up Transaction SIGnatures (TSIG) based transaction security in BIND. It describes changes to the configuration file as well as what changes are required for different features, including the process of creating transaction keys and using transaction signatures with BIND.

BIND primarily supports TSIG for server to server communication. This includes zone transfer, notify, and recursive query messages. Resolvers on other platforms which are based on newer versions of BIND 8 have limited support for TSIG.

TSIG might be most useful for dynamic update. A master server for a dynamic zone should use access control to control updates, but IP-based access control is insufficient. Key-based access control is far superior. The **nsupdate** program supports TSIG via the -k and -y command line options.

Generate shared keys for each pair of hosts

A shared secret is generated to be shared between host1 and host2. An arbitrary key name is chosen: ″host1-host2.″. The key name must be the same on both hosts.

Automatic generation: The following command will generate a 128-bit (16-byte) HMAC-MD5 key as described above. Longer keys are better, but shorter keys are easier to read. Note that the maximum key length is 512 bits; keys longer than that will be digested with MD5 to produce a 128 bit key.

```
dnssec-keygen -a hmac-md5 -b 128 -n HOST host1-host2.
```

The key is in the file Khost1-host2.+157+00000.private. Nothing directly uses this file, but the base-64 encoded string following ″Key:″ can be extracted from the file and used as a shared secret:

```
Key: La/E5CjG9O+os1jq0a2jdA==
```

The string ″La/E5CjG9O+os1jq0a2jdA==″ can be used as the shared secret.

Manual generation: The shared secret is simply a random sequence of bits, encoded in base-64. Most EBCDIC strings are valid base-64 strings (assuming the length is a multiple of 4 and only valid characters are used), so the shared secret can be manually generated. Also, a known string can be run through *mmencode* on another platform (such as Linux), or through a similar program, to generate base-64 encoded data.

Copying the shared secret to both machines

This is beyond the scope of DNS. A secure transport mechanism should be used. This could be secure FTP, ssh, telephone, etc.

Informing the servers of the key's existence

Imagine *host1* and *host2* are both servers. The following is added to each server's named.conf file:

```
key host1-host2. {
  algorithm hmac-md5;
  secret "La/E5CjG9O+os1jq0a2jdA==";
};
```

The algorithm, hmac-md5, is the only one supported by BIND. The secret is the one generated above. Since this is a secret, it is recommended that either named.conf be non-world readable, or the key directive be added to a non-world readable file that is included by named.conf.

At this point, the key is recognized. This means that if the server receives a message signed by this key, it can verify the signature. If the signature succeeds, the response is signed by the same key.

Instructing the server to use the key

Since keys are shared between two hosts only, the server must be told when keys are to be used. The following is added to the named.conf file for host1, if the IP address of host2 is 10.1.2.3:

```
server 10.1.2.3 {
  keys { host1-host2. ;};
};
```

Multiple keys may be present, but only the first is used. This directive does not contain any secrets, so it may be in a world-readable file.

If host1 sends a message that is a request to that address, the message will be signed with the specified key. host1 will expect any responses to signed messages to be signed with the same key. A similar statement must be present in host2's configuration file (with host1's address) for host2 to sign request messages to host1.

TSIG key based access control

BIND allows IP addresses and ranges to be specified in ACL definitions, and in access control directives such as allow-query, allow-transfer, and allow-update. This has been extended to allow TSIG keys also. The above key would be denoted **key host1-host2**. An example of an allow-update directive would be:

```
allow-update { key host1-host2. ;};
```

This allows dynamic updates to succeed only if the request was signed by a key named ″**host1-host2**.″.

Errors

The processing of TSIG signed messages can result in several errors. If a signed message is sent to a non-TSIG aware server, a FORMERR will be returned, since the server will not understand the record. This is a result of misconfiguration, since the server must be explicitly configured to send a TSIG signed message to a specific server.

If a TSIG aware server receives a message signed by an unknown key, the response will be unsigned with the TSIG extended error code set to BADKEY. If a TSIG aware server receives a message with a signature that does not validate, the response will be unsigned with the TSIG extended error code set to BADSIG. If a TSIG aware server receives a message with a time outside of the allowed range, the response will be signed with the TSIG extended error code set to BADTIME, and the time values will be adjusted so that the response can be successfully verified. In any of these cases, the message's rcode is set to NOTAUTH.

DNSSEC

Cryptographic authentication of DNS information is possible through the DNS Security (DNSSEC) extensions, defined in RFC 2535. This section describes the creation and use of DNSSEC signed zones.

The set of *dnssec-* tools rely on a /dev/random device for the entropy it needs to generate cryptographically strong keys. If RSA keys are used, only *dnssec-keygen* requires random data. z/OS UNIX does not include such a device, but the tools provide alternate methods of providing them with random data. The user can specify a file containing random data or can provide random data via the keyboard. To specify a file, use the **-r** *random data file* option on the tool command line. The *dnssec-* tools use the timing between keystrokes as the source of entropy. As such, TN3270 terminal emulation is not the ideal interface. Setting up a VT100 terminal session is a better solution. Refer to the "Configuring the z/OS UNIX Telnet server (otelnetd)" on page 531 for more information on setting up otelnetd.

To set up a DNSSEC secure zone, there are a series of steps which must be followed. z/OS ships with several tools that are used in this process, which are explained in more detail below. In all cases, the -h option prints a full list of parameters. Note that the keyset and signedkey files created by some DNSSEC tools must be put in the name server working directory before another DNSSEC tool is used for signing a master zone file.

There must also be communication with the administrators of the parent and/or child zone to transmit keys and signatures. A zone's security status must be indicated by the parent zone for a DNSSEC capable resolver to trust its data.

For other servers to trust data in this zone, they must be statically configured with either this zone's zone key or the zone key of another zone above this one in the DNS tree, using the *trusted-keys* statement in the configuration file.

Generating keys

The **dnssec-keygen** program is used to generate keys.

A secure zone must contain one or more zone keys. The zone keys will sign all other records in the zone, as well as the zone keys of any secure delegated zones. Zone keys must have the same name as the zone, a name type of **ZONE**, and must be usable for authentication. On z/OS, you should use the RSA algorithm for DNSSEC if the zone you will sign with the key will be a dynamic zone (that is, one maintained with nsupdate). You can also use the DSA algorithm, provided you have the 'random-device' option coded in the named.conf file.

The following command will generate a 768-bit RSA key for the child.example zone:

```
dnssec-keygen -a RSA -b 768 -n ZONE child.example
```

Two output files will be produced: Kchild.example.+001+12345.key and Kchild.example.+001+12345.private (where 12345 is an example of a key tag). The key file names contain the key name (child.example.), algorithm (3 is DSA, 1 is RSA, etc.), and the key tag (12345 in this case). The private key (in the .private file) is used to generate signatures, and the public key (in the .key file) is used for signature verification.

To generate another key with the same properties (but with a different key tag), repeat the above command.

The public keys should be inserted into the zone file with **$INCLUDE** statements, including the .key files.

Creating a key set

The **dnssec-makekeyset** program is used to create a key set from one or more keys.

Once the zone keys have been generated, a key set must be built for transmission to the administrator of the parent zone, so that the parent zone can sign the keys with its own zone key and correctly indicate the security status of this zone. When building a key set, the list of keys to be included and the TTL of the set must be specified, and the desired signature validity period of the parent's signature may also be specified.

The list of keys to be inserted into the key set may also include non-zone keys present at the top of the zone. dnssec-makekeyset may also be used at other names in the zone.

The following command generates a key set containing the above key and another key similarly generated, with a TTL of 3600 and a signature validity period of 10 days starting from now.

```
dnssec-makekeyset -t 3600 -e +8640 Kchild.example.+001+12345
Kchild.example.+001+23456
```

One output file is produced: keyset-child.example. This file should be transmitted to the parent to be signed. It includes the keys, as well as signatures over the key set generated by the zone keys themselves, which are used to prove ownership of the private keys and encode the desired validity period.

Signing the child's key set

The **dnssec-signkey** program is used to sign one child's key set.

If the child.example zone has any delegations which are secure, for example, grand.child.example, the child.example administrator should receive key set files for each secure subzone. These keys must be signed by this zone's zone keys.

The following command signs the child's key set with the zone keys:

```
dnssec-signkey keyset-grand.child.example. Kchild.example.+001+12345
Kchild.example.+001+23456
```

One output file is produced: signedkey-grand.child.example.. This file should be both transmitted back to the child and retained. It includes all keys (the child's keys) from the key set file and signatures generated by this zone's zone keys.

Signing the zone

The **dnssec-signzone** program is used to sign a zone.

Any **signedkey** files corresponding to secure subzones should be present, as well as a signedkey file for this zone generated by the parent (if there is one). The zone signer will generate NXT and SIG records for the zone, as well as incorporate the zone key signature from the parent and indicate the security status at all delegation points.

The following command signs the zone, assuming it is in a file called zone.child.example. By default, all zone keys which have an available private key are used to generate signatures.

```
dnssec-signzone -o child.example zone.child.example
```

One output file is produced: zone.child.example.signed. This file should be referenced by named.conf as the input file for the zone.

Configuring servers

Data is not verified on load in BIND 9, so zone keys for authoritative zones do not need to be specified in the configuration file. The public key for any security root must be present in the configuration file's **trusted-keys** statement.

IPv6 support in BIND 9

BIND 9 fully supports all currently defined forms of IPv6 name to address and address to name lookups.

For forward lookups, BIND 9 supports both A6 and AAAA records. The use of AAAA records is recommended, as A6 records have been made experimental by RFC 3363.

For IPv6 reverse lookups, BIND 9 supports the standard *nibble* label format, as well as the experimental *bitstring* format. Both formats are used under the ip6.arpa domain, although some resolvers and applications use the nibble format under the deprecated ip6.int domain.

Address lookups using AAAA records

The AAAA record is a parallel to the IPv4 A record. It specifies the entire address in a single record. For example:

```
$ORIGIN example.com.
host  3600 IN AAAA 3ffe:8050:201:1860:42::1
```

Address lookups using A6 records

A6 records are supported, but have been moved to experimental status by RFC 3363. The use of AAAA records is strongly recommended.

The A6 record can be used to form a chain of A6 records, each specifying part of the IPv6 address. It can also be used to specify the entire record as well. For example, this record supplies the same data as the AAAA record in the previous example:

```
$ORIGIN example.com.
host  3600 IN A6 0 3ffe:8050:201:1860:42::1
```

For more information on A6 records and A6 chaining, see RFC 2874.

Synthetic IPv6 responses

Synthetic IPv6 responses were previously enabled with the allow-v6-synthesis option statement. This option is now obsolete.

Address to name lookups using nibble format

When looking up an address in nibble format, the address components are simply reversed, just as in IPv4, and ip6.arpa. is appended to the resulting name. For example, the following would provide reverse name lookup for a host with address 3ffe:8050:201:1860:42::1.

```
$ORIGIN 0.6.8.1.1.0.2.0.0.5.0.8.e.f.f.3.ip6.arpa.
1.0.0.0.0.0.0.0.0.0.0.0.2.4.0.0   14400 IN  PTR
host.example.com.
```

Address to name lookups using bitstring format

Bitstring labels can start and end on any bit boundary, rather than on a multiple of 4 bits as in the nibble format.

To replicate the previous example using bitstrings:

```
$ORIGIN \[x3ffe805002011860/64].ip6.arpa.
\[x0042000000000001/64]  14400  IN  PTR
host.example.com.
```

Using DNAME for delegation of IPv6 reverse addresses

DNAME is supported, but is considered experimental.

In IPv6, the same host may have many addresses from many network providers. Since the trailing portion of the address usually remains constant, **DNAME** can help reduce the number of zone files used for reverse mapping that need to be maintained.

For example, consider a host which has two providers (example.net and example2.net) and therefore two IPv6 addresses. Since the host chooses its own 64 bit host address portion, the provider address is the only part that changes:

```
$ORIGIN example.com.
host   A6 64 ::1234:5678:1212:5675
cust1.example.net.
```

```
   A6 64 ::1234:5678:1212:5675
subnet5.example2.net.
$ORIGIN example.net.
cust1   A6 48 0:0:0:dddd::
ipv6net.example.net.
ipv6net   A6 0 aa:bb:cccc::
$ORIGIN example2.net.
subnet5   A6 48 0:0:0:1::
ipv6net2.example2.net.
ipv6net2  A6 0 6666:5555:4::
```

This sets up forward lookups. To handle the reverse lookups, the provider example.net would have:

```
$ORIGIN \[x00aa00bbcccc/48].ip6.arpa.
\[xdddd/16]  DNAME  ipv6-rev.example.com.
```

and example2.net would have:

```
$ORIGIN \[x666655550004/48].ip6.arpa.
\[x0001/16]  DNAME  ipv6-rev.example.com.
```

example.com needs only one zone file to handle both of these reverse mappings:

```
$ORIGIN ipv6-rev.example.com.
\[x1234567812125675/64] PTR  host.example.com.
```

Advanced BIND 4.9.3–Name server topics

Connection optimization in a sysplex domain

This section describes *connection optimization*, a technique that uses DNS for balancing IP connections and workload in a sysplex domain. It also dynamically manages the active and available list of IP addresses for resources in a sysplex domain.

The name server can perform Connection Optimization (DNS/WLM) only when running in BIND 4.9.3 mode. Those wishing to do sysplex load balancing while running the name server in BIND 9 mode are encouraged to use the Sysplex Distributor function as an alternative.

You may run a BIND 9 and a BIND 4.9.3 name server simultaneously; however, the host's interfaces must be divided among the two. That is, clients wishing to be served by a BIND 4.9.3 name server must send queries to the BIND 4.9.3 addresses and clients wishing to be served by a BIND 9 name server must send queries to the BIND 9 addresses. If you wish to run BIND 4.9.3 and BIND 9 name servers simultaneously while using DNS/WLM (Connection Optimization) with the BIND 4.9.3 name server, you can still make all addresses on the host available to the BIND 4.9.3 name server for Connection Optimization. You do not have to limit that set of addresses to the set of addresses that the BIND 4.9.3 name server is listening on. The mutually exclusive set of addresses that the two name servers are listening on is independent of the addresses that the name servers may return to clients.

In BIND 9 mode, primitive load balancing without Sysplex Distributor can be achieved in DNS using multiple A records for one name, if true sysplex load balancing is not desired through the Sysplex Distributor. In this case, the addresses from the multiple A records will be handed out in a round-robin fashion.

Overview

Connection optimization uses DNS for distributing connections among hosts or server applications within a sysplex domain. A sysplex is a set of MVS systems communicating and cooperating with each other through multisystem hardware and software components.

In DNS terms, a sysplex is a subdomain that is added to the DNS name space. Name servers running within the sysplex perform name resolution. Resolvers query these name servers directly or indirectly through the name server authoritative for the resources in the sysplex domain.

Connection optimization extends the concept of a ″DNS host name″ to clusters, or groups of server applications or hosts. Server applications within the same group are considered to provide equivalent service. Connection optimization utilizes round-robin logic and load-based ordering to determine which addresses to return for a given cluster.

Connection optimization increases overall efficiency by favoring connections to systems with the most available resources and by avoiding unavailable sysplex resources. Addresses of the most available server applications or hosts are returned more frequently than the addresses of loaded server applications or hosts.

A connection-optimized sysplex domain is also scalable—that is, it can add servers and interface addresses dynamically to provide more service capacity. Client applications have dynamic access to the addresses of those servers, with no DNS restart or administration required.

Registration: To ensure maximum availability, server applications register with Workload Manager (WLM), which quantifies the availability of server resources within a sysplex. WLM must be configured in goal mode on all hosts within the sysplex. See “Step 7: Configure WLM in goal mode” on page 673 for a description of this procedure.

TCP/IP stacks also register with WLM. Additionally, they provide the active IP addresses. For a description of how IP addresses are associated with a sysplex domain name, see “Associating IP addresses with the sysplex domain name” on page 663. For a discussion of TCP/IP configuration issues, see “Configuring TCP/IP” on page 669.

When registering, server applications provide the following information:

- *Group name*. This is the name of a cluster of equivalent server applications in a sysplex. It is also the name within the sysplex domain that client applications use to access the server applications. To connect to *any* server application in a group, a client application uses the combination *group_name.sysplex_domain_name*.

  Note: The group name *TCPIP*, the group name for the sysplex domain (for example, *mvsplex*), and the DNS names of the resource records in your sysplex *cluster* zone file are reserved and cannot be used by server applications.
- *Server name*. This is the name of the server application instance. The server name must be unique among all servers that share the same group name. A server application instance can belong to more than one group.
- *Host name*. This is the host name of the TCP/IP stack on which the server application runs.

The sysplex domain name should be registered with the domain name server under a special address, 127.0.0.128, which is used by the ioctl() SIOCGSPLXFQDN. For details, see "Configuring a master (primary) name server" on page 606.

Name resolution: In connection optimization, a name server performs resolution for a name representing a cluster of hosts or server applications. Figure 64 depicts a sysplex domain called `mvsplex.mycorp.com`. The sysplex domain contains only the resources participating in the sysplex. Client applications append the domain name, `mycorp.com`, to all requests for name resolution.

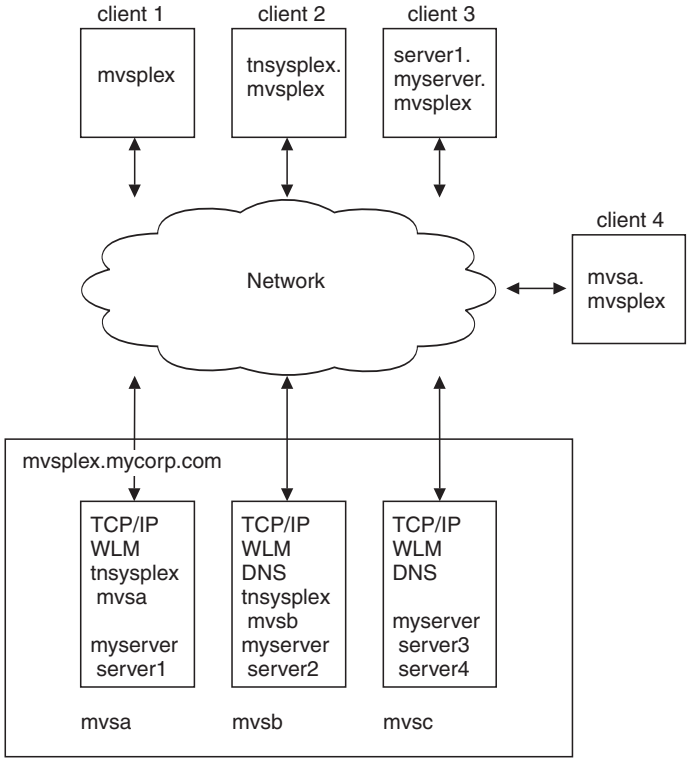


Figure 64. Name resolution to a sysplex

Each host system runs TCP/IP. Hosts `mvsb` and `mvsc` are also running the z/OS UNIX sysplex name servers (`DNS`). `mvsb` is running the master name server for the sysplex subdomain, and `mvsc` is running the slave.

Each host is running one or more `myserver` server applications, and `mvsa` and `mvsb` are also running `tnsysplex` server applications. The group names (`tnsysplex` and `myserver`), the `tnsysplex` server names (`mvsa` and `mvsb`), and the `myserver` server names (for example, `server1`) are known to the name server through WLM registration by the respective application.

Note: `tnsysplex` represents a cluster of TN3270 server applications. Every instance of TN3270 running on a particular host registers using the same server name.

Four types of requests are shown in Figure 64. Client Application 1 requests the services of any host on the system by providing only the name of the sysplex subdomain, `mvsplex`, to DNS for resolution. Client Application 2 requests the services of any `tnsysplex` server instance running in the sysplex. Client

Application 3 requests a particular server instance in the myserver group, and Client Application 4 requests connection to a particular host.

In Figure 64 on page 662, only Client Applications 1 and 2 are candidates for connection optimization. Client Application 1 can connect to either mvsa, mvsb, or mvsc. Client Application 2 can connect to either mvsa or mvsb (but not mvsc). Client Applications 3 and 4, which can connect only to mvsa, are ineligible for connection optimization. They do, however, benefit from the round-robin selection process DNS uses to balance across available network interfaces.

Note: When using connection balancing via MVS clients (such as FTP), to avoid caching of IP addresses, during name resolution set the following environment variable:

```
export _EDC_IP_CACHE_ENTRIES=0
```

Generated names vs. statically defined names: All name servers use *statically defined* names. These are the names in the forward domain data file. As the DNS administrator for a name server using connection optimization, statically define the names of the hosts in the sysplex and the NS and SOA resource records in the forward domain data file. For a description of sysplex forward domain data files, see "Sysplex data files" on page 671.

Note: The host names in the data files must match the host names specified in the stack's resolver configuration file and should be 20 characters or less to ensure server uniqueness.

A name server using connection optimization also uses *generated names*. These are added dynamically to the domain name space as TCP/IP stacks and server applications in the sysplex register with WLM. The name server uses three types of generated names. In Figure 64 on page 662, the following generated names are used:

- The group name that is registered by the server applications (tnsysplex and myserver)
- The server name concatenated with the group name of each server application that registers with WLM in the sysplex (for example, server1.myserver)
- An alias for the sysplex domain name, mvsplex

The generated names become resources (or host names) within the sysplex domain, creating fully qualified domain names:

- Fully qualified group names (these are the connection balanced names):
  - tnsysplex.mvsplex.mycorp.com
  - myserver.mvsplex.mycorp.com
- Fully qualified server names:
  - mvsa.tnsysplex.mvsplex.mycorp.com
  - server1.myserver.mvsplex.mycorp.com
- Fully qualified alias name for the sysplex name mvsplex.mycorp.com:
  - mvsplex.mvsplex.mycorp.com

Associating IP addresses with the sysplex domain name: The sysplex domain name mvsplex.mycorp.com is associated with the *intersection* of the set of statically defined addresses associated with the stacks that are registered with WLM and the set of addresses associated with the adapters that are active on those stacks. The TCP/IP

stack must be registered with WLM for this to occur. Figure 65 depicts this intersection.

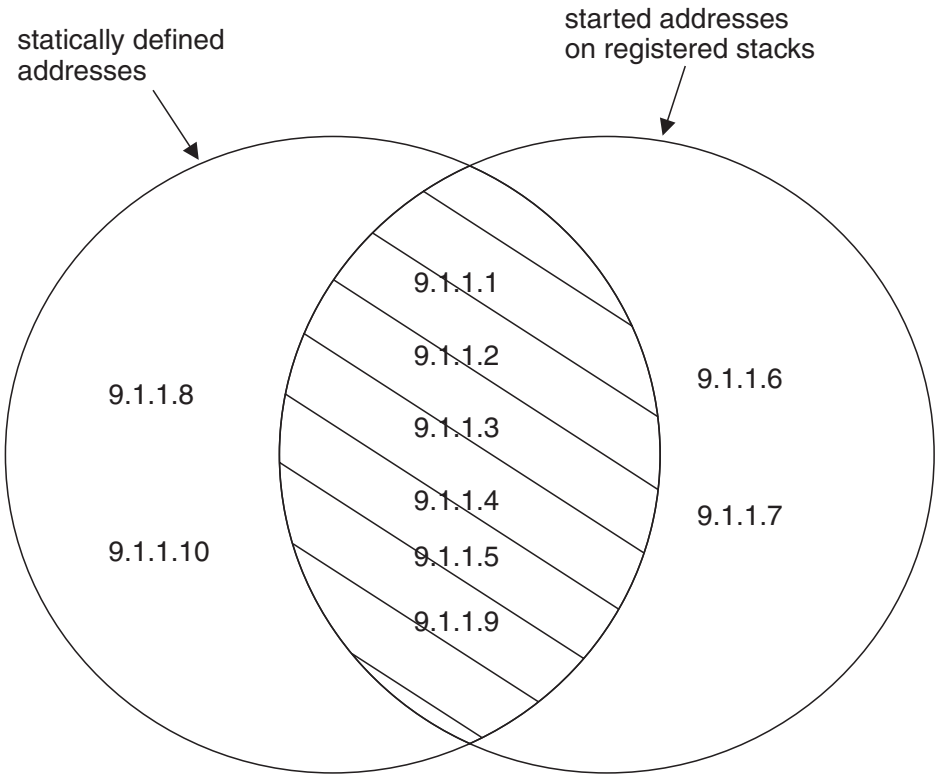


Figure 65. Address association with mvsplex.mycorp.com

The hosts, `mvsa`, `mvsb`, and `mvsc` in the figure have the following addresses in the HOME statement in the PROFILE.TCPIP data set. Only certain adapters are active.

```
mvsa                    mvsb                    mvsc
9.1.1.1 (adapter active) 9.1.1.4 (adapter active)   9.1.1.8
9.1.1.2 (adapter active) 9.1.1.5 (adapter active)   9.1.1.9 (adapter active)
9.1.1.3 (adapter active) 9.1.1.6 (adapter active)   9.1.1.10
                         9.1.1.7 (adapter active)
```

The forward domain data file for the sysplex contains the following statically defined addresses:

```
mvsa            IN      A       9.1.1.1
                IN      A       9.1.1.2
                IN      A       9.1.1.3
mvsb            IN      A       9.1.1.4
                IN      A       9.1.1.5
mvsc            IN      A       9.1.1.8
                IN      A       9.1.1.9
                IN      A       9.1.1.10
```

Using the intersection of two sets lets the domain administrator selectively exclude certain IP addresses from use, such as 9.1.1.6 and 9.1.1.7 on mvsb, while distributing only active addresses to client applications. For instance, 9.1.1.8 would never be used since it is not active.

The process of associating IP addresses with server applications is similar to that for hosts. When a server application registers with WLM, the dynamically generated group name (for example, `myserver`) is added to the sysplex domain. If

the stack on which the server application is running is registered with WLM, then the addresses associated with the group name are the intersection of the statically defined addresses for that stack and those addresses that are associated with adapters that are active. When the server application is replicated on other hosts in the sysplex, the addresses associated with the group name are the union of all intersection sets.

Detailed example: The following example describes in detail how IP addresses become associated with a server application `myserver` running in the sysplex `mvsplex.mycorp.com`. When reading the following section, refer to Figure 66.

The example is described in terms of a *process*, beginning initially with no registered servers applications or stacks. The statically defined addresses associated with the `mvs` hosts and coded in the forward domain data file are the statically defined addresses listed earlier. As server applications and stacks register, the IP addresses associated with `myserver.mvsplex.mycorp.com` change. Figure 66 shows the conclusion of this process.

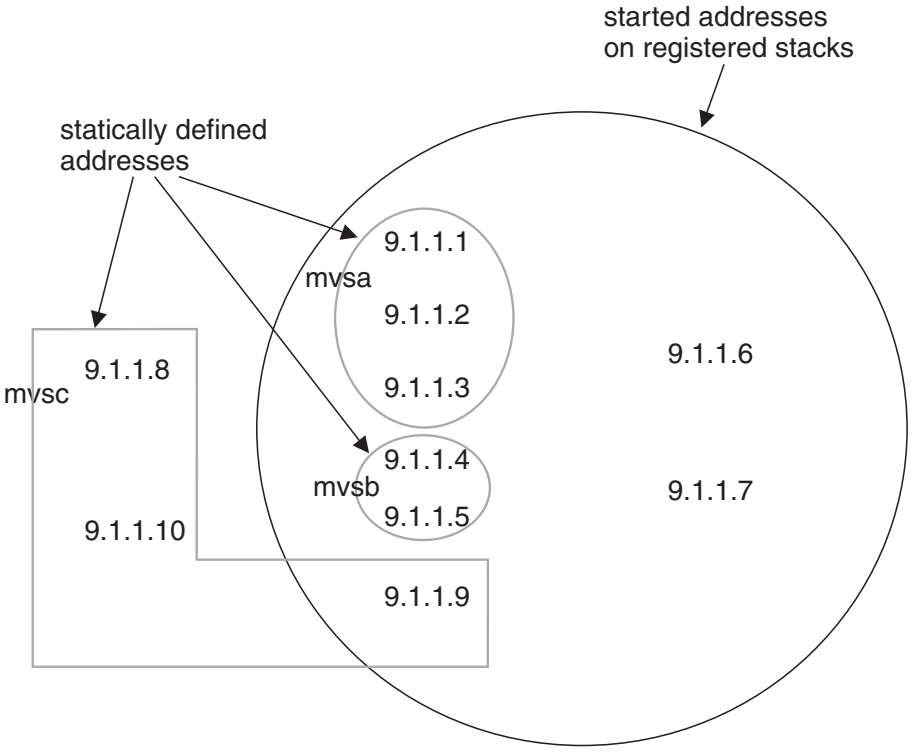


Figure 66. Address association with myserver

Assume initially that a server application which registers with the group name `myserver` is running on `mvsc`, but that the stack on `mvsc` is not currently registered with WLM. The addresses associated with the name `myserver.mvsplex.mycorp.com` are `9.1.1.8`, `9.1.1.9`, and `9.1.1.10`.

Next, assume that the stack on `mvsc` registers with WLM. The addresses associated with the name `myserver.mvsplex.mycorp.com` are reduced to the intersection of the statically defined addresses for the stack and those addresses that are associated with adapters that are active. The only address associated with group name `myserver.mvsplex.mycorp.com` is thus `9.1.1.9`.

Now another instance of an equivalent server application registers with WLM on mvsb, but the stack on mvsb is *not* registered with WLM. All the statically defined addresses associated with mvsb are added to the set of addresses currently associated with the group name myserver.mvsplex.mycorp.com (9.1.1.4, 9.1.1.5, and 9.1.1.9). If the addresses on mvsb are active, the set of addresses associated with the group name does not change.

Similarly, if another instance of an equivalent server application registers with WLM on mvsa and if the addresses on mvsa are active, the set of addresses associated with the server application myserver.mvsplex.mycorp.com are those shown in Figure 66 on page 665 (and Figure 65 on page 664).

Note: The adapters associated with addresses 9.1.1.6 and 9.1.1.7 are not associated with the dynamically generated group name because they are not statically defined in the forward domain data file.

Connecting to a particular server instance: The server name with which the server application registers is unique among all other server instances that share the same group name. Client applications can use the server name to bypass connection optimization. For example, client applications can bypass connection optimization if they are in the middle of a transaction with a server application, and the client/server session fails before the transaction completes. If the client application software has the ability to recognize this situation, the client can reconnect to the same server instance and complete the transaction.

The group name is prefaced with the server name, separated by a dot (server_name.group_name). If the group name is myserver and if the application instance on mvsc registered with WLM as myserver3, then client applications can connect to that particular instance using the name myserver3.myserver.mvsplex.mycorp.com. The address associated with this name (9.1.1.9) is a subset of the addresses currently associated with the group name myserver.mvsplex.mycorp.com that exist on stack mvsc.

Usage considerations in a connection optimized sysplex: Connection optimization extends the concept of a DNS host name to include a name that generically represents (1) all hosts in the sysplex (provided their stacks register with WLM) and (2) names that represent groups of equivalent server applications spread across the sysplex (provided those server applications register with WLM). The maximum benefit from connection optimization is realized when all stacks in the sysplex register with WLM and all TCP/IP server applications register with WLM, if they are capable of doing so.

To take advantage of connection optimization even when registration is not available, consider the following scenarios.

TCP/IP server application does not register: In some cases, you might want to use connection optimization for a particular TCP/IP server application, but it does not register with WLM. If the *stacks* that these applications are running on register with WLM, users can still use connection optimization with server applications by entering the sysplex domain name. A typical invocation might be the following:

```
tn3270 mvsplex
```

where tn3270 is the name of the invoked application and mvsplex is the sysplex domain name.

In this scenario, system administrators must ensure that equivalent instances of the server application are running on each registered stack. Otherwise, connection optimization might result in connecting the client application to a host that is not running the server application. (`mvsc` in Figure 64 on page 662, for example, is not running the TN3270 server.) Depending on the client software, connection timeouts and connection retries might result.

One or more (or all) stacks do not register with WLM: In some cases, you might want to use connection optimization for a particular TCP/IP server application, but one or more (or all) stacks do not register with WLM because they do not support WLM registration or they are not configured to do so. Consider also that the server application does not register with WLM. For the stacks that have not registered, only the statically defined addresses will be used.

In this scenario, a certain degree of connection optimization can occur between hosts whose stacks register with WLM if users enter the sysplex domain name (for example, `mvsplex` in Figure 64 on page 662.) In addition, system administrators must ensure that equivalent instances of the server application are running on each registered stack. Connections will not be made to hosts whose stacks do not register.

Similarly, suppose that the application programmer develops a server application called `ourApp` that registers with the group name of `myserver`. Suppose also that one or more (or all) of the stacks on which the server application runs are not registered with WLM. A typical invocation might be the following:

```
ourApp myserver
```

Since one or more (or all) of the stacks are not registered with WLM, it is possible that an unusable IP address could be returned to the client because a stack has not reported the IP addresses that are active. In this case, only statically defined addresses are used for stacks that are not registered. If an unusable IP address is returned to the client, the connection times out, and depending on the client software, the client application might or might not retry the same or different address returned by the DNS query.

Considerations for connection balancing with multiple instances of TCP/IP on a single host system: Results of connection balancing in a multiple TCP/IP environment are not predictable. Servers such as TN3270 that bind to a single instance of TCP/IP should work as expected as long as they provide the correct host name in the WLM registration. Servers such as FTP which do not bind to a single instance of TCP/IP will have less predictable results. The host name provided on the registration will link each server with a set of IP addresses for a single stack.

Multiple servers on the same port: When running multiple server application instances in a shareport group (that is, all the application instances are sharing the same TCP port on the same TCP/IP stack), only one of the servers should register with WLM.

Caching issues: Proper distribution of server application addresses within a cluster requires DNS queries to be answered by the name server within the sysplex. For this reason, name servers that are located outside the sysplex cannot be configured as master or slave servers for the sysplex domain.

Name servers or resolvers outside the sysplex can prevent client application queries from reaching the sysplex name servers on subsequent requests if they use cached information. This is undesirable for connection optimization since the name

servers and resolvers would not have up-to-date information about capacity and availability of the resources in the sysplex.

To disable other name servers from caching information about sysplex domain resources, a time-to-live (ttl) value of 0 is returned by default. Note that some resolver and name server implementations do not support a ttl value of 0 or anything less than an internally defined minimum (for example, 300 seconds).

Depending on the DNS and network configuration, the number of DNS queries on the network for the sysplex resources might increase. At the expense of reduced availability and load distribution information, administrators can choose a different default ttl for the sysplex resources by using the `-l` option when starting the sysplex name server.

Configuring a sysplex domain for connection optimization

Follow the steps below to configure name servers in a sysplex domain:

1. Identify server applications.
2. Configure server applications for WLM registration.
3. Choose sysplex name and identify name servers in the sysplex.
4. Update parent domain name server.
5. Configure the sysplex name servers.
6. Configure client applications.
7. Configure WLM in goal mode.

Each of these steps is explained below.

Step 1: Identify server applications: Identify the server applications to run in the sysplex. Refer to the product documentation for each application to determine if it supports registration with WLM for connection optimization. It is possible to modify the application to register with WLM. See "Registering your own applications" on page 673.

Candidate applications must have the following attributes:

- Client applications must use DNS for name resolution.
- Server applications must run in a single sysplex.
- Server applications within a specified group provide equivalent functions to their clients. That is, the client application receives the same services from any of the registered server applications.
- To be considered equivalent, all servers registering in a group must be listening on the same port.
- Client applications must use portmapper or a well-known port number to access the server application.

Note: Data Facility Storage Management Subsystem (DFSMS) restrictions do not currently allow sharing of HFS files in write mode.

Maximum benefits are attained when server applications have the following attributes:

- *Registration with WLM.* This feature allows WLM to track the availability of the registering server application and to allow clusters of servers on specified systems within the sysplex. Client application use of the sysplex domain name assumes that the server application is available on all hosts within the sysplex and the stacks running on the sysplex hosts are configured to register with

WLM. Even if a server application does not register with WLM, it might still be able to take advantage of connection optimization under certain circumstances. See "Usage considerations in a connection optimized sysplex" on page 666.

- *Workloads.* The server application has system or network workloads sufficient enough to warrant load distribution. Determination of what is sufficient is subjective, but the value of balancing the incoming connections should outweigh the cost of the extra DNS queries on the network. See "Caching issues" on page 667 for more information.
- *Session Length.* In most cases, server applications selected to use the connection optimization model should have long sessions. Because reduced caching in the network name servers causes additional network traffic, server applications with many short sessions might not be appropriate candidates.

Step 2: Configure server applications for WLM registration: After identifying server applications you want to run in the sysplex, you configure them for WLM registration. Typical configuration involves specification of the group name by which the application will be known. (Do not use ″TCPIP″ or the sysplex domain name as a group name.) See "Registering your own applications" on page 673 for information about how to register the applications you write. The following sections describe how to configure the z/OS Communications Server TCP/IP stack and the applications that are able to register with WLM.

Configuring TN3270: To configure TN3270 servers for registration with WLM, use the WLMCLUSTERNAME and ENDWLMCLUSTERNAME statements to enter a unique sysplex server group name (or names) in the TELNETPARMS section of the PROFILE.TCPIP data set. For a complete description of these statements, refer to *z/OS Communications Server: IP Configuration Reference*.

Unique TN3270 server group names can be used to separate or direct client application requests to only those host systems supporting the target application. For example, a TN3270 server group name of `CICS3270` could be used on host systems actually running the target CICS application that the TN3270 clients access.

If necessary, modify the server group names using the VARY TCPIP,,OBEYFILE command. (The VARY TCPIP,,OBEYFILE command allows changes to the system operation and network configuration without stopping and restarting the TCP/IP address space.) If using the VARY TCPIP,,OBEYFILE command, however, specify all of the TN3270 parameters between the TELNETPARMS and ENDTELNETPARMS statements (not just additions and deletions). For more information on the VARY TCPIP,,OBEYFILE command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Configuring TCP/IP: To register TCP/IP with WLM at startup, add the IPCONFIG SYSPLEXROUTING statement in the PROFILE.TCPIP configuration data set. To register TCP/IP using the VARY TCPIP,,OBEYFILE command, add the IPCONFIG SYSPLEXROUTING statement to the data set referenced by the command. For a complete description of this statement, refer to *z/OS Communications Server: IP Configuration Reference*.

TCP/IP also registers addresses that are active for each stack. These addresses are the active, configured addresses in the `HOME` list for which a `START` device has been processed. WLM registration supports only 15 addresses per TCPIP instance.

To deregister TCP/IP, add IPCONFIG NOSYSPLEXROUTING to the data set referenced by the VARY TCPIP,,OBEYFILE command. (Although it is possible to

add this statement to the PROFILE.TCPIP data set, typical de-registration occurs through VARY TCPIP,,OBEYFILE command processing.) When using the VARY TCPIP,,OBEYFILE command with IPCONFIG NOSYSPLEXROUTING, add IPCONFIG SYSPLEXROUTING on a subsequent VARY TCPIP,,OBEYFILE command to reregister. Note that de-registration occurs automatically when TCP/IP terminates.

Note: Use the IPCONFIG SYSPLEXROUTING and IPCONFIG NOSYSPLEXROUTING statements only when the host is running as part of a sysplex.

Configuring the FTP server: Refer to *z/OS Communications Server: IP Configuration Reference* for more information on WLMCLUSTERNAME statement.

Configuring CICS: See *z/OS Communications Server: IP CICS Sockets Guide*.

Step 3: Choose sysplex name and identify name servers: Identify a unique name with which DNS client applications can access the sysplex. This name can be the same as the configured sysplex name. Names must be 18 characters or less to accommodate WLM restrictions. The sysplex name becomes part of the fully qualified domain name. For example, a sysplex, `mvsplex`, in the domain `mycorp.com` has a fully qualified name of `mvsplex.mycorp.com`. (This is the domain name you specify in the primary directive that contains the `cluster` keyword in the named boot file.)

After selecting a name for the sysplex, identify the master name servers. A sysplex should have only one master name server, and it should have a slave name server to provide redundancy. The slave name server must run in the same sysplex as the master name server. Both the master and slave name servers configured for connection optimization must run on hosts within the sysplex.

Step 4: Update parent domain name server: The parent domain name server is the master name server authoritative for the domain that contains the sysplex subdomain. The parent domain name server can be the same name server as the sysplex name server.

Update the parent domain name server data files by entering the names and IP addresses of the name servers for the sysplex. Use NS (Name Server) resource records to enter the information.

For example, if the name of the domain in which the sysplex is located is `mycorp.com` and the name of the sysplex is `mvsplex` and the master name servers are running on hosts `mvsb` and `mvsc` in the sysplex, add the following records to the forward domain data file of the master domain name server for `mycorp.com`:

```
mvsplex NS mvsb.mvsplex.mycorp.com.
        NS mvsc.mvsplex.mycorp.com.
mvsb.mvsplex.mycorp.com. A  9.67.116.201
                         A  9.67.116.206
                         A  9.67.116.208
mvsc.mvsplex.mycorp.com. A  9.67.116.203
                         A  9.67.116.207
                         A  9.67.116.210
```

This tells the parent domain name server that mvsb and mvsc are master servers for mvsplex. It does not tell the name server that mvsc is slave, only that it is an

additional master name server in the sysplex. The address (A) records are glue records. They enable remote name servers to contact the name servers in the sysplex.

In addition, it might be useful to add CNAME records for resources within the sysplex subdomain to avoid a name change in the client application and to prevent confusion for the end user. See "Step 6: Configure client applications" on page 672.

For example, if client applications use a default domain name of mycorp.com, then requests for `tnsysplex.mycorp.com` can be delegated to the sysplex subdomain with the following resource record:

```
tnsysplex   CNAME   tnsysplex.mvsplex.mycorp.com.
```

Step 5: Configure the sysplex name servers: When the TCP/IP stack is registered with WLM, the list of IP addresses made available to client applications includes the addresses that are common to those provided by TCP/IP and the list of application servers in the forward domain data file of the sysplex name server. See "Generated names vs. statically defined names" on page 663.

Note: The addresses returned to the client application depend upon several factors including whether all or some stacks in the sysplex are registered with WLM, the availability of the adapters assigned to the IP addresses, the names the client application uses for connection, and whether the servers are registered with WLM. See "Usage considerations in a connection optimized sysplex" on page 666.

The steps for configuring sysplex name servers are similar to those for configuring the name servers in an ordinary subdomain:

"Step 2. For BIND 4.9.3–DNS only: specify stack affinity (Multiple stack environment)" on page 609.

"Step 3. Specify port ownership" on page 609.

"Step 4. Update the name server start procedure (Optional)" on page 610.

"Step 5. Create the domain data files (master name server only)" on page 612. See examples below for more information.

"Step 6. Create the hints (root server) file" on page 618.

"Step 7. Create the loopback file" on page 619.

"Step 1a. Create the boot file for BIND 4.9.3–DNS" on page 606.

"Step 11. Start the name server" on page 626.

Sysplex data files: The data files must contain the "A" resource records for the host names internal to the sysplex. The host names must match the host names specified in the stack's resolver configuration file. For a discussion of how IP addresses are associated with a sysplex domain name, see "Associating IP addresses with the sysplex domain name" on page 663.

Note: Consider using VIPA addresses for the MVS hosts. If VIPA addresses are used, they are the only addresses needed to code in the forward domain data file. When Dynamic VIPAs (DVIPAs) are used for VIPA Takeover, code all of the DVIPAs in the sysplex under each host name in the forward domain data file. This will circumvent manual intervention in the DNS data files when a DVIPA is taken over or given back and will not cause any unexpected results in DNS/WLM.

Following is an example of a forward domain data file for a sysplex name server.

```
mvsplex.mycorp.com.  SOA mvsb.mvsplex.mycorp.com.
                              administrator@us.mycorp.com.  (
   1997061300  ; serial
   10800       ; refresh after 3 hours
   1800        ; retry 1/2 hour after failed zone transfer
   3600000     ; expire; length of time secondary keeps data unless refreshed
   259200 )    ; default TTL for static resource records = 3 days
;
;  Define nameservers
          IN  NS  mvsb.mvsplex.mycorp.com.
          IN  NS  mvsc.mvsplex.mycorp.com.
;
;  Define localhost
;
localhost IN  A   127.0.0.0
;
;  Hostnames specified here must match hostnames
;  specified in the stack's resolver configuration file.
mvsa      IN  A     9.1.1.1
          IN  A     9.1.1.2
          IN  A     9.1.1.3
              TXT   "Text record"
              HINFO "z/Series 990" "z/OS V1R7"
mvsb      IN  A     9.1.1.4
          IN  A     9.1.1.5
          IN  A     9.1.1.6
          IN  A     9.1.1.7
mvsc      IN  A     9.1.1.8
          IN  A     9.1.1.9
          IN  A     9.1.1.10
```

Sysplex boot files: The boot file is the main configuration file for a name server. It points the name server to the domain data files, the loopback file, and the hints (root server) file. The contents and format of boot files for sysplex name servers are identical to those for the boot files of ordinary master servers with the exception of an additional keyword, `cluster`. This keyword is used only once in a boot file, at the end of either the primary or secondary directive to identify the sysplex domain.

The following is a sample boot file for the master name server in a sysplex:

```
directory /etc/dnsdata
primary    mvsplex.mycorp.com         named.wlm.for    cluster
primary    113.67.9.in-addr.arpa      named.rev
primary    0.0.127.in-addr.arpa        named.lbk
cache      .                          named.ca
```

The file `named.wlm.for` identifies the forward domain data file for the master name server in the sysplex.

The following is an example of a boot file for a slave name server in a sysplex:

```
directory   /u/usr35/plex/secondary/zone1
secondary   mvsplex.tcp.raleigh.ibm.com   9.67.116.200 mvsplex.bak cluster
secondary   116.67.9.in-addr.arpa         9.67.116.200 mvsplex.rev
primary     0.0.127.in-addr.arpa           db.127.0.0
cache       .                            named.ca
```

Step 6: Configure client applications: Since the sysplex domain is created as a subdomain of an existing domain, resolvers should be reconfigured if it is expected that the clients will use names in the sysplex domain. Otherwise, the client will be forced to use fully qualified domain names to resolve sysplex domain names. For example, if the resolver's default domain was `mycorp.com` before adding the sysplex domain, consider changing the resolver's default domain to

mvsplex.mycorp.com. Also consider adding the sysplex domain to the resolver's search list if the client's resolver supports it.

See "Name resolution" on page 662 for the various names a client can specify.

Step 7: Configure WLM in goal mode: All hosts in a sysplex must operate in goal mode for proper load balancing (splitting). If hosts are not in goal mode, they are treated with equal weight and round-robin selection is applied.

Configure WLM in goal mode on a host by issuing the following MVS command:

```
F WLM,MODE=GOAL
```

Alternatively, IPL in goal mode by omitting the IPS= keyword from your IEASYSxx parmlib member and from your IEASYS00 parmlib member. See *z/OS MVS Programming: Workload Management Services*.

Registering your own applications

Register a server application with WLM using a C interface or an assembler interface. See "Step 1: Identify server applications" on page 668. The C function is invoked as follows:

```
extern long IWMDNREG( char *group_name,
                      char *host_name,
                      char *server_name,
                      char *netid,
                      char *wlm_user_data,
                      long *diag_code);
```

A sample header file, iwmwdnsh.h, comes with the product. Refer to the program directory for its location. The following definitions apply:

- *group_name* is the name client applications use. It can be up to 18 characters.
- *host_name* is the TCP/IP name of the host on which the server is running. See *z/OS XL C/C++ Programming Guide*.
- *server_name* is a unique name that defines a particular instance of the server. It can be up to eight characters.
- *netid* and *wlm_user_data* should be null pointers.

Return values and diag_code values are documented in *z/OS MVS Programming: Workload Management Services* .

To register with a macro, use the IWMSRSRG macro. For a description of this macro, see *z/OS MVS Programming: Workload Management Services*. When using this macro, note the following:

- *Location* contains the *group_name*

  Note: The group name *TCPIP*, the group name for the sysplex domain (for example, *mvsplex*), and the DNS names of the resource records in your sysplex *cluster* zone file are reserved and cannot be used by server applications. If this rule is violated, you will receive message EZZ6649E:

  ```
  WLM group group_name not created
  ```

- *Network_ID* can be blank.
- *LUName* contains the *server_name*.
- *Host* contains the TCP/IP host name.

You can deregister a server application from WLM using a C interface or an assembler interface. Deregister whenever you do not want the server to receive additional client application connections. The C function is invoked as follows:

```
extern long IWMDNDRG( char *group_name,
                      char *host_name,
                      char *server_name,
                      char *netid,
                      long *diag_code);
```

To deregister with a macro, use the IWMSRDRS macro. For a description of this macro, see *z/OS MVS Programming: Workload Management Services*.

For C interfaces for WLM registration and de-registration calls, see the sample registration file, /usr/lpp/tcpip/samples/wlmreg.c.

Dynamic IP

DHCP on z/OS and OS/2 is only compatible with BIND 4.9.3. This section describes the purpose of Dynamic IP and its benefits. Also included is an introduction to the Dynamic IP components and an overview of design concepts.

DHCP on other platforms and Windows 2000 Active Directory can be compatible with BIND 9.

Overview

To add a new workstation to an IP network, several parameters and a variety of information is required to configure the TCP/IP software. Network components, such as a domain name server, are also required. A new TCP/IP host would normally require the following information:

- IP address
- IP subnet mask
- Default router address
- Local host name
- Domain name
- Name server address

Additional parameters, such as other server addresses, time zones, or protocol-specific configurations, might also be necessary.

Keeping track of that information in a large TCP/IP network is not an easy task for network administrators, especially if users or machines or both change their location frequently. IP address lists and domain name server databases have to be updated manually to keep track of any changes in the network.

From a user's point of view, a system administrator would have to be called to provide the necessary information to install a TCP/IP system. If the user moves to another location, this information must not be taken; the user will have to be assigned at least a new IP address if not a new host name as well. Thus, users could potentially cause disorder in a TCP/IP network.

Even if workstations will be automatically installed using software distribution techniques, the TCP/IP configuration parameters have to be pre-assigned per distribution client.

The Bootstrap Protocol (BootP), as described in RFCs 951 and 1497, was introduced to the TCP/IP community in 1985 to provide automatic assignment of some

TCP/IP configuration parameters to a new TCP/IP host. A table has to be maintained at BootP servers to enter information specific to any client that has been planned for installation. Typically, clients are identified by their LAN adapter's hardware address, which has to be known to the system administrator in charge of a BootP server before preparing a new client entry in the database. Even though some manufacturers put the adapter hardware address on a label on the backplane of their LAN adapters, this is a tedious process if many hosts have to be installed in a short period of time.

Objectives and customer benefits of dynamic IP: To lessen the problems of having to manually update any centrally maintained information files and of having a user manually configure a TCP/IP workstation, the Dynamic Host Configuration Protocol (DHCP) has been designed and is described in RFCs 1533, 1534, 1541, and 1542. A DHCP server need not be preconfigured with a workstation's LAN address to submit the necessary TCP/IP configuration to it.

With DHCP in place, the assignment of IP addresses is a lot easier, but one problem persists—how would a domain name server learn about dynamically assigned IP addresses and host names so it can update its database accordingly? This can be solved by the Dynamic Domain Name Services (DDNS).

IBM is actively participating in the designs and implementations of DHCP and DDNS, and it has coined the term *Dynamic IP*. To summarize, the objectives of Dynamic IP and its benefits to TCP/IP system administrators and users are as follows:

- Provides automatic IP network access and host configuration
- Simplifies IP network administration
- Leverages existing IP network products and infrastructure
- Employs only open standards
- Allows customers to administer site-specific host environments
- Enables customized, location-sensitive parameter setups

Note: For further information on the Dynamic Host Configuration Protocol (DHCP) server, see "Configuring the DHCP server for z/OS" on page 679.

Dynamic IP components: Table 30 gives a brief description of the four types of network components that comprise Dynamic IP.

Table 30. DHCP server configuration

System components	Description
Dynamic IP Hosts	Dynamic IP hosts contain DHCP client software and might contain Dynamic DNS client software. Together, they discover and cooperate with their DHCP and Dynamic DNS server counterparts in the network to automatically configure the hosts for network participation.
DHCP Servers	DHCP servers provide the addresses and configuration information to DHCP and BootP clients on the network. DHCP servers contain information about the network configuration and about host operational parameters, as specified by the network administrator. DHCP server can also be configured to be the proxy for the DDNS client and issue the commands to update the Dynamic DNS server.
DDNS Servers	Dynamic DNS servers are a superset of static DNS servers. The dynamic enhancements enable client hosts to dynamically register their name and address mappings in the DNS tables directly, rather than having an administrator manually perform the updates.

Table 30. DHCP server configuration (continued)

System components	Description
BootP Relay Agents (or BootP Helpers)	BootP relay agents can be used in IP router products to pass information between DHCP clients and servers. BootP relays eliminate the need for having a DHCP server on each subnet to service broadcast requests from DHCP clients.

Administering dynamic domains

The DDNS server performs Dynamic DNS database updates in the appropriate domain file as the updates occur. Therefore, **do not edit domain files for dynamic zones while the DDNS server is running.** Furthermore, dynamic domains cannot be dynamically reinitialized with new configuration information using the traditional **nssig -HUP** command.

Also, note that when entering comments into a domain file for a dynamic zone, the comments will be deleted when the first update to the domain is made. Domain file comments are not maintained because they would degrade the performance of the file update process.

Change the configuration information for a dynamic domain in two different ways:

- Manually, by editing the domain files *only after shutting down the DDNS server*
- Dynamically, by using **nsupdate** while the DDNS server is running

For information on how to manually enter configuration information into domain files, refer to "Step 5. Create the domain data files (master name server only)" on page 612.

When first setting up your BIND 4.9.3 dynamic domain, **nsupdate -g** was used to create the zone key RSA key pair; **nsupdate -g** creates an entry in the /etc/ddns.dat file. The public key, the second key in the /etc/ddns.dat file entry, needs to be copied into the appropriate domain file. The zone private key, the first key in the /etc/ddns.dat file entry, is used by **nsupdate** when signing update requests for the administrator of the zone. Note that the zone private key is separated from the public key by a space. The server then examines the signature to identify update requests from the zone administrator versus those from ordinary hosts. The zone key gives the possessor the power to use **nsupdate** to create, modify, and delete any host's record in a dynamic domain.

Once the zone key information is generated and the DDNS server started, the administrator can take the /etc/ddns.dat file with him and administer the zone remotely using **nsupdate**.

Steps for Migrating an existing DNS configuration to BIND 4.9.3 dynamic IP

Before you begin: You need to decide if you want to:

- Leave existing resource records as they are and allow new ones to be created and updated dynamically. This will allow existing systems to keep their host names, but they will not be able to update their resource records dynamically unless a system administrator deletes them.
- Delete all existing resource records and start with a dynamic domain from the beginning.

Perform the following steps to migrate existing DNS server configuration files to Dynamic IP:

1. For a boot file, add the *dynamic secured* or *dynamic presecured* keywords to the *primary* statements for the authoritative DNS server that is being upgraded.

   _____

2. Use the NSUPDATE -g command to create the encryption keys in the etc/ddns.dat file. The command creates a zone private key followed by a public key. The keys are separated by a space. Copy the public key to the zone files.

   _____

3. Start the DDNS server.

   _____

4. Make a new entry in the DDNS.DAT file for each zone the DHCP server will update, if you have a DHCP server configured for DDNS updates.

   _____

You have now migrated existing DNS server configuration files to Dynamic IP.

RSA encryption

Because the z/OS UNIX DDNS server and client products implement not only dynamic DNS but also DNS security functions, below is a brief explanation of cryptography. This section is courtesy of RSA Data Security, Inc., Redwood City, California, and has been modified slightly here for z/OS Communications Server.

Secret key cryptography: This method uses a secret key to encrypt a message. The same secret key must be used again to decrypt the message. This means that the key must be sent along with the message which exposes it to whoever might be eavesdropping on the conversation. Secret keys are very fast in terms of processing, and it is not easy to break them even though they are exposed through the communication process.

Public key cryptography: This method uses a combination of a modulus and a pair of exponents, called the public key and the private key. Exponents and modulus must be used together to encrypt or decrypt a message, but only the modulus and the public exponent are communicated since they are important to everyone who wants to send or receive encrypted messages using this method. The private exponent will never be publicly exposed. This ensures that no one else can decrypt messages that have been intended for a specified recipient, nor can anyone else disguise as that recipient to intercept a message.

Encryption and authentication: Encryption means that a message will be scrambled before it can be sent over a communications link. The plain message itself will never be sent to ensure privacy. Authentication is used to ensure that a message has indeed originated from the source specified in the message, and that the message has not been altered in transit. It additionally serves the purpose of non-repudiation, which means that whoever has digitally signed a message cannot claim later that he or she has not done so. In this case, the plain message itself will be sent since there is no need for privacy. The message will also be used to generate a digital signature by using one of the aforementioned cryptographic methods, preferably public keys.

Hash functions: A hash function is a computation that takes a variable-size input and returns a fixed-size string, which is called the hash value. If the hash function is one-way, that means hard to invert, it is also called a message-digest function, and the result is called a message digest. The idea is that a digest represents concisely the longer message or document from which it was computed; one can think of a message digest as a digital fingerprint of the larger document.

The RSA encryption standard: This standard public key encryption method, along with the MD5 hash function, is used with the IBM DDNS product in z/OS Communications Server. The principle of the RSA algorithm is as follows:

1. Take two large primes, p and q.
2. Find their product n = p * q; n is called the modulus.
3. Choose a number, e, less than n and relatively prime to (p-1) * (q-1).
4. Find its inverse, d, mod (p-1) * (q-1), which means that e * d = 1 mod (p-1) * (q-1).

e and d are called the public and private exponents, respectively. The public key is the pair (n,e); the private key is d. The factors p and q must be kept secret or destroyed.

An example of RSA privacy (encryption) follows. Suppose Alice wants to send a private message, m, to Bob. Alice creates the ciphertext c by exponentiating:

```
c = m^e mod n
```

where e and n are Bob's public key. To decrypt, Bob also exponentiates:

```
m = c^d mod n
```

and recovers the original message, m; the relationship between e and d ensures that Bob correctly recovers m. Since only Bob knows d, only Bob can decrypt.

An example of RSA authentication follows. Suppose Alice wants to send a signed document, m, to Bob. Alice creates a digital signature s by exponentiating:

```
s = m^d mod n
```

where d and n belong to Alice's key pair. She sends s and m to Bob. To verify the signature, Bob exponentiates and checks that the message, m, is recovered:

```
m = s^e mod n
```

where e and n belong to Alice's public key.

Thus encryption and authentication take place without any sharing of private keys: each person uses only other people's public keys and his or her own private key. Anyone can send an encrypted message or verify a signed message, using only public keys, but only someone in possession of the correct private key can decrypt or sign a message.

To make encryption methods secure, a fairly large modulus should be chosen since it becomes increasingly difficult to break a large number into factors to determine the original primes. RSA uses a minimum length of 512 bits for the modulus, which would convert to a number with approximately 155 digits.

Due to security concerns, public key systems that use a key length of more than 512 bits must not be exported from the US.

For encryption, in reality, RSA is combined with a secret-key crypto system, such as DES, to encrypt a message by means of an RSA digital envelope. Data Encryption Standard (DES) is one of the most widely used secret key algorithms and was originally developed by IBM.

Suppose Alice wishes to send an encrypted message to Bob. She first encrypts the message with DES, using a randomly chosen DES key. Then she looks up Bob's public key and uses it to encrypt the DES key. The DES-encrypted message and the

RSA-encrypted DES key together form the RSA digital envelope and are sent to Bob. Upon receiving the digital envelope, Bob decrypts the DES key with his private key, then uses the DES key to decrypt the message itself.

For authentication, in reality, RSA is combined with a hash function, such as MD5.

Suppose Alice wishes to send a signed message to Bob. She uses a hash function on the message to create a message digest, which serves as a digital fingerprint of the message. She then encrypts the message digest with her RSA private key; this is the digital signature, which she sends to Bob along with the message itself. Bob, upon receiving the message and signature, decrypts the signature with Alice's public key to recover the message digest. He then hashes the message with the same hash function Alice used and compares the result to the message digest decrypted from the signature. If they are exactly equal, the signature has been successfully verified, and he can then be confident that the message did indeed come from Alice. If, however, they are not equal, then the message either originated elsewhere or was altered after it was signed, and he rejects the message.

For authentication, the roles of the public and private keys are converse to their roles in encryption, where the public key is used to encrypt and the private key to decrypt. In practice, the public exponent is usually much smaller than the private exponent; this means that the verification of a signature is faster than the signing. This is recommended because a message or document will only be signed by an individual once, but the signature can be verified many times.

Configuring the DHCP server for z/OS

Dynamic Host Configuration Protocol (DHCP) allows clients to obtain IP network configuration information, including IP addresses, from a central DHCP server. The DHCP server controls whether the addresses it provides to clients are allocated permanently or are leased for a specific period. When a client is allocated a leased address, it must periodically request that the server revalidate the address and renew the lease.

The process of address allocation, leasing, and lease renewal are all handled dynamically by the DHCP client and server programs and are transparent to the end user.

DHCP defines three IP address allocation policies:

Dynamic
A DHCP server assigns a temporary, leased IP address to a DHCP client.

Static A DHCP server administrator assigns a static, predefined address reserved for a specific DHCP client.

Permanent
A DHCP server administrator assigns a permanent IP address to a DHCP client. No process of lease renewal is required.

Note: If the network uses routers or gateways, ensure that they can be enabled as DHCP relay agents. Enabling the routers or gateways for DHCP allows the DHCP packets to be sent across the network to other LAN segments.

If there are no routers that can configure to be used as DHCP relay agents, you could:

- Use a UNIX system or RS/6000 system that has the necessary code to be configured to receive limited DHCP broadcasts. Then, forward those broadcast requests to the appropriate host server.
- Use a host server that is located on the same LAN segment as the DHCP client. This would eliminate any need for routers or intermediate UNIX systems to pass on the broadcast requests of the DHCP client.

To configure OSA-Express operating in QDIO mode (for example, gigabit Ethernet) to send and receive broadcast packets, use the IPBCAST parameter on the LINK statement in your TCP/IP profile. The OSA-Express microcode level must support broadcast to use this parameter. Some levels of OSA-Express microcode support multicasting but not broadcasting.

For dynamic address allocation, a DHCP client that does not have a permanent lease must periodically request the renewal of its lease on its current IP address in order to keep using it. The process of renewing leased IP addresses occurs dynamically as part of the DHCP and is transparent to the user.

How does DHCP work?: DHCP allows clients to obtain IP network configuration information, including IP addresses, from a central DHCP server. DHCP servers control whether the addresses they provide to clients are allocated permanently or are *leased* for a specific time period. When a client receives a leased address, it must periodically request that the server revalidate the address and renew the lease.

The DHCP client and server programs handle the processes of address allocation, leasing, and lease renewal.

To further explain how DHCP works, look at some frequently asked questions:

- How is configuration information acquired?
- How are leases renewed?
- What happens when a client moves out of its subnet?
- How are changes implemented in the network?

Acquiring configuration information: DHCP allows DHCP clients to obtain an IP address and other configuration information through a request process to a DHCP server. DHCP clients use RFC-architected messages to accept and use the options served them by the DHCP server. For example:

1. The client broadcasts a message (containing its client ID) announcing its presence and requesting an IP address (DHCPDISCOVER message) and desired options such as subnet mask, domain name server, domain name, and static route.
2. Optionally, if routers on the network are configured to forward DHCP and BOOTP messages (using BOOTP Relay), the broadcast message is forwarded to DHCP servers on the attached networks.
3. Each DHCP server that receives the client's DHCPDISCOVER message can send a DHCPOFFER message to the client offering an IP address (a server that does not want to serve a client can simply ignore the DHCPDISCOVER).

   The server checks the configuration file to see if it should assign a static or dynamic address to this client.

   In the case of a dynamic address, the server selects an address from the address pool, choosing the least recently used address. An address pool is a range of IP addresses to be leased to clients. In the case of a static address, the server uses

a Client statement from the DHCP server configuration file to assign options to the client. Upon making the offer, the IBM DHCP server reserves the offered address.

4. The client receives the offer messages and selects the server it wants to use.
5. The client broadcasts a message indicating which server it selected and requesting use of the IP address offered by that server (DHCPREQUEST message).
6. If a server receives a DHCPREQUEST message indicating that the client has accepted the server's offer, the server marks the address as leased. If the server receives a DHCPREQUEST message indicating that the client has accepted an offer from a different server, the server returns the address it offered to the client to the available pool. If no message is received within a specified time, the server returns the address it offered to the client to the available pool. The selected server sends an acknowledgment which contains additional configuration information to the client (DHCPACK message).
7. The client determines whether the configuration information is valid. Upon receipt of a DHCPACK message, the DHCP client sends an Address Resolution Protocol (ARP) request to the supplied IP address to see if it is already in use. If it receives a response to the ARP request, the client declines (DHCPDECLINE message) the offer and initiates the process again. Otherwise, the client accepts the configuration information.
8. Accepting a valid lease, the client enters a BINDING state with the DHCP server, and proceeds to use the IP address and options.

To DHCP clients that request options, the DHCP server typically provides options that include subnet mask, domain name server, domain name, static route, class-identifier (which indicates a particular vendor), user class, and the name and path of the load image.

However, a DHCP client can request its own, unique set of options. For example, Windows NT 3.5.1 DHCP clients are required to request options. The default set of client-requested DHCP options provided by IBM includes subnet mask, domain name server, domain name, and static route. For option descriptions, see "Specifying DHCP options" on page 698.

Renewing leases: The DHCP client keeps track of how much time is remaining on the lease. At a specified time prior to the expiration of the lease, usually when half of the lease time has passed, the client sends a renewal request, containing its current IP address and configuration information, to the leasing server. If the server responds with a lease offer, the DHCP client's lease is renewed.

If the DHCP server explicitly refuses the request, the DHCP client might continue to use the IP address until the lease time expires and then initiate the address request process, including broadcasting the address request. If the server is unreachable, the client might continue to use the assigned address until the lease expires.

Moving a client out of its subnet: One benefit of DHCP is the freedom it provides a client host to move from one subnet to another without having to know ahead of time what IP configuration information it needs on the new subnet. As long as the subnets to which a host relocates have access to a DHCP server, a DHCP client will automatically configure itself correctly to access those subnets.

For a DHCP client to reconfigure itself to access a new subnet, the client host must be rebooted. When a host restarts on a new subnet, the DHCP client might try to

renew its old lease with the DHCP server which originally allocated the address. The server refuses to renew the request since the address is not valid on the new subnet. Receiving no server response or instructions from the DHCP server, the client initiates the IP address request process to obtain a new IP address and access the network.

Implementing changes in the network: With DHCP, you can make changes at the server, reinitialize the server, and distribute the changes to all the appropriate clients. A DHCP client retains DHCP option values assigned by the DHCP server for the duration of the lease. If you implement configuration changes at the server while a client is already up and running, those changes are not processed by the DHCP client until the client attempts to renew its lease or until it is restarted.

Setting up a DHCP network: The following sections contain information to help you in setting up your DHCP system:

- To create a scoped DHCP network, see "Creating a scoped network."
- To start the DHCP server, see "Starting the DHCP server" on page 683.
- For tips on maintaining a DHCP server, see "Maintaining the DHCP server" on page 684.

The IBM DHCP server provides configuration information to clients based on statements contained in the server's configuration file and based on information provided by the client. The server's configuration file defines the policy for allocating IP addresses and other configuration parameters. The file is a *map* that the server uses to determine what information should be provided to the requesting client.

Before starting the DHCP server, create or modify the DHCP server configuration file.

Once the DHCP server is running, you can also make dynamic changes to the configuration by modifying the configuration file and using the DHCP Server Maintenance program to reinitialize the DHCP server.

Creating a scoped network: You create a hierarchy of configuration parameters for a DHCP network by specifying some configuration values that are served globally to all clients, while other configuration values are served only to certain clients. Serving different configuration information to clients is often based on network location, equipment vendor, or user characteristics.

Depending on your configuration, you can specify subnets, classes, vendors, and clients to provide configuration information to different groups of clients:

- When defined globally, client, vendor or class options are available to DHCP clients regardless of their network location.

  Parameters specified for a subnet, class, or client are considered local to the subnet, class, or client. A client defined within a subnet inherits both the global options and the options defined for that subnet. If a parameter is specified in more than one level in the network hierarchy, the lowest level (which is the most specific) is used.
- Use the Subnet statement to specify configuration parameters for one subnet for a specific location in your network or enterprise.

- Use the Class statement to configure DHCP classes to provide unique configuration information from the server to clients that identify themselves as belonging to that class. For example, a group of clients can all use a shared printer or load image.
- Use a Vendor statement to provide unique configuration information to clients that identify themselves as using a specific vendor's equipment or software. Specially defined options can be served to these clients.
- Use a Client statement in the DHCP server configuration file to serve specified options to a specific client or to exclude that client from service. You can also use a Client statement to exclude IP addresses from service.

For more information on obtaining information for a DHCP client, see "Maintaining the DHCP server" on page 684.

Handling errors in configuration files: Configuring the server incorrectly causes few, if any, warning messages. The DHCP server normally runs even when it encounters errors in the configuration file. The server might ignore the incorrect data and optionally post a message to its log.

Starting the DHCP server:

Note: DHCPSD is installed in the /usr/lpp/tcpip/sbin directory.

To start the DHCP server, use a start procedure (found in the EZATDHDP member of SEZAINST), or the following form of the **dhcpsd** command:

dhcpsd [-q|-v] [-f configFile]

-q Starts the server in **quiet** mode, which means that no banner is displayed when the server starts.

-v Starts the server in **verbose** mode. Causes messages dealing with client communication to print to screen.

-f configFile
Is the name of the DHCP server configuration file. By default, the server searches for a file called dhcpsd.cfg in the directory specified by the ETC environment variable.

When starting the DHCP server with a procedure (proc), the example start proc is found in the DHCP member of the installation partitioned data set SEZAINST.

Stopping the DHCP server: When the DHCP server is started, it stores its process ID (PID) in a Hierarchical File System (HFS) file. If the name of the TCP/IP stack for which the server has affinity can be determined, the process ID file is named /etc/dhcpsd.*tcpipname*.pid, where *tcpipname* is the name of the TCP/IP stack. The name of the TCP/IP stack is determined by calling __iptcpn(), which searches the resolver configuration data set for either of the keywords TCPIPuserid or TCPIPjobname, returning whichever is the last one read. If neither keyword is found in the resolver configuration data set, the character string returned will be a null string and the PID file will be named /etc/dhcpsd.INET.pid.

The DHCP server can be stopped from the UNIX shell or from MVS.

- To stop the server from the UNIX shell, issue the following command, where *pid* is the process ID of the server:

```
kill -s TERM pid
```

The process ID file can be used by issuing the following command, where *tcpipname* is the name of the TCP/IP stack for which the server has affinity:

```
kill -s TERM $(cat /etc/dhcpsd.tcpipname.pid)
```

- To stop the server from MVS, issue the cancel command, where *procname* is the name of the procedure used to start the server:

```
C procname
```

Maintaining the DHCP server:

Note: DADMIN is installed in the /usr/lpp/tcpip/sbin directory.

To maintain a running DHCP server, IBM provides the **dadmin** command to:

- Reinitialize a DHCP server by causing the server to reread its configuration file
- Delete a lease
- Control server tracing
- Display client information
- Display IP address information
- Display server statistics

Note: Verbose mode provides additional information for debugging purposes. Verbose mode is allowed on any of the following **dadmin** command instances. Verbose is shown as a parameter in those instances where additional, more detailed information is of particular value.

Displaying dadmin command syntax: To display information about the command syntax, enter:

dadmin -?

Reinitializing the running server: If you make changes to the configuration file, you will need to reinitialize the running server to implement the changes. To reinitialize the server, use the following form of the **dadmin** command:

dadmin [**-h**]*host*] **-i** [**-v**]

-h Specifies the host

host
The IP address or host name of the DHCP server. If no server is specified, the local server is assumed.

-i Reinitializes the specified server.

-v Executes the command in verbose mode.

Displaying client information: To display information for a client ID, use the following form of the **dadmin** command:

dadmin -c*value* [**-v**]

-c Requests information for one or more clients that match this client ID.

value
The client ID is a MAC address. For example, enter 004ac77150fc. Information is returned for any matching hardware type.

-v Executes the command in verbose mode.

Displaying IP address information: To display information for one IP address, use the following form of the **dadmin** command:

dadmin -q*n.n.n.n* **[-v]**

-q Requests the IP address information.

n.n.n.n
The IP address of the client.

-v Executes the command in verbose mode.

Information returned is an address record for the IP address:

- IP Address - IP address
- Status - Status of the IP Address
  - N/A - Address is not available
  - Free - Address is available
  - Reserved - Address is available, but is reserved for a specific client
  - Leased - Address is currently leased to a client
  - Released - Address has been released by the client
  - Expired - Address lease has expired
  - Used - Address is not available because it is in use in the network
- Lease Time - Length of current lease
- Start Time - Time when address was first leased
- Last Leased - Time of most recent lease
- Proxy - DNS A record update done for the client when lease obtained.
- Client ID - ID for the client associated with this IP address

Querying an address pool: To display information for a pool of IP addresses, use the following form of the **dadmin** command:

dadmin -p*n.n.n.n* **[-v]**

-p Requests the address pool information.

n.n.n.n
The IP address of the address pool.

-v Executes the command in verbose mode.

Information returned is an address record for each IP address in the pool:

- IP Address - IP address
- Status - Status of the IP Address
  - N/A - Address is not available
  - Free - Address is available
  - Reserved - Address is available, but is reserved for a specific client
  - Leased - Address is currently leased to a client
  - Released - Address has been released by the client
  - Expired - Address lease has expired
  - Used - Address is not available because it is in use in the network
- Lease Time - Length of current lease
- Start Time - Time when address was first leased
- Last Leased - Time of most recent lease

- Proxy - DNS A record update done for the client when lease obtained.
- Client ID - ID for the client associated with this IP address

Querying all address pools: To display information for all IP addresses, use the following form of the **dadmin** command:

dadmin -s [**-v**]

-s Requests the address information for all IP addresses.

-v Executes the command in verbose mode.

Returns address records (as described above) for all addresses in the DHCP server's address pools.

Controlling server tracing: To start and stop tracing on the DHCP server, use the following form of the **dadmin** command:

dadmin -t*value* [**-v**]

-t Specifies server tracing.

value
The value is ON to start tracing or OFF to stop tracing.

-v Executes the command in verbose mode.

Displaying server statistics: To display statistics information about the pool of addresses administered by the server, use the following form of the **dadmin** command:

dadmin [[**-h**]*host*] **-n***value* [**-v**]

-h Specifies the host

host
The IP address of the DHCP server. If no host is specified, the local server is assumed.

-n Requests statistics for the server specified as *host*.

value
The value is a decimal integer indicating the number of intervals from 0 to 100. For example, a value of three returns a summary record that includes totals information, the current interval record, and the 3 most recent history records. A value of 0 returns a summary record of activity since the last summary.

-v Executes the command in verbose mode.

Statistics include:
- Discover packets processed
- Discover packets with no response
- Offers made
- Leases granted
- Negative acknowledgments (NAKs)
- Informs processed, including informs plus acknowledgments (ACKs)
- Renewals
- Releases
- BOOTP clients processed

- proxyARec updates attempted
- Unsupported packets
- Monitor requests processed

Deleting leases: If you find that an assigned lease is not being used and you want to make the IP address available for allocation, you can delete the lease. You can only delete one lease at a time. You will be prompted to confirm deletion of the lease. To delete the lease, use the following form of the **dadmin** command:

dadmin [**-f**] [**-v**] [[**-h**]*host*] **-d** *ip_address*

-f Forces deletion of the lease without prompting.

-v Executes the command in verbose mode.

-h

host
Specifies the IP address of the DHCP server. If no server is specified, the local server is assumed.

-d Deletes the lease for the specified IP address.

ip_address
The IP address for the lease to be deleted.

Configuring the DHCP server for the IBM Network Station client: You can configure the DHCP server to be used by an IBM Network Station®. The DHCP server sets up the subnet and specifies the next bootstrap server. The IBM Network Station client can request information. The DHCP server should be configured to provide options that include subnet mask, router, domain name, and boot file name.

Changing the DHCP configuration file

The name of the DHCP server configuration file is dhcpsd.cfg. The default location for dhcpsd.cfg is the \etc directory. In the configuration file, you create a hierarchy of configuration parameters for a DHCP network by specifying some configuration values that are served globally to all clients and other configuration values that are served only to certain clients. The information supplied to the clients is determined by the statements you use and the position of the statements in the configuration file.

Depending on your configuration, you can specify subnets, classes, vendors, and clients to provide configuration information to different groups of clients:

- When defined globally, client, vendor or class options are available to DHCP clients regardless of their network location.

  Parameters specified for a subnet, class, or client are considered local to the subnet, class, or client. A client defined within a subnet inherits both the global options and the options defined for that subnet. If a parameter is specified in more than one level in the network hierarchy, the lowest level (which is the most specific) is used.
- Use the Subnet statement to specify configuration parameters for one subnet for a specific location in your network or enterprise.
- Use the Class statement to configure DHCP classes to provide unique configuration information from the server to clients that identify themselves as belonging to that class. For example, a group of clients can all use a shared printer or load image.

- Use a Vendor statement to provide unique configuration information to clients that identify themselves as using a specific vendor's equipment or software. Specially-defined options can be served to these clients.
- Use a Client statement in the DHCP server configuration file to serve specified options to a specific client or to exclude that client from service. You can also use a Client statement to exclude IP addresses from service.
- A sample DHCP server configuration file is installed in the HFS as /usr/lpp/tcpip/samples/dhcpsd.cfg.

Editing tips: When editing the DHCP server configuration file, keep in mind the following:

- Comments must begin with a # character.
- Class and vendor names that include spaces must be surrounded by double quotes (").
- Statement parameters are dependent on their position. If you omit a required parameter and enter a subsequent required parameter in a statement, the server recognizes that a parameter is missing, writes an error message to a log file, and continues to read the configuration file.
- A continuation character \ indicates that the information is continued on the next line. When used within a comment, the character is treated as part of the comment and is ignored as a continuation character.
- Braces are used to specify statements that are defined within other statements.
- If a parameter is specified in more than one place, the lowest level statement (which is the most specific) is used:
  - Statements specified outside braces are considered global and are used for all addresses served by this server unless the statement is overridden at a lower-defined level.
  - Parameters specified within braces under a statement such as a Subnet statement are considered local and apply only to clients within the subnet.
  - Definition of a parameter in a class takes precedence over definition of the parameter in a subnet.
- Class statements are not allowed inside Client statements.
- Client statements are not allowed inside Option, Vendor, or Class statements.
- Subnet statements are not allowed inside Class or Client statements.
- Vendor statements are always defined at a global level.
- Keywords are not case-sensitive. The capitalization that is used in this documentation is not required in the configuration file. However, use the convention that keywords start with a lowercase letter and subsequent word subparts start with a capital letter. For example, a keyword is proxyARec.

DHCP server statements:

ServerType statement: To specify whether the server will operate only as a standard DHCP server, will perform both normal DHCP operations and PXE proxy DHCP operations, or will act only as a redirection server (PXE proxy DHCP server), use the following statement:

```
ServerType [DHCP | PXEDHCP | PXEPROXY]
```

The default value is `DHCP`, meaning the server will operate only as a standard DHCP server and will not interpret any PXE client extensions. The server will, however, still pass DHCP options and parameters to PXE clients.

PXEDHCP specifies that the server will perform both normal DHCP operations and PXE proxy DHCP operations. PXEPROXY indicates that the server will act only as a redirection server (PXE proxy DHCP server).

ImageServer statement: To specify the siaddr field of the reply packet for a server acting as a PXE DHCP or PXE proxy only server, use the following statement:

```
ImageServer [ip address | hostname]
```

This statement separates the PXE siaddr field (the address of the BINL server) from the DHCP siaddr field. If the server is acting as a PXE DHCP server, it will use the ImageServer value to fill in the siaddr field of the reply packet. If the ImageServer statement is not specified, the siaddr field is left null for the PXE packet. A null value indicates that the BINL server resides on the same machine as the PXE proxy server. ImageServer is a global statement.

Log statements: To enable logging by the server, use the following statements:

- Number of DHCP log files.

  `numLogFiles number_of_log_files`

  number_of_log_files is the maximum number of log files maintained.
- Size of DHCP log file.

  `logFileSize size_of_log_file`

  size_of_log_file is the size of the log file in kilobytes.
- Name of DHCP log file.

  `logFileName name_of_log_file`

  name_of_log_file is the name of the most recent log file.
- Type of log item.

  `logItem type_of_log_item`

  type_of_log_item is the type of the item to be logged. You should specify at least one log item. *type_of_log_item* can be:

  ```
  SYSERR
  OBJERR
  PROTERR
  WARNING
  EVENT
  ACTION
  INFO
  ACNTING
  TRACE
  ```

supportBootP statement: To specify whether the server responds to requests from BOOTP clients, use the following statement:

```
supportBootP [YES | NO]
```

The default value is `NO`.

If this server previously supported BOOTP clients and has been reconfigured not to support BOOTP clients, the address binding for any BOOTP clients that was established before the reconfiguration is maintained until the BOOTP client sends another request (when it is restarting). At that time, the server does not respond, and the binding is removed.

Use this statement outside of braces.

supportUnlistedClients statement: To specify whether the server responds to requests from DHCP clients other than those whose client IDs are specifically listed in this configuration file, use the following statement:

```
supportUnlistedClients [YES | NO]
```

The default is YES. If you specify NO, the server responds only to requests from DHCP clients that are listed (by client ID) in the configuration file.

For example:

```
client 6 10005aa4b9ab ANY
client 6 10a03ca5a7fb ANY
```

If the `supportUnlistedClients` statement is not specified or if you specify YES, the server responds to requests from any DHCP client. Use this option to limit access to addresses that are issued by this DHCP server. Listing the client IDs for all acceptable clients might take a long time.

Use this statement outside of braces.

bootStrapServer statement: To specify whether the DHCP server specifies a bootstrap server for BOOTP clients, use the following statement:

`bootStrapServer` *address_of_bootstrap*

address_of_bootstrap is the IP address of the bootstrap server for the client.

This statement can appear at the global level, or within a Subnet, Class, or Client statement.

Option 67, Boot File Name: For clients who need a boot or must load images to initialize, use the DHCP Option 67 (Boot File Name) for the name of the boot file. The client downloads the image from the BOOTP server. For additional information about Option 67, see "Specifying DHCP options" on page 698.

Lease statements:

- To specify the default lease duration for the leases that are issued by the server, use the following statement:

  `leaseTimeDefault` *amount_of_default_lease_time*

  amount_of_default_lease_time is a decimal integer, followed by a space and a unit of time. The unit of time can be years, months, weeks, days, hours, minutes, or seconds.

  Default interval:
  24 hours (1440 minutes)

  Default unit:
  minute

  Minimum:
  180 seconds

  Maximum:
  -1, which is infinity

  You can use this statement at the global level. To override this statement for a set of clients, use Option 51 (IP Address Lease Time) for a specific client, a class of clients, a subnet, or at the global level.

- To specify the interval at which the lease condition of all addresses in the address pool is examined, use:

  `leaseExpireInterval` *`interval_of_lease_time`*

  interval_of_lease_time is a decimal integer optionally followed by a space and a unit of time, which can be years, months, weeks, days, hours, minutes, or seconds. If you do not specify a unit of time, minutes are assumed. The value that is specified should be less than the value for `leaseTimeDefault` to ensure that expired leases are returned to the pool in a timely manner.

  Default interval:
  1 minute

  Default unit:
  minute

  Minimum:
  15 seconds

  Maximum:
  12 hours
- To specify the maximum amount of time the server holds an offered address in reserve while waiting for a response from the client, use:

  `reservedTime` *`amount_of_time_reserved`*

  amount_of_time_reserved is a decimal integer optionally followed by a space and a unit of time, which can be years, months, weeks, days, hours, minutes, or seconds. If you do not specify a unit of time, minutes are assumed.

  Default interval:
  5 minutes

  Default unit:
  minute

  Minimum:
  30 seconds

  Maximum:
  -1, which is infinity
- To improve the performance of DHCP, limit the times the server accesses the dhcps.ar and dhcps.cr DHCP database files by using the DHCP smart caching feature. Use this feature to reduce the chance of not receiving a DHCP server response to a client request.

  Smart caching can help in cases where the dhcps.ar and dhcps.cr files are large. Smart caching works with the leaseExpireInterval. The default interval is one minute and the maximum interval is 12 hours. Smart caching uses this value as the time interval to access and update dhcps.ar and dhcps.cr files. During this time interval, the DHCP server keeps the new incoming data in memory. The value setting is dependent on the system stability.

  If the DHCP server ends before the leaseExpireInterval, the new data may be lost. Smart caching is enabled at DHCP initialization if the file /etc/dhcp_sc exists in the HFS. The file content is not important, only the existence of the file itself. If the file is located, a DHCP log file message will state this fact and smart caching will be activated.

pingTime statement: For each DHCP client request the DHCP server issues a ping to determine if the IP address chosen for the client is already in use. The server waits the specified time for a ping response before assuming that the IP address is not in use.

```
pingTime  n.n second
```

The value is a time amount. The default time is one second. The use of floating point numbers for the time amount is allowed such as 0.1 seconds to set the ping time to a tenth of a second. The maximum value is 30 seconds. If the amount is set to 0 the DHCP server will not issue a ping.

Subnet statement: Use the Subnet statement to specify configuration parameters for an address pool that is administered by a server. An address pool is a range of IP addresses to be leased to clients. If you configure subnets, you can set the lease time and other options for clients that use the address pool.

- Define a subnet.

  To define a subnet, use the following statement:

  ```
  subnet subnet_address [subnet_mask] subnet_range [(alias=subnet_name ]
  ```

  Note: The DHCP Server Configuration program uses the parameters to the right of a left parenthesis. The DHCP server parses statements to the right of a left parenthesis as comments.

  subnet_address
  The address of this subnet, specified in dotted-decimal notation (for example, 192.67.48.0).

  subnet_mask
  The mask for the subnet, specified in dotted decimal notation or in integer format. A subnet mask divides the subnet address into a subnet portion and a host portion. If no value is entered for the subnet mask, the default is the class mask appropriate for an A, B, or C class network.

  – Class A network - 255.0.0.0
  – Class B network - 255.255.0.0
  – Class C network - 255.255.255.0

  A subnet mask can be expressed either in dotted-decimal notation or as an integer between 8 and 31. For example, you can enter a subnet mask as a dotted decimal notation of 255.255.240.0 or an integer format of 20. In subnet 192.67.48.0, a mask of 255.255.240.0 implies an address range from 192.67.48.001 to 192.67.63.254. The value 20 is the total number of 1's in a mask that is expressed in binary as 11111111.11111111.11110000.00000000.

  Although not required, in most configurations the DHCP server should send Option 1 (Subnet Mask) to DHCP clients. Client operation can be unpredictable if the client does not receive subnet masks from the DHCP server and assumes a subnet mask that is not appropriate for the subnet.

  subnet_range
  All addresses, specified in dotted-decimal notation, to be administered to this subnet. The ranges should not overlap, for example, 192.67.48.1-192.67.48.128.

  Notes:

  1. In the range of addresses, do not include the address of the subnet and the address that is used for broadcast messages. For example, if the subnet address is 192.67.96.0 and the subnet mask is 255.255.240.0, do not include 192.67.96.0 and 192.67.111.255 in the range of addresses.
  2. To exclude an IP address in a range of address, use the Client statement (see step 695).

(alias=*subnet_name*
A symbolic name for ease in identifying a subnet.

- Define a subnet group.

  To define a subnet group, use the following statement:

  `subnet` *`subnet_address`* [*`subnet_mask`*] *`subnet_range`* [`label:`*`value[/priority]`*]

  `label:`*value[/priority]*
  Identifies subnets that are grouped together on the same wire. *value[/priority]* is a string of 1 to 64 alphanumeric characters that identifies the subnet, followed by the priority in which this subnet's address pool is used. Do not uses spaces when specifying the label. More than one subnet can have the same identifier. *priority* is a positive integer, where 1 is a higher priority than 2. If you do not specify a priority, the highest priority is assigned. If two subnets have identical priority, the subnets within a label are processed based on the physical position in the configuration file.

  For example, the following two subnets are on the same wire:

  ```
  inOrder
  subnet 192.67.49.0 255.255.240.0 192.67.49.1-192.67.49.100 label:WIRE1/2
  subnet 192.67.48.0 255.255.240.0 192.67.48.1-192.67.48.50 label:WIRE1/1
  ```

- Serve IP addresses from multiple subnets.

  To serve IP addresses from multiple subnets, use the `inOrder` or `balance` statement. The `inOrder` or `balance` statement is defined at a global level.

  `inOrder` *subnet_labelslist*
  subnet_labelslist is a list of labels in which each label identifies a subnet group. Each listed group is processed in order within that group. The subnet address pool with the highest priority within that group is completely exhausted before the subnet address pool with the next highest priority is used.

  `balance:` *subnet_labelslist*
  subnet_labelslist is a list of labels in which each label identifies a subnet group. The server provides the first IP address from the subnet that is first in the priority list, and subsequent IP addresses from each lesser-priority subnet, repeating the cycle until addresses are exhausted equally from all subnets.

  The following is an example using the `inOrder` statement. Requests for subnet group WIRE1 exhaust addresses in subnet 192.67.48.0 (WIRE1/1) first, followed by subnet 192.67.49.0 (WIRE1/2). WIRE1 and WIRE3 are not related. Requests for subnet group WIRE3 exhaust addresses in subnet 192.67.50.0 (WIRE3/1) first, followed by subnet 192.67.51.0 (WIRE3/2), and then 192.67.50.0 (WIRE3/3), which has the same subnet address as WIRE3/1 but specifies a higher address range:

  ```
  inOrder: WIRE3 WIRE1
  subnet 192.67.49.0 255.255.240.0 192.67.49.1-192.67.49.100 label:WIRE1/2
  subnet 192.67.48.0 255.255.240.0 192.67.48.1-192.67.48.50 label:WIRE1/1
  subnet 192.67.51.0 255.255.240.0 192.67.51.1-192.67.51.50 label:WIRE3/2
  subnet 192.67.50.0 255.255.240.0 192.67.50.1-192.67.50.50 label:WIRE3/1
  subnet 192.67.50.0 255.255.240.0 192.67.50.51-192.67.50.100 label:WIRE3/3
  ```

  The following `balance` statement uses all IP addresses equally in WIRE1/3 and WIRE1/4:

  ```
  balance: WIRE1
  subnet 192.67.49.0 255.255.240.0 192.67.49.101-192.67.49.200 label:WIRE1/3
  subnet 192.67.48.0 255.255.240.0 192.67.48.201-192.67.48.300 label:WIRE1/4
  ```

  A sequence of `inOrder` or `balance` statements is cumulative. For example, the statements:

```
inOrder: WIRE1
inOrder: WIRE3
```

have the cumulative effect of the single statement:

```
inOrder: WIRE1 WIRE3
```

Note: To disable multiple subnets, comment out either the `balance` or `inOrder` processing statement or the priority.

Class statement: Use the Class statement to specify configuration parameters for a user-defined group of clients that are administered by a server. You can use the Class statement at the global or subnet level. When the Class statement is specified within a Subnet statement, the server only serves clients in the class that are located in the specified subnet and request the class.

For example, to create a class that is called "accounting" that allows member hosts to use the LPR server (Option 9) at 192.67.123.2, do the following:

- At the DHCP server, define a class that is called "accounting" and set the LPR server for that class to 192.67.123.2
- At the client, configure the client to identify itself as belonging to the class "accounting".

When the client requests configuration information, the server sees that it belongs to the accounting class and provides configuration information that instructs the client to use the LPR server at 192.67.123.2. DHCP clients use Option 77 to indicate their class to DHCP servers.

To define a class, use the following statement:

`class` *`class_name`* [*`class_range`*]

class_name
The user-defined label that identifies the class. The class name is an ASCII string of up to 255 characters (for example, accounting). If the class name contains spaces, it must be surrounded by double quotes.

class_range
A range of addresses. To specify a range of addresses, enter addresses in dotted-decimal notation, beginning with the lower end of the range, followed by a hyphen, then the upper end of the range, with no spaces between. For example, enter 192.17.32.1-192.17.32.128.

At a global level, a class cannot have a range. A range is allowed only when a class is defined within a subnet. The range can be a subset of the subnet range.

A client that requests an IP address from a class that has used all the addresses from its range is offered an IP address from the subnet range, if available. The client is offered the options associated with the class that has used all the addresses from its range.

To assign configuration parameters such as a lease time for all clients in a class, follow the Class statement with Option statements that are surrounded by braces. For more information about options, see"Specifying DHCP options" on page 698.

Client statement: Use the Client statement to specify a unique set of options for a client. You can assign either a static address and configuration parameters, or

configuration parameters. You can also use the Client statement to exclude an IP address from a range of available IP addresses. You can use the Client statement at the global, subnet, or class level.

- Define a client.

  `client` *hw_type clientID client_ipaddr* `(alias=`*client_name*

  hw_type
  A number that represents the hardware type of the client computer, required to decode the MAC address. For more information about hardware types, see "Hardware types" on page 699.

  clientID
  Either the hexadecimal MAC address, a string such as a domain name, or a name that is assigned to the client such as the host name. If you specify a string, you are required to enclose it in double quotes and specify 0 for the hardware type.

  client_ipaddr
  The DHCP client's IP address in dotted-decimal notation. *client_ipaddr* must contain an address if unlisted clients are not supported.

  (alias=*client_name*
  A symbolic name for ease in identifying the client. Enter **alias=***client_name* immediately after a left parenthesis. This symbolic name appears in the display of the server configuration. If no name is entered, the MAC address is used.

  For example, for a client (10005aa4b9ab), to reserve the static address 192.22.3.149 and also specify a lease time (Option 51) or 12 hours (43200 seconds) and a subnet mask (Option 1), use the following statement:

```
client 6 10005aa4b9ab 192.22.3.149
{
   option 51 43200
   option 1 255.255.255.0
}
```

- Specify options and assign any IP address to a client.

  To specify options for any IP address from the subnet, use the following `client` statement:

  `client` *hw_type clientID* `ANY`

  ANY
  Specifies that any IP address can be assigned to the specific client ID.

- Exclude a client ID.

  To have the DHCP server exclude requests from a particular client ID, use the following `client` statement:

  `client` *hw_type clientID* `NONE`

  NONE
  Specifies no IP address and no options are served to the specified client ID.

  For example:

```
client 6 10005aa4b9ab NONE
```

- Exclude an IP address.

  To exclude one or more IP addresses from the pool of addresses available for lease, use the following statements:

```
client 0 0 192.67.3.123
client 0 0 192.67.3.222
```

In this case, the hardware type and the client ID are 0. IP addresses 192.67.3.123 and 192.67.3.222 are excluded. You must specify a separate statement for each address you want excluded.

- Exclude a range of IP addresses.

  To exclude a range of IP addresses from the pool of addresses available for lease, specify many `client` statements.

  Each range of excluded addresses should contain no more than 10 addresses. Each excluded address results in a separate `client` statement in the configuration file. To exclude larger numbers of addresses, define subnets that do not include the addresses you want excluded. For example, to exclude addresses 50-75 in subnet 192.67.3.0, specify:

```
inOrder: WIRE1
subnet 192.67.3.0 255.255.240.0 192.67.3.1-192.67.3.49 label:WIRE1/1
subnet 192.67.3.0 255.255.240.0 192.67.3.76-192.67.3.100 label:WIRE1/2
```

Vendor statement: To provide vendor configuration information to the DHCP clients in your network:

- Define a vendor and assign the appropriate configuration values. Unlike the Class statement, you cannot control the scope of the Vendor statement by its placement in the file. Use the Vendor statement only at the global level. Vendor statements within Subnet, Class, or Client statements are ignored. Options can be redefined in the vendor class.
- The DHCP client identifies itself to the DHCP server as belonging to a vendor class by sending Option 60 (Class Identifier) with a specific vendor name.
- The DHCP server recognizes that the client has a specific vendor and returns encapsulated Option 43 (Vendor-specific Information), which contains vendor-specific DHCP options and option values.

To define a vendor, use the following statement:

`vendor` *`vendor_name`* `[hex` *`value`*`]`

vendor_name
The user-defined label that identifies the vendor. The vendor name is an ASCII string of up to 255 characters (for example, ″IBM″). If the vendor name contains spaces, it must be surrounded by quotes (″).

[hex *value***]**
value must be specified either as an ASCII string or as hexadecimal in the hexadecimal ASCII string construct. For example:

```
hex "01 02 03"
```

For more information, see descriptions of Option 60 (Class-Identifier) in “Specifying DHCP options” on page 698.

The vendor statement can also be specified in the DHCP server configuration file as a vendor statement followed by a pair of braces containing the options particular to this vendor. Within these braces, the usual option value encoding and decoding rules do not apply.

Statements specifying server information:

- Querying in-use address.

Before the server allocates an IP address, it PINGs (unless pingtime is 0) the address to make sure that it is not already in use by a host on the network. The server places an in-use address in a special pool and allocates a different address.

To specify the amount of time a DHCP server holds an in-use address in a special pool before returning the address to the active pool available for assignment, use the following statement:

`usedIPAddressExpireInterval` *`in_use_time_value`*

in_use_time_value is a decimal integer optionally followed by a space and a unit of time, which can be years, months, weeks, days, hours, minutes, or seconds. The default is 1000 seconds. If you do not specify a unit of time, minutes are assumed.

Default interval:
1000 seconds

Default unit:
minute

Minimum:
30 seconds

Maximum:
-1, which is infinity

- Transform canonical addresses.

  For 802.3 clients, use the canonical keyword to instruct the DHCP server to transform MAC addresses to canonical (byte starts with least significant bit) form. In most cases, you do not want the DHCP server to transform canonical addresses. MAC addresses of 802.3 clients are normally in canonical format on an 802.3 network. When 802.3 MAC addresses are transmitted across a transparent bridge, the bridge reformats the bits that identify an 802.3 client MAC address to a noncanonical (byte starts with most significant bit) form. When the bridge returns the MAC address to an 802.3 network, the bridge again reformats MAC addresses.

  To cause the DHCP server to transform MAC addresses, use the following statement:

  `canonical [YES | NO]`

  NO prevents the DHCP server from transforming MAC addresses. YES causes the DHCP server to transform MAC addresses. NO is the default value.

- Specify statistics snapshots.

  To specify the number of intervals that expire before the DHCP server takes a snapshot of statistics, use the following statement:

  `statisticSnapshot` *`number_of_intervals`*

  The length of each interval is determined by the `leaseExpireInterval` keyword. For example, a value of 3 collects statistics after a span of three intervals, where each interval has a length specified by the `leaseExpireInterval` keyword. If no value is specified, the server takes a snapshot of statistics at the end of every lease expire interval.

Additional options: To assign additional configuration parameters, use the Option statement. All clients inherit all globally-defined options. A client that is defined within a Subnet statement inherits global options and options that are defined for that address pool. To assign configuration parameters for all clients in a subnet, follow the Subnet statement with Option statements that are surrounded by braces.

A sample DHCP server configuration files: The following is a sample DHCP configuration file:

```
logFileName dhcpsd.log
logFileSize 100
numLogFiles 4
logItem SYSERR
logItem ACNTING
logItem OBJERR
logItem EVENT
logItem PROTERR
logItem WARNING
logItem INFO
logItem TRACE
logItem ACTION
supportBootP yes
supportUnlistedClients NO
bootStrapServer 192.168.1.4
option 211 "nfs"
option 212 "10.1.1.2
option 213 "/usr/lpp/nstation/standard/configs/
option 214 "nfs"
option 67 /usr/lpp/nstation/standard/kernel

leaseTimeDefault 12 HOURS

option 15  mycompany.com

# Addresses 8.67.112.24 through 8.67.112.25 do not inherit
# options defined for 8.67.112.26 through 8.67.112.30

subnet 192.168.1.00 255.255.255.0 192.168.1.1-192.168.1.100
 {
  option 1 255.255.255.0
  option 3 10.1.1.1

  client 0 0 192.168.1.4
  client 0 0 192.168.1.5
 }
```

Specifying DHCP options: DHCP allows you to specify options, also known as BOOTP vendor extensions, to provide additional configuration information to the client. RFC 2132 defines the options that you can use. Each option is identified by a numeric code.

Architected Options 0 though 127 and Option 255 are reserved for definition by the RFC. The DHCP server and the DHCP client use options in this set. The administrator can modify some architected options. Other options are for exclusive use by the client and server. The administrator cannot or should not configure the following options at the DHCP server:

- 52 (Option Overload)
- 53 (DHCP Message Type)
- 54 (Server Identifier)
- 55 (Parameter Request List)
- 56 (Message)
- 57 (Maximum DHCP Message Size)
- 60 (Class Identifier)

Options 128 through 254 represent options that can be defined by administrators to pass information to the DHCP client to implement site-specific configuration parameters. Additionally, IBM provides a set of IBM-specific options, such as Option 192 (TXT RR).

The format of user-defined options is:

`option` *code value*

code can be any option code 1 through 254. *value* must always be a string. At the server, it can be an ASCII string or a hexadecimal string. At the client, however, it always appears as a hexadecimal string that is passed to the processing program.

The server passes the specified value to the client. You must, however, create a program or command file to process the value. For more information about data formats for the DHCP options, refer to *z/OS Communications Server: IP Configuration Reference*.

Option categories: The DHCP options are divided into the following categories:

- Base Options
- IP Layer Parameters per Host Options
- IP Layer Parameters per Interface Options
- Link Layer Parameters per Interface Options
- TCP Parameter Options
- Application and Service Parameter Options
- DHCP Extensions Options
- Load Balancing Options
- IBM-Specific Options

Hardware types: The following is a list of hardware types you can specify on the Client statement:

Type	Description
0	Unspecified. If you specify a symbolic name for the client ID, specify 0 for the hardware type.
1	Ethernet (100MB)
6	IEEE 802 Networks (which include 802.5 token ring)

Configuring DHCP server as DDNS client proxy

The DDNS proxy A-record update feature supports clients that do one of the following:

- Use DHCP option 81 to communicate their host name information to a DHCP server.
- Send some or all of their host name information in DHCP option 12 and option 15.
- Have their host name information defined to the DHCP server by an administrator.

The proxyARec update feature allows non-Dynamic IP clients (clients which have DHCP client capability but not DDNS capability) to effectively participate in the Dynamic IP system. This alternative is well-suited for network environments where client hosts are not mobile or do not change administrative domains.

New keywords in the DHCP server configuration file instruct the DHCP server how to interact with the DDNS server using A records on behalf of DHCP clients. The new keywords are:

proxyARec
For more information on enabling DHCP server A record updates, see "Enabling dynamic A record updates."

updateDNSA
For more information on enabling A record updates, see "Specifying the keyword for A record updates" on page 701.

updateDNSP
For more information on enabling PTR record updates, see "Specifying the keyword for PTR record updates" on page 702.

Enabling dynamic A record updates: Once the DHCP server reads the configuration file and is instructed how and when to perform a DDNS proxy A record update, the following rules of precedence are used to derive the host name data to be used:

1. As a first priority, option 81 is included in the DHCP client request and indicates that an A record update should be performed at the DDNS server on behalf of the client.
2. Either option 12 or option 15, or both, are included in a DHCP client request.
3. Either option 12 or option 15, or both, are defined in the DHCP server configuration file.

The data used in the DDNS update packet is derived using one or more of the above means until a fully qualified host name data is established. For example, in the case of Microsoft Windows 3.11+ and Windows 95 clients, only option 12, the Microsoft Windows client computer name, is included in DHCP lease requests. In this case, the administrator can define an option 15, domain name, to use with the host name provided by a client's option 12 to obtain the host's fully qualified name.

The proxy A-record update feature works with any client which is capable of including host name info in option 81 and/or options 12 and 15. Otherwise, the DHCP client host, as identified by its LAN adapter's MAC address, and its host name information can be specified entirely in the DHCP server's configuration file.

Note: If you use DDNS proxy A record support for subnets which include some hosts capable of updating their own DDNS A-records such as IBM OS/2 Warp Dynamic IP clients, you might want to disable those client's own DDNS update feature by commenting out updateDNSA and UpdateDNSTxt keywords in the client's DHCPCD.CFG configuration file. If you do not disable the dynamic A record updates at the clients, the proxy A record updates attempted at the DHCP server for those clients will fail because the DHCP client, rather than the DHCP server, will own the client's entry on the DDNS database.

Specify the proxyARec keyword to enable the DHCP server to dynamically update an A record mapping (host name-to-IP-address) on a DDNS server for clients unwilling or unable to update their A records. The format of the proxyARec keyword is:

```
proxyARec
```

The DHCP server performs updates of the client's A record regardless of the client's client ID value.

The proxyARec keyword is required to be enabled globally or within the subnet, class, or client level. The updateDNSA keyword must also be specified. If the proxyARec keyword is not specified, clients will not have their A records updated.

Specify the proxyARec statement within braces to indicate the information applies only to the subnet, class, or clients that meet the criteria of the preceding conditional statement. To globally assign proxyARec, specify the keyword outside any braces.

Specifying the keyword for A record updates: To specify how the server updates A records for a non-IBM IP client, use keyword:

```
updateDNSA command_string
```

The following is a typical updateDNSA string issued by the DHCP server:

```
updateDNSA "nsupdate -f -h%s -s"d;a;*;a;a;%s;s;%s;3110400;q" -q"
```

This default string is normally adequate. Typical changes to the command string are to modify the value of an expiration time (such as 3110400 seconds) beyond the A record expiration. Assuming proxyARec is specified, this example instructs the DDNS name server to delete all A records for this host name and add a record that maps the host name to the specified IP address for the specified lease time.

The command string includes:

nsupdate
The name of the DDNS client program, which updates the DDNS database.

-f The request originates from a DHCP server.

-h Client host name (value is substituted by the DHCP server).

-s Indicates that the command should be run in command-line mode and all subcommands are included within the quotation marks that follow. The following command string is appropriate for most cases. The string contains:

d;a;*; Delete all A (host name-to-IP-address) records for this host.

a;a;%s;
Add an A record using an IP address (%s) provided by the server, where %s indicates string substitution.

s;%s; Send a lease time (%s) (value provided by the server for the IP address), where %s indicates string substitution.

3110400;
The effect of this value is to reserve the host name for an additional time interval after the A record expiration. In this case, an interval of 36 days (3110400 seconds) is added to the A record expiration time. This value works to preserve a user's host name during times when the name-to-address mapping might expire, such as holidays, vacation, or other times of inactivity.

q Quit delimiter for the command string.

-q Nsupdate will process the request in quiet mode.

For more information on client control of A record updates for non-dynamic DHCP clients, see client option 81 in Specifying DHCP Options.

Releasing a client A record: Use the releaseDNSA keyword at a global level as a template which tells the DHCP server how to release a client record, when the client requests release. The template is adequate for normal purposes, but can be modified if necessary.

The keyword is:

```
releaseDNSA string
```

The following is an example of a releaseDNSA string issued:

```
releaseDNSA "nsupdate -f -h%s -s"d;a;%s;s;%s;0;q" -q"
```

The command string includes:

nsupdate
The name of the DDNS client program, which updates the DDNS database.

-f The request originates from a DHCP server.

-h%s
Client host name (value is substituted by the DHCP server).

-s Information in the command string releases the DHCP client's A record at the DDNS server. The command string is appropriate for most cases. The string contains:

d;a;%s;
Delete the A (host name-to-IP-address) record for this host.

s;%s; Send a lease time (%s) (value provided by the server for the IP address).

0; An additional time interval added to the normal expiration of the host name beyond the expiration of the A resource record. In this case, the value is 0 because the A resource record is deleted.

q Quit delimiter for the command string.

-q Nsupdate will process the request in quiet mode.

Specifying the keyword for PTR record updates: DHCP servers own the PTR records and can update DDNS servers with a PTR record (resource records that map an IP address to a host name) for each address allocated to a host that identifies itself with DHCP options 12 (host name) and 15 (domain name), or with option 81, Client DHCP-DNS.

To enable DHCP server PTR record updates, use:

```
updateDNSP command_string
```

The following is a typical updateDNSP string issued by the DHCP server:

```
updateDNSP "nsupdate -f -r%s -s"d;ptr;*;a;ptr;%s;s;%s;0;q" -q"
```

This default string is normally adequate. Typical changes to the command string are to modify the name expiration extension time such as 0. This example instructs the DDNS name server to delete all PTR records for this address and add a record that maps the IP address to the new host name for the specified lease time. The updateDNSP command is issued when the DHCP client is assigned an address.

The command string includes:

nsupdate
The name of the DDNS client program, which updates the DDNS database.

-f The request originates from a DHCP server.

-r%s
The IP address which is to be mapped to the new host name.

-s The command should be run in command-line mode and all subcommands are included within the quotation marks that follow. The command string is appropriate for most cases. The string contains:

d;ptr;*;
Delete all PTR (IP-address-to-host name) records for this IP address.

a;ptr;%s;
Add a PTR record which maps the IP address to the host name (%s) provided by the server.

s;%s; Send a lease time (%s) (value provided by the server).

0 The effect of this value is to reserve the IP address entry this many seconds beyond when the IPaddress-to-host name mapping expires. In this case, no additional time is added.

q Quit delimiter for the command string.

-q Nsupdate will process the request in quiet mode.

Note: The updateDNSP keyword is equivalent to the updateDNS keyword in earlier releases. The updateDNS keyword continues to be supported.

Releasing a client PTR record: Use the releaseDNSP keyword at a global level as a template which tells the DHCP server how to release the client PTR record, when the client requests release. The template is adequate for normal purposes, but can be modified if necessary. For example, this DHCP server keyword enables immediate release of a PTR record when a mobile client's lease is terminated.

The DHCP server command is:

```
releaseDNSP string
```

The following is an example of a releaseDNSP string:

```
releaseDNSP "nsupdate -f -r%s -s"d;ptr;%s;s;%s;0;q" -q"
```

The command string includes:

nsupdate
The name of the DDNS client program, which updates the DDNS database.

-f The request originates from a DHCP server.

-r%s
The IP address substituted by the DHCP server.

-s Information in this command string releases the DHCP client's PTR (IP-address-to-host-name mapping) record at the DDNS server. The command string is appropriate for most cases. The string contains:

d;ptr;%s;
Delete the PTR record for this address.

s;%s; Send a lease time (%s) (value provided by the server for the IP address).

0 An additional time interval added to the normal expiration of the PTR resource record beyond the expiration of the PTR resource record. In this case, the value is 0 because the record is deleted.

q Quit delimiter for the command string.

-q Nsupdate will process the request in quiet mode.

Defining DHCP proxy authority

Defining DDNS support also includes ensuring the DHCP domain is recognized at the DDNS server and setting up the necessary security for making updates to the DDNS server. To enable A record and PTR record dynamic update support in your DHCP server:

- Ensure that the administrator of the DDNS server configures dynamic reverse domains (*.in-addr.arpa) for all addresses for which the DHCP server will be making dynamic PTR updates.
- Create key file entries in the DDNS.DAT file for each dynamic reverse domain to be updated. If you use DDNS in your network, you must also create a DDNS.DAT file, which contains the security keys used to process IP address and host name updates with a DDNS server.

Defining DDNS key files for the PTR record: The DDNS.DAT file contains the security key pairs to be used for processing IP address-to-host name and host name-to-IP-address updates for the specified addresses which are sent to the corresponding primary DDNS server. The file must contain at least one entry per primary DDNS server.

Run the nsupdate command on the server to create key entries to the /etc/ddns.dat file for DHCP supported clients. The format of the nsupdate command for creating security key pairs for updating PTR records in the DDNS.DAT file is:

```
nsupdate -f -hinverse_ip_address -g -pprimary_ddns_server
```

The command string includes:

-f The request originates from a DHCP server.

-h The inverse IP addresses for which keys are generated.

inverse_ip_address
The format of the inverse_ip_address is the IP address in reverse with in-addr.arpa appended. For example, 9.67.149.111 as an inverse_ip_address entry is 111.149.67.9.in-addr.arpa. You can use a wildcard to expand the scope of the entry. For example, if you want this entry to apply to all hosts whose IP address begins with 9.67, you could specify *.67.9.in-addr.arpa. If you want this entry to apply to all hosts, you could specify *.in-addr.arpa.

-g A key pair should be created for this entry.

-p The primary DDNS server.

primary_ddns_server
The host name or IP address of the primary DDNS server for the specified addresses.

For example:

```
nsupdate -f -g -h *.33.37.9.in-addr.arpa -p ns-updates.company.com
```

The wildcard (*) allows aggregate entries such as *.33.37.9.in-addr.arpa to represent all addresses beginning with the three octets 9.37.33.

The format of the entry in the DDNS.DAT file for a DHCP server is:

```
inverse_ip_address  primary_ddns_server  private_key  public_key
```

The inverse_ip_address and primary_ddns_server are explained above.

The private_key and public_key are used to provide fail-save authentication for updates submitted by this DHCP server. As stored in the DDNS.DAT file, the private key is encrypted.

Defining DDNS key files for the A record: The format of the nsupdate command for creating security key pairs for updating A records in the DDNS.DAT file is:

```
nsupdate -f -g -hfully_qualified_host_name -pprimary_ddns_server
```

The command string includes:

-f The request originates from a DHCP server.

-h The host names for which keys are generated.

fully_qualified_host_name
The format of the fully_qualified_host_name is the host name followed by the fully qualified domain name or a wildcard to allow aggregate entries.

-g A key pair should be created for this entry.

-p The primary DDNS server.

primary_ddns_server
The host name or IP address of the primary DDNS server for the specified addresses.

For example:

```
nsupdate -f -g -h *.city.company.com -p ns-updates.company.com
```

You can specify a specific host, such as myhost, or use a wildcard to allow aggregate entries such as *.city.company.com to represent all hosts in the zone.

Configuring the BINL server

The BINL server (binlsd) is similar to the DHCP server and is started and configured similarly. Port 4011 must be reserved for the BINL server. To reserve port 4011 for binlsd, add the following line to the PORT statement of the TCPIP profile:

```
4011 UDP BINLSD
```

Also, if the system is running in a multiple stack (common INET) environment, the INADDRANYPORT range cannot include 4011.

Following are the command line options that can be used with binlsd:

Option Description

-? Display the help message.

-b Display the program banner.

-q Execute in quiet mode.

-v Execute in verbose mode.

-f Override default configuration file location.

The BINL configuration file, binlsd.cfg by default, is searched for first in the current working directory. If not found in the current working directory, the /etc directory is searched. Following is a sample BINL configuration file shipped as /usr/lpp/tcpip/samples/binlsd:

```
######################################################################
#
#   binlsd.cfg -- BINL (Image Server) Configuration File
#
#   This file contains the directives that can be specified by the
#   server's administrator to configure the server and enforce
#   policies. This file is only a sample.  The finished file must be
#   placed in the directory specified by the ETC environment variable.
#
#   Do not put any long line without spaces in this file.
#
#   A line starting with a '#' character is a comment and is ignored.
#   A '#' on a line which is not part of a quoted string indicates
#   that anything to the right of this character is a comment and should
#   be ignored.
#
#   A continuation character of '\' is supported.  It must be
#   the last non-whitespace character on the line prior to any comments.
#
#   The directives are specified in the form of
#   <keyword> <value1> ... <valueN>.
#
#   Here is a partial list of keywords whose value can be specified
#   in this file:
#
#   Keyword             Effect
#   -------------       ------------------------------------------------------------
#   sa                  Specifies the system architecture.
#   nit                 Specifies the network interface type.
#   lsa1nic             Specifies the lsa1 nic type.
#   tftp                Specifies the address of the tftpd server.
#   bpname              Specifies the install filename given to clients.
#   numLogFiles         The number of log files desired.
#   logFileSize         The size of log files in kilobytes.
#   logFileName         The name of the most recent log file.
#   logItem             An item to be logged.
#   call                Server exit definition.
#   servername          The LCM server name for LSA-1 clients. Not used for LSA-2 clients.
#   serverdomainname    The LCM domain name for LSA-1 clients. Not used for LSA-2 clients.
#
#   client              Definition of a set of options for a specific client
#                       or a definition of a client not to be serviced.
#
#   pxevendor           Configuration delimiter to indicate pxe options to follow.
#
#   pxeoption           A pxe configuration option value to pass to clients.
#
#   The scope of a keyword is limited by a pair of curly brackets ({, })
#   within which the keyword is located.  If a keyword is located outside
#   of any pair of curly brackets, its scope is applicable to all the
#   clients served by this server.  The curly brackets must appear alone on
#   a line.
#
#  Log files.  This set of parameters specifies the log files that will be
#  maintained by this server.  Each parameter is identified by a keyword
#  and followed by its value.
#
#  Keyword      Value           Definition
#  --------     ------------    ----------------------------------------
#  numLogFiles  0 to 99         number of log files.  If 0 is specified,
#                               no log file will be maintained and no log
#                               message is displayed anywhere.  When a log
#                               file reaches maximum size, a new log file
#                               is created, until the maximum number of
#                               log files have been created.  Only the most
```

```
#                               recent n log files are kept/
#
#  logFileSize  in K bytes      maximum size of a log file.  When the size
#                               of the most recent log file reaches this
#                               value, it is renamed and a new log file is
#                               created.
#
#  logFileName  file path       name of the most recent log file.  Less
#                               recent log files have the number 1 to
#                               n-1 appended to their names; the larger
#                               the number, the less recent the file.
#
#  logItem                      An item that will be logged.
#               SYSERR          System error, at the interface to the platform.
#               OBJERR          Object error, in between objects in the process.
#               PROTERR         Protocol error, between client and server.
#               WARNING         Warning, worth of attention from the user.
#               EVENT           Event occurred to the process.
#               ACTION          Action taken by the process.
#               INFO            Information that might be useful.
#               ACNTING         Accounting information on clients served.
#               TRACE           Code flow, for debugging.
#
#

#  call <dll name> <function name> [options]
#                               If present, the server will load the dll given in dllname,
#                               and load the function given in function name and any options
#                               it may have.

#  servername <server name>     If present this name is sent to LSA-1 clients in the offer
#                               packet.  If this option in not present the server will send
#                               its name to the client as the LCM server.

#  serverdomainname  <domain name>  <workgroup | domain>
#                               Specifies the domain name of the LCM server that is sent to
#                               LSA-1 clients in the offer packet.
#
#                                <domain name> the name of the domain.
#
#                                The last parameter can be the string "workgroup" to indicate
#                                that <domain name> is a workgroup or the string "domain" to
#                                indicate that <domain name> is a domain.

#  client   <id_type>  <id_value>  <exclude | blank>
#                               Definition of a client record.
#
#                               <id_type> is one of the hardware types defined
#                               in RFC 1340 (e.g. 1 for 10 megabit Ethernet,
#                               6 for 802.5 Token Ring.)  The type may be
#                               0, in which case the hardware type is not
#                               specified and the
#                               id_value may be a string of any format.
#
#                               <id_value> is a character string if the type
#                               is 0. Typically, this would be a domain name.
#                               For a non-zero <id_type>, the <id_value> is
#                               a hexadecimal string representing
#                               the hardware address of the client.
#
#                               The last paramter can be blank to indicate that
#                               the client matching <id_type> and <id_value>
#                               should get the specific parameters defined in
#                               its scope.
#
#                               The last parameter can be the string "exclude" to
#                               indicate that the client matching
#                               <id_type> and <id_value> should
#                               not be serviced by this server.
#
#       The client statement may be immediately followed
#       by a pair of curly brackets, in which the options
#       particular to this client can be specified.
#
#       Note: All clients inherit all globally defined options.
#
```

```
#
#
#  Pxeoption.  This keyword identifies an option statement.  The options assigned
#  are defined in the "Wired for Management Baseline Version 1.1" specification
#  authored by the Intel Corporation.
#
#  An option is specified by the "pxeoption" keyword followed by the option code
#  of this option and its data field, in a single line.  One or more options
#  may be specified.
#
#  The scope within which an option applies is delimited by a pair of curly
#  brackets ({, }) surrounding the option statement.
#
#  If two or more options with the same option code are specified, the
#  one with the most specific scope is used.  This allows, for example,
#  an option specified at the sa and nit scope to override that same option
#  specified at the global scope.  If two or more options
#  with the same option code are specified within the same scope,
#  the first one read by the server will be the one used, (subject to
#  its being overridden by the same option in a more specific scope).
#
#  All options specified will be sent to the client encapsulated in DHCP option 43.
#

#  Sa.  This parameter specifies the system architecture.  At the global level sa is
#  considered to be "unspecified" and any sa which is received will match if a more
#  specific sa is NOT explicitly configured.  A specific sa may be configured and values
#  for boot path name, TFTP server, and options maybe scoped under this specific specification.
#
#
#  NIT.  This parameter specifies the network interface type.  Works in a similiar manner as
#  the system architecture (Sa.)  It can be configured to more specifically qualify a SA
#  (e.g sa 0 nit 1) or if configured alone will be considered to more specifically qualify
#  the unspecified SA.
#
#  LSA1NIC.  This parameter specifies the network interface type for LSA-1 clients.  Works
#  in a similiar manner as the system architecture (Sa.) and network interface type (NIT.)
#  Specific lsa-1 nic types to provide values for boot path name and TFTP server.  No
#  additional options are sent to the LSA-1 client.
#
#
#  TFTP.   This parameter specifies the address of the tftp server that contains
#  the install image.
#
#  Keyword             Value                 Definition
#  -----------   ----------------------    -------------------------
#  tftp          [ipaddress|hostname]       If "ipaddress" actual address of image server.
#
#                                             If "hostname" dns lookup is done to resolve
#                                               ipaddress of image server.
#
#
#  bpname.  The fully qualified pathname of the install image file.
#
#  Keyword           Value
#  -----------     -------------------
#  bpname         [filename]
#
#
################################################################################

# The remaining portion of this file is an sample configuration file.
# Comments are added to assist in understanding the configuration file.
# Further information and detail is found in the online user documentation.

# Setup of the log file information.  This includes the size and name of the
# logfile along with number of logfiles maintained and type of information that
# will be logged.
numLogFiles     10
logFileSize     1000
logFileName     binlsd.log
logItem         SYSERR
logItem         OBJERR
logItem         PROTERR
logItem         WARNING
logItem         EVENT
```

```
logItem         ACTION
logItem         INFO
logItem         ACNTING
logItem         TRACE

# default TFTP server
tftp   9.33.44.55

# Fully qualified boot Image pathnmae
bpname c:\tftp\lccm\lccm.1

# Global Options
pxevendor
{
  pxeoption 2 555
  pxeoption 3 544
  pxeoption 4 5
  pxeoption 5 5
}

#Excluded client list
client 1 08005aa4bea9 exclude
client 1 08005aa4beaa exclude
client 6 08005aa4beab exclude

#Server exit call
call dllname1 dllfunction1 optionlist

sa 0 nit 1
{
  tftp 9.180.71.79
  bpname c:\tftp\lccm\lccm.1

  #Server exit call
  call dllname1 dllfunction2 optionlist
  #Options specific to this sa and nit values
  pxevendor
  {
    pxeoption 1 224.1.1.1
    pxeoption 2 1758
    pxeoption 3 1759
    pxeoption 4 1
    pxeoption 5 2
  }
}

nit 2
{
  tftp 9.180.71.79
  bpname c:\tftp\lccm\lccm.1

  #Client excluded from service if it falls into this category.
  client 6 08005aa4beac exclude

  # Special configuration for client following into this nit2 type
  client 1 08005aa4bead
  {
    tftp 9.37.56.2
    bpname D:\intel\nit2\nt40s
  }
}

lsa1nic 53
{
  tftp 9.180.71.79
  bpname  d:\lcm\system\images\53.X
  #Server exit call
  call dllname2 dllfunction1 optionlist

  # Special configuration for client following into this nic type
  client 1 08005aa4bdff
  {
    tftp 9.37.56.2
    bpname D:\intel\nic2\nt40s
  }
```

```
  # Deny this client LSA1 service
  client 6 08005aa4bc00 exclude

}

#
# end of binlsd.cfg
#
```

Configuring a DDNS server (BIND 4.9.3 only)

There is no explicit configuration utility for the DDNS server as there is for the DHCP server. You can either create new DDNS server configuration files, or you can migrate an existing DNS configuration to dynamic DNS server configuration files.

There are three ways to use the DDNS server:

- Static DDNS server
- Dynamic secured DDNS server
- Dynamic presecured DDNS server

When used as a static DDNS server, there is nothing you have to do but use your existing DNS configuration files with the DDNS server. It will then work exactly the same way as the previous DNS server.

When used in dynamic secured mode, the DDNS server will allow clients to update their resource records dynamically using encryption keys that have been created by the clients themselves.

When used in dynamic presecured mode, the DDNS server will only allow those clients to update their records to which an encryption key has been provided that has been generated by the administrator.

Creating a new DDNS server configuration: The files required for a minimum configuration are:

- The nameserver boot file contains the path and file names for any other configuration files. The default for this file is /etc/named.boot. It will be examined by the DDNS server at startup.
- The nameserver domain file contains information about the zones for which this server will be authoritative, and all mappings from names to IP addresses (ordinary or forward name resolution). The file name is defined in the nameserver boot file.
- The nameserver reverse file contains information about the mappings from IP addresses to names (inverse or reverse name resolution). The file name is defined in the nameserver boot file.

Use the following steps to create DDNS server configuration files from scratch:

1. Create the DDNS configuration files.

   A nameserver boot file might look as follows:

   ```
   ;
   ; boot file for name server configuration.
   ;
   directory /test/dns/zones/
   ;
   ; type       domain                      source file or host
   ;
   ```

```
primary  test.itsc.raleigh.ibm.com        test.data       dynamic secured
;
primary  200.200.200.in-addr.arpa        db.200.200.200  dynamic secured
;
```

On the primary statements, you can specify if you want to use the DDNS server in dynamic secured or in dynamic presecured mode by using either the *dynamic secured* or the *dynamic presecured* keywords. A nameserver domain file might look as follows:

```
;
;*******************************
;*  Start of Authority Records  *
;*******************************
;
@   IN  SOA  ns-updates.test.itsc.raleigh.ibm.com.
             ns-updates.test.itsc.raleigh.ibm.com. (
        95111601 ; Serial number for this data (yymmdd##)
        86400    ; Refresh value for secondary name servers
        300      ; Retry value for secondary name servers
        86400    ; Expire value for secondary name servers
        3600     ; Minimum TTL value
         300   ) ; dynamic update increment time
    IN  NS   ns-updates.test.itsc.raleigh.ibm.com.
;
localhost    IN A     127.0.0.1
;
ns-updates   IN A     200.200.200.2
martin       IN CNAME ns-updates
;
BPClient     IN A     200.200.200.14
;
```

A nameserver reverse file might look as follows:

```
;
;*******************************
;*  Start of Authority Records  *
;*******************************
;
200.200.200.in-addr.arpa   IN  SOA  ns-updates.test.itsc.raleigh.ibm.com.
                                    ns-updates.test.itsc.raleigh.ibm.com. (
        95111601 ; Serial number for this data (yymmdd##)
        86400    ; Refresh value for secondary name servers
        300      ; Retry value for secondary name servers
        86400    ; Expire value for secondary name servers
        3600     ; Minimum TTL value
         300   ) ; dynamic update increment time
200.200.200.in-addr.arpa.  IN  NS   ns-updates.test.itsc.raleigh.ibm.com.
;
;
; Addresses for the canonical names
;
2          IN PTR martin.test.itsc.raleigh.ibm.com.
14         IN PTR BPClient.test.itsc.raleigh.ibm.com.
;
```

2. After you have created the files, use the NSUPDATE -g command to create the encryption key pairs for the zone resource records in the domain and reverse files.

   After the administrator has copied the public key into the files, they might look as follows:

   - Domain Name file:

```
;
;*******************************
;*  Start of Authority Records  *
;*******************************
;
@   IN  KEY     80 0 1 AQPOzUYWvAUyZhYxogDcrtxOZOH33V31Tmrs1Db1WYiyI4Y
                       7Mmoz6Vm3XY/QTMHOyeHcVAMKmuba+rW4/+IkMeP3
@   IN  SOA  ns-updates.test.itsc.raleigh.ibm.com.
             ns-updates.test.itsc.raleigh.ibm.com. (
        95111601 ; Serial number for this data (yymmdd##)
        86400    ; Refresh value for secondary name servers
        300      ; Retry value for secondary name servers
        86400    ; Expire value for secondary name servers
        3600     ; Minimum TTL value
         300   ) ; dynamic update increment time
    IN  NS   ns-updates.test.itsc.raleigh.ibm.com.
;
localhost    IN A     127.0.0.1
;
ns-updates   IN A     200.200.200.2
martin       IN CNAME ns-updates
;
BPClient     IN A     200.200.200.14
;
```

- Reverse Name file:

```
;
;*******************************
;*  Start of Authority Records  *
;*******************************
;
200.200.200.in-addr.arpa   IN  KEY     80 0 1 AQPR+3ObXCgcjmBfKSnN4fD6v
                                              VH/AUIwincGNeD1MAuz2BTQSQ
                                              /bJkXLA3nxfV+HxKfxWptkRck
                                              wzxEk1DD3DSB
200.200.200.in-addr.arpa   IN  SOA  ns-updates.test.itsc.raleigh.ibm.com.
                                    ns-updates.test.itsc.raleigh.ibm.com. (
        95111601 ; Serial number for this data (yymmdd##)
        86400    ; Refresh value for secondary name servers
        300      ; Retry value for secondary name servers
        86400    ; Expire value for secondary name servers
        3600     ; Minimum TTL value
         300   ) ; dynamic update increment time
200.200.200.in-addr.arpa.  IN  NS   ns-updates.test.itsc.raleigh.ibm.com.
;
;
; Addresses for the canonical names
;
2          IN PTR martin.test.itsc.raleigh.ibm.com.
14         IN PTR BPClient.test.itsc.raleigh.ibm.com.
;
```

The NSUPDATE -g command will also create the DDNS.DAT file that contains the private encryption keys to sign any updates to the zone resource records in the domain and reverse files. This is shown in the following example:

```
test.itsc.raleigh.ibm.com ns-updates.test.itsc.raleigh.ibm.com
Pb7bySIfzXcW...

200.200.200.in-addr.arpa ns-updates.test.itsc.raleigh.ibm.com
KlexSRMP/q/k...
```

3. Start the DDNS server.
4. If you have a DHCP server configured for DDNS updates, you need to add an entry into the DDNS.DAT file using NSUPDATE -g that represents a wildcard entry for the zones that DHCP will update (*.test.itsc.raleigh.ibm.com).

Note: KEY and SIG resource records, as well as encryption keys, always use a single line. The examples were indented for illustration purposes only.

Configuring for presecured mode operation: The dynamic DNS server supports two modes of securing updates to a dynamic DNS zone. In the default mode, called *dynamic secured*, any host that complies with the DDNS protocol can create resource records in a zone declared as dynamic. After created, these records can only be updated by the administrator or by the host that created them.

In presecured mode, only hosts that have been preauthorized by a DDNS administrator may create resources in a particular dynamic zone. After created, these records can only be updated by the administrator or by the host that has been preauthorized.

Functionally, the difference in these modes is whether or not the DDNS server will allow the creation of a KEY RR or whether it must already have a KEY RR defined for a resource in the zone. In the case of presecured mode, the KEY RR must be already defined in the zone before an update is accepted. Specifically, the administrator must enter the KEY RR data in the domain file for each client that will be making updates.

To configure a particular DNS dynamic zone for presecured mode operation, you must:

1. Specify **dynamic presecured** on the primary zone statement in the DDNS server boot file and create the corresponding domain file.
2. Run the nsupdate -g command to create an /etc/ddns.dat with the zone key information. Create a zone KEY RR.
3. Start the DDNS server by entering **named** at a z/OS UNIX command prompt.
4. For each client host:
   a. Use **nsupdate** with the -g parameter, which will generate a key for the host and save it in /etc/ddns.dat.
   b. Use **nsupdate** with the -a parameter to dynamically register and save a host's KEY RR in the DDNS server domain file.
   c. Manually extract the /etc/ddns.dat file key entry for the host into a separate DDNS.DAT file and distribute it to the end user for installation into their /etc subdirectory.

The following example is a scenario demonstrating an administrator using **nsupdate** to preregister an end user in a **dynamic presecured** domain. Administrator input is highlighted in bold:

```
#nsupdate -g -h newuser.dynozone.sandbox -p netadmin.dynozone.sandbox
--- NSUPDATE Utility ---
---
Key Gen ...... succeeded ...


#nsupdate -a -h newuser.dynozone.sandbox -p netadmin.dynozone.sandbox
--- NSUPDATE Utility ---

Enter Action (Add,Delete,Exists,New,TTL,Send,Quit)
> a
..rrtype (A,PTR,CNAME,MX,KEY,HINFO): key
DDNSUpdate_KEY (Add Flags 0000 Protocol 0 Algid 1
      Keylen 64 Key[0-15]:       AQO/3Ah6986cXDhR  ...        succeeded

Enter Action (Add,Delete,Exists,New,TTL,Send,Quit)
> s
```

```
..sig Expiration (secs from now, ENTER for 3600):
..sig KEY pad (ENTER for default of 3110400):
DDNSSignUpdate ...succeeded
DDNSFinalizeUpdate ...succeeded
DDNSSendUpdate ...succeeded

Enter Action (Add,Delete,Exists,New,TTL,Send,Quit)
> q

#
```

DNS-related RFCs

The following RFCs contain basic information about the DNS:

974 *Mail Routing and the Domain System*, C. Partridge

1033 *Domain Administrators Operations Guide*, M. Lottor

1034 *Domain Names—Concepts and Facilities*, P.V. Mockapetris

1035 *Domain Names—Implementation and Specification*, P.V. Mockapetris

Proposed standards

1995 *Incremental Zone Transfer in DNS*, M. Ohta

1996 *A Mechanism for Prompt Notification of Zone Changes*, P. Vixie

2136 *Dynamic Updates in the Domain Name System*, P. Vixie, S. Thomson, Y. Rekhter, and J. Bound

2181 *Clarifications to the DNS Specification*, R. Bush Elz

2308 *Negative Caching of DNS Queries*, M. Andrews

2845 *Secret Key Transaction Authentication for DNS (TSIG)*, P. Vixie, O. Gudmundsson, D. Eastlake, 3rd, and B. Wellington

Proposed standards still under development

1886 *DNS Extensions to support IP version 6*, S. Thomson and C. Huitema

2065 *Domain Name System Security Extensions*, D. Eastlake, 3rd and C. Kaufman

2137 *Secure Domain Name System Dynamic Update*, D. Eastlake, 3rd

Other important RFCs about DNS implementation

1535 *A Security Problem and Proposed Correction With Widely Deployed DNS Software*, E. Gavron

1536 *Common DNS Implementation Errors and SUggested Fixes*, A. Kumar, J. Postel, C. Neuman, P. Danzig, and S. Miller

1982 *Serial Number Arithmetic*, R. Elz and R. Bush

Resource record types

1183 *New DNS RR Definitions*, C.F. Everhart, L. A. Mamakos, R. Ullmann, and P. Mockapetris

1706 *DNS NSAP Resource Records*, B. Manning and R. Colella

1876 *A Means for Expressing Location Information in the Domain Name System*, C. Davis, P. Vixie, T., and I. Dickinson

2052 *A DNS RR for Specifying the Location of Services*, A. Gulbrandsen and P. Vixie

2163 *Using the Internet DNS to Distribute MIXER Conformant Global Address Mapping*, A. Allocchio

2168 *Resolution of Uniform Resource Identifiers using the Domain Name System*, R. Daniel and M. Mealling

2230 *Key Exchange Delegation Record for the DNS*, R. Atkinson

DNS and the Internet

1101 *DNS Encoding of Network Names and Other Types*, P. V. Mockapetris

1123 *Requirements for Internet Hosts - Application and Support* Braden

1591 *Domain Name System Structure and Delegation*, J. Postel

2317 *Classless IN-ADDR.ARPA Delegation*, H. Eidnes, G. de Groot, and P. Vixie

DNS operations

1537 *Common DNS Data File Configuration Errors*, P. Beertema

1912 *Common DNS Operational and Configuration Errors* D. Barr

2010 *Operational Criteria for Root Name Servers*, B. Manning and P. Vixie

2219 *Use of DNS Aliases for Network Services*, M. Hamilton and R. Wright

Other DNS-related RFCs

1464 *Using the Domain Name System To Store Arbitrary String Attributes,* R. Rosenbaum

1713 *Tools for DNS Debugging* A. Romao

1794 *DNS Support for Load Balancing*, T. Brisco

2240 *A Legal Basis for Domain Name Allocation*, O. Vaughan

2345 *Domain Names and Company Name Retrieval*, J. Klensin, T. Wolf, and G. Oglesby

2352 *A Convention For Using Legal Names as Domain Names*, O. Vaughan

Chapter 14. Policy-based networking

Businesses typically define goals for network behavior in human terms. Network implementations provide a wide variety of controls for priority treatment of traffic, bandwidth management, security, and control of network behavior. The link between the high level business goals and network implementations is defined as policy-based networking and is provided by policies. Policies are an administrative means to define controls for a network decision or a resource balancing decision. To be effective, consistent mechanisms need to be used throughout the network to classify and differentiate traffic, and to provide a consistent implementation of policy decisions. To achieve this, policies are usually defined in a central policy repository, accessed by all hosts and routers that need to make policy decisions (Policy Decision Point, or PDP), or implement such decisions (Policy Enforcement Point, or PEP).

For details on the types of policies supported, see "Policy components overview."

Policy components overview

Policy Agent

Network administrators can use the z/OS Communications Server Policy Agent to define policies for their users. Superuser authority is required.

The policies supported by the Policy Agent can be used for any of the following purposes:

- Quality of service (See Chapter 15, "Quality of service (QoS)," on page 751)
- Intrusion Detection Services (See Chapter 16, "Intrusion Detection Services (IDS)," on page 789)
- IP filtering, and manual and dynamic VPN tunnels, collectively referred to as IPSec policies (See Chapter 17, "IP security," on page 819)
- Application Transparent Transport Layer Security (AT-TLS, see Chapter 18, "Application Transparent Transport Layer Security (AT-TLS) data protection," on page 987)

The common Policy Agent functions are further described in this chapter.

The Policy Agent supports QoS functions other than reading and installing policies, such as Sysplex Distributor policy performance monitoring, and mapping IPv4 Type of Service (ToS) byte or IPv6 Traffic Class values to outbound interface and virtual LAN (VLAN) user priorities. The QoS specific Policy Agent functions are further described in "QoS-specific Policy Agent functions" on page 754.

A Policy API (PAPI) is provided to allow access to policy information by external user applications. The PAPI interface can be used by policy performance monitoring applications to retrieve policy performance data. For more information on PAPI, refer to *z/OS Communications Server: IP Programmer's Guide and Reference*.

The Policy Agent runs in the z/OS environment and reads policy definitions from local configuration files, a central repository that uses the Lightweight Directory

Access Protocol (LDAP), or both. The Policy Agent also installs policies in one or more z/OS Communications Server stacks. It can be used to replace existing policies or update them as necessary.

Notes:

1. Policies defined on an LDAP server use the configuration files and mechanisms provided by the LDAP server product. The definition of the elements of policies is known as the *schema*. z/OS Communications Server provides the schema definition for policies that may be defined on an LDAP server in a set of sample files. The sample files are provided in both LDAP protocol version 2 and version 3 format (see "Policy sample files" on page 721 for the names of these samples). These sample files must be installed on the LDAP server as the schema definition. Policy Agent uses the z/OS Integrated Security Server LDAP Client library to communicate with an LDAP server. Refer to *z/OS Integrated Security Services LDAP Server Administration and Use* for more information about LDAP and *z/OS Integrated Security Services LDAP Client Programming* for more information about the z/OS Communications Server LDAP Client support.
2. A copy of the policy schema definition files that define the policy schema for LDAP protocol version 2 are also available in the Appendix of the *z/OS Communications Server: IP Configuration Reference*. Different files are used to define the schema in LDAP protocol version 3 syntax, but are not shown because the protocol version 2 syntax is easier to read and also because the schema definitions are essentially the same.

QoS services

It is important to be able to monitor the performance of policies as implemented by the network. This leads to the need to define service level management information kept by the PDP and PEP, which can then be analyzed for a variety of network services, from traffic trend analysis to network capacity planning and dynamic QoS-level tuning.

Together, service differentiation, policies, and service-level performance monitoring form an integral part of the service-level agreements (SLA) function that is becoming more and more important to network administrators as a means of controlling network traffic flows.

QoS policies are used by the following functions:

- RSVP agent
- Network SLAPM MIBs

RSVP agent

The z/OS Communications Server RSVP agent provides integrated services functions, such as communicating with RSVP agents on other hosts or routers and reserving resources on certain types of outbound interfaces. The RSVP agent queries the Policy Agent for QoS policies that relate to RSVP processing.

Network SLAPM MIBs

This section contains information on the z/OS CS Network SLAPM2 subagent (nslapm2) and the SLA subagent (pagtsnmp). The subagent queries Policy Agent for QoS policies. Using the Network SLAPM2 subagent is recommended for the following reasons:

- The Network SLAPM2 subagent has better and more relevant information on network and TCP related performance data (for example, loss, delays, 64-bit count value).

- The Network SLAPM2 subagent reduces system overhead, due to the restructure of the MIB and the interface for the subagent to get performance data.
- The Network SLAPM2 subagent provides transparency in supporting either IPv4 or IPv6, since this MIB no longer keeps track of IP addresses.
- The SNMP networking SLA subagent will be discontinued in the next release of z/OS. This decision has been made due to extensive improvements in the SNMP Network SLAPM2 subagent in several areas, such as policy performance monitor information, simplicity, and less platform overhead for data collection. It is strongly recommended that management applications be moved to the SNMP Network SLAPM2 MIB as soon as possible.

SNMP Network SLAPM2 subagent

The z/OS CS Network SLAPM2 subagent (nslapm2) allows network administrators to retrieve data and determine if the current set of Network SLAPM2 policy definitions are performing as needed or if adjustments need to be made. The Network SLAPM2 subagent supports the Network Service Level Agreement Performance Monitor (NETWORK-SLAPM2) MIB. Refer to usr/lpp/tcpip/samples/slapm2.mi2 for more information about the Network SLAPM2 MIB.

SNMP SLA subagent

The z/OS Communications Server SLA subagent (pagtsnmp) allows network administrators to retrieve data and determine if the current set of SLA policy definitions are performing as needed or if adjustments need to be made. The SLA subagent supports the Service Level Agreement Performance Monitor (SLAPM) MIB. Refer to RFC 2758 for more information about the SLAPM MIB.

IPSec services

IPSec services include IP filtering and support for manual and dynamic VPN tunnels. The Internet Key Exchange (IKE) daemon works with the stack to provide IPSec support. IPSec policy can be defined for IP filtering (including manual VPN tunnels), key exchange, and dynamic VPN tunnels.

Restriction: You can define IPSec policies only in configuration files. You cannot define them on LDAP servers.

Application Transparent Transport Layer Security (AT-TLS) services

AT-TLS provides invocation of system SSL in the TCP transport layer of the stack. This support is policy driven and can be deployed transparently underneath many existing sockets applications, without modification to the application. System SSL provides support for the TLS, SSLv3, and SSLv2 protocols. AT-TLS policy provides system SSL configuration for connections that are to use AT-TLS. The application continues to send and receive clear text over the socket, but data sent over the network is protected by system SSL.

IOCTL support is provided for applications that need to be aware of AT-TLS for status or to control the negotiation of security.

Restriction: You can define AT-TLS policies only in configuration files. You cannot define them on LDAP servers.

Intrusion Detection Services

Intrusion Detection Services (IDS) support is available to detect and report on network intrusion events. The Traffic Regulation Management (TRM) support has been extended and incorporated into the IDS support. IDS policy regulates the types of events to report and provides the definition of several types of events. IDS policy may be defined for scans, attacks and traffic regulation for both TCP and UDP ports.

Results:

1. Policy Scope TR policies found in a Policy Agent configuration file are compatibly transformed into IDS TR TCP policies by Policy Agent.
2. **pasearch** will display the transformed policy.

Restriction: You can define IDS policies only on an LDAP server.

Policy Agent overview

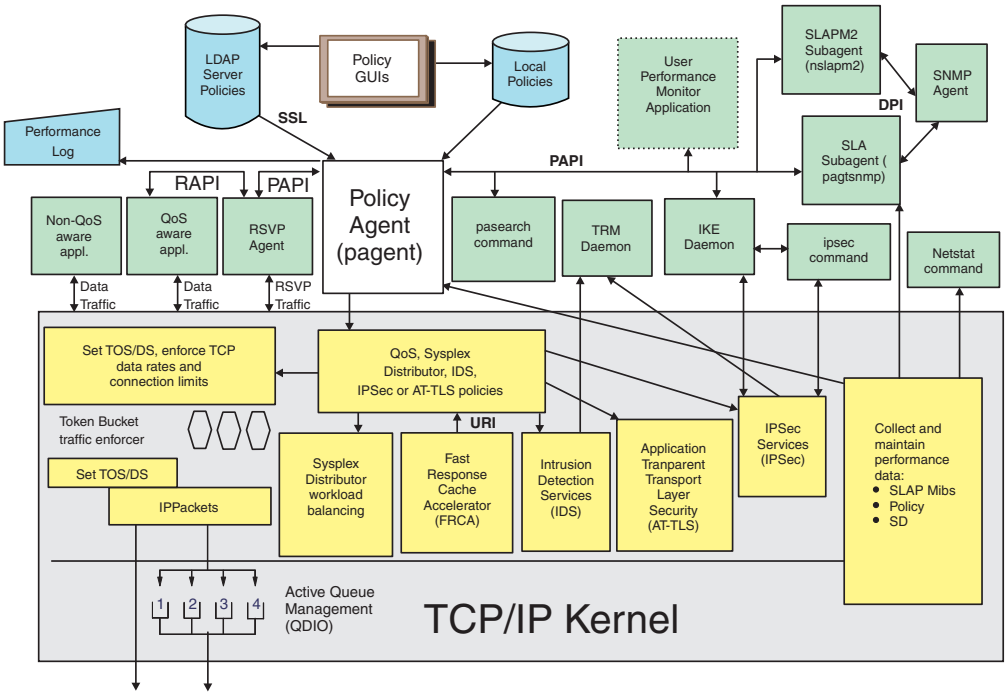


Figure 67. Policy Agent overview

Figure 67 shows the big picture for Policy Agent. The policy GUIs shown include the following:

zQoS Manager
GUI to build QoS LDAP policies

eServer™ IDS Configuration Manager
GUI to build IDS LDAP policies

z/OS Network Security Configuration Assistant
GUI to build the configuration flat-file for IPSec and AT-TLS policies.

Policy sample files

A set of sample files is shipped with z/OS CS that provides several functions. The first sample file provides an example of policy definitions in a Policy Agent configuration file.

/usr/lpp/tcpip/samples/pagent.conf
This file contains overall policy definition rules, syntax and semantics for defining policies in a configuration file, and examples of such policy definitions.

The next set of sample files provide sample IPSec policy definitions.

/usr/lpp/tcpip/samples/pagent_CommonIPSec.conf
This file contains sample common IPSec policy definitions. These can be referenced and reused by multiple stack-specific IPSec configuration files.

/usr/lpp/tcpip/samples/pagent_IPSec.conf
This file contains sample stack-specific IPSec policy definitions. Some of these refer to common definitions in /usr/lpp/tcpip/samples/pagent_CommonIPSec.conf.

The following file provides sample AT-TLS policy definitions.

/usr/lpp/tcpip/samples/pagent_TTLS.conf
This file contains sample AT-TLS policy definitions. These definitions can either be in a common or stack-specific AT-TLS file. If these definitions are in a common AT-TLS file, they can be referenced and reused by multiple stack-specific AT-TLS configuration files. If these definitions are in a stack-specific AT-TLS file, they are used only by that specific stack.

The following set of sample files provides an example of policy definitions in LDAP. For more information on using these sample files, see "Using the sample LDAP objects" on page 740.

/usr/lpp/tcpip/samples/pagent.ldif
This file contains the top level directory structure for the set of sample QoS and IDS policies.

/usr/lpp/tcpip/samples/pagent_starter_QOS.ldif
This file contains the starter set sample of LDAP definitions of QoS objects. This file requires the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_starter_IDS.ldif
This file contains the starter set sample of LDAP definitions of IDS objects. This file requires the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_advanced_QOS.ldif
This file contains the advanced set sample of LDAP definitions of QoS objects. This file requires the objects defined in the QoS starter set sample file pagent_starter_QOS.ldif and the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_advanced_IDS.ldif
This file contains the advanced set sample of LDAP definitions of IDS objects. This file requires the objects defined in the IDS starter set sample file pagent_starter_IDS.ldif and the directory structure defined in sample file pagent.ldif.

The next set of samples is the definition of the schemas in LDAP protocol version 2 format. They must be installed in the LDAPv2 server's configuration as included files. See "Installing the schema definition on the LDAP server" on page 738 for more information.

/usr/lpp/tcpip/samples/pagent_oc.conf
This file contains the schema object class definitions.

/usr/lpp/tcpip/samples/pagent_at.conf
This file contains the schema attribute definitions.

The next set of samples is the definition of the schemas in LDAP protocol version 3 format. They must be installed on the LDAPv3 server in the proper order as an object in the server's database, rather than as configuration information. This process is known as schema publication. Refer to RFCs 1804 and 2251. The files need to be specified on ldapmodify commands to modify the cn:schema entry in the server's database. See "Installing the schema definition on the LDAP server" on page 738 for more information.

/usr/lpp/tcpip/samples/pagent_schema.ldif
This file contains the schema version 2 core and QoS schema object class and attribute definitions.

/usr/lpp/tcpip/samples/pagent_v3schema.ldif
This file contains the schema version 3 additions to the schema version 2 core and QoS schemas.

/usr/lpp/tcpip/samples/pagent_schema_updates.ldif
This file contains changes to the schema version 2 core and QoS schema definitions in support of schema version 3, for z/OS V1R2 and z/OS V1R4.

/usr/lpp/tcpip/samples/pagent_idsschema.ldif
This file contains the schema version 3 IDS schema definitions, for z/OS V1R2 and z/OS V1R4.

/usr/lpp/tcpip/samples/pagent_qosschema.ldif
This file contains the schema version 2 and 3 core and QoS schema definitions, for z/OS V1R5.

/usr/lpp/tcpip/samples/pagent_r5idsschema.ldif
This file contains the schema version 3 IDS schema definitions, for z/OS V1R5.

/usr/lpp/tcpip/samples/pagent_schema_r5updates.ldif
This file contains changes to the schema version 2 and 3 core, QoS, and IDS schema definitions, for z/OS V1R5.

/usr/lpp/tcpip/samples/pagent_r6qosschema.ldif
This file contains the schema version 3 QoS schema definitions, for z/OS V1R6.

/usr/lpp/tcpip/samples/pagent_schema_r6updates.ldif
This file contains additions to the schema version 3 core and QoS schema definitions, for z/OS V1R6.

The next set of samples contain the text of draft documents used to develop the Version 3 schema.

/usr/lpp/tcpip/samples/pagent_pcim.txt
This file contains the draft version of the proposed Policy Core Information Model (PCIM) used as the basis for the support in this release. PCIM is described by RFC 3060 but this file is an earlier draft level.

/usr/lpp/tcpip/samples/pagent_core.txt
This file contains the draft version of the proposed Policy Core LDAP Schema Internet Draft used in z/OS V1R2 and later releases.

Note: This file is provided as-is and there are some differences between the draft and the implementation. The intent of this file is to provide background information on the level of the supported schema.

/usr/lpp/tcpip/samples/pagent_cond.txt
This file contains the draft version of the proposed Policy Conditions Internet Draft used in z/OS V1R2 and later releases.

Note: This file is provided as-is and there are some differences between the draft and the implementation. The intent of this file is to provide background information on the level of the supported schema.

The following files include sample C applications that can be used to develop policy performance monitoring applications.

/usr/lpp/tcpip/samples/pagent/README
This file contains instructions for compiling and running the following sample C applications.

/usr/lpp/tcpip/samples/pagent/pCollector.c
This file is a sample C application (pCollector) that uses the Policy API (PAPI) interfaces to access policy performance data. It can be used as the base for an application that provides near real-time policy performance monitoring.

/usr/lpp/tcpip/samples/pagent/pCollector.h
This file is a header file for the pCollector sample application.

/usr/lpp/tcpip/samples/pagent/pLogReader.c
This file is a sample C application (pLogReader) that reads the policy performance log file to access policy performance data. It can be used as the base for an application that provides offline policy performance monitoring.

Note: This documentation refers to Versions 1, 2, and 3 when defining policies. Version 1 refers to policy definitions defined with the ServicePolicyRules and ServiceCategories statements or LDAP objects. Versions 2 and 3 refer to policy definitions defined with other policy statements or LDAP objects. The primary difference between Version 2 and 3 is in the definition of the LDAP schema.

Policy object model overview

Policies consist of several related objects. The main object is the *policy rule*. A policy rule object refers to one or more *policy condition*, *policy action*, or *policy time period condition* objects, and also contains information on how these objects are to be used. Policy time period objects are used to determine when a given policy rule is active. Active policy objects are related in a way that is analogous to an 'IF' statement in a program. For example:

```
IF condition THEN action
```

In other words, when the set of conditions referred to by a policy rule are TRUE, then the policy actions associated with the policy rule are executed.

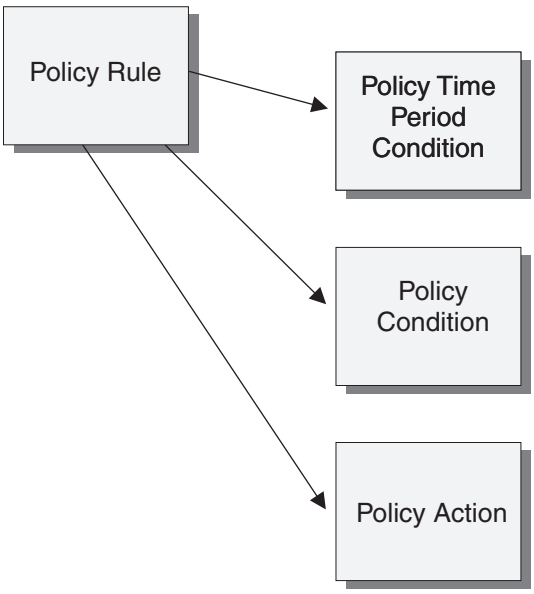


Figure 68. Basic policy objects

Policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a *simple* rule, while one with more conditions is known as a *complex* rule. Complex policy rules can have their conditions evaluated according to one of two different methods. The first is Disjunctive Normal Form (DNF), which means an ORed set of ANDed conditions. The second is Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. In order to accomplish these evaluations, individual policy conditions are assigned an arbitrary group number, and also an indication of whether or not the condition is negated. For example, consider the following set of conditions for a policy rule:

```
C1:  Group Number = 1, Condition Negated = FALSE
C2:  Group Number = 1, Condition Negated = TRUE
C3:  Group Number = 1, Condition Negated = FALSE
C4:  Group Number = 2, Condition Negated = FALSE
C5:  Group Number = 2, Condition Negated = FALSE
```

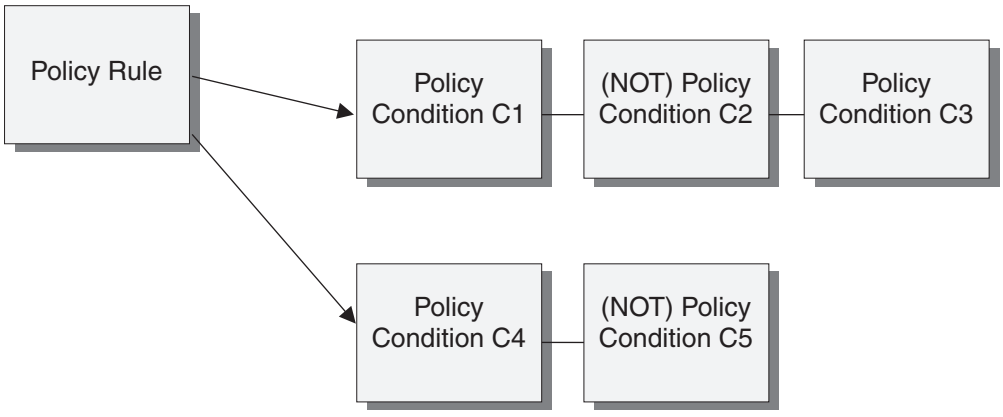


Figure 69. Complex policy conditions

If the conditions are to be evaluated using DNF, then the overall condition for the policy rule is:

```
(C1 AND (NOT C2) AND C3) OR (C4 AND C5)
```

On the other hand, if CNF is used to evaluate the conditions, then the overall condition for the policy rule is:

```
(C1 OR (NOT C2) OR C3) AND (C4 OR C5)
```

Complex rules can be exploded into multiple simple rules. Negated conditions are not allowed in a rule if explosion is to be performed. Consider the following set of conditions for a policy rule:

```
C1:  Group Number = 1, Condition Negated = FALSE
C2:  Group Number = 1, Condition Negated = FALSE
C3:  Group Number = 1, Condition Negated = FALSE
C4:  Group Number = 2, Condition Negated = FALSE
C5:  Group Number = 2, Condition Negated = FALSE
```

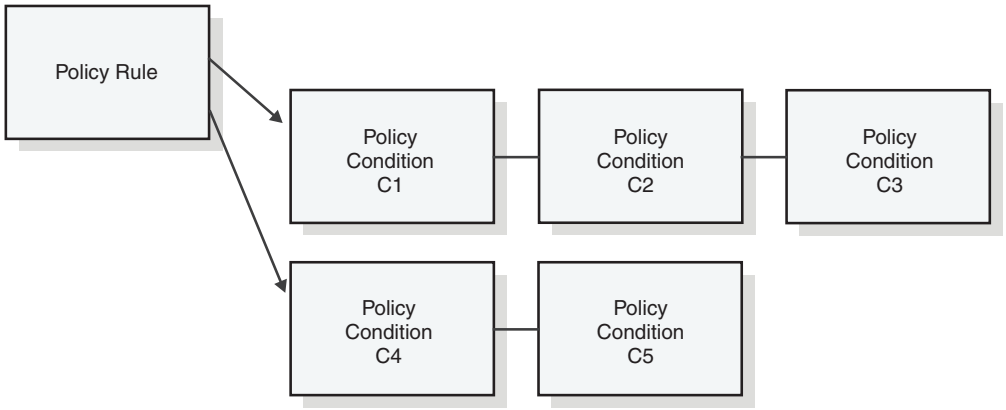


Figure 70. Complex policy conditions before explosion

If DNF is used to evaluate the conditions, exploding the complex rule produces the following simple rules:

```
Simple Rule 1:  C1 AND C2 AND C3
Simple Rule 2:  C4 AND C5
```

If CNF is used to evaluate the conditions, exploding the complex rule produces the following simple rules:

```
Simple Rule 1:  C1 AND C4
Simple Rule 2:  C1 AND C5
Simple Rule 3:  C2 AND C4
Simple Rule 4:  C2 AND C5
Simple Rule 5:  C3 AND C4
Simple Rule 6:  C3 AND C5
```

Policy actions specify actions to take when the set of conditions for a policy rule evaluate to TRUE. The policy model allows multiple actions for a policy rule. Many policy rules typically use only a single action, but multiple actions make sense for some policy types.

Policy conditions and actions can either be specific to a single rule, or be reusable among several policy rules. To allow either type of conditions and actions, and to specify related information such as condition group number and negation indicator, several other policy objects are required. First are *policy condition association* and *policy action association* objects. These objects contain condition and action related attributes, respectively, and may directly contain policy conditions and actions (rule-specific).

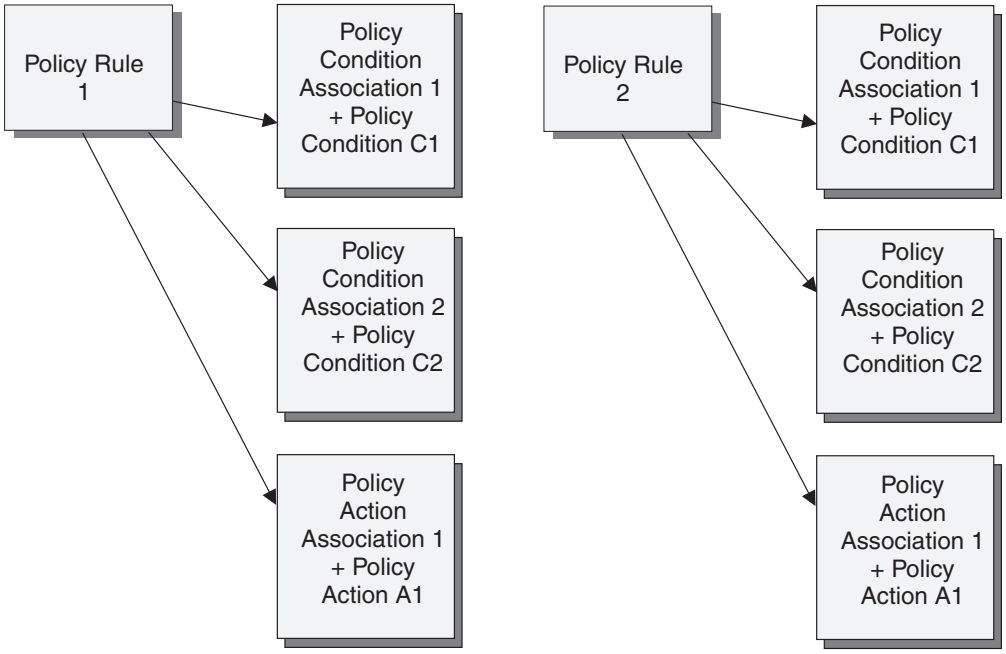


Figure 71. Rule-specific conditions and actions

The policy association objects alternatively may refer to conditions and actions (reusable). *Policy condition instance* and *policy action instance* objects are used to represent reusable policy conditions and actions, respectively.

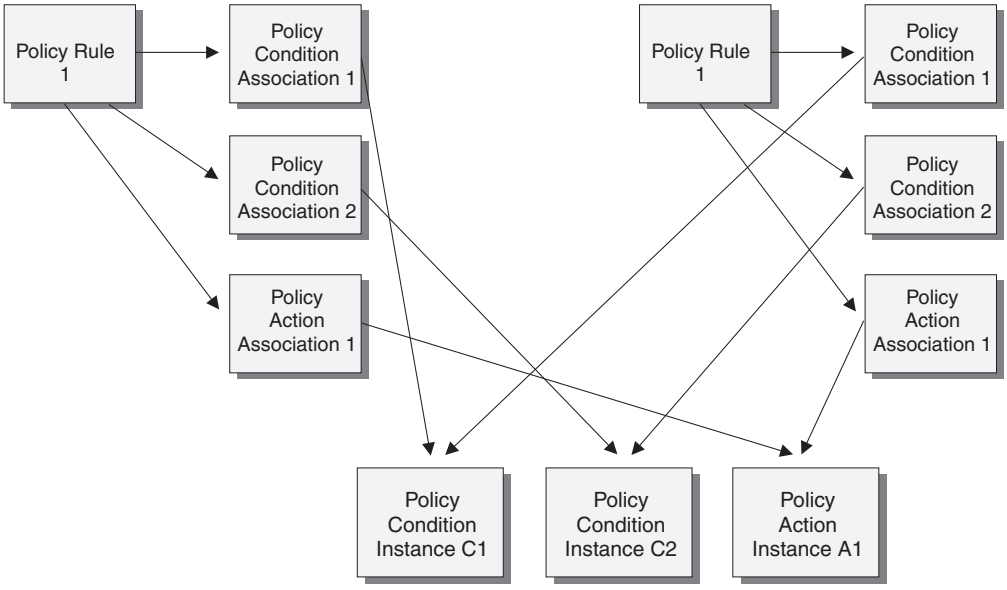


Figure 72. Reusable conditions and actions

Primarily for administrative grouping of policy rules, the *policy group* object is used. Policy groups can refer either to policy rules or to policy groups. This allows related policy rules to be grouped together, and also allows policy groups to be grouped to any needed level of nesting.

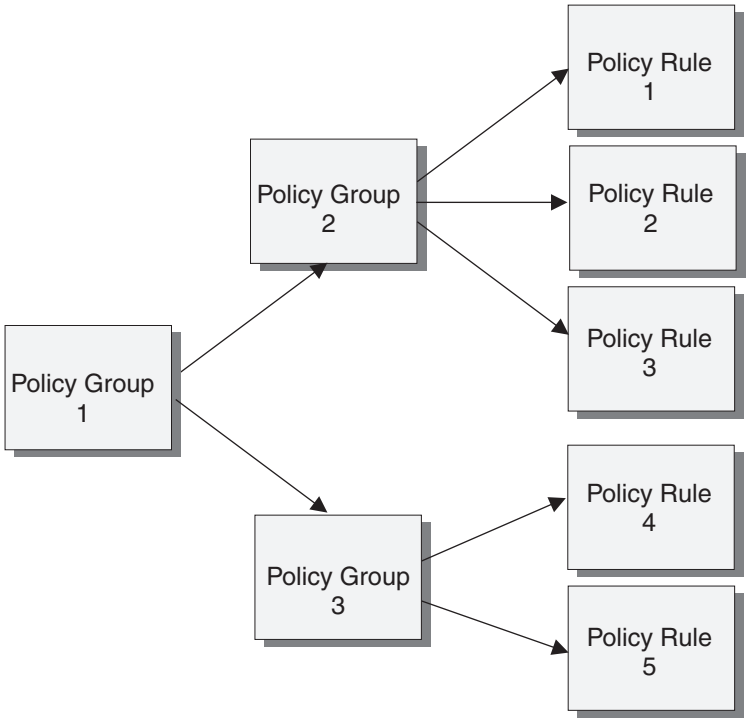


Figure 73. Policy groups

What kind of policy do you want?

The Policy Agent supports the following types of policies. Each policy type is referred to as a discipline.

- Quality of service (QoS) policies
  - Differentiated Services (DS) policies
  - Integrated Services (RSVP) policies
  - Sysplex Distributor (SD) policies
- Intrusion Detection Services (IDS) policies
  - Scan policies
  - Attack policies
  - Traffic Regulation policies
- IP security (IPSec) policies
  - IP filtering policies
  - Key exchange policies
  - Local dynamic VPN policies
- Application Transparent Transport Layer Security (AT-TLS) policies

These policy types are defined using different policy schemas. They use a common rule, but have separate conditions and actions. None of the different policy types can be mixed in a given policy object. All policy rules can contain time-related information that indicates when the policy rule should be considered active or inactive. For the QoS, IDS, and AT-TLS types, active policy rules are installed in the TCP/IP stack, so they can be applied as traffic filters, while inactive policy rules exist only in the Policy Agent. For the IPSec type, both active and inactive IP

filtering policies are installed in the TCP/IP stack. However, only manual VPN tunnels that are active due to time condition are installed in the stack.

The Policy Agent supports all of the above policy types, installing them into one or more TCP/IP stacks as configured.

QoS policy

Policy conditions consist of a variety of selection criteria that act as traffic filters. Traffic can be filtered based on source and destination IP addresses, source and destination ports, protocol, inbound and outbound interfaces, application name, application specific data or application priority. Only packets that match the filter criteria are selected to receive the accompanying action. Policy rules can refer to several policy actions, but only one policy action is executed per policy scope. A given policy action may be referred to by several policy rules.

The type of policy defined is in general controlled by the policy scope value defined for the policy action. SD policies are an exception. SD policies are a subset of DS policies, so use the DS scope.

Although RSVP policies are installed into the TCP/IP stack, they are only used for collecting policy statistics. For policy use and limit enforcement, these policies are requested from the Policy Agent by the RSVP Agent, to apply to RSVP reservation requests from RSVP applications.

IDS policy

Policy conditions primarily determine the portion of IDS function that is being configured. A given IDS policy rule refers to a single IDS policy action. A given IDS policy action may be referred to by several policy rules of the same IDS type. See Chapter 16, “Intrusion Detection Services (IDS),” on page 789 for more details.

Tip: For compatibility, QoS Scope TR policies can also be defined in a Policy Agent configuration file. However, they are transformed internally by Policy Agent into IDS TR policies.

IPSec policy

Policy conditions consist of a variety of selection criteria that act as filters for IP filtering rules. Traffic can be filtered based on source and destination IP addresses, source and destination ports, protocol, direction, routing information, and security class. For other types of IPSec policies, policy conditions contain information about dynamic key exchange filters or dynamic VPN tunnels. For more details, see Chapter 17, “IP security,” on page 819.

IP filter rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, while one with more conditions is known as a complex rule. Complex IP filter rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details on CNF, see “Policy object model overview” on page 723.

Complex IP filter rules (rules that contain groupings, or sets, of individual conditions) are exploded to produce multiple simple rules to be installed in the TCP/IP stack. The conditions in the IpFilterRule statement that can make a filter rule complex are:

- IpSourceAddr

If multiple source addresses (or address ranges) are specified in a rule, the rule is considered complex. Multiple source addresses can be specified by referencing a set or group of addresses from the rule (IpSourceAddrGroupRef).

- IpDestAddr

  If multiple destination addresses (or address ranges) are specified in a rule, the rule is considered complex. Multiple destination addresses can be specified by referencing a set or group of addresses from the rule (IpDestAddrGroupRef).

- IpService

  If multiple IpService statements are specified in a rule, the rule is considered complex. Multiple IpService statements can be specified either inline or by referencing a group of IpService statements (IpServiceGroupRef).

- IpService Direction

  If the Direction parameter in an IpService statement is configured as BiDirectional, the rule is considered complex.

For more details on these IPSec policy configuration statements and parameters, refer to *z/OS Communications Server: IP Configuration Reference*.

The **pasearch** command displays IP filter rules as complex rules, and not exploded as installed in the TCP/IP stack.

For IP filter rules, the condition level summaries are not applicable and are always displayed as all zeros.

AT-TLS policy

Policy conditions consist of a variety of selection criteria that act as filters for AT-TLS rules. Traffic can be filtered based on local addresses, remote addresses, local port range, remote port range, job name, user identification, and direction. For more details, see Chapter 18, "Application Transparent Transport Layer Security (AT-TLS) data protection," on page 987.

AT-TLS policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, while one with more conditions is known as a complex rule. Complex AT-TLS policy rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details on CNF, see "Policy object model overview" on page 723.

When AT-TLS rules are read and parsed, Policy Agent creates the rule as a complex rule. For example, consider the following TTLSRule statement:

```
TTLSRule ttlsRule1
{
  LocalAddrGroupRef    addrGroup1
  RemoteAddrGroupRef   addrGroup2
  LocalPortGroupRef    portGroup1
  RemotePortGroupRef   portGroup2
  Jobname              jobABC
  Userid               user1
  Direction            Outbound
  TTLSGroupActionRef   ttlsAction7
}

IpAddrGroup addrGroup1
{
  IpAddr
  {
    Addr 9.1.1.1
```

```
    }
    IpAddr
    {
      Addr 10.1.1.1
    }
}

IpAddrGroup addrGroup2
{
  IpAddr
  {
    Addr 200.1.1.1
  }
  IpAddr
  {
    Addr 201.1.1.1
  }
}

PortGroup portGroup1
{
  PortRange
  {
    Port 21
  }
  PortRange
  {
    Port 23
  }
}

PortGroup2
{
  PortRange
  {
    Port 10
  }
  PortRange
  {
    Port 15
  }
}
```

This rule is represented as a CNF rule with the following condition levels (levels are ANDed together):

- Level 1 = local address 9.1.1.1 OR local address 10.1.1.1
- Level 2 = remote address 200.1.1.1 OR remote address 201.1.1.1
- Level 3 = local port 21 OR local port 23
- Level 4 = remote port 10 OR remote port 15
- Level 5 = job name jobABC, user ID user1, direction outbound

The **pasearch** command displays the AT-TLS policy as complex rules.

Where do you want to define your policies?

Policies can be defined in the Policy Agent Configuration file, in the LDAP server, or both. Policies from both sources are combined into a single list. Note that this requires unique policy object names per type (QoS, IDS, IPSec, and AT-TLS). For policies defined on the LDAP server, the Distinguished Name (DN) must be unique, but the user-friendly name does not have to be unique (although it is recommended). The Policy Agent appends a unique suffix if required to make LDAP user-friendly names unique within the scope of policies defined on the

LDAP server. When policies from a configuration file are combined with LDAP defined policies, the LDAP user-friendly names must be unique with respect to the names defined in the configuration file. Any policy objects with duplicate names at this point are discarded by the Policy Agent. The following table shows some of the advantages and disadvantages of both.

Type	Advantages	Disadvantages
Configuration File	• Policies are easy to define and change. • Definitions are local - no connection needed. • Very little overhead when policies do not change. • IPSec policy is available only in configuration files. • AT-TLS policy is available only in configuration files.	• Full power of QoS policy definitions not available (for example can only define simple rules). • Each host must maintain its own policy definitions. • IDS policy is not available.
LDAP Server	• Central policy definitions for many hosts. • Policy definitions can be shared and reused between different platforms and hosts. • Full power of QoS policy definitions available (for example can define complex rules). • Enables robust hierarchical policy definition (Sysplex, system, TCP/IP image). • IDS policy is available only in LDAP.	• Policy definitions are more complex. • LDAP server must be queried at each refresh interval to check for new or changed policies. • IPSec policy is not available. • AT-TLS policy is not available.

LDAP server

Lightweight Directory Access Protocol (LDAP) is a fast-growing technology for accessing common directory information. LDAP has been embraced and implemented in most network-oriented middleware. As an open, vendor-neutral standard, LDAP provides an extendable architecture for centralized storage and management of information that needs to be available for today's distributed systems and services.

After a fast start, it can be assumed that LDAP has become the de facto access method for directory information, much the same as the Domain Name System (DNS) is used for IP address lookup on almost any system on an intranet and on the Internet.

Note: If the z/OS LDAP server is used, a DB2 backend is required.

Overview of the object classes

Policies defined on an LDAP server are comprised of one or more objects, each with a defined object class, and a unique name. Object names are in the form of LDAP Distinguished Names (DNs), which are text strings composed of individual

parts known as Relative Distinguished Names (RDNs). The structure of the naming defines a hierarchical tree, in a manner similar to directories in a hierarchical file system. For example, consider the following set of objects:

```
Object 1, DN: o=IBM, c=US
Object 2, DN: cn=group_1, o=IBM, c=US
Object 3, DN: cn=group_5, o=IBM, c=US
Object 4, DN: cn=group_1_sub_A, cn=group_1, o=IBM, c=US
```

This set of objects can be viewed as a tree, with Object 1 as the root. Objects 2 and 3 are branches under the root, with Object 4 a branch under Object 2. The names used are attributes of the objects they define. For example, Object 2, whose name starts with "cn=group_1" contains a cn attribute with the value group_1. Refer to *z/OS Integrated Security Services LDAP Server Administration and Use* for more information on LDAP naming.

Object class names define the type of each LDAP object. The *top* object class is predefined and is the root of all other object classes.

There are three types of object classes.

Abstract object classes
: Used to define broad concepts, such as policy and to define basic attributes that apply to all subclasses.

Structural object classes
: Basic building blocks, and are the only type of object class that can be instantiated as real objects on an LDAP server.

Auxiliary object classes
: Used to define attributes that are shared among different structural object classes, or are used to extend the basic set of objects.

Attributes from auxiliary classes are *attached* to structural objects by including them in the structural objects, and also by including the auxiliary object class as one of the values of the objectClass attribute in the structural object.

The following object classes are recognized by the Policy Agent. The indentation defines subclasses. For example, ibm-policyGroup is a subclass of ibm-policy, and therefore inherits all of the attributes defined for ibm-policy.

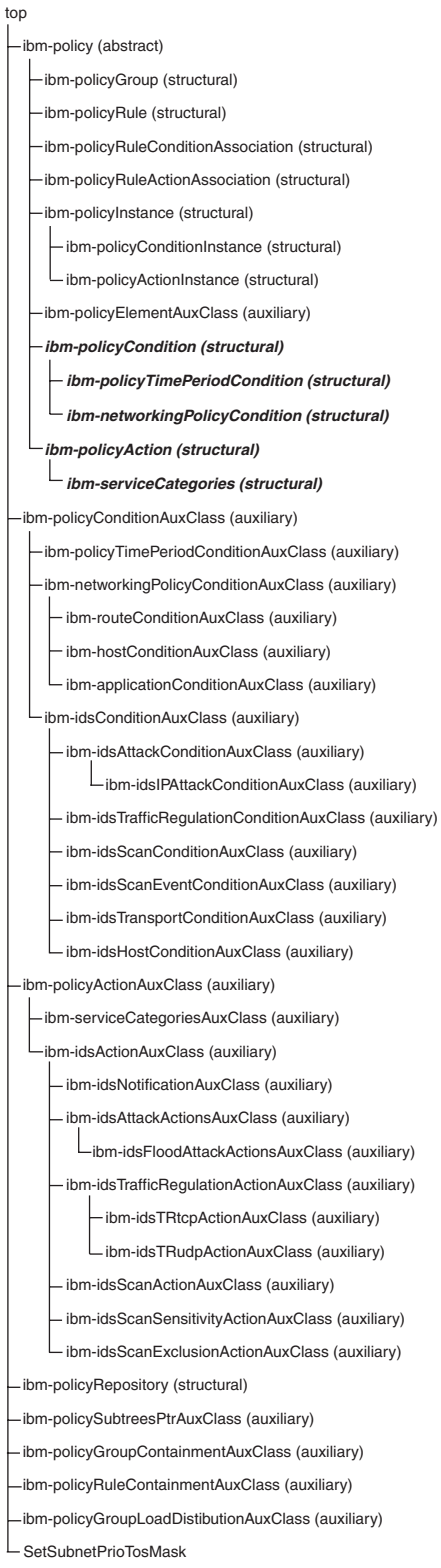


Figure 74. LDAP schema object class hierarchy

Note: The classes identified in italics in the diagram above are for schema version 2 and are mutually exclusive with the schema version 3 classes with similar names.

Object class name	Purpose of object
Top	Used to anchor the LDAP hierarchical tree root.
ibm-policy	Used as the root for all policy objects.
ibm-policyGroup	Defines a policy group object.
ibm-policyRule	Defines a policy rule object.
ibm-policyRuleConditionAssociation	Defines an association between a policy rule object and a policy condition.
ibm-policyRuleActionAssociation	Defines an association between a policy rule object and a policy action.
ibm-PolicyInstance	Defines an instance of a reusable policy object.
ibm-PolicyConditionInstance	Defines an instance of a reusable policy condition object.
ibm-PolicyActionInstance	Defines an instance of a reusable policy action object.
ibm-PolicyElementAuxClass	Defines an auxiliary class that can be used to *tag* non-policy objects as though they were policy objects.
ibm-policyCondition	Defines a policy condition object. (**schema version 2** — supported for migration)
ibm-policyTimePeriodCondition	Defines an auxiliary class to represent time periods during which a policy rule is considered to be active. (**schema version 2** — supported for migration)
ibm-networkingpolicyCondition	Defines a subclass of ibm-PolicyCondition used to define networking policy conditions. (**schema version 2** — supported for migration)
ibm-policyAction	Defines a policy action object. (**schema version 2** — supported for migration)
ibm-serviceCategories	Defines an auxiliary class to represent a set of QoS attributes for a policy action. (**schema version 2** — supported for migration)
ibm-policyConditionAuxClass	Defines an auxiliary class for generic policy conditions.
ibm-policyTimePeriodConditionAuxClass	Defines an auxiliary class to represent time periods during which a policy rule is considered to be active.
ibm-networkingPolicyConditionAuxClass	Defines an auxiliary class used to define networking policy conditions.
ibm-routeConditionAuxClass	Defines an auxiliary class to represent QoS routing conditions for a policy rule.
ibm-hostConditionAuxClass	Defines an auxiliary class to represent QoS host (end-point) conditions for a policy rule.

Object class name	Purpose of object
ibm-applicationConditionAuxClass	Defines an auxiliary class to represent QoS application and transport conditions for a policy rule.
ibm-idsConditionAuxClass	Defines an auxiliary class to represent generic IDS conditions.
ibm-idsAttackConditionAuxClass	Defines an auxiliary class to represent IDS attack conditions.
ibm-idsIPAttackConditionAuxClass	Defines an auxiliary class to represent IDS IP attack conditions.
ibm-idsTrafficRegulationConditionAuxClass	Defines an auxiliary class to represent IDS Traffic Regulation conditions.
ibm-idsScanConditionAuxClass	Defines an auxiliary class to represent IDS global scan conditions.
ibm-idsScanEventConditionAuxClass	Defines an auxiliary class to represent IDS scan event conditions.
ibm-idsTransportConditionAuxClass	Defines an auxiliary class to represent IDS transport conditions.
ibm-idsHostConditionAuxClass	Defines an auxiliary class to represent IDS host conditions.
ibm-policyActionAuxClass	Defines an auxiliary class for generic policy actions.
ibm-serviceCategoriesAuxClass	Defines an auxiliary class to represent a set of QoS attributes for a policy action.
ibm-idsActionAuxClass	Defines an auxiliary class to represent generic IDS actions.
ibm-idsNotificationAuxClass	Defines an auxiliary class to represent notification options for IDS actions.
ibm-idsAttackActionsAuxClass	Defines an auxiliary class to represent attack attributes for IDS actions.
ibm-idsFloodAttackActionsAuxClass	Defines an auxiliary class to represent flood-specific attack attributes for IDS actions.
ibm-idsTrafficRegulationActionAuxClass	Defines an auxiliary class to represent generic Traffic Regulation attributes for IDS actions.
ibm-idsTRtcpActionAuxClass	Defines an auxiliary class to represent Traffic Regulation TCP attributes for IDS actions.
ibm-idsTRudpActionAuxClass	Defines an auxiliary class to represent Traffic Regulation UDP attributes for IDS actions.
ibm-idsScanActionAuxClass	Defines an auxiliary class to represent global scan attributes for IDS actions.
ibm-idsScanSensitivityActionAuxClass	Defines an auxiliary class to represent scan sensitivity attributes for IDS actions.
ibm-idsScanExclusionActionAuxClass	Defines an auxiliary class to define scan exclusion lists for IDS actions.

Object class name	Purpose of object
ibm-policyRepository	Defines a repository for generic reusable policy objects.
ibm-policySubtreesPtrAuxClass	Defines an auxiliary class to represent pointers to subtrees in the LDAP directory tree to be retrieved by the Policy Agent. This allows entire subtrees to be retrieved at once, improving retrieval performance in some situations.
ibm-policyGroupContainmentAuxClass	Defines an auxiliary class for binding a policy group object to another policy group.
ibm-policyRuleContainmentAuxClass	Defines an auxiliary class for binding a policy rule object to another policy group.
ibm-policyGroupLoadDistributionAuxClass	Defines an auxiliary class to represent load distribution attributes for policy groups. The load distribution attributes are applied to all policy rules that are pointed to by groups to which this auxiliary class has been attached.
SetSubnetPrioTosMask	Defines a mapping of outbound IPv4 packet Type of Service (ToS) byte or IPv6 packet Traffic Class values to QDIO device priorities and Virtual LAN (VLAN) user priorities.

Policy objects are used to accomplish the following objectives:

- Group related objects together. Policy groups can contain related policy rules, and can also contain other policy groups. This allows objects to be grouped in various administrative ways. If the resulting objects will be retrieved by any Policy Agent prior to z/OS Communications Server V1R2, then the object should not include the values ibm-policyGroupContainmentAuxClass, ibm-policyRuleContainmentAuxClass or ibm-policyGroupLoadDistributionAuxClass for the objectClass attribute.
- Specify conditions for a policy rule. The conditions are used to filter traffic packets, and can specify attributes such as source and destination port, application name, protocol, and so on. Policy rules can be either simple or complex, depending on the nature of the specified conditions. When a single condition is specified, the rule is a simple rule. This single condition can be composed of any of the condition attributes, but only one instance of each. For example, only a single destination port range can be specified in a simple rule. Complex rules specify more than one condition. The specified conditions are organized into one or more levels, and each level is composed of one or more conditions. Each condition can be composed of one instance of any of the condition attributes. The conditions can thus be thought of as a two-dimensional array. Any individual condition can be negated. Two types of processing are applied to the conditions, depending on the specified condition list type:
  - Disjunctive Normal Form (DNF). DNF conditions are logically ANDed at each level, and ORed between levels.
  - Conjunctive Normal Form (CNF). CNF conditions are logically ORed at each level, and ANDed between levels.

For example, suppose five conditions are specified, two at one level and three at another:

```
Level 1: C1, NOT C2
Level 2: C3, C4, C5
```

If DNF is specified, the conditions are evaluated as:

```
(C1 AND NOT C2) OR (C3 AND C4 AND C5)
```

CNF evaluation of the same conditions is:

```
(C1 OR NOT C2) AND (C3 OR C4 OR C5)
```

This allows a wide variety of conditional logic to be defined for policy rules.

- Specify time periods during which policy rules are active. Active policy rules are those that are installed in a TCP/IP stack by the Policy Agent. A wide variety of attributes are available to specify time periods, and up to 25 time periods can be specified for any policy rule. The policy rule is active if any of the specified time intervals include the current time.
- Specify actions to take on behalf of traffic that maps to active policy rules, based on the evaluation of its conditions. QoS Actions contain a scope attribute that indicates the type of policy being defined, namely Differentiated Services, RSVP, or both Differentiated Services and RSVP. Up to four actions can be specified for each rule, but only one action per scope can be active at a time. IDS actions contain an action type that indicates the type of policy being defined, namely Attack, TR, Scan Global, or Scan Event. Only one IDS action can be specified for each rule. QoS and IDS actions (or conditions) can't be mixed within a single policy rule.

Considerations for defining LDAP objects

LDAP objects can refer to other objects, using the DN of the referenced object. For example, a policy rule can be separated from its conditions and time periods, with those objects being referenced by the rule object.

Each LDAP object is composed of a number of attributes. Some of the attributes are generic LDAP attributes that apply to all LDAP objects. Other attributes are used only for Version 1 policy definitions. All other Version 2 and later policy attributes must begin with a unique prefix:

```
ibm-
```

When defining complex policy rules (those with more than one condition or action), two mutually exclusive methods can be used to associate the conditions or actions with the rule:

- The ibm-policyConditionListDN and ibm-policyActionListDN attributes can be omitted from the rule. In this case, the condition and action association objects MUST be created as subordinate objects to the policy rule, in other words, *under* the rule in the directory subtree. This is known as Directory Information Tree (DIT)-containment.
- The ibm-policyConditionListDN and ibm-policyActionListDN attributes can be specified in the rule. In this case, the condition and action association objects SHOULD be created as subordinate objects to the policy rule, in other words, *under* the rule in the directory subtree. However, this is not a requirement, only a recommendation. The objects can actually exist anywhere in the DIT.

Policy Agent retrieval of LDAP objects

The design of the LDAP object tree should be carefully thought out. The Policy Agent uses a variety of mechanisms to search for and retrieve objects from an LDAP server:

- An initial search is done for a subtree of objects based on the SearchBaseDN parameter on the ReadFromDirectory statement.
- If any objects retrieved by this initial search contain subtree pointer references (using the ibm-policySubtreesAuxContainedSet attribute) then a search is done for all such subtrees. This is a recursive search: additional objects retrieved might also contain subtree pointer references.
- The above searches use a filter to only retrieve certain object classes. For LDAP protocol version 3, the default is to only scan for the ibm-policy object class. This is an abstract object class from which all other policy object classes are derived. Most LDAPv3 servers implement abstract and auxiliary classes such that this search will properly retrieve policy, and only policy, object classes. However, some LDAPv3 servers do not honor abstract/auxiliary object classes as a search filter. For these servers, specify LDAP_AbstractPolicy NO on the ReadFromDirectory statement. This causes the searches to use a filter that retrieves ALL object classes. This same filter to retrieve all object classes is used for LDAPv2 servers.
- All of the above searches may be scoped, or filtered, using keywords specified on the ReadFromDirectory statement parameters SearchPolicyKeyword, SearchPolicyGroupKeyword, or SearchPolicyRuleKeyword. The LDAP server only returns objects with any matching keywords.
- Some objects retrieved using the above searches may contain DN pointer references to additional objects. These objects are individually retrieved. If the object to be retrieved is a policy rule, then a subtree search is performed, using the keywords specified on the ReadFromDirectory statement. All other objects are retrieved as single objects, using the DN pointers (no keywords are used on the search).
- All policy rule objects retrieved using the above searches are further filtered using the PolicyRole parameter on the ReadFromDirectory statement. Any rules that do not match policy roles specified on the ReadFromDirectory statement are discarded.

Therefore, it is possible to design an LDAP tree such that a minimal set of objects is initially retrieved, followed by many additional individual LDAP retrievals. If the total set of objects is large, there is a performance impact to retrieving objects in this manner. If possible, try to design the tree and the ReadFromDirectory parameters to retrieve the largest set of objects initially, to achieve the best performance, or to use subtree pointer references to retrieve larger sets of objects in one operation.

Installing the schema definition on the LDAP server

The files that define the schema supported by the Policy Agent are shipped as a set of sample files. You need to modify the configuration of the LDAP server to include these schema definition files. How this is accomplished depends on the protocol supported by the server.

For LDAP protocol version 2, include the pagent_oc.conf and pagent_at.conf files (located in the /usr/lpp/tcpip/samples directory) in the server's configuration file. For example:

```
include  /usr/lpp/tcpip/samples/pagent_at.conf
include  /usr/lpp/tcpip/samples/pagent_oc.conf
```

For LDAP protocol version 3, the schema definition is shipped in *ldif* format and installed on the LDAP server as a modification to the generic schema entry, known as a subschema. The existing schema entry must be modified to include the supported schema as a subschema, by using the ldapmodify command. There are

several schema definition files that must be installed in the proper order. All of these files are located in the /usr/lpp/tcpip/samples directory, although not all files must be installed for a given release:

- pagent_schema.ldif
- pagent_v3schema.ldif
- pagent_schema_updates.ldif
- pagent_idsschema.ldif
- pagent_qosschema.ldif
- pagent_r5idsschema.ldif
- pagent_schema_r5updates.ldif
- pagent_r6qosschema.ldif
- pagent_schema_r6updates.ldif

Note: This process is supported for the z/OS LDAP server.

The files that you need to install depend on the current schema definition being used by the LDAP server. When migrating to a later release from a previous release, Table 31 shows the files that should already be installed from the previous release, and the files that you need to install to migrate to the later release. Table 31 also shows the files that need to be installed when installing the complete schema for the first time.

Table 31. Files needed to install the schema definition on the LDAP server

Migrating to this z/OS release	Migrating from this z/OS release	Files that should be already installed	Files you need to install to migrate
V1R4	Not applicable; you are installing the schema for the first time.	None	pagent_schema.ldif pagent_v3schema.ldif pagent_schema_updates.ldif pagent_idsschema.ldif
V1R5	V1R4	pagent_schema.ldif pagent_v3schema.ldif pagent_schema_updates.ldif pagent_idsschema.ldif	pagent_schema_r5updates.ldif
V1R5	Not applicable; you are installing the schema for the first time.	None	pagent_qosschema.ldif pagent_r5idsschema.ldif
V1R6 or V1R7	V1R4	pagent_schema.ldif pagent_v3schema.ldif pagent_schema_updates.ldif pagent_idsschema.ldif	pagent_schema_r5updates.ldif pagent_schema_r6updates.ldif
V1R6 or V1R7	V1R5	pagent_qosschema.ldif pagent_r5idsschema.ldif	pagent_schema_r6updates.ldif
V1R6 or V1R7	Not applicable; you are installing the schema for the first time.	None	pagent_r6qosschema.ldif pagent_r5idsschema.ldif

To install the schema definitions, use commands like those in the following example:

```
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_schema.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_v3schema.ldif
```

```
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_schema_updates.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_idsschema.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_qosschema.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_r5idsschema.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_schema_r5updates.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_r6qosschema.ldif
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_schema_r6updates.ldif
```

Refer to the TDBM backend information in *z/OS Integrated Security Services LDAP Server Administration and Use* for more details.

Using the sample LDAP objects

There are 5 sample files that provide examples of policy definitions in LDAP:

- pagent.ldif
- pagent_starter_IDS.ldif
- pagent_starter_QOS.ldif
- pagent_advanced_IDS.ldif
- pagent_advanced_QOS.ldif

For brief descriptions of these files, see "Policy sample files" on page 721. You can either use some or all of these predefined policies in the starter and advanced sets, or modify them as needed.

The recommended way to create customized policies is to copy the sample policies you want to change to the custom portion of the pagent.ldif file (under the appropriate cn=custom root, QoS or IDS), modify them as needed, and then point to the custom set as the search base on the ReadFromDirectory statement.

For example, the pagent.ldif file has the following hierarchical structure [this shows the relevant parts of the Distingushed Name (DN) for each object]:

```
o=IBM, c=US (root object)
    cn=repository (root of all reusable policy conditions and actions)
    ou=policy (root of all policy groups and rules)
       cn=groups (root of sample groups)
       cn=starter (root of simple starter set of policies)
           cn=IDS (IDS starter set - actually defined in file pagent_starter_IDS.ldif)
           cn=QoS (QOS starter set - actually defined in file pagent_starter_QOS.ldif)
       cn=advanced (root of advanced set of policies)
           cn=IDS (IDS advanced set - actually defined in file pagent_advanced_IDS.ldif)
           cn=QoS (QOS advanced set - actually defined in file pagent_advanced_QOS.ldif)
       cn=custom (root of customer-defined set of policies)
           cn=IDS (root of customer-defined IDS policies (empty))
           cn=QoS (root of customer-defined QOS policies (empty))
```

To obtain only the customized policies, specify the top custom policy group object as the search base on the ReadFromDirectory statement as follows:

```
ReadFromDirectory {
...
SearchPolicyBaseDN   dn:cn=custom, ou=policy, o=IBM, c=US
...
}
```

Note: If your LDAP server has a root structure other than ″o=IBM, c=US″, be sure to change the root structure in all the files you want to use by replacing every instance of ″o=IBM, c=US″ with the appropriate root used on your LDAP server.

zQoS Manager for z/OS Communications Server

Quality of Service Manager for z/OS Communications Server (zQoS Manager) enables networking QoS policy management across data centers, assigning a variety of QoS service labels to different traffic. Network administrators use this tool to produce a file that an LDAP server can process. The LDAP server can then be read by the Policy Agent to configure policies. For more information, refer to:

`http://www.ibm.com/software/network/commserver/downloads/`

eServer IDS Configuration Manager

eServer IDS Configuration Manager enables centralized configuration of IDS policy for z/OS using LDAP as a policy repository. eServer IDS Configuration Manager provides a user-friendly interface with help panels, to isolate network administrators from having to know LDAP policy schema and the complexity of directly writing to an LDAP server. For more information, refer to:

`http://www.ibm.com/software/network/commserver/downloads/`

Configuring the Policy Agent

Step 1: General configuration

The Policy Agent is responsible for reading policies from configuration files, an LDAP server, or both. Before defining policies, some basic operational characteristics of the Policy Agent need to be configured. Follow these steps to configure these items.

1. Define the TcpImage or PEPInstance statements in the main Policy Agent configuration file.

   The PEPInstance statement is a synonym for TcpImage, and either can be used. PEPInstance refers to policy enforcement point (PEP), the component of a policy system that enforces policies, which for z/OS is the TCP/IP stack.

   The Policy Agent can be configured to install policies on one or more TCP/IP stacks, or images. Each TCP/IP stack is configured using a TcpImage statement in the main configuration file. A secondary configuration file can be defined for any given stack, a set of stacks can share configuration information in the main configuration file, or a combination of these techniques can be used.

   To install different sets of policies to different stacks, configure each image with a different secondary configuration file. In this case, each image can be configured with a different policy refresh interval if desired. The refresh interval used for the main configuration file will be the smallest of the values specified for the different stacks.

   Result: When the main configuration file is an MVS data set, it is reread at each refresh interval (which is the smallest of the individual stack refresh intervals), regardless of whether it has actually been changed or not. Because Policy Agent restarts all stack-related processing when the main configuration file is reread, this effectively makes the refresh interval for all stacks the same as this smallest configured interval.

   To install a common set of policies to a set of stacks, do not specify secondary configuration files for each image. In this case, there is only one configuration file (the main one) and the policy information contained in it is installed to all

of the configured stacks. Different refresh intervals can also be configured for each image, but would probably be less useful in this case.

In either case, it is possible that TCP/IP stacks configured to the Policy Agent are not started or even defined. The Policy Agent will fail when trying to connect to those stacks and log appropriate error messages.

Rule: To dynamically add a TCP/IP stack to the Policy Agent configuration and have active policies automatically installed, in addition to adding the TcpImage statement to the configuration file, further action might be necessary as follows:

- If the Policy Agent was started with the -i startup option, no further action is necessary. Active policies will be automatically installed to the stack when it becomes active.
- If the Policy Agent was not started with the -i startup option, do one of the following:
  - Issue the MODIFY REFRESH or MODIFY UPDATE command after the stack becomes active. If you issue the MODIFY REFRESH or MODIFY UPDATE command before the stack becomes active, policies will not be automatically installed.
  - Wait on the next update interval to check for configuration changes. If the stack is not active, policies will not be automatically installed.

The Policy Agent does not end when any (or all) stacks end. When the stacks are restarted, active policies are automatically reinstalled.

FLUSH or NOFLUSH processing:

- For QoS and IDS policy types, this is configured on the TcpImage or PEPInstance statement.
- For AT-TLS policy type, this is configured on the TcpImage or PEPInstance statement, or the TTLSConfig statement.
- For IPSec policies, this option has no effect.

Results: When FLUSH is specified, policies are deleted at the following times for policy types that honor FLUSH processing:

- When Policy Agent is starting.
- When a new TcpImage or PEPInstance statement is processed for the first time.
- When a MODIFY REFRESH command is entered.

PURGE or NOPURGE processing:

- For QoS and IDS policy types, this is configured on TcpImage or PEPInstance statement.
- For AT-TLS policy type, this is configured on the TcpImage or PEPInstance statement, or the TTLSConfig statement.
- For IPSec policies, this option has no effect.

Results: When PURGE is specified, policies are deleted at the following times for policy types that honor PURGE processing:

- When a TcpImage or PEPInstance statement is deleted from the configuration file.
- When Policy Agent terminates.

The TcpImage statement specifies a TCP/IP image and its associated configuration file to be installed to that image. The following example installs the policy control file */tmp/TCPCS.policy* to the *TCPCS* TCP/IP image, after flushing the existing policy control data:

```
TcpImage TCPCS /tmp/TCPCS.policy FLUSH
```

2. Define the appropriate logging level.

   The LogLevel statement is used to define the amount of information to be logged by the Policy Agent. The default is to log only event, error, console, and warning messages. This might be appropriate for a stable policy configuration, but more information might be required to understand policy processing or debug problems when first setting up policies or when making significant changes. Specify the LogLevel statement with the appropriate logging level in the main configuration file.

   Note: The maximum logging level (511) can produce a significant amount of output, especially with large LDAP configurations. This is not a concern if an HFS log file is used, because Policy Agent uses a set of log files with a finite size in a round-robin configuration (the number and size of these files is controllable with the PAGENT_LOG_FILE_CONTROL environment variable). But when using the syslog daemon as the log file, the amount of log output produced should be taken into consideration.

3. Provide the following security authorizations. Because the Policy Agent can affect system operation significantly, the following security authorizations are required.
   - A user starting Policy Agent must be a superuser.
   - Security product authority (for example, RACF(R)) is required to start the Policy Agent. For sample commands needed to create the profile name and permit users to it, refer to the EZARACF sample in SEZAINST.

4. If Policy Agent's PAPI clients, including the **pasearch** command, are not defined as a superuser, to retrieve policies you must define security product authority in the SERVAUTH class for that client. These profiles can be defined by TCP/IP stack (TcpImage) and policy type (ptype = QoS, IDS, IPSec, or TTLS). Wildcarding of profile names is allowed.

   `EZB.PAGENT.`*`sysname`*`.`*`TcpImage`*`.`*`ptype`*

   where:
   - *sysname* - System name defined in sysplex
   - *TcpImage* - Tcp name for policy information that is being requested
   - *ptype* - Policy type that is being requested:
     - QOS - Policy QoS
     - IDS - Policy IDS
     - IPSec - Policy IPSec
     - TTLS - Policy AT-TLS

   Tip: Wildcarding is allowed on segments of the profile name.

   Policy Agent will check all client's requests to verify SERVAUTH class is active and the profile exists for the TcpImages and policy types in the request. If a client's request is for ALL TcpImages and policy types defined then Policy Agent will only return information for any information for which permission is granted. For example, if the request is for ALL policy types, and QoS, IDS, IPSec, and AT-TLS policy types are defined, Policy Agent returns only information for which permission is granted.

   If SERVAUTH class is absent (not RACLIST) or profiles are absent for client's request (TcpImage, policy type) then permission is denied and data is NOT returned.

   If SERVAUTH class is active and profiles are present for client's request (TcpImage and policy type) and MVS user is defined for all profiles then permission is granted and data is returned.

If SERVAUTH class is active and profiles are present for client's request (TcpImage and policy type) and MVS user (that is, Policy Agent client) is not defined for all profiles then permission is refused and data is not returned.

Refer to the EZARACF sample in SEZAINST for sample commands needed to create the profile name and permit users to it.

Step 2: Configure policies in Policy Agent configuration files

Policies can be defined in any referenced Policy Agent configuration file. For more information, see the appropriate information for each policy type:

- Quality of Service (See Chapter 15, "Quality of service (QoS)," on page 751)
- Intrusion Detection Services (See Chapter 16, "Intrusion Detection Services (IDS)," on page 789)
- IP filtering, and manual and dynamic VPN tunnels, collectively referred to as IPSec policies (See Chapter 17, "IP security," on page 819)
- Application Transparent Transport Layer Security (See Chapter 18, "Application Transparent Transport Layer Security (AT-TLS) data protection," on page 987)

Step 3: Configure Policy Agent to use LDAP server via the ReadFromDirectory statement

The ReadFromDirectory statement in the Policy Agent configuration file initializes the Policy Agent as an LDAP client. The policies are downloaded from the LDAP server, along with the policies specified in the Policy Agent configuration files.

When configuring the ReadFromDirectory statement, first specify the name (or IPv4 address) and port of the primary server and the same for the backup server (if one is used).

Notes:

1. The LDAP client library used to connect to the LDAP server does not support IPv6.
2. When using the z/OS LDAP server, the server listens on a separate port for SSL connections. This means that you should specify the correct port depending on whether or not SSL is used.

Next, configure other connection attributes. The Policy Agent (as an LDAP client) must log in to the LDAP server. *The userid and password for logging in must be configured on the ReadFromDirectory statement.*The userid is also known as Distinguished Name for userid, and it is in the form of an LDAP DN. If the userid and password are not specified, the Policy Agent uses anonymous login to connect to the server.

The LDAP server might be running either Version 2 or Version 3 of the LDAP protocol, which must also be configured on the ReadFromDirectory statement. This statement also configures the version of the schema to be retrieved from the server.

Finally, configure attributes to indicate how to search the LDAP server for policies. Policy roles allow one or more roles, or role-combinations, to be assigned to policy rules using the ibm-policyRoles attribute. These roles represent the intended usage of the policy rules. For example, a role of "East Coast WAN" might be used to represent policies for the wide area network on the US East coast for an enterprise. Policy role values are not standardized; they are simply values used to assign roles to policies. When an entity that requires policies (such as Policy Agent) requests policies from an LDAP server, it can filter out policy rules that do not match the

roles that it plays. Although similar to policy keywords, which also allow search scoping, policy roles are a bit more sophisticated. Specifically, role-combinations are allowed, which take the form of a specification like ″roleA && roleB″, meaning both roleA AND roleB. Since the ibm-policyRoles attribute is multi-valued, a form of CNF/DNF logic can be used for policy roles: the roles in a role-combination are ANDed, and the roles or role-combinations specified on different values of this attribute are ORed.

For the Version 1 schema, a base DN to start the search, and a *selector tag* value are configured. The selector tag is used to match against the SelectorTag attribute in the policy objects. For Version 1, the Policy Agent also automatically includes the stack name when searching for policies; this value is matched against the TcpImageName attribute in the policy objects. For the Version 2 schema, a base DN to start searching is also configured. This DN can specify a single LDAP object, a policy group, or an LDAP subtree containing many objects. For filtering the search, three keywords can be configured:

- SearchPolicyKeyword matches against the ibm-policyKeywords attribute in any policy object.
- SearchPolicyGroupKeyword matches against the ibm-policyGroupKeywords attribute in policy group objects.
- SearchPolicyRuleKeyword matches against the ibm-policyRuleKeywords attribute in policy rule objects.

The following example does the following:

- Connects to the LDAP server at IP address 9.100.1.1, using the default port 389.
- Specifies a userid and password to log in to the server.
- Specifies LDAP protocol version 3.
- Specifies schema version 3.
- Starts searching at the DN ou=policy, o=IBM, c=US object/subtree.
- Only selects policy objects that contain either the ″POLICY″ or ″EASTERN″ keywords.

```
ReadFromDirectory
{
LDAP_Server 9.100.1.1
LDAP_DistinguishedName cn=root, o=IBM, c=US
LDAP_Password 4qr56jb
LDAP_ProtocolVersion 3
LDAP_SchemaVersion 3
SearchPolicyBaseDN ou=policy, o=IBM, c=US
SearchPolicyKeyword POLICY
SearchPolicyKeyword EASTERN
}
```

Step 4: Optionally add SSL to the Policy Agent connection to LDAP

The Secure Sockets Layer (SSL) protocol begins with a handshake. During the handshake, the client authenticates the server, the server optionally authenticates the client, and the client and server agree on how to encrypt and decrypt information.

Server Authentication: When using SSL to secure communications, the SSL authentication mechanism known as server authentication is used. With server authentication, the LDAP server must have a digital certificate which authenticates the server to the Policy Agent client. The server supplies the client with the

certificate during the initial SSL handshake. If the client validates the server's certificate, then a secure communication channel is established between the LDAP server and the Policy Agent client.

For server authentication to work, the LDAP server must have a private key and associated server certificate in the server key ring file.

To conduct commercial business on the Internet, you might use a widely known Certificate Authority (CA), such as VeriSign, to get a high assurance certificate. For a relatively small private network within your own enterprise or group, you can issue your own certificates, called self-signed certificates, for your own use.

Client Authentication: When using SSL Client Authentication, the client passes a digital certificate to the server as part of the SSL handshake. To pass authentication, the Certificate Authority (CA) that signed the client certificate must be considered trusted by the server.

Self-signed Server Certificates: Normally, a server certificate should be obtained from a known CA. However, for testing, an installation might use a self-signed server certificate. Because the clients will not know about the issuer of the self-signed server certificate, in most cases it is necessary to add the server's self-signed certificate to the client's signer certificates.

The gskkyman utility is used to create public/private key pairs and certificate requests, receive certificate requests into a key ring, and manage keys in a key ring. The gskkyman utility is documented in *z/OS Crytographic Service System Secure Sockets Layer Programming*. The gskkyman utility is shipped with z/OS in System SSL, which is part of the cryptographic services base element of z/OS.

The Policy Agent connection to LDAP can be secured using SSL by tailoring the following parameters on the ReadFromDirectory statement listed below. This allows for protection of policy retrieval from an LDAP server.

- LDAP_SSLKeyringFile
- LDAP_SSLKeyringPassword
- LDAP_SSLName

For more detail about these ReadFromDirectory parameters, refer to *z/OS Communications Server: IP Configuration Reference*. Additional information about the concepts of cryptography and SSL can be found at the following Web sites:

- http://home.netscape.com/eng/ssl3/
- http://www.verisign.com/repository/crptintr.html

Starting and stopping the Policy Agent

You can start the Policy Agent from the z/OS shell, as a started task, or by using the TCP/IP AUTOLOG statement. If you use the shell, the Policy Agent should be started in a background shell session, by specifying a trailing & on the command line.

If a procedure in the AUTOLOG list also has a PORT statement reserving a TCP or UDP port but does not have a listening connection on that port, TCP/IP periodically attempts to cancel that procedure and start it again.

Therefore, if Policy Agent is being started with AUTOLOG and the PolicyPerfMonitorForSDR statement is not being used (no listening connection on the pagentQosListener port), it is important to do one of the following:

- Ensure that the pagentQosListener port (1700) is not reserved by the PORT statement in the TCP/IP profile.
- Add the NOAUTOLOG parameter to the PORT statement in the TCP/IP profile. For example:

```
PORT
 1700 TCP PAGENT NOAUTOLOG
```

In addition, when the PolicyPerfMonitorForSDR statement is being used, if the pagentQosCollector port (1701) is reserved in the PORT list it should always be specified with the NOAUTOLOG parameter, because this port is never used as a listening port. For example:

```
PORT
 1701 TCP PAGENT NOAUTOLOG
```

Tip: When the PolicyPerfMonitorForSDR statement is not being used, the autostart feature of AUTOLOG can be used as described above. However, the monitoring and auto-restart features of AUTOLOG are unavailable due to the AUTOLOG dependence on listening to a TCP or UDP connection, which does not exist without the PolicyPerfMonitorForSDR statement.

If you fail to take one of the above actions, Policy Agent will be periodically canceled and restarted by TCP/IP.

The Policy Agent requires access to one or more DLLs at run time. The LIBPATH environment variable needs to be set to include the /usr/lib directory, which normally includes all the required DLLs.

For policy time specifications to be properly acted upon, the TZ environment variable needs to be set to local time. You can set the LIBPATH and TZ environment variables as follows:

- When starting from the z/OS shell:

  Export the LIBPATH and TZ environment variables before starting the Policy Agent. This is best accomplished in /etc/profile or in .profile in the HOME directory. For example, if you are in the Eastern time zone in the Unites States:

  ```
  export LIBPATH=/usr/lib
  export TZ=EST5EDT4
  ```

- When starting as a started task:
  - Specify LIBPATH and TZ using the ENVAR parameter on the PARM statement in the started procedure. For example:

    ```
    // PARM=('POSIX(ON) ALL31(ON)',
    // 'ENVAR("LIBPATH=/usr/lib"',
    // '"TZ=EST5EDT4")/')
    ```

  - Export the LIBPATH and TZ environment variables in a file specified with the STDENV DD statement. For example:

    ```
    //STDENV DD PATH='/etc/pagent.env',PATHOPTS=(ORDONLY)
    ```

    In the /etc/pagent.env file:

    ```
    LIBPATH=/usr/lib
    TZ=EST5EDT4
    ```

For more information on specifying run-time options, refer to *z/OS Language Environment Programming Guide*. For details on setting the LIBPATH and TZ environment variables, refer to *z/OS UNIX System Services Command Reference*.

Although the /etc/pagent.conf is the default configuration file, a specific search order is used when starting the Policy Agent. The following order is used:

1. File or data set specified with the -c startup option
2. File or data set specified with the PAGENT_CONFIG_FILE environment variable
3. /etc/pagent.conf

The syntax for an HFS file is *'/dir/file'*, and the syntax for an MVS data set is *"//'mvs.dataset.name'"*.

If the Policy Agent cannot successfully parse the start options, log output is written to the syslog daemon (syslogd).

At initialization, the Policy Agent creates an HFS file called /tmp/*tcpname*.Pagent.tmp. This occurs for every TCP/IP stack defined on a TcpImage statement.

In this HFS file, tcpname is the name of a TCP/IP stack from a TcpImage statement. During TCP/IP stack initialization, the TCP/IP stack will attempt to modify a file by this name to notify the Policy Agent that the stack has been reactivated. This causes the Policy Agent to automatically attempt to reinstall the existing policies to this stack.

To ensure that only one Policy Agent is started, the Policy Agent uses the following enqueue:

- Enqueue name is TCP_TCPI
- Resource name is TCPIP.PAGENT

When starting from the shell, note that the Policy Agent executable resides in /usr/lpp/tcpip/sbin. There is also a link from /usr/sbin. Make sure your PATH statement contains either /usr/sbin or /usr/lpp/tcpip/sbin.

For example, the following command starts Policy Agent with these characteristics:

```
pagent -c /u/user10/pldap.conf -l SYSLOGD &
```

- Policy Agent uses the configuration file /u/user10/pldap.conf
- Policy Agent logs output to the syslog daemon (SYSLOGD). Note that "SYSLOGD" must be specified in uppercase to obtain this behavior

Use the S PAGENT command on an MVS console or SDSF to start the Policy Agent as a started task. A sample procedure is shipped in member EZAPAGSP in SEZAINST.

You can stop the Policy Agent by:

- Using the operator command STOP
- Using the kill command in the z/OS shell
- Using the operator command CANCEL. Use the CANCEL command only as a last resort if the STOP or kill commands do not completely stop the Policy Agent.

Result: When the Policy Agent is shut down normally (KILL or STOP), if the PURGE option is configured, all QoS, IDS, and AT-TLS policies are purged from this stack. IPSec policies are not automatically purged.

The following kill command with the TERM signal will enable Policy Agent to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

where pid is the Policy Agent process ID.

The Policy Agent process ID can be obtained using the following z/OS UNIX command:

```
ps -A
```

It can also be obtained from the /tmp/pagent.pid file. The /tmp/pagent.pid file is a temporary file created by the Policy Agent. It contains the process ID of the current (or last) invocation of the Policy Agent.

Refreshing policies

The MODIFY command may be used to interactively cause the Policy Agent to reread the configuration information and, if requested, download objects from the LDAP server. In addition to this, the Policy Agent will also accept SIGHUP signals to perform the refresh function. Refer to the *z/OS Communications Server: IP System Administrator's Commands* for more detailed information on the MODIFY command.

Verification

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies defined correctly to the LDAP server?
- Are the policies defined correctly to the Policy Agent?

The following sections provide more details about these considerations.

Are the policies defined correctly to the LDAP server?

Refer to the documentation appropriate for the LDAP server which you are using. LDAP servers usually allow you to install multiple files (LDIF), each containing different objects in the LDAP hierarchy. Structural objects higher in the directory tree must be installed before objects that are contained below them. Check for any error messages as each LDIF is installed. Some LDAP servers interpret two consecutive blank lines as end of file. Ensure that all of the objects in the LDIF have been installed by the LDAP server.

Are the policies defined correctly to the Policy Agent?

When starting the Policy Agent, first check for any error messages issued to the console. Message EZZ8434I indicates something is wrong with the Policy Agent environment. Message EZZ8438I indicates a syntax or semantic error in the policy definitions. Messages EZZ8439I and EZZ8440I indicate a problem with the LDAP server configuration or the server itself. For more information on diagnosing Policy Agent problems, refer to *z/OS Communications Server: IP Diagnosis Guide*. Use the UNIX **pasearch** command to display policy definitions. The output from this command indicates whether or not policy rules are active, and shows the parsed results of the policy definition attributes. One thing to note is that the Policy Agent

is designed to ignore unknown attributes, so misspelled attributes will result in default values being used. The **pasearch** output can be used to verify that policies are correctly defined.

Chapter 15. Quality of service (QoS)

Applications and users of TCP/IP networks have different requirements for the service they receive from those networks. A network that treats all traffic as best effort does not meet the needs of such users. *Service differentiation* is a mechanism to provide different service levels to different traffic types based on their requirements and importance in an enterprise network. For example, it might be critical to provide Enterprise Resource Planning (ERP) traffic better service during peak hours than that of FTP or web traffic. The overall service provided to applications or users, in terms of elements such as throughput and delay, is termed *Quality of service* (QoS). Network service providers that need to provide different QoS levels express their business goals in *Service Level Agreements* (SLAs). There are two types of service in TCP/IP networks that relate to QoS. The first is *Differentiated Services*, which provides QoS to broad classes of traffic or users, for example all outbound web traffic accessed by a particular subnet. The second is *Integrated Services*, which provides end-to-end QoS to an application, by reserving resources along a data path. For z/OS Communications Server, Integrated Services is largely provided by the RSVP Agent, which implements the Resource ReserVation Protocol.

Workload distribution also relates to QoS, in terms of the throughput and delay characteristics of a given server in a sysplex. The ability to dynamically monitor server performance and affect sysplex workload distribution is an important part of the overall QoS of a sysplex. Also important is the ability to limit the set of target systems considered for sysplex routing based on network selection criteria, such as source subnet.

Differentiated Services (DS) policies

Policies for Differentiated Services are used to select and control DS traffic for selected IP servers, such as FTP server traffic. The policy administrator selects the IP traffic to be controlled by defining policy rules. These policy rules include several attributes that can be specified to identify the traffic to be managed. These attributes fall into 2 categories, general attributes and application specified attributes. General attributes can be used to identify the IP traffic of most IP applications using a variety of information, such as:

- The source or destination IPv4 or IPv6 addresses or subnets
- The source and destination ports used by the application
- The IP protocol the application is using (TCP or UDP)
- The network interface selected for the outgoing traffic
- The jobname of the application

Application specified attributes allow policy administrators to identify outgoing application IP traffic based on information that is provided and defined by an application. For example, the IBM HTTP Server provides the TCP/IP stack with the URI (Universal Resource Identifier) associated with any outgoing data being sent to a client. This allows the policy administrator to define rules that identify traffic related to specific URIs and policy actions with unique DS controls for this traffic. For example, an installation can define a policy that specifies preferential treatment of outgoing traffic related to the servicing of any URIs beginning with */product/placeorder*. For more information on defining policy rules for the IBM HTTP

Server based on URIs, refer to *z/OS HTTP Server Planning, Installing, and Using* and the policy configuration file section in *z/OS Communications Server: IP Configuration Reference*.

Any IP application using the TCP protocol can provide application specified attributes using extensions to the sendmsg() socket API. For more information, refer to the programming interfaces appendix in the *z/OS Communications Server: IP Programmer's Guide and Reference*. Application provided attributes can be specified in 2 forms:

- Application defined data: This allows applications to provide free-form text data that can be used to classify the application's outgoing traffic in terms that should be familiar to the application's administrator (for example, URIs are provided by the IBM HTTP server).
- Application specified priority: This allows applications to associate an application priority on the outgoing IP traffic. This application priority in itself does not automatically cause the application's traffic to get preferential treatment. In order to make use of these application specified priorities the policy administrator needs to define policy rules that map these priorities to policy actions that will govern the outgoing traffic of each priority level.

Applications can pass both application defined data and application specified priorities to the TCP/IP stack. When both are specified, the administrator is free to use either or both criteria in their service policy rules. However, it is strongly recommended that any policy rules defined using the application specified attributes should also include at least one general attribute that uniquely identifies the application instance. For example, when defining rules for the HTTP server using URIs, you can help further identify the application by specifying the source port for the server or the HTTP Server's jobname. This will help insure that unauthorized applications cannot exploit policy actions intended for the HTTP Server.

Several aspects of connection and throughput control can be specified with DS policies, including the following:

- TCP connection limits
- Maximum and minimum TCP connection rates, TCP maximum delay
- Committed access bandwidth (mean rate and peak rate) control/enforcement, also known as token bucket traffic shaping
- IPv4 type of service (ToS) byte or IPv6 traffic class setting, and mapping to zSeries Queued Direct I/O (QDIO) device priorities and VLAN user priorities.

The above DS service attributes are enforced by the TCP/IP stack in which the DS policies are installed. For additional information on the enforcement of these attributes, refer to *z/OS Communications Server: IP Configuration Reference*.

Token bucket traffic shaping is defined using the following parameters:

- DiffServInProfileRate is the average or mean rate that is desired to be transmitted over time. For example, 256 kilobits per second.
- DiffServInProfileTokenBucket is the *burst* size. This is how much data is allowed to be sent from the application to TCP/IP and still be allowed to be transmitted at the mean rate above. It is suggested, if the application is not policing itself, that this burst size be at least one second's worth of data. Otherwise, if the application is sending large amounts of data at one time to TCP/IP, TCP/IP will slow that application down via congestion windows, and the mean rate may not be achieved.

- DiffServInProfilePeakRate is the highest rate that is allowed to be transmitted for a shorter interval of time. For example, though a customer may only want on average 256 Kb of data per second, they may allow a peak of 512 Kb of data for 1/4 second. The peak rate is used to control the spacing of outbound packets on the transmission line. By having a smaller peak rate, there will be longer spacing between packets, and thus less burstiness of traffic and increased efficiency. Higher peak rates result in shorter spacing and increased burstiness, which can result in lower link utilization. However, some applications may require it, such as real time or video data.
- DiffServInProfileMaxPacketSize is the amount of data that will be policed at the peak cell rate. For example, if the peak rate is 512 Kb per second, and the maximum packet size is 120 Kb, TCP/IP will only allow about 10 packets of size 1492 bytes to be transmitted every .23 seconds. Again, if an application is sending large amounts of data at one time to TCP/IP, TCP/IP will enforce the peak rate, and anytime more than 10 packets are sent within .25 seconds, TCP/IP will begin to slow this TCP connection. The peak rate can be achieved over a longer period of time if the maximum packet size is entered in larger multiples of packets. However, this will cause greater burstiness as described above. For example, if the maximum packet size is entered as 240 Kb, TCP/IP will allow 20 packets in a .23 second range before enforcing slowdown.

  Note that the peak rate cannot be enforced without mean rate policing. However, you can enforce mean rate without peak rate. Also, setting of these parameters depends on the type of applications and the network that carries it.

Integrated Services (RSVP) policies

RSVP policies are used to set limits on certain parameters requested by RSVP applications. These applications interact with the RSVP Agent to establish resource reservations along a network path, using the RSVP API (RAPI). The reservation requests are in the form of an entity known as a Traffic Specification, or Tspec, which consists of the following values:

- Token bucket mean rate (r)
- Token bucket depth (b)
- Peak rate (p)
- Minimum policed unit (m)
- Maximum packet size (M)

RSVP policies can be used to limit the values requested for (r) and (b), as well as limiting the total number of RSVP reserved flows. The RSVP service attributes are enforced by the RSVP agent which gets RSVP policies from the Policy Agent. For additional information on the enforcement, refer to *z/OS Communications Server: IP Configuration Reference* or RFC 1363.

Sysplex Distributor (SD) policies

Sysplex Distributor policies are used to specify a set of SD target nodes for a given set of traffic. For example, all traffic destined to a given port or application from a specified subnet can be assigned one group of SD target nodes, while traffic for the same port or application from another subnet can be assigned to a different group of target stacks. These policies be can used in conjunction with other Sysplex Distributor controls to assist in load balancing. For more information, see “Policy interactions” on page 391.

QoS-specific Policy Agent functions

In addition to supporting the various types of policies, the Policy Agent performs functions related to the Sysplex Distributor. The Policy Agent can be configured to collect network QoS performance data relevant to SD on behalf of policies defined for a target port or application, and assign a default QoS weight fraction to such policy traffic. This weight is then used by SD (in conjunction with weights assigned by the Workload Manager) to assist in load balancing decisions. This function is performed by the Policy Agent on SD target nodes within the sysplex.

The Policy Agent also supports load distribution by service level. Performance data is kept for each Policy Action (service level) that a target's DVIPA port or application supports. A Policy Action weight fraction is generated. If available, this weight is used (instead of the default QoS weight fraction) in conjunction with the Workload Manager weight to assist in load distribution decisions for traffic assigned to this service level. If the Policy Action weight fraction is unavailable, the Sysplex Distributor will continue to use the default QoS weight fraction.

Another function related to policy performance is the performance collection function. When so configured, the Policy Agent collects policy performance data from the stack and caches it. This performance data is then made available to user applications through the Policy API (PAPI), for near real-time performance monitoring applications. The data are also optionally logged to a performance log file for offline performance monitoring. Sample C applications are provided to show how to use the PAPI interfaces to access performance data, and how to access and read the performance log file.

Policy performance data collected is affected by the FLUSH or NOFLUSH parameter on the Policy Agent TcpImage statement that defines the corresponding stack that is collecting the data. When FLUSH is specified, policies are deleted at the following times:

- When a new TcpImage statement is processed for the first time, including Policy Agent starting. This should not be a concern in most cases.
- When a MODIFY REFRESH command is entered.

As a result, all previously collected metrics start again from 0 when the policies are reinstalled. Conversely, policies are never deleted when NOFLUSH is specified, so performance metrics are never reset to 0.

Sysplex Distributor policy performance monitoring and policy performance collection are similar in some respects but distinctly different in others:

- Sysplex Distributor policy performance monitoring is actively performed by the Policy Agent. This performance monitoring is only used to assist with load balancing in a Sysplex Distributor environment.
- Policy performance collection is performed without regard to whether or not the Policy Agent is running in a Sysplex Distributor environment. Also, the Policy Agent does not actively participate in performance monitoring, only making the performance data available to user applications that perform the actual monitoring.

Policy performance data might not immediately change when changes are made to policy definitions. Some of the performance metrics are average values, and some are smoothed over several sampling intervals. As a result, when making changes to policies, some period of time will need to elapse before a new steady state is achieved.

Another function supported by the Policy Agent is to map IPv4 type of service (ToS) byte or IPv6 traffic class values to outbound interface priority values for outbound traffic. The ToS byte is also referred to as the Differentiated Services (DS) byte as an alternative definition (refer to RFC 2474). Note that outbound interface priority values are only meaningful for QDIO interfaces. A set of mappings can be defined to cover various ToS byte or traffic class values and map them to an appropriate interface priority. All outbound packets over the associated interfaces with a given ToS byte or traffic class value will then be assigned the corresponding priority value. ToS byte or traffic class values can also be mapped to Virtual LAN (VLAN) user priorities for propagation over LANs directly connected through OSA-Express.

Note: Coding the virtual LAN (VLAN) user priority causes a frame to be sent out based on the IEEE 802.1Q specification, which establishes a standard method for tagging Ethernet frames with VLAN priority and membership information. Specifically, a VLAN priority-tagged frame is used to convey packet priority to the switches; it has a value of NULL for VLANID. A full VLAN-tagged frame contains both the priority and non-null VLANID. If you have switches in your network that do not support the IEEE 802.1Q standard or that are not properly configured for these types of frames, the frames might be dropped by the switch.

Sysplex Distributor policy performance monitoring configuration

Before activating the Sysplex Distributor policy performance monitoring function, see "Policy interactions" on page 391 for information on workload balancing and policy interactions with Sysplex Distributor.

The following example illustrates how to activate the policy performance monitoring function for Sysplex Distributor.

Note: This function is activated on SD target servers and is used to monitor the performance of outbound traffic being serviced by the target servers. The goal is to detect TCP traffic that exceeds defined thresholds for dropped packets or time-outs, and derive a default QoS weight fraction for the target server. This default QoS weight fraction is then used to reduce the WLM weight assigned to the target servers, so that the SD distributing stack can take QoS performance into account.

The following statements apply to the example in this section:

- The policy performance monitoring sampling interval is 60 seconds.
- Policy Agent assigns a loss ratio weight fraction of 25% when the TCP loss ratio (dropped packets to total packets) starts to exceed 2%.
- The loss ratio weight fraction is increased to 50% when the loss ratio starts to exceed 4%, continuing in this manner up to the maximum loss ratio weight fraction of 95%.
- In a similar manner, a TCP timeout weight fraction of 50% is assigned when the timeout ratio starts to exceed 5%, increasing up to a maximum timeout weight fraction of 100%.
- The loss ratio weight fraction and TCP timeout weight fraction are added together to form a single default QoS weight fraction for the target server, up to a maximum of 100%. When the Traffic Regulation policy connection limit

reaches constrained threshold (90%), the default QoS weight fraction is set to 100% and forces SD to route requests to other target nodes with better routing weights.

- The default QoS weight fraction is used at the SD distributing stack to reduce the WLM weight. For example, if the WLM weight is 40, a weight fraction of 50% results in the weight being reduced to 20.
- The traffic to be monitored must be represented by at least one Differentiated Services policy defined for the target application (in this example a policy is defined for Telnet).
- An additional Policy Action weight fraction is calculated for a target's DVIPA/Port if there are any active connections to the target using that service level.

  The Policy Action weight fraction is calculated as the largest of three fractions:

  – The number of active connections to this target DVIPA/Port will be compared with the maximum connections allowed for this Policy action.
    - When the number of active connections reaches 50% of maximum connections, then the Policy Action weight fraction will be set to MAX (50%, current calculated value).
    - When the number of active connections reaches 65% of maximum connections, then the Policy Action weight fraction will be set to MAX (85%, current calculated value).
    - When the number of active connections reaches 80% of maximum connections, then the Policy Action weight fraction will be set to 100%.
  – The throughput rate for this timer interval will be calculated and compared to the DiffServ mean rate of this Policy action. If the throughput rate is greater than 85% of the DiffServ mean rate, the average throughput rate per connection will be calculated. If the throughput rate per connection is less than the DiffServ min rate, the minimum throughput requirement per connection is not being met and the Policy Action weight fraction will be set to 100%.
  – The default QoS weight fraction.
- Only one policy rule and policy action are defined here.

  As a result, only Telnet QoS performance information is monitored by the Policy Agent for Sysplex Distributor to route incoming Telnet connections to this target node relative to other target nodes which presumably can also accept Telnet requests.

```
PolicyPerfMonitorForSDR  enable
{
   samplinginterval 60
   LossRatioAndWeightFr 20 25
   TimeoutRatioAndWeightFr 50 50
   LossMaxWeightFr 95
   TimeoutMaxWeightFr 100
   MaxConnWeightFr 50 65 80
}
policyAction telnetGold
{
MinRate 500 # Provide minimum rate of 500 Kbps.
OutgoingTOS 10100000 # the TOS value of outgoing telnet packets.
}

policyRule targetelnet
{
```

```
  ProtocolNumberRange 6
  SourcePortRange 23
  policyactionreference telnetGold
  }
```

Policy performance collection configuration

The following example shows how to activate the policy performance collection function. Policy performance data for all active policies is maintained by the TCP/IP stack. This function allows this data to be collected and made available for policy performance monitoring applications. The following statements apply to the example in this section:

- Performance data is collected for both rules and actions. Action data is an aggregate of the data for the rules that refer to the action. For policies that have a single action per rule, the performance data will be the same for both the rule and action.
- The default minimum sampling interval is 30 seconds. This is the minimum value accepted for the acceptableCachedTime parameter on the PAPI papi_get_perf_data() function that gets performance data.
- Performance data will be logged to the file /u/user10/perflog, based on a sampling interval of 60 seconds.
- The number of performance log files maintained is 10, each of which is the default 300 kilobytes in size. In this example, the log files will be named assuming a stack name of TCPCS:

  ```
  /u/user10/perflog.TCPCS
  /u/user10/perflog.TCPCS.1
  /u/user10/perflog.TCPCS.2
  ⋮
  /u/user10/perflog.TCPCS.9
  ```

- Each performance data record is 232 bytes, so a file size of 300 kilobytes can contain 1324 records. Since 10 files are maintained, the total set of log files can contain 13240 records. Assuming that 50 policy rules and actions exist in the configuration, this means that the set of log files will wrap in approximately 4.4 hours, according to the following formula:

  ```
  number of records (13240) / number of policies (50) = number of refresh cycles (264)

  number of refresh cycles (264) * refresh interval in minutes (1) = 264 minutes worth of data
  ```

```
PolicyPerformanceCollection Enable
{
DataCollection           Rule Action
LogSamplingInterval      60
PerformanceLogFile       /u/user10/perflog
NumberOfLogFiles         10
}
```

IPv4 type of service (ToS) or IPv6 traffic class mapping configuration

There are two mappings provided by the SetSubnetPrioTosMask statement:

- IPv4 ToS or IPv6 traffic class to device priority

  Quality of service (QoS) support in z/OS Communications Server allows the IPv4 ToS byte, also known as the Differentiated Services (DS) field, or the IPv6 traffic class to be set for outbound IP packets according to defined policies managed by the z/OS Communications Server UNIX Policy Agent. When IP packets are sent out over QDIO devices, the ToS/DS or traffic class value is mapped to a QDIO priority value. Device priority values are 1-4, where 1 is the highest priority.

- IPv4 ToS or IPv6 traffic class to VLAN user priority

  ToS/DS or traffic class values can be mapped to user priorities for directly attached LANs using OSA-Express in QDIO mode. VLAN user priority values are 0-7, where 7 is the highest priority. This allows assigned user priorities to be propagated through such networks, resulting in no loss of priority information for data being served by z/OS.

Refer to *z/OS Communications Server: IP Configuration Reference* for more detail on these statements.

The following example shows a mapping of various ToS byte or traffic class values to associated interface priority values. Note that the mapping can be applied to individual interfaces or all interfaces:

- The first example defines a mapping for a specific interface. Note that the specified interface must be a valid interface specified in the HOME list. The second example shows a different mapping for all other interfaces.
- The subnet mask defines the bits in the ToS byte or traffic class that are significant. These examples use the leftmost 3 bits.
- The first example shows a set of mappings defining the complete set of ToS byte or traffic class values and the device and VLAN user priorities to be assigned for each value.
- The second example shows a set of mappings defining the complete set of ToS byte or traffic class values and the device priority to be assigned for each value.

```
SetSubnetPrioTosMask
{
  SubnetAddr         10.10.1.5
  SubnetTosMask      11100000
  PriorityTosMapping 1 11100000 7
  PriorityTosMapping 1 11000000 7
  PriorityTosMapping 2 10100000 6
  PriorityTosMapping 2 10000000 5
  PriorityTosMapping 2 01100000 5
  PriorityTosMapping 3 01000000 3
  PriorityTosMapping 4 00100000 2
  PriorityTosMapping 4 00000000 0
}
SetSubnetPrioTosMask
{
  SubnetTosMask      11100000
  PriorityTosMapping 1 11100000
  PriorityTosMapping 1 11000000
  PriorityTosMapping 1 10100000
  PriorityTosMapping 1 10000000
  PriorityTosMapping 2 01100000
  PriorityTosMapping 2 01000000
  PriorityTosMapping 3 00100000
  PriorityTosMapping 4 00000000
}
```

Defining policies in a Policy Agent configuration file

Configure the following statements in the configuration file to define policies:

- PolicyAction
- PolicyRule

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about these statements.

The following sections contain examples of these tasks.

Note: These examples are for illustrative purposes only. The policies deliberately use a wide variety of attributes, and they do not necessarily represent real-world usage. Some examples show continued and indented statements that were modified to fit on the page and therefore are not an actual representation of proper syntax.

Differentiated Services policy examples

The goal of this Differentiated Services policy is to map a subset of the traffic outbound from an FTP server.

This policy is identified as a Differentiated Services policy by the PolicyScope DataTraffic attribute on the PolicyAction statement, as well as the use of several DS-only attributes.

The following statements apply to the example in this section:

- The policy rule selects traffic originated by ports in the range 20-21 (FTP outbound data connection uses port 20) from the source address 200.50.23.11.
- The policy rule is active on weekdays between 6 a.m. and 10 p.m. local time, between the dates 7/1/2000 and 7/1/2005..
- The policy action specifies that the ToS byte be set to '10000000' for traffic that conforms to this policy.
- The action establishes a *token bucket* traffic conditioner with a mean rate of 256 kilobits per second, a peak rate of 512 kilobits per second, and a burst size of 64 kilobytes. Any traffic that exceeds these specifications will be sent as *best effort*, with an accompanying ToS byte of '00000000'.

```
PolicyRule                              diffServ
    {
       ProtocolNumberRange                  6
       SourceAddressRange                   200.50.23.11
       SourcePortRange                      20-21
       PolicyActionReference                tokenbucket
       PolicyRulePriority                   10
       ConditionTimeRange                   20000701000000:20050630235959
       DayOfMonthMask                       11111111111111111111111111111111
       DayOfWeekMask                        0111110
       TimeOfDayRange                       06:00-22:00
    }
    PolicyAction                            tokenbucket
    {
       PolicyScope                          DataTraffic
       OutgoingTOS                          10000000
       DiffServInProfileRate                256       # 256 Kbps
       DiffServInProfileTokenBucket         512       # 512 Kbits
       DiffServInProfilePeakRate            512       # 512 Kbps
       DiffServInProfileMaxPacketSize       120       # 120 Kbits
       DiffServOutProfileTransmittedTOSByte 00000000
       DiffServExcessTrafficTreatment       BestEffort
    }
```

The goal of this policy is to ensure that outgoing data that match the specified attributes will be assigned a QoS service level defined in action "interactive1".

The following statements apply to the example in this section:

- This rule will only match traffic on TCP connections (protocol 6) with a source port of 80 (i.e. HTTP server) and application defined data beginning with the string "/catalog".

- Since we are dealing with HTTP traffic, this rule is basically indicating that all outgoing traffic associated with a URI that begins with ″/catalog″ should be managed using the DS characteristics specified in the ″interactive1″ policy action.

```
PolicyRule          web-catalog       # web catalog traffic
{
     protocolNumberRange  6
     SourcePortRange    80
     ApplicationData /catalog
     policyActionReference  interactive1

}

PolicyAction      interactive1
{
    policyScope DataTraffic
    outgoingTOS 10000000
}
```

RSVP policy example

The goal of this RSVP policy is to establish limits on resource reservations requested by RSVP applications using the RSVP API (RAPI) interface. The policy is identified as an RSVP policy by the PolicyScope attribute on the PolicyAction statement, as well as the use of RSVP-only attributes.

The following statements apply to the example in this section:

- The policy rule selects traffic from source ports in the range 8000 to 8001, with a protocol ID of 6 (TCP).
- The DataTraffic policy action specifies that the ToS byte be set to 01100000 for differentiated services traffic that conforms to this policy. Essentially, any traffic sent by the target application without an RSVP reservation in place will use this policy action. Once an RSVP reservation is in place, the RSVP action gets used.
- The RSVP policy action specifies that the ToS byte be set to 01100000 while an RSVP reservation is in place. It also limits the type of RSVP service requested by RSVP applications to Controlled Load. Applications requesting Guaranteed service are downgraded to using Controlled Load service. In addition, the action limits the mean rate and token bucket size to 50000 bytes per second and 6000 bytes, respectively. These values are requested by RSVP applications in the traffic specification, or Tspec.
- The action also limits the number of active RSVP flows that map to this policy to 10.

```
PolicyRule                              intserv
     {
        SourcePortRange                      8000 8001
        ProtocolNumberRange                  6
        PolicyActionReference                intserv1
        PolicyActionReference                intserv2
     }
     PolicyAction                            intserv1
     {
        PolicyScope                          DataTraffic
        OutgoingTOS                          01100000
     }
     PolicyAction                            intserv2
     {
        PolicyScope                          RSVP
        OutgoingTOS                          01100000
        FlowServiceType                      ControlledLoad
```

```
        MaxRatePerFlow                          400    # 50000 bytes/second
        MaxTokenBucketPerFlow                   48     # 6000  bytes
        MaxFlows                                10
     }
```

Sysplex Distributor policy example

The goal of this Sysplex Distributor policy is to limit the number of SD target stacks for inbound Telnet traffic. The policies are identified as SD policies by the ForLoadDistribution TRUE attribute on the PolicyRule statement. The corresponding policy on the target is also shown.

The following statements apply to the example in this section:

- Separate policies are defined on the Sysplex Distributor distributing and target stacks.
- The policy rules select incoming Telnet connection requests.
- The selected target stack will be based on WLM information and QoS information if activated at the target stacks.
- The rule (disttelnet) is coded on the distributing stack to select inbound traffic destined to the Telnet server.
- The rule (targtelnet) is coded on the target stack to select outbound data from the Telnet server.
- If none of the specified target stacks is available to service incoming requests (either the node is down or the Telnet server is not active), then Sysplex Distributor will distribute the requests to any available target stack.

  Note: If the OutboundInterface 0.0.0.0 statement were not present, and neither of the defined target stacks were available, Sysplex Distributor would reject the request.

```
policyAction        telnetGold
    {
        MinRate        500       # Provide minimum rate of 500 Kbps.
        OutgoingTOS    10100000  # the TOS value of outgoing telnet packets.
        outboundinterface  129.100.11.1
        outboundinterface  129.100.21.1
        outboundinterface  129.200.12.1
        outboundinterface  129::1B0D:13F0
        outboundinterface  0.0.0.0
        outboundinterface  ::
    }

    policyRule          disttelnet
    {
        ProtocolNumberRange     6
        DestinationPortRange    23
        PolicyRulePriority      20
        policyactionreference   telnetGold
        ForLoadDistribution     TRUE
    }

    policyRule          targtelnet
    {
        ProtocolNumberRange     6
        SourcePortRange         23
        PolicyRulePriority      20
        policyactionreference   telnetGold
        ForLoadDistribution     FALSE
    }
```

Notes:

1. The ApplicationName attribute is only valid for a target rule and should not be coded on a distributor rule because the application name determined for inbound traffic (which is always the case on a distributor) will always be the stack's TCP jobname.
2. If you are using Telnet with multiple stacks in conjunction with the Sysplex Distributor, see Chapter 10, "Accessing remote hosts using Telnet," on page 441 for more information.

Defining policies using LDAP

This section contains information on Quality of service Manager for z/OS Communications Server (zQoS Manager), as well as some example policy definitions.

zQoS Manager for z/OS Communications Server

Quality of service Manager for z/OS Communications Server (zQoS Manager) enables networking QoS policy management across data centers, assigning a variety of QoS service labels to different traffic. Network administrators use this tool to produce a file that an LDAP server can process. The LDAP server can then be read by the Policy Agent to configure policies. For more information, refer to:

`http://www.ibm.com/software/network/commserver/downloads/`

Differentiated Services policy example

The goal of this Differentiated Services policy is to map a subset of the traffic outbound from an FTP server.

This policy is identified as a Differentiated Services policy by the ibm-PolicyScope:DataTraffic attribute in the ibm-PolicyActionInstance object.

The following statements apply to the example in this section:

- The policy rule selects traffic originated by ports in the range 20-21 (FTP outbound data connection uses port 20) from the source address 200.50.23.11 or 200.50.33.14 or 202::B055:1.
- The policy rule is active on weekdays between 6 a.m. and 10 p.m. local time, between the dates 7/1/2000 and 7/1/2005.
- The policy action specifies that the ToS byte be set to '10000000' for traffic that conforms to this policy.
- The action establishes a *token bucket* traffic conditioner with a mean rate of 256 kilobits per second, a peak rate of 512 kilobits per second, and a burst size of 64 kilobytes. Any traffic that exceeds these specifications will be sent as *best effort*, with an accompanying ToS byte of '00000000'.

```
dn:cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:diffserv-rule
ibm-policyRuleName:diffserv-rule
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListType:2
ibm-policyRuleConditionListDN:cn=condassoc1, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc2, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3a, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3b, cn=diffserv-rule, cn=QoS, cn=advanced,
```

```
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3c, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc1, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:10
ibm-policyRuleMandatory:TRUE
ibm-policyRuleSequencedActions:1
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Differentiated Services rule

dn:cn=condassoc1, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:diffserv-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable CNF condition at level 1 - TCP

dn:cn=condassoc2, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:diffserv-condition2
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=ftpdPorts, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable CNF condition at level 2 - ftpd ports

dn:cn=condassoc3a, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3a
ibm-policyConditionName:diffserv-condition3a
ibm-policyConditionGroupNumber:3
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host1, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents first reusable CNF condition at level 3 - host1

dn:cn=condassoc3b, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3b
ibm-policyConditionName:diffserv-condition3b
ibm-policyConditionGroupNumber:3
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host2, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents second reusable CNF condition at level 3 - host2

dn:cn=condassoc3c, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3c
ibm-policyConditionName:diffserv-condition3c
ibm-policyConditionGroupNumber:3
```

```
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host3, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents third reusable CNF condition at level 3 - host3

dn:cn=actassoc1, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:diffserv-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=tokenbucket, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - token bucket

dn: cn=tokenbucket, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:tokenbucket
ibm-policyActionName:tokenbucket-action
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:10000000
ibm-DiffServInProfileRate:256
ibm-DiffServInProfilePeakRate:512
ibm-DiffServInProfileTokenBucket:512
ibm-DiffServInProfileMaxPacketSize:120
ibm-DiffServOutProfileTransmittedTOSByte:00000000
ibm-DiffServExcessTrafficTreatment:BestEffort
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS diffserv token bucket action

dn:cn=ftpdPorts, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:ftpdPorts
ibm-policyConditionName:ftpdPorts-condition
ibm-sourcePortRange:20-21
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS ftpd ports condition

dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition

dn:cn=Host1, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host1
```

```
ibm-policyConditionName:Host1-condition
ibm-SourceIPAddressRange:3-200.50.23.11
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 1 condition

dn:cn=Host2, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host2
ibm-policyConditionName:Host2-condition
ibm-SourceIPAddressRange:3-200.50.33.14
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 2 condition

dn:cn=Host3, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host3
ibm-policyConditionName:Host3-condition
ibm-SourceIPAddressRange:5-202::B055:1
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 3 condition

dn:cn=period1, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period1
ibm-policyInstanceName:WeekdayPrime-time
ibm-ptpConditionTime:20000701000000:20050630235959
ibm-ptpConditionMonthOfYearMask:111111111111
ibm-ptpConditionDayOfMonthMask:11111111111111111111111111111111
ibm-ptpConditionDayOfWeekMask:0111110
ibm-ptpConditionTimeOfDayMask:060000:220000
ibm-ptpConditionLocalOrUtcTime:1
ibm-policyKeywords:POLICY
description:Active weekdays 6am - 10pm local time, 7/1/2000 to 7/1/2005
```

The goal of this policy is to ensure that outgoing data that match the specified attributes will be assigned a QoS service level defined in action ″interactive1″.

The following statements apply to the example in this section:

- This rule will only match traffic on TCP connections (protocol 6) with a source port of 80 (i.e. HTTP server) and application defined data beginning with the string ″/catalog″.
- Since we are dealing with HTTP traffic, this rule is basically indicating that all outgoing traffic associated with a URI that begins with ″/catalog″ should be managed using the DS characteristics specified in the ″interactive1″ policy action.

```
dn:cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:web-catalog-rule
ibm-policyRuleName:web-catalog-rule
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListType:1
ibm-policyRulePriority:10
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
```

```
ibm-policyKeywords:POLICY
description:QOS Web catalog rule

dn:cn=condassoc1, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:web-catalog-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition - TCP

dn:cn=condassoc2, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:web-catalog-condition2
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=webPort, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition - web port

dn:cn=condassoc3, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc3
ibm-policyConditionName:web-catalog-condition3
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-applicationData:/catalog
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific condition - web catalog pages

dn:cn=actassoc1, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:web-catalog-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=interactive1, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - interactive 1

dn: cn=interactive1, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:interactive1
ibm-policyActionName:interactive1-action
ibm-policyScope:DataTraffic
ibm-outgoingTOS:10000000
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS interactive 1 action (TOS 100)

dn:cn=webPort, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
```

```
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:webPort
ibm-policyConditionName:webPort-condition
ibm-sourcePortRange:80
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS web port condition

dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition
```

RSVP policy example

The goal of this RSVP policy is to establish limits on resource reservations requested by RSVP applications using the RSVP API (RAPI) interface. The policy is identified as an RSVP policy by the ibm-PolicyScope:RSVP attribute in the ibm-PolicyActionInstance object.

The following statements apply to the example in this section:

- The policy rule selects traffic from source ports in the range 8000 to 8001, with a protocol ID of 6 (TCP).
- One policy action specifies that the ToS byte be set to 01100000 for traffic that conforms to this policy.
- The RSVP policy action limits the type of RSVP service requested by RSVP applications to Controlled Load. Applications requesting Guaranteed service are downgraded to using Controlled Load service. In addition, the action limits the mean rate and token bucket size to 50000 bytes per second and 6000 bytes, respectively. These values are requested by RSVP applications in the traffic specification, or Tspec.
- The action also limits the number of active RSVP flows that map to this policy to 10.

```
dn:cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:intserv-rule
ibm-policyRuleName:intserv-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleConditionListDN:cn=condassoc1, cn=intserv-rule, cn=QoS, cn=advanced,
     ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc1, cn=intserv-rule, cn=QoS, cn=advanced,
     ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc2, cn=intserv-rule, cn=QoS, cn=advanced,
     ou=policy, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period2, cn=time, cn=repository, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period3, cn=time, cn=repository, o=IBM, c=US
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Integrated Services rule

dn:cn=condassoc1, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
```

```
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc1
ibm-policyConditionName:intserv-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-ProtocolNumberRange:6
ibm-SourceIPAddressRange:1
ibm-SourcePortRange:8000-8001
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific condition - all local IP addresses, application TCP ports

dn:cn=actassoc1, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:actassoc1
ibm-policyActionName:intserv-action1
ibm-policyActionOrder:1
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:01100000
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific action - set TOS

dn:cn=actassoc2, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:actassoc2
ibm-policyActionName:intserv-action2
ibm-policyActionOrder:2
ibm-PolicyScope:RSVP
ibm-OutgoingTOS:01100000
ibm-FlowServiceType:ControlledLoad
ibm-MaxRatePerFlow:400
ibm-MaxTokenBucketPerFlow:48
ibm-MaxFlows:10
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific action - RSVP limitations

dn:cn=period2, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period2
ibm-policyConditionName:EndOfMonth-time
ibm-ptpConditionDayOfMonthMask:00000000000000000000000000000001
     00000000000000000000000000000000
ibm-ptpConditionLocalOrUtcTime:2
ibm-policyKeywords:POLICY
description:Active last day of the month (in UTC)

dn:cn=period3, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period3
ibm-policyConditionName:PacificNight-time
ibm-ptpConditionTimeOfDayMask:190000:030000
```

```
ibm-ptpConditionTimeZone:-08
ibm-policyKeywords:POLICY
description:Active 7pm - 3am local time, Pacific Time Zone (no daylight savings)
```

Sysplex Distributor routing policy example

The goal of this Sysplex Distributor policy is to limit the number of SD target servers for inbound Telnet traffic. The policies are identified as SD policies by the ibm-policyGroupForLoadDistribution:TRUE attribute in the ibm-PolicyGroup object.

The following statements apply to the example in this section:

- Separate policies are defined on the Sysplex Distributor distributing and target servers.
- The policy rules select incoming Telnet connection requests.
- The selected target server will be based on WLM information and QoS information if activated at the target servers.
- The rule (disttelnet) is coded on the distributing stack to select inbound traffic destined to the Telnet server.
- The rule (targtelnet) is coded on the target server to select outbound data from the Telnet server.
- If none of the specified target servers are available to service incoming requests (either the node is down or the Telnet server is not active), then Sysplex Distributor will distribute the requests to any available target server.

Note: If the ibm-Interface:1–0.0.0.0 attribute were not present, and none of the defined target servers were available, Sysplex Distributor would reject the request.

```
dn:cn=sysplex, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyGroup
objectclass:ibm-policyRuleContainmentAuxClass
objectclass:ibm-policyGroupLoadDistributionAuxClass
cn:sysplex
ibm-policyGroupName:QoSadvancedsysplex-Group
ibm-policyRulesAuxContainedSet:cn=disttelnet-rule,cn=QoS,cn=advanced,ou=policy,
     o=IBM,c=US
ibm-policyGroupForLoadDistribution:TRUE
ibm-policyKeywords:Sysplex
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description: QoS advanced examples sysplex group.

dn:cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:disttelnet-rule
ibm-policyRuleName:disttelnet-rule
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:20
ibm-policyKeywords:Sysplex
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Sysplex Distributor telnet rule

dn:cn=condassoc1, cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc1
ibm-policyConditionName:disttelnet-condition
```

```
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-ProtocolNumberRange:6
ibm-DestinationPortRange:23
ibm-policyKeywords:Sysplex
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific condition - telnet inbound SD traffic

dn:cn=actassoc1, cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:disttelnet-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Sysplex
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - telnet Gold Service

dn:cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:targtelnet-rule
ibm-policyRuleName:targtelnet-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:20
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Sysplex Target telnet rule

dn:cn=condassoc1, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:targtelnet-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable condition at level 1 - TCP

dn:cn=condassoc2, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:targtelnet-condition2
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=telnetdPort, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable condition at level 2 - telnetd port

dn:cn=actassoc1, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:targtelnet-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
```

```
description:Represents reusable action - telnet Gold Service

dn: cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:telnetGold
ibm-policyActionName:telnetGold-action
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:10100000
ibm-MinRate:500
ibm-Interface:1--129.100.11.1
ibm-Interface:1--129.100.21.1
ibm-Interface:1--129.200.12.1
ibm-Interface:1--0.0.0.0
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS telnet Gold Service action

dn:cn=telnetdPort, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:telnetdPort
ibm-policyConditionName:telnetdPort-condition
ibm-sourcePortRange:23
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS telnetd port condition

dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition

dn:cn=period1, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period1
ibm-policyInstanceName:WeekdayPrime-time
ibm-ptpConditionTime:20000701000000:20050630235959
ibm-ptpConditionMonthOfYearMask:111111111111
ibm-ptpConditionDayOfMonthMask:11111111111111111111111111111111
ibm-ptpConditionDayOfWeekMask:0111110
ibm-ptpConditionTimeOfDayMask:060000:220000
ibm-ptpConditionLocalOrUtcTime:1
ibm-policyKeywords:POLICY
description:Active weekdays 6am - 10pm local time, 7/1/2000 to 7/1/2005
```

RSVP

Resource ReSerVation Protocol (RSVP) is a protocol that provides a mechanism to reserve resources in support of Integrated Services. The z/OS UNIX RSVP agent provides the following services on behalf of applications that want to use Integrated Services:

- An RSVP API (RAPI) that allows applications to explicitly request RSVP services. Using RAPI, applications indicate their intent to send or receive data, describe the characteristics of the data traffic and request that RSVP reserve resources along the data path to provide a given QoS to one or more traffic flows. For more information about RAPI, refer to *z/OS Communications Server: IP Programmer's Guide and Reference*.
- Mapping of IP ToS settings to RSVP traffic, using policies defined for RSVP.
- Establishment of resource reservations on ATM interfaces by use of reserved SVC connections.

  Note: Resource reservations cannot be made on interfaces other than ATM for outbound traffic on z/OS. However, RSVP-capable routers in the network can still reserve resources, and the ToS byte can be set for RSVP traffic to provide further means of prioritizing traffic.
- Support for VIPA addresses as well as real IP addresses.
- Communication with other RSVP agents on hosts and routers in the network to communicate application resource reservation requests.

Network administrators can use the z/OS UNIX Policy Agent to define RSVP-specific policies. These policies can be used to limit the parameters of application-requested resource reservations, provide ToS mappings for RSVP traffic, and limit the number of traffic flows that can use RSVP services simultaneously.

RSVP is designed to be implemented on both end systems (hosts) and routers. Different functions are provided by RSVP in these two environments. The z/OS RSVP agent is supported as a host RSVP implementation only. It can communicate with router RSVP implementations, but is not itself supported as such. For more information about RSVP, refer to RFC 2205.

Configuring the RSVP agent

To configure the RSVP agent, update the configuration file to specify RSVP agent operational parameters using the LogLevel, TcpImage, Interface and RSVP statements. Refer to *z/OS Communications Server: IP Configuration Reference* for detailed information about the statements.

To start the RSVP agent, you must first authorize the RSVP Agent using the security product. See SEZAINST(EZARACF) for SAF considerations for started tasks.

The following is an example of an RSVP configuration file.

This example:

- Runs the RSVP Agent on the stack selected using the standard resolver search order, because a TcpImage statement is not configured.
- Disables RSVP processing on interface 10.11.12.13, while enabling it for all other interfaces.
- Disables traffic control on interface 200.1.1.1. This means that no reservations will be made on this interface.
- Allows a maximum of 50 active RSVP flows per interface.

```
Interface 10.11.12.13 Disabled
{}
Interface 200.1.1.1 Enabled
{
TrafficControl Disabled
```

```
}
Interface Others Enabled
{}
Rsvp All Enabled
{
MaxFlows 50
}
```

Starting and stopping RSVP

RSVP can be started from the z/OS shell or as a started task.

The RSVP agent uses the following search order to locate the configuration file (highest priority is listed first):

- HFS file or MVS data set specified by the -c startup option. The syntax for an HFS file is '/dir/file', and the syntax for an MVS data set is "//'MVS.DATASET.NAME'".
- HFS file or MVS data set specified with the RSVPD_CONFIG_FILE enviroment variable.
- /etc/rsvpd.conf HFS file.
- 'hlq.RSVPD.CONF' MVS data set.

  Note: If this file is not present, RSVP is enabled on all network interfaces with default parameters.

When starting from the shell, note that the RSVP executable resides in /usr/lpp/tcpip/sbin. There is also a link from /usr/sbin. Make sure your path statement (in the profile) contains either /usr/sbin or /usr/lpp/tcpip/sbin.

Use the S RSVPD command on an MVS console or SDSF to start RSVP as a started task. A sample procedure is shipped in member EZARSVPP in SEZAINST.

RSVP can be stopped using the cancel command (C RSVPD) or using the kill command in the z/OS shell. The following kill command with the TERM signal will enable RSVP to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

where *pid* is the RSVP process ID.

The RSVP process ID can be obtained using the following z/OS UNIX command:

```
ps  -A
```

It can also be obtained from the /tmp/rsvpd.pid.imagename file. Refer to *z/OS Communications Server: IP Configuration Reference* for more information.

Network service level agreement performance monitor MIB subagents

Following are the SNMP network service level agreement performance monitor MIB subagents:

- SNMP Network SLAPM2 subagent (nslapm2)

  Provides information about defined service policies and performance data for applications that are mapped to those policies. Statistics are retrieved by this subagent and monitored for possible Network SLAPM2 performance deviations. For more information about the Network-SLAPM2 MIB, refer to usr/lpp/tcpip/samples/slapm2.mi2.

- SNMP SLA subagent (pagtsnmp)

  Provides information about defined service policies and performance data for applications which are mapped to those policies. Statistics are retrieved by this subagent and monitored for possible SLA performance deviations. Refer to RFC 2758 for more information about the SLAPM MIB.

Configuring the Network SLAPM2 subagent

The z/OS CS Network SLAPM2 subagent allows network administrators to retrieve data and determine if the current set of Network SLAPM2 policy definitions are performing as needed or if adjustments need to be made. Before starting the Network SLAPM2 subagent, some basic operational characteristics of the Network SLAPM2 subagent need to be configured using the following steps:

1. Configure the Network SLAPM2 subagent security authorization.

   For the subagent to retrieve performance monitor data from the Policy Agent, the Network SLAPM2 subagent must have superuser authority or security product authority in the SERVAUTH class.

   These profiles can be defined by TCP/IP stack and policy type as follows, where *sysname* is the system name defined in the sysplex, *TcpImage* is the TCP name for policy information that is being requested, and *ptype* is the policy type (*QOS*) that is being requested:

   ```
   EZB.PAGENT.sysname.TcpImage.ptype
   ```

   Wildcarding is allowed on segments of the profile name.

   If SERVAUTH class is absent (not RACLIST) or profiles are absent for the subagent's request (TcpImage and policy type), permission is denied and data is not returned.

   If SERVAUTH class is active, profiles are present for the subagent's request (TcpImage and policy type), and an MVS user is defined for all profiles, permission is granted and data is returned.

   If SERVAUTH class is active, profiles are present for the subagent's request (TcpImage and policy type), and an MVS user is not defined for all profiles, permission is refused and data is not returned.

   For the sample commands needed to create the profile name and permit users to it, refer to the EZARACF sample in SEZAINST.

2. Configure the Policy Agent as follows:
   - Configure QoS policy rules and QoS policy actions in Policy Agent. The Network SLAPM2 subagent only keeps statistics for active QoS policies. For details, see "Configuring the Policy Agent" on page 741.
   - Configure PolicyPerformanceCollection in Policy Agent. The PolicyPerformanceCollection statement for rules needs to be enabled to retrieve performance monitoring information from Policy Agent for the Network SLAPM2 Subagent. For details on the PolicyPerformanceCollection statement, see "Policy performance collection configuration" on page 757.
3. Configure and start the SNMP agent. For details, see "Step 1: Configure the SNMP agent (OSNMPD)" on page 1056.

Starting and stopping the Network SLAPM2 subagent

Before you start the Network SLAPM2 subagent, the following applications need to be started and initialized:

- Policy Agent - For details, see "Starting and stopping the Policy Agent" on page 746.

- SNMP agent - For details, see "Start the SNMP agent (OSNMPD)" on page 1066.

The Network SLAPM2 subagent can be started from the z/OS shell or as a started task.

- When starting from the shell:

  The Network SLAPM2 subagent executable (nslapm2) resides in /usr/lpp/tcpip/bin. There is also a link from /bin. Make sure your PATH statement (in the profile) contains either /bin or /usr/lpp/tcpip/bin.

  The Network SLAPM2 subagent requires access to one or more DLLs at run time. The LIBPATH environment variable needs to be set to include the /usr/lib directory, which normally includes all the required DLLs.

  Export the LIBPATH environment variable before starting the subagent. This is best accomplished in /etc/profile or in .profile in the HOME directory. For example:

  ```
  export LIBPATH=/usr/lib
  ```

  Following is an example command:

  ```
  nslapm2 -d 3 -t 1800 -c special -P 5000
  ```

  The command above starts the Network SLAPM2 subagent with the following characteristics:

  - Connect to the SNMP agent using a community name of *special* and a port of *5000*.
  - The debugging level is set to *3*, to log the following debugging messages to syslogd:
    - Trace Network SLAPM2 subagent error and system console messages
    - Trace Network SLAPM2 subagent warning messages
  - The MIB table cache time is set to 30 minutes.

- When starting as a started task:

  Use the S NSLAPM2 command on an MVS console or SDSF. A sample procedure is shipped in member EZAPAGSB in SEZAINST.

  - Specify LIBPATH using the ENVAR parameter on the PARM statement in the started procedure. For example:

    ```
    // PARM=('POSIX(ON) ALL31(ON)',
    // 'ENVAR("LIBPATH=/usr/lib")/')
    ```

  - Export the LIBPATH environment variable in a file specified with the STDENV DD statement. For example:

    ```
    //STDENV DD PATH='/etc/nslapm2.env',PATHOPTS=(ORDONLY)
    ```

    In the /etc/nslapm2.env file:

    ```
    LIBPATH=/usr/lib
    ```

For more information about specifying run-time options, refer to *z/OS Language Environment Programming Guide*. For details on setting the LIBPATH environment variable, also refer to *z/OS UNIX System Services Command Reference*.

The Network SLAPM2 subagent can be stopped using the stop command (P NSLAPM2), or using the kill command in the z/OS shell. For example, the following kill command with the TERM signal, where *pid* is the nslapm2 process ID, enables the Network SLAPM2 subagent to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

The nslapm2 process ID can be obtained using the following z/OS UNIX command:

```
ps  -A
```

Starting and stopping the SLA subagent

The SLA subagent can be started from the z/OS shell or as a started task.

When starting from the shell, note that the SLA subagent executable resides in /usr/lpp/tcpip/sbin. There is also a link from /usr/sbin. Make sure your PATH statement (in the profile) contains either /usr/sbin or /usr/lpp/tcpip/sbin.

For example, the following command starts the SLA subagent with these characteristics:

```
pagtsnmp -d 1 -t 1800 -c special -P 5000
```

- To connect to the SNMP Agent, a community name of ″special″ and a port of 5000 are used.
- The debugging level is set to 1, meaning internal debugging messages are written to syslogd.
- The MIB table cache time is set to 30 minutes.

Use the S PAGTSNMP command on an MVS console or SDSF to start the SLA subagent as a started task. A sample procedure is shipped in member EZAPAGSN in SEZAINST.

The SLA Subagent can be stopped using the stop command (P PAGTSNMP) or using the kill command in the z/OS shell. The following kill command with the TERM signal will enable the SLA subagent to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

where *pid* is where pid is the pagtsnmo process ID.

The pagtsnmp process ID can be obtained using the following z/OS UNIX command:

```
ps  -A
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about the MIB objects supported by the SLA subagent.

Verification

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies installed in the TCP/IP stacks?
- Is the expected traffic mapping to the correct policies?
- Are the Sysplex Distributor policy functions working correctly?
- Does anything need to be tuned?

The following sections provide more details about these considerations.

Are the policies installed in the TCP/IP stacks?

Use the NETSTAT SLAP or onetstat -j command to display QoS policy statistics. This command only displays statistics for active QoS (installed) policies, so it can be used to verify the correct policies are installed, even if the statistics are all 0. Since the Policy Agent can install policies on multiple stacks, issue this command on each stack to verify the correct set of QoS policies is installed.

Is the expected traffic mapping to the correct QoS policies?

While connections are active, use the NETSTAT ALL or onetstat -A command to display details about the active connections. One piece of information displayed is the policy rule name. If this name is blank, then the traffic is not mapped to any active rule. Also, use the NETSTAT SLAP or onetstat -j command to display QoS policy statistics. The output shows the time that each policy was last mapped to traffic, and accumulated statistics for each policy. Monitor these values over time to verify that new traffic is mapping as expected.

Note: The values displayed by the NETSTAT SLAP and onetstat -j commands can wrap around to 0. If some of the values do not seem correct (for example, total out bytes less than total out bytes in profile), then wrapping has probably occurred.

Are the Sysplex Distributor policy functions working correctly?

To verify that the distributor is using the expected service levels when deciding how to distribute traffic to each DVIPA/Port target, use the Netstat VDPT DETAIL or onetstat -O DETAIL command on the distributing stack. The following QoS related information will be displayed for each DVIPA/Port target:

- WLM weight unmodified by QoS
- Modified WLM/QoS aggregate weight, identified by *DEFAULT*
- Modified WLM/QoS service level weights, identified by service level name

To verify that active connections distributed to DVIPA/Port targets are using the expected service level, use the Netstat VCRT DETAIL command on the distributing stack. This will display the following policy related information:

- PolicyRule: the policy rule that the distributor used in selecting the policy action for this connection.
- PolicyAction: the policy action that this connection is currently using. If PolicyAction is specified by *NONE*, then the distributor is using the *DEFAULT* fraction to distribute this connection.

Refer to *z/OS Communications Server: IP Diagnosis Guide* for more information.

Does anything need to be tuned?

Poor performance, such as low throughput, long response times, and so on, might be suddenly and consistently experienced by a certain set of users or applications. Also, traps might be generated by one of the Network SLAPM subagents (SLA or Network SLAPM2 subagent). When this happens, the problem might be the way the QoS policy is defined for the corresponding set of users or applications.

For example, the IPv4 ToS/DS or IPv6 traffic class value might be set incorrectly to a lower QoS level than is intended, for example, medium or low priority instead of high priority. It is important to remember that given a fixed amount of network

resources, changing some traffic demand from a lower to higher QoS level will mean that other traffic demands will be affected. Therefore, use care to ensure that in attempting to meet one set of QoS requirements, different or worse problems do not result.

Another cause for poor performance might be in the way the bandwidth allocation defined via the DiffServ token bucket parameters, or TCP maxrate or minrate, is not adequate to accommodate the traffic demand. Yet another possibility might be that either network or the server capacity is not adequate to handle the traffic demand. This is evident when a majority of users or applications do not have their QoS requirements met. When this happens, the network planning process must be revisited.

For more information, see "Using the Network SLAPM MIBs to monitor policies."

Using pasearch

Use the **pasearch** command to display policy details. This command displays both active (installed in the stack) and inactive policies. Various parameters can be specified to filter the results, for example to display only policies for certain stacks, only QoS policies, only policy names, or only a single policy specified by name. Refer to *z/OS Communications Server: IP System Administrator's Commands* for the complete syntax and sample output for **pasearch**.

Using the Network SLAPM MIBs to monitor policies

The Network SLAPM2 and SLA subagents provide information about service policies and performance data for applications mapped to those policies.

Network SLAPM2 subagent performance monitoring

The Network SLAPM2 subagent provides information about service policies and performance data for applications mapped to those policies through the slapm2PolicyRuleStatsTable.

Note: The Network SLAPM2 subagent can be used to monitor Differentiated Services policies.

slapm2PolicyRuleStatsTable
Provides statistics on a per policy rule basis.

The Network SLAPM2 subagent also supports performance monitoring using the slapm2PRMonTable object. Entries are created in the monitor table to establish the desired criteria for monitoring. The following level of monitoring is provided:

Aggregate
Monitoring is performed based on the aggregate of all TCP or UDP applications that are mapped to one or more service policies.

Three types of monitoring are provided for measuring application performance:

TCP round-trip time
The current TCP round-trip time of applications are compared to the threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

TCP packets retransmit ratio
The current TCP packets retransmit ratios of applications are compared to

the threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

Average accept queue delay
The current average accept queue delay of applications are compared to threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

For more details about how to make the various monitoring calculations, refer to the NETWORK-SLAPM2-MIB in the sample file slapm2.mi2 in the /usr/lpp/tcpip/samples directory.

When SNMP traps are enabled, and a *not achieved* trap is sent as described above, a corresponding *okay* trap is sent when the traffic once again conforms to the boundaries established in the monitor table entry.

For example, suppose the slapm2PRMonTcpRttDelayHigh value is set to 2 seconds and the slapm2PRMonTcpRttDelayLow value is set to 1 second. If the TCP round-trip delay rises above 2 seconds, a *not achieved* trap is sent. If the TCP round-trip delay then drops below 1 second, an *okay* trap is sent to indicate the problem has been resolved. However, if the row becomes inactive before conforming to the established boundaries, an *okay* trap is never sent, since this removes monitoring for this entry.

The following traps are used to monitor table administration:

Policy deleted
A slapm2PolicyRuleDeleted trap is sent when an entry is deleted from the slapm2PolicyRuleStatsTable, if slapm2PolicyDeletedTrapEnabled is enabled(1).

Monitor deleted
A slapm2PolicyRuleMonDeleted trap is sent when a slapm2PRMonEntry is deleted, if the value of slapm2PolicyDeletedTrapEnabled is enabled(1).

Creating monitor table entries and enabling SNMP traps: Several MIB objects are used when establishing monitor table entries and when configuring whether and how often traps are sent. To establish monitor table entries, set the following MIB object variables. Most of these objects have default values, so you might be able to achieve the desired monitoring using only a subset of the objects.

slapm2PRMonTcpRttDelayHigh, slapm2PRMonTcpRttDelayLow
Establishes the threshold values for the average TCP round-trip time. The minimum and maximum rates are in units of milliseconds.

slapm2PRMonTcpReXmitHigh, slapm2PRMonReXmitLow
Establishes the threshold values for the TCP packets retransmit ratio. The minimum and maximum rates are in tenths of a percent units.

slapm2PRMonTcpAcceptQDelayHigh, slapm2PRMonTcpAcceptQDelayLow
Establishes the threshold values for the average accept queue delay. The minimum and maximum rates are in units of milliseconds.

slapm2PRMonRowStatus
This object allows entries to be created and deleted in the slapm2PRMonTable.

In addition, the following MIB objects are used to control the generation of traps:

slapm2PRMonTrapEnable
Indicates whether slapm2PolicyRuleMonNotOkay and slapm2PolicyRuleMonOkay notifications should be generated for this conceptual row.

slapm2PRMonTrapFilter
The purpose of this object is to suppress excessive slapm2PolicyRuleMonNotOkay notifications. A monitored quantity must exceed its high threshold for the number of consecutive intervals indicated by this object for a notification to be generated. The length of the intervals is specified by the slapm2PolicyMonInterval object.

Creating the monitor table index: When you create monitor table entries, specify the appropriate index value. The index is composed of the following:

- slapm2PRMonOwnerIndex
- slapm2PolicyRuleIndex

The OwnerIndex is expressed in the following format, where *character* is in ASCII decimal form:

```
length.character.character...
```

For example, the value *u1* is expressed as 2.117.31. The PolicyRuleIndex that maps to the policy name value is the index into the slapm2PolicyRuleTable.

The Network SLAPM2 subagent creates an entry in the slapm2PolicyRuleTable to represent a policy rule. The index value for this entry is arbitrary and assigned by the subagent. Corresponding entries in the other MIB tables, including the monitor table, contain the index value that maps to the entry in the name table.

To assist you in creating the index for the monitor table entries, note that the index value used in the slapm2PolicyRuleStatsTable entries consist of the last value used in the monitor table index, namely the PolicyRuleIndex. Thus, you can walk through the policy statistics table using the following command:

```
osnmp -v walk slapm2PolicyRuleStatsTable
```

Then, cut and paste the index value from the PolicyRuleStatsTable and add an OwnerIndex of your choosing at the beginning of the index.

For the above example, the complete index using an OwnerIndex of *u1* is:

```
2.117.31.3
 |       +--- name table index value (PolicyRuleIndex)
 +----------- length + "u1" (OwnerIndex)
```

Monitor table examples: If you are going to change any of the monitor table object values for an existing table entry or row, you must take the row out of service to make the changes. To do this, set the value of slapm2PRMonRowStatus to 2. After your changes are made, set the row status to a value of 1 to put it back in service.

The following examples show how to create monitor table entries to monitor.

- This example assumes SNMP version 1 security and no SNMPD.CONF file.
  1. Enable traps. The snmptrap.dest file should contain the IP address and protocol of an entity to receive traps:

     ```
     /etc/snmptrap.dest   contains: 9.67.191.5        UDP
     /etc/pw.src          contains: public 0.0.0.0    0.0.0.0
     ```

In this example, use the osnmp command running in the background to receive traps:

```
osnmp trap > /tmp/trap.output &
```

2. Change status to notInService:

   ```
   osnmp set slapm2PRMonRowStatus.index 2
   ```

3. Enable monitoring for slapm2PolicyRuleMonNotOkay and slapm2PolicyRuleMonOkay (traps):

   ```
   osnmp set slapm2PRMonTrapEnable.index 1
   ```

4. If desired, change default thresholds:
   - TCP round-trip, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonTcpRttDelayLow.index l
     osnmp set slapm2PRMonTcpRttDelayHigh.index h
     ```

   - TCP retransmit ratio, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonTcpReXmitDelayLow.index l
     osnmp set slapm2PRMonTcpReXmitDelayHigh.index h
     ```

   - Accept Queue delay ratio, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonAcceptQDelayLow.index l
     osnmp set slapm2PRMonAcceptQDelayHigh.index h
     ```

5. Make row active:

   ```
   osnmp set slapm2PRMonRowStatus.index 1
   ```

- Evaluate the following fields to determine why the slapm2PolicyRuleMonNotOkay trap was generated:
  - If the maxTcpRttDelayExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonTcpRTTCurrentDelay to determine the average round-trip time over the most recent interval for all outgoing TCP packets affected by this policy rule.
  - If the maxTcpReXmitRatioExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonTcpCurrentTcpReXmit to determine the TCP retransmit ratio over the most recent interval for all outgoing TCP packets affected by this policy rule.
  - If the maxAcceptQueueDelayExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonAcceptQCurrentDelay to determine the smoothed average accept queue delay over the most recent interval for all flows affected by this policy rule.
- Evaluate the following fields to determine why the slapm2PolicyRuleMonOkay trap was generated:

- If the maxTcpRttDelayExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonTcpRTTCurrentDelay to determine the average round-trip time over the most recent interval for all outgoing TCP packets affected by this policy rule.
- If the maxTcpReXmitRatioExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonTcpCurrentTcpReXmit to determine the TCP retransmit ratio over the most recent interval for all outgoing TCP packets affected by this policy rule.
- If the maxAcceptQueueDelayExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonAcceptQCurrentDelay to determine the smoothed average accept queue delay over the most recent interval for all flows affected by this policy rule.

SLA subagent performance monitoring

The SLA subagent provides information about service policies and performance data for applications mapped to those policies via two tables.

Note: The SLA subagent can be used to monitor Differentiated Services policies, and RSVP reservations if a corresponding RSVP policy is defined.

slapmPolicyRuleStatsTable
Provides information about defined service policies and aggregate performance data for mapped applications.

slapmSubcomponentTable
Provides information about individual TCP or UDP applications and application-specific performance data.

The SLA Subagent also supports performance monitoring via the slapmPRMonTable object. Entries are created in the monitor table to establish the desired criteria for monitoring. Two levels of monitoring are provided:

Aggregate
Monitoring is performed based on the aggregate of all TCP or UDP applications that are mapped to one or more service policies.

Subcomponent
Monitoring is performed based on a single TCP or UDP application.

Three types of monitoring are provided for measuring application performance:

MinRate
The current input/output rates of the applications are compared to threshold values established in the monitor table entry. If the current rates are less than the threshold, an SNMP trap is sent if traps are enabled.

MaxRate
The current input/output rates of the applications are compared to threshold values established in the monitor table entry. If the current rates are greater than the threshold, an SNMP trap is sent if traps are enabled.

MaxDelay
The current delay rates of the applications are calculated by using TCP round trip time (RTT). For aggregate monitoring, the RTT of all TCP applications are averaged. The delay rates are compared to threshold values established in the monitor table entry. If the current rates are greater than the threshold, an SNMP trap is sent if traps are enabled.

Note: MaxDelay monitoring is only available for TCP applications.

Refer to RFC 2758 for the SLAPM MIB for more details about how to make the various monitoring calculations.

When SNMP traps are enabled, and a *not achieved* trap is sent as described above, a corresponding *okay* trap is sent when the traffic once again conforms to the boundaries established in the monitor table entry.

For example, you can establish MaxDelay monitoring and use a max delay low value of 50 and a max delay high value of 75. If the RTT of the applications rises above 75, a *not achieved* trap is sent. If the RTT then drops below 50, an *okay* trap is sent to indicate the problem has been resolved. However, if the applications end before conforming to the established boundaries, an okay trap may not be sent naturally. For application level traps, the okay trap is never sent because the corresponding subcomponent table entry is deleted when the applications end, which also removes the application level monitoring.

For aggregate traps, when the applications end, MaxRate and MaxDelay okay traps will be sent, because a value of 0 for each of these should fall below the minimum values established in the monitor table entry. Conversely, for MinRate monitoring, an aggregate okay trap is not sent, because a value of 0 will never be greater than the maximum value established in the monitor table.

In addition to the traps used to measure application performance, two additional traps are used to monitor table administration:

Policy Deleted
Trap is sent when an entry is deleted from the slapmPolicyRuleStatsTable.

Monitor Deleted
Trap is sent when an entry is deleted from the slapPRMonTable.

Creating monitor table entries and enabling SNMP traps: Several MIB objects are used when establishing monitor table entries and for configuring whether and how often traps are sent. First, to establish monitor table entries, set the following MIB object variables.

Note: Because most of these objects have default values, you might be able to achieve the desired monitoring using only a subset of the objects.

slapmPRMonControl	Controls what levels and types of monitoring are in effect.
slapmPRMonInterval	Sets the interval for calculating input/output and delay rates and checks those values against the monitor table thresholds.

slapmPRMonMinRateLow, slapmPRMonMinRateHigh, slapmPRMonMaxRateLow, slapmPRMonMaxRateHigh, slapmPRMonMaxDelayLow, slapmPRMonMaxDelayHigh
: Establishes the threshold values for MinRate, MaxRate, and MaxDelay monitoring. The min and max rates are in units of kilobits per second, and the max delay is in units of milliseconds.

slapmPRMonRowStatus
: Controls the status of a monitor table entry, for instance whether or not the entry is active.

In addition, the following MIB objects are used to control the generation of traps:

slapmPolicyTrapEnable
: Enables or suppresses generation of Policy Deleted and Monitor Deleted traps.

slapmPolicyTrapFilter
: Establishes the number of times a given MinRate, MaxRate, or MaxDelay event must be encountered before a trap is generated.

slapmPolicyPurgeTime
: Establishes a timeout value for Policy Deleted traps. After a service policy is deleted, the amount of time indicated by this object must expire before a Policy Deleted trap is generated.

slapmPRMonControl
: Controls whether or not aggregate and/or subcomponent traps are enabled.

Creating the monitor table index: When you create monitor table entries, specify the appropriate index value. The index is composed of the following

- slapmPRMonOwnerIndex
- slapmPRMonSystemAddress
- slapmPRMonIndex

The OwnerIndex is expressed in the following format:

```
length.character.character...
```

where *character* is in ASCII decimal form.

For example, the value *u1* is expressed as 2.117.31. The SystemAddress value will always be 0. The Index value is the index into the slapmPolicyNameTable that maps to the policy name.

For example, assume the following service policy is defined:

```
PolicyAction                ElaineCat
     {
          PolicyScope             RSVP
          MaxRatePerFlow          640
          MaxTokenBucketPerFlow   440
          Maxdelay                1
     }
     PolicyRule                   ElainePol
     {
          ProtocolNumberRange     6
          SourcePortRange         8000
          PolicyActionReference   ElaineCat
     }
```

The SLA Subagent will create an entry in the slapmPolicyNameTable to represent the policy rule. The index value for this entry is arbitrary and assigned by the subagent. Corresponding entries in the other MIB tables, including the monitor table, contain the index value that maps to the entry in the name table.

To assist you in creating the index for the monitor table entries, note that the index value used in the slapmPolicyRuleStatsTable entries consists of the last two values used in the monitor table index, namely the SystemAddress and Index. Thus, you can walk through the policy statistics table using the following command:

```
osnmp -v walk slapmPolicyRuleStatsTable
```

Then, cut and paste the index value from the PolicyRuleStatsTable and add an OwnerIndex of your choosing at the beginning of the index.

For the above example, the complete index using an OwnerIndex of ″u1″ is:

```
2.117.31.0.3
|        | +--- name table index value (Index)
|        +----- no SystemAddress (SystemAddress)
+-------------- length + "u1" (OwnerIndex)
```

Monitor table examples:

Note: If you are going to change any of the monitor table object values for an existing table entry or row, you must take the row out of service to make the changes. To do this, set the value of slapmPRMonRowStatus to 2.

After your changes are made, set the row status to a value of 1 to put it back in service.

Two of the monitor table objects, monitor control and monitor status fields, are important in setting up the entries and understanding why traps are generated. Both of these fields have the SNMP data type BITS, which means they are bit strings, where bit 0 is the low order bit. Any combination of bits can be set into these objects.

Table 32 shows the meaning of the various bits:

Table 32. Monitor control and monitor status object bit values

slapmPRMonControl Object	**xx54 3210**
0 - monitor MinRate	0000 0001 = 01
1 - monitor MaxRate	0000 0010 = 02
2 - monitor MaxDelay	0000 0100 = 04
3 - enable aggregate traps	0000 1000 = 08
4 - enable subcomponent traps	0001 0000 = 10
5 - monitor subcomponent	0010 0000 = 20
slapmPRMonStatus Object	xx98 7654 3210
0 - slaMinInRateNotAchieved	0000 0000 0001 = 001
1 - slaMaxInRateExceeded	0000 0000 0010 = 002
2 - slaMaxDelayExceeded	0000 0000 0100 = 004
3 - slaMinOutRateNotAchieved	0000 0000 1000 = 008
4 - slaMaxOutRateExceeded	0000 0001 0000 = 010
5 - monitorMinInRateNotAchieved	0000 0010 0000 = 020

Table 32. Monitor control and monitor status object bit values (continued)

6 - monitorMaxInRateExceeded	0000 0100 0000 = 040
7 - monitorMaxDelayExceeded	0000 1000 0000 = 080
8 - monitorMinOutRateNotAchieved	0001 0000 0000 = 100
9 - monitorMaxOutRateExceeded	0010 0000 0000 = 200

The following examples show how to create monitor table entries to monitor at various levels and types. You can also create other combinations using the monitor control object.

This example assumes SNMP version 1 security and no SNMPD.CONF file.

First, enable traps, assuming Version 1 security and no SNMPD.CONF file. The snmptrap.dest file should contain the IP address and protocol of an entity to receive traps. In this example, use the osnmp command running in the background to receive traps.

```
/etc/snmptrap.dest   contains: 9.67.191.5    UDP
  /etc/pw.src          contains: public 0.0.0.0    0.0.0.0

  osnmp set slapmPolicyTrapEnable.0 1
  osnmp set slapmPolicyTrapFilter.0 1
  osnmp set slapmPolicyPurgeTime.0 60
  osnmp trap > /tmp/trap.output &
```

To monitor for MinRate at an aggregate level:

```
osnmp set slapmPRMonControl.index \'00000009\'h
           where  9 is aggregate monitor and trap for MinRate
                  1 is aggregate monitor for MinRate

  osnmp set slapmPRMonMinRateLow.index  l
  osnmp set slapmPRMonMinRateHigh.index h
           where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MinRate occurs, then look at the monitor status object in the trap, or:

```
  osnmp get slapmPRMonStatus.index
           should be 1 if inbound MinRate not achieved
                     8 if outbound MinRate not achieved
```

To monitor for MaxRate at an aggregate level:

```
osnmp set slapmPRMonControl.index \'0000000a\'h
           where  a is aggregate monitor and trap for MaxRate
                  2 is aggregate monitor for MaxRate

  osnmp set slapmPRMonMaxRateLow.index  l
  osnmp set slapmPRMonMaxRateHigh.index h
           where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MaxRate occurs, then look at the monitor status object in the trap, or:

```
osnmp get slapmPRMonStatus.index
           should be 2 if inbound MaxRate not achieved
                    10 if outbound MaxRate not achieved
```

To monitor for MaxDelay at an aggregate level:

```
osnmp set slapmPRMonControl.index \'0000000c\'h
          where  c is aggregate monitor and trap for MaxDelay
                 4 is aggregate monitor for MaxDelay

  osnmp set slapmPRMonMaxDelayLow.index  l
  osnmp set slapmPRMonMaxDelayHigh.index h
          where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MaxDelay occurs, then look at the monitor status object in the trap, or:

```
osnmp get slapmPRMonStatus.index
          should be 4 if MaxDelay exceeded
```

To monitor for MinRate at an application level:

```
osnmp set slapmPRMonControl.index \'00000031\'h
          where 31 is subcomponent monitor and trap for MinRate
                21 is subcomponent monitor for MinRate

  osnmp set slapmPRMonMinRateLow.index  l
  osnmp set slapmPRMonMinRateHigh.index h
          where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MinRate occurs, then look at the monitor status object in the trap, or:

```
osnmp get slapmPRMonStatus.index
          should be 20 if inbound MinRate not achieved
                   100 if outbound MinRate not achieved
```

To monitor for MaxRate at an application level:

```
osnmp set slapmPRMonControl.index \'00000032\'h
          where 32 is subcomponent monitor and trap for MaxRate
                22 is subcomponent monitor for MaxRate

  osnmp set slapmPRMonMaxRateLow.index  l
  osnmp set slapmPRMonMaxRateHigh.index h
          where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MaxRate occurs, then look at the monitor status object in the trap, or:

```
osnmp get slapmPRMonStatus.index
          should be 40 if inbound MinRate not achieved
                   200 if outbound MinRate not achieved
```

To monitor for MaxDelay at an application level:

```
osnmp set slapmPRMonControl.index \'00000034\'h
          where 34 is subcomponent monitor and trap for MaxDelay
                24 is subcomponent monitor for MaxDelay

  osnmp set slapmPRMonMaxDelayLow.index  l
  osnmp set slapmPRMonMaxDelayHigh.index h
          where l is the lower boundary and h is the upper boundary

  osnmp set slapmPRMonRowStatus.index 1
```

If the MaxDelay occurs, then look at the monitor status object in the trap, or:

```
osnmp get slapmPRMonStatus.index
          should be 80 if MaxDelay exceeded
```

Chapter 16. Intrusion Detection Services (IDS)

It is becoming increasingly important to not just protect systems from attacks but to detect patterns of usage that might indicate impending attacks. Many attacks follow a sequence of information gathering, unauthorized access to resources (information, applications, storage) and denial of service. It can be difficult, or at times, impossible to determine the originator of denial of service attacks. Correlating information gathering activities with access violation may help identify an intruder before they succeed.

Intrusion Detection Services provides support for:

- Scan detection and reporting
- Attack detection, reporting and prevention
- Traffic regulation for TCP connections and UDP receive queues

Each of these is described in detail.

IDS policies are used to specify what events are to be detected under what circumstances and what action to take. All IDS policies support logging events to a specified message priority level in syslogd and/or the system console. Most IDS policies support discarding packets when a specified limit is reached. Most IDS policies support writing statistics records to the INFO message level of Syslogd on a specified time interval, optionally only if exceptional events have occurred. All IDS policies support tracing all or part of the triggering packet to an IDS specific CTRACE facility, SYSTCPIS. IDS assigns a correlator value to each event. Messages written to the system console and syslogd and records written to the IDS trace all use this correlator. A single detected event may involve multiple packets. The correlator value identifies which messages and packets are related to each other. Each IDS policy has additional attributes that are specified either in conditions or in the action.

Scan policies

Scans are recognized as the result of multiple information gathering events from a single source IP within a defined period of time. Scanning in and of itself is not harmful. However, many serious attacks, especially access violation attacks, are preceded by information gathering scans. Because scans by their nature must use reliable source IP addresses, they can be interesting events to monitor.

The IDS support defines a scanner as a source host that accesses multiple unique resources (ports or interfaces) over a specified period of time. The number of unique resources (Threshold) and the time period (Interval) can be specified via policy. Two categories of scans are supported:

- Fast scan
  - many resources rapidly accessed in a short time period (usually less than 5 minutes and program driven)
- Slow scan
  - different resources intermittently accessed over a longer period of time (many hours). This could be a scanner trying to avoid detection.

Sample scanners:

- Source host A has a program that loops through all low ports and tries to connect to each port on target host X (fast scan). Note: each port is considered a unique resource.
- Source host B manually does pings to each interface on target host X and then tries to access well-known ports on target host X (most likely a slow scan) . Note: each interface accessed by the ping is considered a unique resource and each port accessed is considered a unique resource.

Not a scanner:

- Source host C starts 20 connections to port 23 . Since these connections are to the same port, only one unique resource has been accessed. Therefore, host C is not considered a scanner.

Certain scans may not be detected by IDS:

- source host E issues pings to addresses 9.1.1.1 through 9.255.255.255. Since host X only collects data for the pings directed to X's interfaces, this is not detected by host X as a scan. Network IDS may detect this as wide scan.

Scan policy provides the ability to:

- Control the parameters that define a scan:
  - Fast scan time interval
  - Slow scan time interval
  - Fast scan threshold
  - Slow scan threshold
  - Exclude well-known legitimate scanners via an exclusion list
  - Specify a sensitivity level by port or portrange (to reduce performance impacts)
  - Notify the installation of a detected scan via console message or syslogd message
  - Trace potential scan packets

The individual packets used in a scan can be categorized as normal, possibly suspicious or very suspicious. To control the performance impact and analysis load of scan monitoring, it will be useful to have a mechanism for adjusting our interest level in potential scan events. For information gathering we will provide sensitivity levels of High, Medium and Low to control recognizing countable events for normal, possibly suspicious and very suspicious packets.

The following table shows how the policy-specified sensitivity affects the counting of scan events. The event suspicion level is determined by the stack.

Sensitivity (from policy)	Normal event	Possibly suspicious event	Very suspicious event
Low			count
Medium		count	count
High	count	count	count

To help reduce or eliminate false positives, IDS will allow policy-specified source IP addresses, subnet masks, and (optionally) source port numbers to be excluded

from scan detection. For UDP and TCP port scans, scan detection can be limited to specified destination port ranges. The sensitivity (high, medium and low) may be specified by these port ranges.

Another way IDS will reduce false positives is by counting only unique events from a specific source IP address within a scan interval. An event is considered unique if the IP Protocol, Destination IP Address and Destination Port (UDP, TCP) or Type (ICMP) have not been seen before within this scan interval.

IDS scan policy supports a fast scan interval and threshold and a slow scan interval and threshold. A fast scan will be recognized if more than the fast scan threshold-specified unique events are received. A slow scan will be recognized if more than the slow scan threshold-specified unique events are received. Counting of scan events will be done on an internal interval no greater than half of the fast scan interval to avoid missing scans that occur within the fast scan interval but spread across two reporting intervals. Within an internal interval, once the number of unique events reaches the slow scan threshold, IDS knows that a scan has been detected and it is not necessary to continue to save information about additional related events in storage. This saves both storage and processing overhead. These events, however, will be traced if requested by policy via the ibm-idsTraceData attribute in the action.

Note: When system resources are constrained, scan detection can be temporarily suspended.

Scan events come from the categories listed below. Any countable scan event will count against an origin source IP address. The total number of countable events from all categories is compared to the policy thresholds. When an origin source IP address has exceeded the policy-defined fast or slow threshold an event may be sent to the TRMD for logging to SYSLOG. Additionally, a console message may be issued and the packet may be logged to the IDS packet trace depending on the notification options in the action. When an origin source IP address has exceeded the policy-defined fast or slow threshold an event will be sent to the TRMD for logging to SYSLOG or console. Once a scan event is logged for a particular source IP address, no further scan events will be reportable within the specified fast interval. The intervals and thresholds for fast and slow scan are global, that is, only one definition of them is allowed across all event categories.

- ICMP Scans

  ICMP requests (Echo, Information, Timestamp, Subnet Mask) are used to map network topology. Any request sent to a subnet base or broadcast address will be treated as a very suspicious event. Echo Requests (ping) and Timestamp Requests are very common and will be treated as normal events when they do not include the IP Options for Record Packet Route or Record Timestamp. These options are intended to be used only with ICMP Echo Request packets. The stack ignores them on any other type of packet. The other types of requests are uncommon and will be treated as possibly suspicious events.

Request type	Destination address	Event classification
any	subnet base or broadcast	very suspicious
Information or Subnet Mask	single host	possibly suspicious
Echo with IP Option Record Route or Record Timestamp	single host	possibly suspicious
Echo or Timestamp	single host	normal

- UDP Port Scans

  Because UDP is stateless, the stack is unable to differentiate between a client port and a server port. A scanner sending messages to many ephemeral ports looks very similar to a DNS server sending replies to many clients on ephemeral ports. TCP/IP configuration allows UDP ports to be RESERVED, therefore restricting a port so that it cannot be used. Any datagram received for a restricted port will be treated as a highly suspicious event. Datagrams received for unbound but unrestricted ports will be treated as possibly suspicious events and datagrams received for bound ports will be treated as normal events. Event generation can also be scoped to specific port ranges.

Socket state	Event	Event classification
Restricted (RESERVED to no one)	recv any packet	very suspicious
Unbound, not restricted	recv any packet	possibly suspicious scanner or application temporarily down
Bound	recv any packet	normal

- TCP Port Scans

  Because TCP is a stateful protocol, there are many different events that may be classified as normal, possibly suspicious or highly suspicious. The identified conditions are listed in the table that follows. TCP/IP configuration allow TCP ports to be RESERVED, therefore restricting a port so that it cannot be used. Event generation can also be scoped to specific port ranges.

Socket state	Event	Event classification
Any state	recv unexpected flags (that is, SYN+FIN)	very suspicious
Restricted (RESERVED to no one)	recv any packet	very suspicious
Unbound, not restricted	recv any packet	possibly suspicious scanner or application temporarily down
Listen	recv standalone SYN	no event (classification deferred)
Half open connection	recv ACK	normal - connection handshake completed
Half open connection	recv RST	possibly suspicious peer covering tracks
Half open connection	final time out	very suspicious peer abandoned handshake
Any connected state	seq# out of window	normal perhaps duplicate packet
Any connected state	recv standalone SYN	normal perhaps peer reboot
Any connected state	final timeout	possibly suspicious peer abandoned connection

Attack policies

An attack can be a single packet designed to crash or hang a system. An attack can also consist of multiple packets designed to consume a limited resource causing a network, system or application to be unavailable to its intended users (that is, denial of service). IDS attack policy allows you to turn on attack detection for one or more categories of attacks independently of each other. In general, the types of actions that can be specified for an attack policy are event logging, statistics gathering, packet tracing and discarding of the attack packets.

Most attack checking is done for inbound packets destined for this stack. The IDS categories of attacks are:

- Malformed packets events

  There are numerous attacks designed to crash a system's protocol stack by providing incorrect or partial header information. These packets are always discarded when received regardless of IDS policy. The source IP address is rarely reliable for these attacks.

  You can use IDS policy to provide notification of malformed packet attacks.
- Inbound fragment restrictions

  Many attacks are the result of fragment overlays in the IP or transport header. This support allows you to protect your system against future attacks by detecting fragmentation within the first 256 bytes of a datagram.

  You can use IDS policy to provide notification of a packet that results from a datagram being fragmented in the first 256 bytes, as well as to discard the packet.
- IP protocol restrictions

  While there are 256 possible valid IP protocols, only a handful are in common usage today. This support allows you to protect your system against future attacks by prohibiting those protocols that you are not actively and intentionally using.

  You can use IDS policy to provide notification of a packet with a restricted IP protocol, as well as to discard the packet.
- IP option restrictions

  As with IP protocols, there are 256 possible IP options, with only a small number currently in common use. This support allows you to prevent misuse of options you are not intentionally using. Note that checking for restricted IP options is performed on all inbound packets, even those forwarded to another system.

  You can use IDS policy to provide notification of a packet with a restricted IP option, as well as to discard the packet.
- UDP perpetual echo

  Some UDP applications unconditionally respond to every datagram received. In some cases, such as Echo, CharGen or TimeOfDay, this is a useful network management or network diagnosis tool. In other cases it may be polite application behavior to send error messages in response to incorrectly formed requests. If a datagram is inserted into the network with one of these applications as the destination and another of these applications spoofed as the source, the two applications will respond to each other continually. Each inserted datagram will result in another perpetual echo conversation between them. This support allows you to define the application ports that exhibit this behavior.

  You can use IDS policy to provide notification of a perpetual echo packet, as well as to discard the packet.

- ICMP redirect restrictions

  ICMP redirect packets can be used to modify your routing tables. The IGNOREREDIRECT statement in the TCPIP profile disables ICMP Redirects. You can use IDS policy to provide notification of attempts to modify your routing tables in this manner.

  You can also use IDS policy to disable ICMP Redirects. ICMP Redirect packets will be ignored or discarded if either IGNOREREDIRECT is specified in the TCPIP profile or if IDS policy is active for ICMP redirect attacks and the associated policy action requests that the packet be discarded (ibm-idsTypeActions:LIMIT).
- Outbound raw restrictions

  Most network attacks require the ability to craft packets that would not normally be built by a proper protocol stack implementation. This support allows you to detect and prevent many of these crafting attempts so that your system is not used as the source of attacks on other systems. As part of this checking, you can restrict the IP protocols allowed in an outbound RAW packet. It is recommended that you restrict the TCP protocol (6) on the outbound raw rule.

  You can use IDS policy to provide notification of an outbound raw packet that is considered an attack, as well as to discard the packet.
- Flood events

  Two types of floods are currently detected:

  – TCP SYN floods

    A popular denial of service attack is to flood a public server with connection requests from incorrect or nonexistent source IP addresses. The intent is to use up the available slots for connection requests and thereby deny legitimate access from completing. z/OS CS provides protection from this attack regardless of IDS policy.

  – Interface floods

    If a large number of discards are occurring in proportion to the number of inbound packets, a malicious user might be attempting a denial of service attack. If this percentage of discards for an interface exceeds a specified percentage, this is considered an interface flood. The default percentage of discards used is 10%. This default can be overridden by specifying the ibm-idsIfcFloodPercentage action in the ibm-idsFloodAttackActionsAuxClass.

    To prevent the false detection of an interface flood condition when there is a low volume of inbound traffic on an interface, to qualify as a flood, a minimum number of discards must occur in a one minute period. The default minimum is 1000 discards per minute. This default can be overridden by the ibm-idsIfcFloodMinDiscard action in the ibm-idsFloodAttackActionsAuxClass.

    Once an interface flood condition is reported for an interface, the discard rate for the interface is evaluated for each subsequent one minute interval. An interface flood condition end is reported when the number of discards for the one minute interval falls below ibm-idsIfcFloodMinDiscard or the discard percentage falls below 50% of the ibm-idsIfcFloodPercentage.

    If the interface flood continues for more than 5 minutes, an interface flood continues record is logged at 5 minute intervals while the interface flood conditions exist, if logging was requested by policy (ibm-idsTypeActions specifies LOG and ibm-idsNotification specifies SYSLOG or SYSLOGDETAIL). This log data contains additional information about the discarded packets for the interface.

    Because it can be difficult to distinguish between a malicious user trying to flood a system, unusual spikes in traffic, and problems that could be caused

by setup problems, it is possible for an interface flood condition to be reported when the source of the problem is not actually a flood. For example, if enough storage was not configured to handle the inbound traffic, a large percentage of the inbound packets could be discarded and cause the ibm-idsIfcFloodPercentage to be exceeded.

You can use IDS policy to provide notification of an attack so that you may address the situation with your network administrators and service providers in a timely manner. Notification of a flood can include flood start and flood end event messages and tracing of the first 100 packets discarded due to the flood.

For each attack category (for example, restricted IP protocol) the single highest priority rule is mapped at policy change.

One or more notification options can be specified in the action to provide the desired documentation of detected attacks.

For IDS attack policy the ibm-idsNotification attribute allows attack events to be logged to syslogd and/or the system console. Note that the console messages provide a subset of the information provided in the syslogd messages. For all attack categories except flood, a single packet triggers an event. To prevent message flooding to the system console, you can specify the maximum number of console messages to be logged per attack category within a 5 minute interval (ibm-idsMaxEventMessage). If you specify logging to the console in your IDS policy it is recommended that you specify a maximum event message; there is no default. To prevent message flooding to syslogd, a maximum of 100 event messages per attack category will be logged to syslogd within a 5 minute interval.

For IDS attack policy the statistics action provides a count of the number of attack events detected during the statistics interval. The count of attacks is kept separately for each category of attack (for example, malformed) and a separate statistics record is generated for each. If you want to turn on statistics for attacks, it is recommended that you specify exception statistics (ibm-idsTypeActions:EXCEPTSTATS). With exception statistics, a statistics record will only be generated for the category of attack if the count of attacks is nonzero. If statistics is requested (ibm-idsTypeActions:STATISTICS) a record will be generated every statistics interval regardless of whether an attack has been detected during that interval or not. An exception to this recommendation is when you want to provide overrides to the interface flood action attributes, ibm-idsIfcFloodPercentage and ibm-idsIfcFloodMinDiscard. In this case, run for a period with statistics (ibm-idsTypeActions:STATISTICS) to collect data to help determine the desired policy action attribute values. Once the policy values are determined, the above recommendation to specify exception statistics applies.

For IDS attack policy the ibm-idsTraceData and ibm-idsTraceRecordSize attributes indicate if packets associated with attack events are to be traced. For all attack categories except flood, a single packet triggers an event and the packet is traced. To prevent trace flooding, a maximum of 100 attack packets per attack category will be traced within a 5 minute interval. For the flood category, the first 100 hundred packets discarded during a SYN flood will be traced. In the case of an interface flood, the flood is detected on an interface basis and the trace limit is applied on an interface basis.

For IDS attack policy, specifying ibm-idsTypeActions:LIMIT indicates that packets associated with attack events should be discarded. It is applicable to all attack categories. However, Malformed and flood packets are always discarded regardless of this setting.

An action can be unique to a specific category of attack (for example, malformed) or shared by one or more categories of attacks. If an action is shared, statistics data is still kept separately for each type of attack. Also, the maximum console message limit is enforced individually for each category of attack.

Traffic Regulation (TR) policies

IDS TR policies are used to limit memory resource consumption and queue delay time during peak loads.

TR TCP

IDS TR policies for TCP ports limit the total number of connections an application has active at one time. This can be used to limit the number of address spaces created by forking applications such as FTPD and otelnetd. A *fair share* algorithm is also provided based on the percentage of remaining available connections already held by a source IP address.

The percentage is applied against the number of available connections for the port. Therefore, as fewer connections become available, each host is allowed fewer new connections. The percentage is applied against the number of available connections, rather than the total number of connections allowed, in order to allow access to a larger number of different hosts when resources are low.

When a host requests a connection, the number of connections it currently holds for the port is compared to the percentage applied to the connections currently available for the port. If the number currently held is less than the percentage of currently available connections, the host is allowed to open an additional connection. If equal or greater, the host is not allowed to open further connections until more connections are freed up. All connection requesters for the port are regulated by this mechanism. If a host does not currently have any connections open on the port and unused connections are available, a host will always be allowed at least 1 connection. Multi-user source IP addresses may be allowed a larger number of connections by specifying a QoS policy with a higher number of connections (MaxConnections) than allowed by the TR policy. TR will honor the QoS differentiated services policy if the port is not in a *constrained state*. A QoS exception is made only when QoS differentiated service policy is applied for the specific source server port and specific outbound client destination IP address; if either of these attributes specify a range or are null, the QoS exception will not be made.

TR TCP generates a Constrained Event when a port reaches about 90% of its Connection Limit. An Unconstrained Event is generated when the port falls below about 88% of its limit. An IDS correlator is assigned for the duration of each constrained state. If tracing is requested in the policy, the first 100 packets that exceed the limit in each constrained state are traced along with the correlator. TR TCP also generates events for each connection allowed because of a QoS override policy and for each connection denied for exceeding either the application's connection limit or the percent available limit.

To prevent possible flooding of syslog, TR TCP limits the number of connection refused, would have been refused, or QOS exception log records written in a five minute interval. For a listening port, a maximum of 100 of these log records are written within a five minute interval. Globally, TR TCP writes a maximum of 1000 of these log records within a five minute interval. If a log record was not written due to these limits, the count of refused or would have been refused connection

log records that were not logged is recorded in the `EZZ8660I TRMD TCP connection log records suppressed` log message after the five minute interval ends. Similarly, the count of QOS exception records that were not written is recorded in the `EZZ8661I TRMD TCP QOS exception log records suppressed` log message.

TR UDP

Previously, control over UDP based applications consisted of application priority management and the TCP/IP profile parameter UDPQueueLimit ON | OFF. Inbound datagrams for bound UDP ports are accepted and queued until the queue limit is reached or buffer memory is exhausted. If UDPQueueLimit is set to OFF, any single bound port under a flood attack or with a stalled application could consume all available buffer storage. It is recommended that UDPQueueLimit always be set to ON. This limits the amount of storage that can be consumed by inbound datagrams for any single bound port. Sockets that use the Pascal API, have a limit of 160 KB in any number of datagrams. Sockets that use other APIs, have a limit of 2000 datagrams or 2880 KB.

IDS TR policies for UDP ports specify one of four abstract queue sizes for specified bound IP addresses and ports. The four abstract sizes are VERY_SHORT, SHORT, LONG and VERY_LONG. The actual queue sizes associated with these abstract values are internal values subject to change. Most UDP applications have timeout values based on human perceptions of responsiveness. These values tend to stay constant while system processing speeds and network delivery speeds continue to advance rapidly. This may require the physical sizes of these queues to change over time. The initial implementation uses the values of 16, 256, 2048 and 8192 (2**4, 2**8, 2**11, 2**13) for the number of datagrams and an average datagram size of 2 KB to calculate the byte sizes (32 KB, 512 KB, 4 MB, 16 MB). For performance reasons, sockets that use the Pascal API will only enforce the byte limit. Sockets that use other APIs will enforce both limits. Sockets without a policy specified for their port will use the existing UDPQueueLimit mechanism.

For applications that can process datagrams at a rate faster than the average arrival rate, the queue acts as a speed matching buffer that shifts temporary peak workloads into following valleys. The more that the application processing rate exceeds the average arrival rate and the larger the queue, the greater the variation in arrival rates that can be absorbed without losing work. Very fast applications with very bursty traffic patterns may benefit from LONG or VERY_LONG queue sizes.

For applications that consistently receive datagrams at a higher rate than they are able to process them, the queue acts to limit the effective arrival rate to the processing rate by discarding excess datagrams. In this case the queue size only influences the average wait time of datagrams in the queue and not the percentage of work lost. In fact, if the wait time gets too large, the peer application may have given up or retransmitted the datagram before it is processed. Slow applications with consistently high traffic rates may benefit from SHORT queue sizes.

In general, client side applications will tend to have lower system priority giving them lower datagram processing rates. They also tend to have much lower datagram arrival rates. Giving them SHORT or VERY_SHORT queue sizes may reduce the risk to system buffer storage under random port flood attacks with little impact on percentage of datagrams lost.

TR UDP generates a Constrained Event when a port reaches about 90% of its Queue Limit. An Unconstrained Event is generated when the port falls below

about 88% of its limit. An IDS correlator is assigned for the duration of each constrained state. If tracing is requested in the policy, the first 100 packets that exceed the limit in each constrained state are traced along with the correlator.

Defining TR TCP policies using the Policy Agent

The TCP Traffic Management policy in IBM Communications Server for OS/390 V2R10 was defined as part of the QoS policy with a PolicyScope of TR. It could only be defined in the Policy Agent configuration file; the LDAP server was not supported. In z/OS Communications Server, the TR function was incorporated into the IDS function. The full range of IDS policy is only available in LDAP. For upward compatibility, TR policy that existed previously will continue to be supported in the Policy Agent configuration file and will work unchanged, with the exception that it will be displayed differently by the pasearch command (as an IDS policy instead of a QoS policy with policy scope TR). However, if you want to use any of the expanded Intrusion Detection Services Traffic Regulation support, the expanded functions are only available through LDAP policy. You can use a combination of the LDAP policy for the new functions and keep the TCP traffic regulation policy in the configuration file, although this is not recommended.

Defining IDS policies using LDAP

IDS policies are stored in a server that supports LDAP, processed by Policy Agent and installed into a z/OS Communications Server TCP/IP stack. You should be familiar with the information in Chapter 14, “Policy-based networking,” on page 717 on running Policy Agent and LDAP before creating IDS policies.

A conservative approach to defining IDS policy will avoid unexpected application outages and excessive rule processing. The following examples describe policies provided in the sample files shipped with the system. (Refer to “Policy sample files” on page 721).

eServer IDS Configuration Manager

eServer IDS Configuration Manager enables centralized configuration of IDS policy for z/OS using LDAP as a policy repository. eServer IDS Configuration Manager provides a user-friendly interface with help panels, to isolate network administrators from having to know LDAP policy schema and the complexity of directly writing to an LDAP server. For more information, refer to:

`http://www.ibm.com/software/network/commserver/downloads/`

IDS policy definition considerations

IDS policies can be defined with different ibm-idsConditionType values. Each IDS policy must define exactly one ibm-idsConditionType. Specification of other additional conditions beyond those listed below will cause the rule to not be found. The supported types are:

SCAN_GLOBAL
: This policy is searched by ibm-idsConditionType only. The single highest priority SCAN_GLOBAL rule is mapped at policy change and cached. The action defines the FastScan and SlowScan parameters as well as reporting and tracing actions to take when a scan is detected. The Limit and Statistics actions are ignored.

SCAN_EVENT
: These policies are searched by ibm-idsConditionType and a protocol condition of ICMP, TCP or UDP. For protocols TCP and UDP the policy

search includes local destination port and bound IP address as well. For ICMP, the single highest priority SCAN_EVENT rule is mapped at policy change and cached. The TCP and UDP rules are mapped when a potentially countable event occurs. If the event is associated with a bound socket, the rule is cached. The actions associated with these rules define the sensitivity level to use for counting events towards the scan thresholds and source exclusion list to use for the mapped events. Packet tracing occurs if the action associated with the SCAN_GLOBAL rule activates tracing and the sensitivity in the action associated with the SCAN_EVENT rule sensitivity indicates the event is countable.

ATTACK

There are several Attack Types. The conditions supported on each are defined below. For each attack type, the single highest priority rule is mapped at policy change and cached. The reporting, tracing and statistics actions are supported for all attack types. Other supported actions are defined below. The supported attack types are:

MALFORMED_PACKET

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. The LIMIT action is ignored; malformed packets are always discarded by the stack, with or without policy.

FLOOD

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. The LIMIT action is ignored; excessive half-open TCP connections are always discarded by the stack, with or without policy.

ICMP_REDIRECT

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. ICMP redirect packets are discarded if either this policy specifies action LIMIT or the TCPIP PROFILE specifies IGNOREREDIRECT.

IP_FRAGMENT

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. If this policy specifies action LIMIT, datagrams that are fragmented within the first 256 bytes are discarded.

RESTRICTED_IP_OPTIONS

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. This attack type condition is expected to be ANDed with a list of conditions defining the IP options to disallow. If no IP options are specified, the rule will not accomplish anything. IP option 0 (end of list) and 1 (NO-OP) may not be disallowed and are ignored if specified. If this policy specifies action LIMIT, packets containing a disallowed option are discarded.

RESTRICTED_IP_PROTOCOL

This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. This attack type condition is expected to be ANDed with a list of conditions defining the IP protocols to disallow. If no IP protocols are specified, the rule will not accomplish anything. IP protocols 1 (ICMP), 6 (TCP), and 17 (UDP)

may not be disallowed and are ignored if specified. If this policy specifies action LIMIT, packets containing a disallowed protocol are discarded.

OUTBOUND_RAW
This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. This attack type condition may optionally be ANDed with a list of conditions defining the IP protocols to disallow. If this policy specifies action LIMIT, any packet written to a RAW socket that has a source IP address not in the stack's home list, that is fragmented by the application, that specifies one of the ICMP reply types or that specifies a disallowed protocol are discarded.

PERPETUAL_ECHO
This policy is searched by ibm-idsConditionType and ibm-idsAttackType only. This attack type condition must be specified in a complex rule using CNF and multiple condition levels. The attack type condition is at one of the condition levels. There must be a list of conditions defining the Local Port list at a second level. There must be a list of conditions defining the Remote Port list at a third level. Each of the port lists is limited by the stack to the first 20 ports specified. The negated flag is ignored by the stack on Port list conditions. Destination port is always checked against Local Port list. Source port is checked against the appropriate port list based on whether the source IP address is in the stack's home list. If this policy specifies action LIMIT, UDP packets with both ports in checked port lists are discarded.

TR These policies may optionally be ANDed with any combination of conditions defining protocol (TCP or UDP), local destination port or local destination IP address. TCP rules are mapped when a local application does a listen on a socket or when an inbound connection handshake completes. UDP rules are mapped when an inbound packet arrives at a local bound socket. UDP TR policy supersedes the TCPIP PROFILE setting of UDPQUEUELIMIT for covered ports. Mapped rules are cached and associated with the bound socket.

For TCP, the action defines the total number of allowed connections, the percentage of remaining available connections any single source IP may acquire and whether these limits are applied globally across all applications using this port number or applied individually to each application using an instance of this port number. For UDP, the action defines which of the four available queue sizes is applied to each application using this port number. TR actions define the reporting, statistics and tracing actions for covered ports. If the policy specifies action LIMIT, connections or packets that exceed the limits are discarded.

Notes:

1. For TCP, a total connection limit or percentage available limit of zero, with an action of LIMIT effectively quiesces the application.
2. For TCP, ibm-idsLocalHostIPAddress cannot be specified in any conditions if ibm-idsTRtcpLimitScope PORT specified.
3. For UDP, a policy for a port without an action of LIMIT effectively makes the application unlimited.
4. Each IDS TR action must specify at least one ibm-idsTypeActions.

IDS scan policy example

The goal of scan policy is to detect all scanners with potentially malicious intent while avoiding large numbers of false positives. You can make this process more efficient by reserving all unused low ports in the TCPIP profile. This will allow you to use the low sensitivity setting on scans for these ports. As you investigate the scans detected, you will initially find your own network management tools. These can be explicitly excluded. If you include UDP ephemeral ports in a high sensitivity policy, you will discover that your DNS servers show up as scanners. You can explicitly exclude these as well.

The following scan rules are defined:

- Scan Global

  Defines a global set of parameters for detecting scans, and also defines reporting parameters for scan events.
  - A Fast Scan is defined as 5 unique events in 2 minutes from a single source IP address.
  - A Slow Scan is defined as 10 unique events in 480 minutes (8 hours).
  - The first 200 bytes of the packet associated with each countable event will be traced.
  - When a scan is detected an event will be written to syslog warning level, along with a detailed list of all the unique events included in the scan.
  - No message will be written to the console.
  - Statistics records will not be written to syslog.
- Scan Event Low

  Defines a set of traffic for which low sensitivity scan detection will be performed. Inbound traffic to all TCP and UDP ports between 1 and 1023 will be monitored. It is recommended that unused low ports be RESERVED in the TCPIP Profile.
- Scan Event Medium

  Defines a set of traffic for which medium sensitivity scan detection will be performed. ICMP inbound traffic will be monitored.

```
dn:cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scanglobal-rule
ibm-policyRuleName:ScanGlobal-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS global scan rule

dn:cn=condassoc1, cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsScanConditionAuxClass
cn:condassoc1
ibm-policyConditionName:ScanGlobal-condition
ibm-idsConditionType:SCAN_GLOBAL
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition

dn:cn=actassoc1, cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
```

```
objectclass:ibm-idsScanActionAuxClass
objectclass:ibm-idsNotificationAuxClass
cn:actassoc1
ibm-policyActionName:ScanGlobal-action
ibm-idsActionType:SCAN_GLOBAL
ibm-policyActionOrder:1
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsNotification:SYSLOGDETAIL
ibm-idsLoggingLevel:4
ibm-idsTraceData:RECORDSIZE
ibm-idsTraceRecordSize:200
ibm-idsFSInterval:2
ibm-idsFSThreshold:5
ibm-idsSSInterval:480
ibm-idsSSThreshold:10
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - Fast scan = 5 in 2 minutes, Slow scan = 10 in 8 hours

dn:cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scaneventlow-rule
ibm-policyRuleName:ScanEventLow-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS scan event rule for low sensitivity on TCP and UDP Low Ports

dn:cn=condassoc2, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:ScanEventLow-condition2
ibm-policyConditionDN:cn=ScanTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable CNF condition level 2 - scan TCP low ports

dn:cn=condassoc3, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3
ibm-policyConditionName:ScanEventLow-condition3
ibm-policyConditionDN:cn=ScanUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable CNF condition level 2 - scan UDP low ports

dn:cn=actassoc1, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsScanSensitivityActionAuxClass
objectclass:ibm-idsScanExclusionActionAuxClass
cn:actassoc1
ibm-policyActionName:ScanEventLow-action
ibm-idsActionType:SCAN_EVENT
ibm-policyActionOrder:1
ibm-idsSensitivity:LOW
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - low sensitivity

dn:cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scaneventmedium-rule
ibm-policyRuleName:ScanEventMedium-rule
```

```
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS scan event rule for medium sensitivity on ICMP

dn:cn=condassoc1, cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:condassoc1
ibm-policyConditionName:ScanEventMedium-condition
ibm-idsConditionType:SCAN_EVENT
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-idsProtocolRange:1
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - ICMP protocol

dn:cn=actassoc1, cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsScanSensitivityActionAuxClass
cn:actassoc1
ibm-policyActionName:ScanEventMedium-action
ibm-idsActionType:SCAN_EVENT
ibm-policyActionOrder:1
ibm-idsSensitivity:MEDIUM
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - medium sensitivity

dn:cn=ScanTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:SCAN_EVENT
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:ScanTcpLowPorts
ibm-policyConditionName:ScanTcpLowPorts-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS Scan TCP Low Ports condition

dn:cn=ScanUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:SCAN_EVENT
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:ScabUdpLowPorts
ibm-policyConditionName:ScanUdpLowPorts-condition
ibm-idsProtocolRange:17
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS Scan UDP Low Ports condition
```

IDS attack policy examples

The goal of attack policy is to help protect your system from both known and unknown attacks and to give you timely notification when attacks do occur. Malformed packet policy covers many known attacks designed to cause system *crashes*. These packets are always discarded and rarely have legitimate source

address information. Many malformed packet attacks use fragmentation to overlay header fields. The IDS fragment restriction policy is intended to protect you from unknown attacks of this type by disallowing fragmentation in the first 256 bytes of any datagram. Unless you know you need ICMP redirect, you should disallow it with policy. There are several types of flood attacks. IDS can identify TCP SYN floods. IDS policy should be used to notify you when a flood occurs. You will need to work with your network administrators and service providers to track the flood backwards, one physical hop at a time, to locate the sources.

The IP protocol restrictions and IP option restrictions provide additional protection against future unknown attacks. The philosophy behind them is to disallow anything that you do not have a known reason to allow. The outbound raw policy is intended to help you detect someone using your system as the base for an attack. It looks for several behaviors associated with *spoofed* packets.

Attack rules define the set of conditions that define what constitutes an attack for a given attack type. The highest priority rule of each attack type is used. The action associated with an attack rule defines reporting and logging options for a detected attack.

The following types of attack rules are defined:

- Malformed Packet: Various types of known attacks based on malformed packets.
- Flood: TCP SYN flood and interface flood attacks.
- ICMP Redirect: Disallows ICMP redirect receipts.
- IP Fragment: Disallows Fragmentation within first 256 bytes of datagrams.
- IP Protocol: Defines disallowable IP protocols.
  - Uses complex conditions to disallow everything except ICMP, TCP and UDP.
  - Uses DNF (condition list type 1) to evaluate complex conditions.
  - Conditions in the same group are ANDed together, groups are ORed.
- Outbound Raw Restrictions: Validity checking for outbound packets using RAW sockets.
  - Uses complex conditions to disallow everything except ICMP, UDP, IGMP and OSPFIGP.
- Several reusable Protocol conditions are defined that can be shared between the IP Protocol Restriction rule and the Outbound Raw rule.
- A single reusable attack action is defined and shared among all the attack rules.
  - Events are written to syslog ALERT level.
  - Events are not written to the system console.
  - The first 200 bytes of packets associated with an attack are traced.
  - Statistics are evaluated every 60 minutes and only written if an attack occurred.
  - Limit was not specified, so packets associated with IP Protocol Restrictions, IP Fragment Restriction and Outbound Raw Restrictions will not be deleted.

```
dn:cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackMalformed-rule
ibm-policyRuleName:AttackMalformed-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Malformed Packets
```

```
dn:cn=condassoc1, cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackMalformed-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:MALFORMED_PACKET
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackMalformed-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackFlood-rule
ibm-policyRuleName:AttackFlood-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Floods

dn:cn=condassoc1, cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackFlood-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:FLOOD
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsAttackActionsAuxClass
objectclass:ibm-idsFloodAttackActionsAuxClass
cn:actassoc1
ibm-policyActionName:attackFlood-action
ibm-policyActionOrder:1
ibm-idsActionType:ATTACK
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:1
ibm-idsTypeActions:EXCEPTSTATS
ibm-idsStatInterval:60
ibm-idsTraceData:RECORDSIZE
ibm-idsTraceRecordSize:200
ibm-idsIfcFloodPercentage:10
ibm-idsIfcFloodMinDiscard:1000
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
```

```
description:Rule-specific action - LOG(SYSLOG(1) NOCONSOLE) EXCEPTSTATS(60) TRACE(200)


dn:cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackICMPRedirect-rule
ibm-policyRuleName:AttackICMPRedirect-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for ICMP Redirect

dn:cn=condassoc1, cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackICMPRedirect-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:ICMP_REDIRECT
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackICMPRedirect-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackIpFragment-rule
ibm-policyRuleName:AttackIpFragment-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for IP fragment restriction

dn:cn=condassoc1, cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackIpFragment-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:IP_FRAGMENT
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackIpFragment-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
```

```
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackIpProt-rule
ibm-policyRuleName:AttackIPprot-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for restricted protocol

dn:cn=condassoc1, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:AttackIPprot-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=attackIpProtcond1, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition at level 1

dn:cn=condassoc1a, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1a
ibm-policyConditionName:AttackIPprot-condition1a
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition at level 1 (negated) allow ICMP

dn:cn=condassoc1b, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1b
ibm-policyConditionName:AttackIPprot-condition1b
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtTCP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents third reusable DNF condition at level 1 (negated) allow TCP

dn:cn=condassoc1c, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1c
ibm-policyConditionName:AttackIPprot-condition1c
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fourth reusable DNF condition at level 1 (negated) allow UDP

dn:cn=actassoc1, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:AttackIPprot-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
```

```
cn:attackOutboundRaw-rule
ibm-policyRuleName:AttackOutboundRaw-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Outbound Raw restrictions

dn:cn=condassoc1, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:AttackOutboundRaw-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=attackOutboundRawcond1, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition at level 1

dn:cn=condassoc1a, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1a
ibm-policyConditionName:AttackOutboundRaw-condition1a
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition at level 1 (negated) allow ICMP

dn:cn=condassoc1b, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1b
ibm-policyConditionName:AttackOutboundRaw-condition1b
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents third reusable DNF condition at level 1 (negated) allow UDP

dn:cn=condassoc1c, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1c
ibm-policyConditionName:AttackOutboundRaw-condition1c
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtIGMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fourth reusable DNF condition at level 1 (negated) allow IGMP

dn:cn=condassoc1d, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1d
ibm-policyConditionName:AttackOutboundRaw-condition1d
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtOSPFIGP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fifth reusable DNF condition at level 1 (negated) allow OSPFIGP

dn:cn=actassoc1, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:AttackOutboundRaw-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
```

```
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn: cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsAttackActionsAuxClass
cn:attackact1
ibm-policyActionName:AttackLog-action
ibm-idsActionType:ATTACK
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:1
ibm-idsTypeActions:EXCEPTSTATS
ibm-idsStatInterval:60
ibm-idsTraceData:RECORDSIZE
ibm-idsTraceRecordSize:200
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS common attack action - LOG(SYSLOG(1) NOCONSOLE) NOLIMIT
description:IDS common attack action - EXECPTSTATS(60) TRACE(200)

dn:cn=attackIpProtcond1, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:attackIpProtcond1
ibm-policyConditionName:AttackIPprot-condition1
ibm-idsConditionType:ATTACK
ibm-idsAttackType:RESTRICTED_IP_PROTOCOL
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS attack condition 1 for restricted IP protocol

dn:cn=attackOutboundRawcond1, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:attackOutboundRawcond1
ibm-policyConditionName:AttackOutboundRaw-condition1
ibm-idsConditionType:ATTACK
ibm-idsAttackType:OUTBOUND_RAW
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS attack condition 1 for Outbound Raw restrictions

dn:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtICMP
ibm-policyConditionName:IpProtICMP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:1
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol ICMP

dn:cn=IpProtIGMP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtIGMP
ibm-policyConditionName:IpProtIGMP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
```

```
description:Reusable IDS condition for IP protocol IGMP

dn:cn=IpProtTCP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:6
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol TCP

dn:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtUDP
ibm-policyConditionName:IpProtUDP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:17
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol UDP

dn:cn=IpProtOSPFIGP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtOSPFIGP
ibm-policyConditionName:IpProtOSPFIGP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:89
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol OSPFIGP
```

Traffic Regulation (TR) policy examples

The goal of TR policy is to protect your system from usage spikes. A phased approach to determine the correct policy for your system is recommended.

To gather baseline statistics, an installation will first need to run in Statistics mode, with the traffic regulation daemon (TRMD) running. In statistics mode, the following information is provided for the port on a policy defined interval:

- Total number of connections requested during the interval
- Total number of connections closed during the interval
- The IP address of the host that requested a connection during the interval and held the highest number of concurrent connections during the interval, and the highest number of concurrent connections held by this IP address
- A suggested value for TotalConnections based on this interval
- A suggested value for Percentage based on this interval

While the baseline statistics records provide suggested policy values for the interval, the installation should evaluate data from multiple intervals. The values suggested are those that would avoid denying any of the connections in the interval. Choose lower values if the interval represents a workload larger than you want to allow.

After the installation determines the policy values to use, try running with the Log action specified. Specifying the Log action, without the Limit action, basically tests out the policy. The connections that would have been denied (if the Limit action

was specified) are logged, but the connection is allowed to occur. After the installation is satisfied with the experimental policy, the policy action can be set to Limit.

The following traffic regulation TCP rules are defined:

- TR TCP: Defines TCP baseline STATISTICS gathering for the low port range.
  - This rule provides statistics reports to determine normal traffic patterns for several applications.
- TR TCP WEB: Defines Application limits and Host percentage limits for a single application.
  - This rule enforces set limits.
  - The rule has a higher priority than the TR TCP rule.
  - The rule is limited to a single server application that is bound to a specific IP address.

```
dn:cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trtcp-rule
ibm-policyRuleName:TRtcp-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trtcp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trtcp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRulePriority:2
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR TCP rule

dn:cn=condassoc1, cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRtcp-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR TCP low ports

dn:cn=actassoc1, cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRtcp-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trtcpact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR TCP action 1

dn:cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trtcpWeb-rule
ibm-policyRuleName:trtcpWeb-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trtcpWeb-rule,cn=IDS,cn=advanced,
     ou=policy, o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trtcpWeb-rule,cn=IDS,cn=advanced,
```

```
     ou=policy,o=IBM,c=US
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:7
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR TCP rule with limit

dn:cn=condassoc1, cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRtcpWeb-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrTcpWebPort, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR TCP web port

dn:cn=actassoc1, cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRtcpWeb-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trtcpact2, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR TCP action 2

dn: cn=trtcpact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRtcpActionAuxClass
cn:trtcpact1
ibm-policyActionName:TRtcpLog-action
ibm-idsActionType:TR
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:STATISTICS
ibm-idsStatInterval:60
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR TCP action TCP(64K,100%) LOG(SYSLOG(4) NOCONSOLE) NOLIMIT
description:TRACE(HEADER) STATISTICS(60)

dn:cn=trtcpact2, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRtcpActionAuxClass
cn:trtcpact2
ibm-policyActionName:TRtcpLimit-action
ibm-idsActionType:TR
ibm-idsTypeActions:LIMIT
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:EXCEPTSTATS
ibm-idsStatInterval:60
ibm-idsTRtcpTotalConnections:1000
ibm-idsTRtcpPercentage:10
ibm-policyKeywords:TR
```

```
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR TCP action TCP(1K,10%) LOG(SYSLOG(4) NOCONSOLE) LIMIT
description:TRACE(HEADER) EXCEPTSTATS(60)

dn:cn=TrTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrTcpLowPorts
ibm-policyConditionName:TrTcpLowPorts-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR TCP Low Ports condition

dn:cn=TrTcpWebPort, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsHostConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrTcpWebPort
ibm-policyConditionName:TrTcpWebPort-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:80
ibm-idsLocalHostIPAddress:3-10.14.243.87
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR TCP Web Port condition
```

The following traffic regulation UDP rule is defined:

- TR UDP: Defines UDP queue size for the low port range.
  - This rule provides statistics reports to determine normal traffic patterns for several applications.

```
dn:cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trudp-rule
ibm-policyRuleName:TRudp-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trudp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trudp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRulePriority:2
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR UDP rule

dn:cn=condassoc1, cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRudp-condition1
ibm-policyConditionGroupNumber:7
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
```

```
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR UDP low ports

dn:cn=actassoc1, cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRudp-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trudpact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR UDP action 1

dn: cn=trudpact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRudpActionAuxClass
cn:trudpact1
ibm-policyActionName:TRudpLog-action
ibm-idsActionType:TR
ibm-idsTypeActions:LOG
ibm-idsTypeActions:LIMIT
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:STATISTICS
ibm-idsStatInterval:60
ibm-idsTRudpQueueSize:LONG
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR UDP action UDPQ(LONG) LOG(SYSLOG(4) NOCONSOLE) LIMIT
description:TRACE(HEADER) STATISTICS(60)


dn:cn=TrUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrUdpLowPorts
ibm-policyConditionName:TrUdpLowPorts-condition
ibm-idsProtocolRange:17
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR UDP Low Ports condition
```

Verification

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies active?
- Is the expected traffic mapping to the correct policies?
- Are the IDS Policy functions working correctly?

The following sections provide more details about these considerations.

Are the correct policies active?

Check your LDAP server log or command output for errors encountered when your policies were loaded into LDAP. Some LDAP servers treat consecutive blank lines in an LDIF file as end of file; ensure that all of the policy objects in your LDIF files are acknowledged by LDAP.

Check your Policy Agent log file for errors while processing your policy.

Use the pasearch command to verify that the intended policies are active and have the expected attributes for the target stack.

Is the expected traffic mapping to the correct policies?

Use the netstat -k SUMmary command to ensure that the intended policy has been mapped for each of the Attack types, Scan-Global type and the Scan-Event type for protocol ICMP. Refer to *z/OS Communications Server: IP System Administrator's Commands* for more information on the netstat command.

These IDS functions each select the single highest priority policy for their respective types at each policy change.

Use the netstat -k PROTOcol TCP and netstat -k PROTOcol UDP commands to ensure that the intended Scan-Event and TR policies have been mapped to the intended local sockets.

These IDS functions select the highest priority policy for the Protocol, local Port and local IP address when there is relevant activity against the socket.

For TCP this usually entails either a listen or the completion of an inbound connection handshake. For UDP this usually entails either a bind or an inbound datagram. Scan policies are also selected on some inbound error paths.

Are the IDS policy functions working correctly?

IDS policies that include TypeAction:STATISTICS or TypeAction:LOG and Notification:SYSLOG cause the stack to make log record information available to TRMD. If TRMD is running you may run the IDS report generator TRMDSTAT against the appropriate log files to produce reports on the area of interest.

TRMD

TRMD runs as an APF-authorized program and requires some RACF setup (TRMD must be able to run as a started task and have superuser authority). See the EZARACF member of SEZAINST for sample RACF commands.

The resolver configuration file is used to determine the stack that TRMD will use. Ensure that the RESOLVER_CONFIG environment variable is correctly set before starting TRMD. A separate instance of TRMD must be run for each TCP/IP stack.

The Log records written by TRMD contain 2 timestamps:

- A timestamp generated when the event was detected by the stack. This timestamp is generated by the stack and is always Coordinated Universal Time (UTC).
- A timestamp that is generated when the syslogd record ID is created. This timestamp is dependent on the setting of the TZ environment variable at the

time that TRMD is started. If the installation wants this timestamp to be based on UTC, then ensure the TZ environment variable is properly set (for example, export TZ=0) before starting TRMD.

If running multiple instances of TRMD, consider using the syslogd -u option when starting syslogd. The -u option causes the jobname of the application writing the log record to be included in the log record.

The TCP/IP stack must be running before TRMD can be started.

As described below, TRMD can be started from the z/OS shell or as a started task.

Running TRMD as a started task

A sample procedure is shipped in member EZATRMDP in SEZAINST. Follow the instructions in the sample member to define your environment.

The offset from Coordinated Universal Time (UTC) of the syslog time in the timestamp of TRMD messages is determined by the TZ environment variable. If the timestamp is required in UTC and has not been set by the TZ environment variable, specify the following in the TRMD procedure:

```
//  PARM=('POSIX(ON) ALL31(ON)',
//  'ENVAR("LIBPATH=/usr/lib"',
//  '"TZ=0")/-d 1')
```

To start TRMD as a started task, use the *S TRMD* command from the MVS console or SDSF. TRMD issues a fork, and in some cases the job name will be the original job name with a number appended. For example, *S TRMD* might result in the TRMD started task running under the job name TRMD1, whereas *S TRMDTASK* would result in the TRMD started task running under the job name TRMDTASK. Use the *D A,TRMD** command to verify the job name that TRMD is running under.

If running as a started task, issue *P jobname* to stop TRMD.

To automatically start TRMD when the TCP/IP stack is started, add TRMD to the AUTOLOG statement in the TCP/IP profile as follows:

```
AUTOLOG
    TRMD  JOBNAME TRMD
ENDAUTOLOG
```

Running TRMD from the z/OS UNIX shell

Only a superuser can run TRMD from the z/OS UNIX shell.

Ensure that the RESOLVER_CONFIG environment variable is correctly set before starting trmd.

The offset from Coordinated Universal Time (UTC) of the syslog time in the timestamp of TRMD messages is determined by the TZ environment variable. If the timestamp is to appear in Coordinated Universal Time (UTC), change the TZ specification in /etc/profile or export TZ=″0″ before starting TRMD.

After the proper environment is set up, issue the following to start TRMD:

```
trmd
```

Stopping TRMD

To stop TRMD, issue the following kill command :

```
kill -s TERM pid
```

where pid is the TRMD process ID

To obtain the TRMD process ID, issue the following z/OS UNIX command:

```
ps  -A
```

Debug options can also be specified when starting TRMD. Refer to *z/OS Communications Server: IP Configuration Reference* for more information.

TRMDSTAT

Trmdstat is a utility program that runs from the z/OS UNIX shell. Trmdstat reads a log file, analyzes the log records generated by TRMD, and provides summary or detailed reports based on the options specified.

The following reports can be requested:

- Overall summary of logged connection events
- IDS summary of logged events
- Reports of logged connection events
- Reports of logged intrusions defined in the ATTACK policy
- Reports of logged intrusions defined in the TCP policy
- Reports of logged intrusions defined in the UDP policy
- Reports of statistics events

Refer to *z/OS Communications Server: IP System Administrator's Commands* for the TRMDSTAT command and samples of the reports generated by TRMDSTAT.

Chapter 17. IP security

This chapter contains a description of IP filtering, IPSec-protected traffic, and preparing and configuring a z/OS system for IP security. Various business configurations are explained, including host-to-host, host-to-gateway, gateway-to-gateway, and gateway-to-host.

Terms and concepts

The following terms and concepts are used in this chapter:

3DES Also known as triple DES, this encryption method uses three DES operations on a single data block with three different keys. Provides greater security than single DES.

Active Used in three ways:

- Describes the filter policy that is in effect (default or Policy Agent).
- Describes the state of the rules or actions that are defined in Policy Agent. These rules or actions can be active or inactive due to a time condition.
- Describes the state of a manual tunnel installed in the TCP/IP stack. A manual tunnel can be active (available for use) or inactive (not available for use).

Active IPSec policy
The policy that is in effect, either the default filter policy or the IP security filter policy.

Asymmetric encryption
Also known as public/private key encryption, this type of encryption is performed between two parties using pairs of encryption and decryption keys.

Authentication Header (AH)
An IP protocol (51) used with an IPSec security association to provide authentication of IP packets.

Autoactivation
The process by which a dynamic tunnel is activated when IP security policy is installed into the TCP/IP stack, either as the result of a user action, or the result of TCP/IP or IKED initialization.

Certificate authority
A trusted third party that verifies information that is contained in an X.509 digital certificate.

Command-line activation
The process of activating a tunnel through the use of the **ipsec** command. Both manual and dynamic tunnels can be activated from the z/OS UNIX command line.

Data encryption standard (DES)
A block cipher with 64-bit blocks and a 56-bit key.

Default IP filter policy
Used until the IP security filter policy is installed by Policy Agent. The default IP filter policy includes both the filter rules you define in the

TCP/IP profile and the implicit default filter rules that the stack generates. The implicit default filter rules deny all traffic that does not match any configured filter rule.

Dynamic tunnel
An IPSec tunnel whose security parameters are negotiated and whose encryption keys are generated dynamically using IKE.

Encapsulating Security Payload (ESP)
An IP protocol (50) used with an IPSec security association to provide authentication and encryption of IP packets.

Hashed message authentication code (HMAC)
A one-way hash function that combines the contents of a message and a secret key to produce a hash value; used for authentication.

HMAC_MD5
HMAC using the MD5 algorithm.

HMAC_SHA
HMAC using the SHA algorithm.

IKE negotiation
A process by which two communicating IKE-enabled peers agree on a set of parameters that are used to protect traffic between them. This set of parameters is collectively known as a security association. One peer acts as the initiator of the negotiation, the other as the responder.

IKE tunnel
A tunnel that protects IKE phase 2 messages.

Internet Key Exchange (IKE)
A protocol for the secure generation and management of encryption keys over an existing IP network.

Internet Security Association and Key Management Protocol (ISAKMP)
Defines procedures and packet formats to establish, negotiate, modify, and delete security associations.

IP filter rule
A configured rule that defines the action applied to an IP traffic pattern that is encompassed by the rule. The possible actions include permit, deny, and permit with IPSec protection.

IP filter table
An ordered list of IP filter rules. When IP filtering is active on a host, the table is consulted for each IP packet that is sent or received. The action of the matching IP filter rule is enforced by the TCP/IP stack.

IPSec A suite of protocols and standards defined by the Internet Engineering Task Force (IETF) for secure communication over an existing IP network.

IPSec tunnel
A tunnel that protects IP traffic between two endpoints using one or both of the IPSec protocols. Manual and dynamic tunnels are both instances of an IPSec tunnel.

IP security filter policy
The policy that is installed by the Policy Agent. It includes the filter rules you define in the Policy Agent configuration files and an implicit deny all rule that is generated by Policy Agent.

IP traffic pattern
The set of IP traffic attributes that can be used as input to an IP filter table query. Typically, this includes IP source address, IP destination address, source port, destination port, protocol, and direction (inbound or outbound).

Manual tunnel
An IPSec tunnel whose security parameters and encryption keys are statically configured and must be manually managed by a security administrator.

Message authentication code (MAC)
A tag derived from the contents of a message and a secret key. The tag can be used to authenticate the integrity of a message as well as the source of the message.

Message digest algorithm 5 (MD5)
An algorithm that produces a 128-bit hash.

NAT traversal (NATT)
Traversal of IPSec traffic through a NAT device.

Network address/port translation (NAPT)
A technique where multiple internal IP addresses are translated into a single public IP address. As part of this translation process, the TCP and UDP ports in the packets are translated. NAPT is sometimes referred to as port address translation (PAT).

Network Address Translation (NAT)
A technique where a 1-to-1 address translation function is performed, translating a single internal IP address to a single public IP address.

On-demand
The process by which a dynamic tunnel is activated by outbound traffic flow without user intervention.

Phase 1
The first stage of an IKE negotiation, in which an ISAKMP security association is established between two IKE-enabled peers.

Phase 2
The second stage of an IKE negotiation, in which an IPSec security association is established between two IKE-enabled peers.

Rivest Shamir Adleman (RSA)
An asymmetric key encryption method, in which the key that is used to encrypt data is different than the key that is used to decrypt the data. RSA can be used for encryption, or to authenticate a digital signature.

Secure hash algorithm 1 (SHA1)
A MAC algorithm similar to MD5, but more secure in that it produces a 160-bit hash value.

Security association
An agreement between two IPSec-enabled hosts that describes the type of data to protect and the methods that are used to protect the data. IKE creates a phase 1 security association to protect IKE messages (also known as the ISAKMP security association), and a phase 2 security association to protect data traffic (also known as the IPSec security association).

Symmetric encryption
Encryption that is performed between two parties sharing the same encryption key. Also known as secret key encryption.

Transport mode encapsulation
A process used to construct IPSec packets by inserting one or more additional IPSec headers between the IP header to be protected and the IP payload of the packet to be protected.

Tunnel
A secure logical connection or channel that is defined by a collection of security associations that define the security parameters protecting traffic between two endpoints.

Tunnel activation
The process by which a tunnel becomes active or usable. For dynamic tunnels, this process involves initiating an IKE negotiation.

Tunnel mode encapsulation
A process used to construct IPSec packets by creating a new IP header with an IP payload consisting of the entire IP packet being protected, and then inserting one or more additional IPSec headers between the new IP header and its IP payload (that is, the original IP packet).

UDP encapsulation
A process used to construct IPSec packets by first applying tunnel mode encapsulation or transport mode encapsulation to an IP packet to be protected by the ESP protocol, and then inserting a UDP header between the IP header and the ESP header.

Virtual private network (VPN)
A logical network of connected network nodes that communicate through secure channels (tunnels), typically by using the IPSec protocols (AH and ESP).

X.500 distinguished name
A collection of X.509 values, such as common name, host name, organization, organizational unit, and so on, that is stored in an X.509 digital certificate. An X.500 distinguished name is used as a globally unique identifier for the owner.

X.509 digital certificate
A set of information in the X.509 standard containing various attributes about an entity, including identity information and a public key that is used for encrypted communications with that entity.

Terminology conventions

The following terminology conventions are used throughout this chapter when referring to z/OS IP security:

IP security
The z/OS Communications Server function.

IPSec The protocol suite.

ipsec The action associated with an IP filter action, or the z/OS UNIX System Services command.

IPSEC The statement in the TCP/IP profile.

IPSECURITY
The parameter on the IPCONFIG statement in the TCP/IP profile.

Commands used to administer IP security

The following commands are used to administer IP security. For more information on these commands, refer to *z/OS Communications Server: IP System Administrator's Commands*.

ipsec Use the z/OS UNIX System Services **ipsec** command to display information about active filters and security associations, and to control aspects of security association negotiation. The **ipsec** command is used to:

- Display filters that are active in the stack
- Revert to default IP filter policy, as defined in the TCP/IP profile
- Reload IP security policy, as defined in the Policy Agent configuration files
- Activate security association negotiations
- Display existing phase 1 security associations
- Display existing phase 2 security associations
- Display remote port mappings used with NAT traversal configurations
- Refresh existing phase 1 security associations
- Refresh existing phase 2 security associations
- Deactivate existing phase 1 security associations
- Deactivate existing phase 2 security associations
- Test for a filter rule match for a given set of IP traffic characteristics

Authority to use the **ipsec** command is controlled through RACF. There are two distinct SERVAUTH profiles that define access to the **ipsec** command, one for display capabilities and one for control capabilities.

Tip: Many of the tasks, examples, and references in this document assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

For the steps to configure access control to the **ipsec** command, see Appendix E, "Steps for preparing to run IP security," on page 1215.

For detailed syntax and usage, and how to control access of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

pasearch
Use the **pasearch** command to display Policy Agent information that is defined in the Policy Agent configuration files, including IP security and other types of policies. The options that are related to IP security include the ability to view IP security policy rules and actions, both active and inactive, for any TCP/IP stack for which policies have been defined and that is IPSECURITY-enabled.

If the user of the **pasearch** command is not a superuser, authority is controlled through RACF.

For detailed syntax and usage of the **pasearch** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

MODIFY

Use the MODIFY console command to have:

- The IKE daemon reread the IKED configuration file
- Policy Agent reread the Policy Agent configuration files

For detailed syntax and usage of the MODIFY command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Netstat

Use the Netstat command to display the following:

- IPSECURITY enablement for a particular stack (Netstat CONFIG/-f)
- SecurityClass (SECCLASS) for a specific interface (Netstat DEVLINKS/-d)

For detailed syntax and usage of the Netstat command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Overview of using IP security

z/OS Communications Server provides the ability to control and monitor network traffic on one or more TCP/IP stacks on a z/OS system. With z/OS Communications Server IP security, the Communications Server provides the necessary function to support IP filtering, IPSec, and Internet Key Exchange (IKE), without requiring Integrated Security Services Firewall Technologies. Key advantages of IP security are easier configuration, greater scalability, improved performance, and enhanced serviceability over the capabilities that are provided by Firewall Technologies. Functionally, IP security provides support for NAT traversal in common server configurations, while Firewall Technologies does not. This function is critical in many environments where private addressing is used in a portion of the network's topology and IPSec protection is also required. For more information regarding the use and configuration of Firewall Technologies, refer to *z/OS Integrated Security Services Firewall Technologies*.

IP security policy can be used for the following:

- Protect a secure host on an internal network from unwanted network traffic
- Provide protection for traffic between business partners over connected networks
- Allow secure sending of data over the Internet by providing IPSec virtual private network (VPN) support

These features are implemented in the IP layer on a per packet basis, and thus are available to any network application without requiring any special modifications. Applications can also implement their own additional security features as necessary, on top of the underlying IP security.

IP security policy is enabled, enforced, managed, and monitored through a coordinated effort of several z/OS Communications Server components:

- Policy Agent

  The Policy Agent is used to configure IP security on a z/OS system. It reads the configuration files that contain the IP security policy configuration statements, checks them for errors, and installs them into the IKE daemon and the TCP/IP stack.
- Internet Key Exchange daemon (IKED)

  The Internet Key Exchange daemon is responsible for retrieving IP security policy from Policy Agent, and dynamically managing keys that are associated with dynamic IPSec VPNs.

- TCP/IP stack

  The stack maintains a list of currently active IP filters and IPSec security associations, actively filters network traffic, controls encryption and decryption of network data, and maintains counters that are associated with an IPSec security association lifetime.
- Traffic Regulation Manager daemon (TRMD)

  The Traffic Regulation Manager daemon is responsible for logging IP security events that are detected by the stack, including IP filter events, updates to IP security policy, and the creation, deletion, and refresh of IPSec security associations.
- System logging daemon (syslogd)

  The system logging daemon manages the logging of messages and events for all of the other components, including where the log messages are written.

These components provide a combination of technologies that form the basis of IP security:

- IP filtering
- IP filter logging
- Data encryption and authentication

Options for configuring IP security

z/OS IP security is configured using an extensive set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the set of IP security requirements for each TCP/IP stack. IBM provides two alternatives for creating the Policy Agent configuration files.

Option 1: Manual configuration

You can manually create the IP security policy configuration files by coding all of the required statements in an HFS file or MVS data set. There are a large number of powerful configuration options provided by IP security policy statements that permit advanced users to carefully fine-tune IP security policy on a per-stack basis. The following sections describe the procedure for creating an IP security policy by manually creating and editing the configuration files. There are examples that step through the process of creating an IpSecConfig file that includes zones corresponding to various security models. For details on the configuration statements and parameters, refer to the Policy Agent and policy applications chapter of *z/OS Communications Server: IP Configuration Reference*.

Option 2: Use the z/OS Network Security Configuration Assistant

IBM provides a configuration GUI that you can use to generate the Policy Agent and IKE daemon configuration files. The z/OS Network Security Configuration Assistant is a standalone application that runs under the Windows operating system and requires no network connectivity or setup. You can download the GUI from the Communications Server family downloadable tools Web page.

Through a series of wizards and online help panels, you can use the GUI to create IP security configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the GUI, there are three types of reusable objects:

- Traffic Descriptors that define the IP traffic type, such as TCP or UDP.
- Security Levels that define the level of security, such as the encryption level.
- Requirement Maps that map Traffic Descriptors to Security Levels. A single Requirement Map should contain a complete set of security requirements that will govern the level of security for multiple IP traffic types.

For each TCP/IP stack, you create a set of Connectivity Rules that indicate the data endpoints and indicate which Requirement Map will govern security between the data endpoints.

The configuration GUI comes with a number of IBM-supplied Traffic Descriptors, Security Levels, and Requirement Maps that are easily applied to an existing network topology, or the IBM-supplied definitions can be used as the basis for your own set of reusable objects.

The GUI can dramatically reduce the amount of time that is required to create IP security policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security, use of the GUI is encouraged to ensure that you have a consistent and easily manageable interface for implementing IP security.

If you plan to configure Policy Agent manually, you cannot take advantage of the GUI. This chapter primarily targets those who plan to use option 1, manual configuration. However, if using the GUI, reading this chapter will help you understand security concepts and the relationship between Policy Agent and IP security function.

IP filtering

IP filtering controls the flow of network traffic. An administrator can deny or allow any given network packet into or out of a z/OS system with an IP security policy.

Filter rules and actions

The IP security policy enables a z/OS system to classify any IP packet that comes across a network interface and take specific action according to a predefined set of rules. The set of properties that identify a packet, together with the action to be performed on it, is known as an IP filter rule. The rule can be used to filter out unwanted packets from the network stream, while allowing others. The collection of all filter rules comprise the IP filter table. The IP filter table contains all of the IP filter rules in the order in which they were configured. IP filter rules are configured using the IpFilterRule statement in an IP security policy configuration file. For more details about the IpFilterRule statement, refer to *z/OS Communications Server: IP Configuration Reference*.

For example, a simple filter table might have the following set of rules:

1. Allow Telnet traffic from IP address A.
2. Allow FTP traffic from IP address B and IP address C.
3. Allow any traffic from subnet D, but ensure that it is encrypted.
4. Allow outbound connections to anywhere.
5. Deny anything that does not match the above four rules.

This set of rules would be considered a filter table comprised of five rules.

The filter table that is configured for a particular installation reflects the security needs for that site. The rules can be restrictive or permissive, as the security policy allows. Normally the rules would deny anything not explicitly permitted, a configuration known as a default-deny policy, in which rules are added as necessary to allow only crucial network traffic. In a default-deny environment, the absence of any IP filter rules essentially isolates the system from the network. The alternative, a default-allow policy, allows all network traffic in the absence of any

configured rules. Specific rules can be added as needed to deny unwanted or potentially malicious traffic. A default-deny policy is considered to be much more secure.

Rule: A z/OS Communications Server TCP/IP stack that is configured for IP security follows a default-deny policy by default, in the absence of any configured filter rules.

IP filter tables can grow very complex, and in many implementations are difficult to maintain. However, the configuration mechanism that is provided by IP security enables you to attach meaningful descriptors to rules, hosts, and other configured items, which makes keeping track of complex filter tables an easier task.

To protect data between hosts, hosts must agree on what type of traffic to protect, and how to protect that traffic. These IP traffic pattern definitions are stored in the locally configured security policy and installed in the IP filter table, which is consulted for each IP packet that enters or leaves the system. When a packet matches one of the rules in the IP filter table, the policy determines what action is taken for that packet. IP filter actions are configured using the IpGenericFilterAction statement in an IP security policy configuration file. For more details about the IpGenericFilterAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

On a z/OS stack that has IPCONFIG IPSECURITY configured and an active IP security policy, there are three possible actions:

- Deny the packet.
- Permit the packet.
- Permit the packet with IPSec protection.

If the action that is associated with the filter rule is an ipsec action, the packet is subject to the application of IPSec authentication and encryption before it is received or sent. Any packet that matches a filter rule with an ipsec action is processed using the IPSec protocols, either Authentication Header (AH), Encapsulating Security Payload (ESP), or both, depending on the locally configured policy. z/OS IP security requires that data authentication be done if the filter rule specifies an ipsec action.

Filtering criteria in an IP packet

IP packets match IP filters based on a number of selection criteria. There are five primary pieces of information that are gathered from the IP packet, commonly referred to as a 5-tuple:

- Source address

  An IP packet can be filtered based on the source address located in the IP header of the packet.
- Destination address

  An IP packet can be filtered based on the destination address located in the IP header of the packet.
- Protocol

  An IP packet can be filtered based on the protocol in the IP header of the packet.
- Source port

  If the protocol in an IP packet is TCP or UDP, the packet can be filtered based on the source port in the TCP/UDP header of the data portion of the packet.

- Destination port

  If the protocol in an IP packet is TCP or UDP, the packet can be filtered based on the destination port in the TCP/UDP header of the data portion of the packet.

Additional filtering criteria based on protocol

Following are some additional filtering criteria that are based on protocol:

- ICMP type and code

  If the protocol in an IP packet is ICMP, the packet can be filtered based on the ICMP type and code located in the ICMP header of the data portion of the packet.

- OSPF type

  If the protocol in an IP packet is OSPF, the packet can be filtered based on the OSPF type located in the OSPF header of the data portion of the packet.

Additional filtering criteria based on network attributes

Some filtering criteria are inferred from the external characteristics of the IP traffic, rather than being found in the actual IP packet. The following additional attributes are used to distinguish IP packets:

- Direction

  The direction of an IP packet is either inbound or outbound.

- Routing

  The routing attribute of an IP packet is either local or routed. IP packets are considered local if either of the following is true:

  – The packet is inbound and the destination address in the IP packet exists on the IPSECURITY stack.

  – The packet is outbound and the source address in the IP packet exists on the IPSECURITY stack.

  Otherwise the IP packet is considered routed.

- Security class

  Each non-virtual interface on a z/OS system is assigned a security class. The security class of an interface is determined by the SECCLASS parameter that is coded on either the LINK statement or the DYNAMICXCF parameter of the IPCONFIG statement in the TCP/IP profile. The value of SECCLASS is a number in the range 1-255. If SECCLASS is not specified for an interface, the interface has a default security class of 255.

  Each IP packet entering or leaving the system inherits the security class of the interface that it traverses:

  – For inbound traffic, this is the interface on which the packet arrived.

  – For outbound traffic, this is the interface over which the packet is sent.

  The value for SECCLASS has no inherent meaning. A SECCLASS of 255 is no more or less secure than a SECCLASS of 1. You can optionally assign a SECCLASS value either to uniquely identify an interface, or to group interfaces with similar security requirements, based on site policy. Consequently, this attribute can be used as an additional criterion for IP filtering. Security class can be used to define broad filter rules that encompass all of the IP traffic that uses a group of interfaces without explicit knowledge of network addressing, or to ensure that an IP packet arrived on a valid interface.

  For example, as shown in Figure 75 on page 829, consider a z/OS system with three physical interfaces that are assigned the following security classes:

```
I1: SECCLASS 10
I2: SECCLASS 20
I3: SECCLASS 20
```

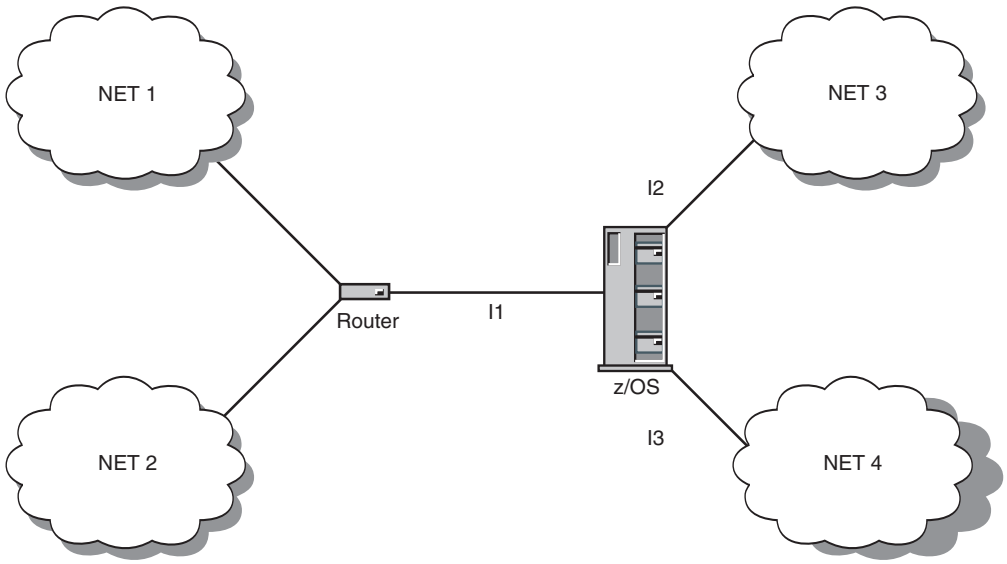


Figure 75. Using SECCLASS to identify interfaces

All IP packets from NET1 and NET2 have the same security class (10), and all IP packets from NET3 and NET4 have the same security class (20). Even though each interface on the z/OS system can reach multiple networks, you can configure a single IP filter rule that matches all of the IP traffic from NET3 and NET4 without explicitly identifying any attributes of the IP packets, other than security class. For instance, if NET3 and NET4 are trusted internal networks in which you want to allow all network traffic, you can configure one IP filter rule permitting all traffic with a security class of 20, without regard to IP address.

An IP filter using security class as a criterion is not limited to scenarios in which the IP addresses are ignored. To counter IP address spoofing, IP filter rules can take into account both security class and IP address. IP address spoofing involves the creation of an IP packet whose source address has been modified to reflect an address other than that of the originator. Because routers ignore the source address when making routing decisions, the modified IP packet can still reach its destination. An IP packet whose source IP address has been spoofed is usually not legitimate and is often used in a malicious manner. IP filtering can counter an attempt to spoof a source IP address by ensuring that an IP packet arrived on a valid network interface. For instance, any inbound packet that has a source address of an internal network node, but entered the system though an external interface, is probably spoofed and should be denied.

Rule: The SecurityClass parameter cannot be used on an IpFilterRule statement with an ipsec action. The IKE protocols do not provide for the negotiation of SecurityClass, and therefore it cannot be used as a selection criterion for a security association. The SecurityClass parameter can be configured only for an IpFilterRule statement that has an action of either permit or deny. Coding a SecurityClass parameter with a value other than 0 on an ipsec rule is considered a configuration error.

Guideline: If there are multiple interfaces by which a packet can reach its destination, the SECCLASS values for the interfaces and the filter rules should be consistent. If a packet could reach its intended destination by traversing one

of a number of interfaces, the SECCLASS values and the related filter rules should account for the possibility of alternate routing.

For more details about the SecurityClass parameter on the IpService statement, the SECCLASS parameter on the LINK statement, and the SECCLASS keyword on the DYNAMICXCF parameter of the IPCONFIG statement, refer to *z/OS Communications Server: IP Configuration Reference*.

IP traffic patterns

The information that is gathered from an IP packet can be used to identify a specific type of network traffic, based on common Internet protocols. For example:

- Web traffic to a server consists of TCP packets inbound to and outbound from port 80.
- TN3270 traffic consists of TCP packets inbound to and outbound from port 23.
- IKE traffic consists of UDP packets inbound from and outbound to port 500. In addition, if NAT traversal is allowed, IKE traffic can use port 4500.
- Ping consists of ICMP packets outbound with type 8, code 0 (echo request), and inbound with type 0, code 0, (echo reply).

The IpService statement groups these attributes into definitions of common traffic patterns that can subsequently be incorporated in an IP filter rule. The sample configuration file /usr/lpp/tcpip/samples/pagent_CommonIPSec.conf provides definitions for many of the common traffic patterns that are seen on a typical z/OS server.

For more details about the IpService statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Conditionally controlling IP filters

When IP filters from IP security policy files are installed into the stack, they are always active by default. Depending on the business need, this might not always be desirable. For instance, you might want to restrict certain types of IP traffic to a specific time, day, or week. IP security provides this flexibility by allowing you to specify a time condition for an IP filter rule. The IpTimeCondition statement specifies when an IP filter or a manual IPSec tunnel is active. You can prescribe not only what traffic is allowed, but when that traffic is allowed, in a way that is not disruptive and without having to edit the IP security policy files.

For more details about the IpTimeCondition statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Default IP filter policy and IP security policy

Policy Agent provides IP filter policy to the stack, as defined by the IP security policy configuration files. For recognition as an IPSECURITY-enabled stack, the TCP/IP profile must have IPCONFIG IPSECURITY coded. For the stack to receive configured policy, the Policy Agent must be active. IP filter policy is installed when Policy Agent and the stack are active. If Policy Agent is not active when an IPSECURITY-enabled stack initializes, the stack cannot be provided with IP filter rules from that policy. Therefore, in the interest of network security, the stack provides a default IP filter policy when an IP filter policy is unavailable from Policy Agent. The default IP filter policy effectively denies all network traffic, with the exception of some select ICMP messages that are necessary for internal stack function. These deny rules are not explicitly coded, but rather are always implicitly added at any time that the default IP filter policy is in effect. The default IP filter

policy can be in effect at times other than stack initialization. Default IP filter policy is in effect in all of the following cases:

- An IPSECURITY-enabled TCP/IP stack has been started but Policy Agent has not been started, or an IPSECURITY-enabled TCP/IP stack has been started but Policy Agent has not completely initialized.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but no IP security configuration file exists.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but the IP security configuration file contains errors.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but no IpFilterPolicy statement exists in either IP security configuration file.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, and IP filter policy has been provided to the stack, but the stack is using the default IP filter policy because the **ipsec -f default** command was issued.

This default behavior ensures that network security is not compromised in the event that IP filter policy is not installed, and is consistent with a secure default-deny policy. However, you can modify the default IP filter policy by coding an IPSECRULE statement in the TCP/IP profile. The IPSECRULE statement describes the attributes of the IP traffic that is allowed when the default policy is active. Because the default behavior is to deny all network traffic, IPSECRULE statements are always permit rules that denote exceptions to the default-deny policy.

Rule: In the interest of network security, routed traffic is not allowed when the default policy is in effect, even if the default IP filter policy has been modified using the IPSECRULE statement.

It is also important to note that neither the default IP filter policy nor a modified default IP filter policy provides authentication and encryption capabilities such as those provided by a complete IP security policy; it only offers the stack the ability to perform simple IP filtering in the absence of an IP security policy.

Modifying the default IP filter policy

The default IP filter policy is in effect when the IP security policy, as configured in the Policy Agent, is not available. The default IP filter policy denies all network traffic, unless you modify the TCP/IP profile.

Two IP filters are created by the default IP filter policy, one denying inbound traffic and the other denying outbound traffic. For example, assuming that the default IP filter policy is active, the following sample of the **ipsec -f display** command shows that SYSDEFAULTDENYRULE was added to the filter table:

```
ipsec -f display

CS V1R7 ipsec  TCPIP Name: TCPCS  Fri Mar 18 11:52:41 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Profile   Scope:    Current           TotAvail: 2
Logging:  Yes             Predecap: No                DVIPSec:  Yes
NatKeepAlive:  20

FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
```

```
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Deny
|                       Scope:                      Both
|                       Direction:                  Outbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    All
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Packet
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Packet
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Packet
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 SYSDEFAULTDENYRULE
|                       FilterNameExtension:        2
|                       GroupName:                  n/a
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Deny
|                       Scope:                      Both
|                       Direction:                  Inbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    All
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Packet
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Packet
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Packet
|                       OrigRmtConnPort:            n/a
```

```
RmtIDPayload:                 n/a
*************************************************************************

2 entries selected
```

Use IPSECRULE for permit rules that denote exceptions to the default-deny policy. When the default IP filter policy is active, these permit rules appear in the default IP filter table before the SYSDEFAULTDENYRULE entries. Typically, these exceptions are few in number and are used for administrative access to the system in the event that IP security policy is unavailable. For instance, a sample default set of IPSECRULE entries might include rules to allow the following:

- Administrative access
- Basic network services, such as DNS and OSPF routing advertisements
- Use of ping to test for server availability

IPSECRULE entries are coded in the IPSEC block of the TCP/IP profile. They describe the attributes of the IP traffic that is allowed when the default IP filter policy is active. These rules can specify source address, destination address, protocol, source port, destination port, and security class.

Unlike IP filter rules that are defined in Policy Agent, which allow direction to be specified, an IPSECRULE is always bidirectional. This means that for any IPSECRULE entry that specifies a source and destination address or port, an outbound rule is created with that source and destination address and port, along with an inbound rule with the source and destination addresses and ports reversed. (This equates to the use of the bidirectional keyword in an IpFilterRule statement.)

IPSECRULE entries always have an action of permit; there is no action specification for deny or permit with IPSec protection. These rules are always applied to local traffic only. Unlike IP filter rules that are defined in the Policy Agent, an IPSECRULE cannot permit routed traffic.

Assuming the administrative machine has an IP address of 9.1.1.2 and connects through an interface whose security class is 100, the following sample IPSECRULE entries would allow the z/OS system to communicate DNS queries and OSPF routing advertisements to anyone, while giving blanket access to the administrator. An asterisk (*) is the default and represents all, indicating that any packet matches this attribute.

```
IPSEC LOGENable
; Rule      SrcAddr DstAddr   Logging Protocol   SrcPort    DestPort     Secclass

; OSPF protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO OSPF

; IGMP protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO 2

; DNS queries to UDP port 53
  IPSECRule *       *         NOLOG   PROTO UDP  SRCPort *  DESTport 53

; Administrative access
  IPSECRule *       9.1.1.2   LOG                                       SECCLASS 100

ENDIPSEC
```

Any IPSECRULE entries that are coded are given a system-generated name for purposes of display. These rules are prefixed with the string SYSDEFAULTRULE. The IPSECRULE entries that are configured in the previous example are reflected

in the **ipsec -f display** command as shown in the following example. Notice that there are two rules created for each IPSECRULE configured, one for each direction. The four rules that are configured in the example IPSEC block have been expanded into eight IP filters. Two additional filters have been created for the default-deny behavior. Each rule is given a unique filter name comprised of SYSDEFAULTRULE.*number*, where *number* is a numerical extension that indicates the relative order of the rule in the IP filter table. Since each IPSECRULE statement results in multiple filter rules with the same name (inbound and outbound), a FilterNameExtension value is assigned by the system to uniquely identify each filter.

```
ipsec -f display

CS V1R7 ipsec  TCPIP Name: TCPCS  Fri Mar 18 11:55:57 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Profile   Scope:    Current           TotAvail: 10
Logging:  Yes             Predecap: No                DVIPSec:  Yes
NatKeepAlive:  20

FilterName:                   SYSDEFAULTRULE.1
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
Protocol:                     OSPF(89)
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     All
TCPQualifier:                 n/a
ProtocolGranularity:          Packet
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        Packet
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          Packet
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*******************************************************************************
FilterName:                   SYSDEFAULTRULE.1
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Permit
Scope:                        Local
```

```
|                       Direction:                  Inbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    None
|                       Protocol:                   OSPF(89)
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   All
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Packet
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 n/a
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Packet
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   n/a
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Packet
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 SYSDEFAULTRULE.2
|                       FilterNameExtension:        1
|                       GroupName:                  n/a
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Permit
|                       Scope:                      Local
|                       Direction:                  Outbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    None
|                       Protocol:                   IGMP(2)
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Packet
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Packet
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Packet
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 SYSDEFAULTRULE.2
|                       FilterNameExtension:        2
|                       GroupName:                  n/a
```

```
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     IGMP(2)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
|                       ProtocolGranularity:          Packet
|                       SourceAddress:                0.0.0.0
|                       SourceAddressPrefix:          0
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   All
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Packet
|                       DestAddress:                  0.0.0.0
|                       DestAddressPrefix:            0
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Packet
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ************************************************************************
|                       FilterName:                   SYSDEFAULTRULE.3
|                       FilterNameExtension:          1
|                       GroupName:                    n/a
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Outbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     UDP(17)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
|                       ProtocolGranularity:          Packet
|                       SourceAddress:                0.0.0.0
|                       SourceAddressPrefix:          0
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   All
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Packet
|                       DestAddress:                  0.0.0.0
|                       DestAddressPrefix:            0
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     53
|                       DestPortRange:                n/a
```

```
DestPortGranularity:          Packet
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*********************************************************************
FilterName:                   SYSDEFAULTRULE.3
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
Protocol:                     UDP(17)
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Packet
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   53
SourcePortRange:              n/a
SourcePortGranularity:        Packet
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Packet
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*********************************************************************
FilterName:                   SYSDEFAULTRULE.4
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                100
Logging:                      All
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Packet
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
```

```
|                       SourcePortGranularity:        Packet
|                       DestAddress:                  9.1.1.2
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Packet
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ************************************************************************
|                       FilterName:                   SYSDEFAULTRULE.4
|                       FilterNameExtension:          2
|                       GroupName:                    n/a
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                100
|                       Logging:                      All
|                       Protocol:                     All
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
|                       ProtocolGranularity:          Packet
|                       SourceAddress:                9.1.1.2
|                       SourceAddressPrefix:          n/a
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   All
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Packet
|                       DestAddress:                  0.0.0.0
|                       DestAddressPrefix:            0
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Packet
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ************************************************************************
|                       FilterName:                   SYSDEFAULTDENYRULE
|                       FilterNameExtension:          1
|                       GroupName:                    n/a
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Deny
|                       Scope:                        Both
|                       Direction:                    Outbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      All
|                       Protocol:                     All
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
```

```
ProtocolGranularity:          Packet
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Packet
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Packet
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      All
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Packet
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Packet
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Packet
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*************************************************************************

10 entries selected
```

For an IPSECURITY-enabled stack, one of the two security policies is always in effect, either the default policy or the IP security policy as defined in Policy Agent. The policy that is in effect at any given time is considered to be the active policy. The source field of the report header for the **ipsec -f display** command can be used to determine which policy is currently active as follows:

```
ipsec -f display

CS V1R7 ipsec  TCPIP Name: TCPCS  Fri Mar 18 11:57:36 2005
```

```
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Profile   Scope:    Current           TotAvail: 10
Logging:  Yes             Predecap: No                DVIPSec: No
NatKeepAlive:  20
```

`Stack Profile` indicates that the default IP filter policy is active; `Stack Policy` would indicate that the IP security policy as defined in Policy Agent is active.

You can also choose which policy is the active policy. The **ipsec** command provides the ability to switch the active policy between the default IP filter policy in the TCP/IP profile and the IP security policy in Policy Agent. Issuing the **ipsec -f default** command causes the default policy to become the active policy, while issuing the **ipsec -f reload** command reloads the IP security policy from Policy Agent, provided that Policy Agent is active and that IP security has been correctly configured.

There are cases in which you might want to switch to the default policy. For instance, in the event of a security breach, the **ipsec -f default** command allows only the network traffic that has been explicitly coded in the TCP/IP profile IPSEC block, which typically permits only administrative access.

It is important to use the **ipsec -f default** command with discretion. Issuing **ipsec -f default** on an operational system can have a dramatic impact and cause packets to be dropped, depending on how the IPSECRULE entries are coded. Consider the use of **ipsec -f default** as equivalent to a shutdown, and do not use it in normal conditions unless the following are true:

- The system is in maintenance mode.
- There is no productive work being done on the system.
- The site policy for IP traffic is default-allow, only local traffic needs to be permitted on this system, and the appropriate IPSECRULE entries were coded.

Malicious use of network resources can be identified by unusual IP traffic conditions, console messages, or log inspection. Should a security concern arise in a production environment, the following steps should be performed quickly to minimize the impact of potentially restricted flow of IP traffic.

1. To secure the system, switch to the default IP filter policy by issuing the **ipsec -f default** command.
2. Analyze system logs to determine the nature and the source of the security concern. If packet logging was active for the time period in question, then inspect the TRMD packet filtering log entries.
3. Update the IP security policy configuration to alleviate the security concern. For example, add an IpFilterRule that denies any suspicious IP address or port use, and activate IP filter logging.
4. Activate the updated IP security policy from the Policy Agent by issuing the **ipsec -f reload** command.
5. Verify that the security concern is eliminated. Monitor the network traffic and inspect the system logs, including any related IP filter log messages.

IP filter logging

Monitoring network events is an important aspect of network security. Logs can be used to verify that policies have been correctly configured and enforced, or to gather statistics on any traffic of interest. For instance, traffic that is persistently denied might be suspect. With IP filter logging, you can inspect any traffic on the system and even fine tune the configuration to show only those entries of interest.

Logging can be controlled at the global level or at the individual rule level, including the ability to specify whether to log permitted traffic, denied traffic, or both. The IP filter log entries provide detailed information about each packet, including the rule that the packet matched and any pertinent IPSec information.

TRMD and syslogd provide the logging service for IP security. If running in the common INET environment, you must configure one instance of TRMD for each stack on a z/OS system.

Guideline: Exhaustive logging of IP traffic can have a negative effect on performance. If logging is excessive, it can be turned off temporarily at the global level while the appropriate logging modifications are made to the individual IP filter rules. IP filter logging is controlled by the IpFilterLogging parameter on the IpGenericFilterAction statement. For more details, refer to *z/OS Communications Server: IP Configuration Reference*.

Data encryption and authentication — IPSec

To participate in a VPN, a host must encrypt and authenticate individual IP packets between itself and another communicating host. IPSec is one of several mechanisms for achieving this, and one of the more versatile.

IPSec is defined by the IPSec working group of the IETF. It provides authentication, integrity, and data privacy between any two IP entities. Management of cryptographic keys and security associations can be either manual or dynamic using an IETF-defined key management protocol called Internet Key Exchange (IKE).

IPSec provides flexible building blocks that can support a variety of configurations. Because an IPSec security association can exist between any two IP entities, it can protect a segment of the path or the entire path. The main advantage of using IPSec for data encryption and authentication is that IPSec is implemented at the IP layer. Consequently, any network traffic that is carried by an IP network is eligible to use IPSec services without any special changes to higher level protocols that are used by applications. However, if the system is using any of these alternate security protocols to secure specific applications, IP filtering can be used to avoid the overhead of multiple security protocols. For example, you might want to exclude Web traffic (based on the well-known secure port of the Web server, port 443) from IPSec coverage because you would like to use SSL.

IPSec enables the creation of virtual private networks (VPN). A VPN enables an enterprise to extend its network across a public network, such as the Internet, through a secure tunnel using security associations. IPSec VPNs enable the secure transfer of data over the public Internet for same-business and business-to-business communications, and protect sensitive data within an enterprise's internal network.

IPSec uses IP filtering to determine which traffic should be protected by IPSec. A special type of filter action specifies to permit the traffic, but only with IPSec protection. The IP filters represent IP security policy to the stack by specifying the traffic that requires IPSec protection. The filters are also used in locating the outbound IPSec security association, and for verifying that inbound traffic is received using the correct security association.

The IETF has standardized the IPSec protocol suite and key management schemes in a series of IPSec RFCs. For more information on these RFCs, see Appendix F, "Related protocol specifications (RFCs)," on page 1223.

IPSec has three major components:

- IP Authentication Header (AH)
- IP Encapsulating Security Payload (ESP)
- Internet Key Exchange (IKE)

AH and ESP protocols

IPSec uses two distinct protocols, Authentication Header (AH) and Encapsulating Security Payload (ESP), which are defined by the IETF.

The AH protocol provides a mechanism for authentication only. AH provides data integrity, data origin authentication, and an optional replay protection service. Data integrity is ensured by using a message digest that is generated by an algorithm such as HMAC-MD5 or HMAC-SHA. Data origin authentication is ensured by using a shared secret key to create the message digest. Replay protection is provided by using a sequence number field with the AH header. AH authenticates IP headers and their payloads, with the exception of certain header fields that can be legitimately changed in transit, such as the Time To Live (TTL) field.

The ESP protocol provides data confidentiality (encryption) and authentication (data integrity, data origin authentication, and replay protection). ESP can be used with confidentiality only, authentication only, or both confidentiality and authentication. When ESP provides authentication functions, it uses the same algorithms as AH, but the coverage is different. AH-style authentication authenticates the entire IP packet, including the outer IP header, while the ESP authentication mechanism authenticates only the IP datagram portion of the IP packet.

Either protocol can be used alone to protect an IP packet, or both protocols can be applied together to the same IP packet. The choice of IPSec protocol is determined by the security needs of your installation, and is configured by the administrator. It does not have to be applied system-wide, and can be configured differently for each set of connection endpoints. For a dynamic tunnel, the choice of IPSec protocol is configured using the IpDataOffer statement in an IP security policy configuration file. For a manual tunnel, the choice of IPSec protocol is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpDataOffer statement and the IpManVpnAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

z/OS IP security requires authentication due to potential security exposures when encryption is used alone. Authentication can be provided by the ESP or AH protocol. The complete list of combinations for authentication and encryption that are provided by z/OS IP security and that can be used for a specific connection are shown in Table 33. There are four combinations that use only authentication and eight that use various combinations of both authentication and encryption.

Table 33. Possible authentication and encryption combinations for a connection

Authentication		Encryption	
Protocol	**Algorithm**	**Protocol**	**Algorithm**
AH	MD5	None	None
AH	SHA	None	None
ESP	MD5	None	None
ESP	SHA	None	None

Table 33. Possible authentication and encryption combinations for a connection (continued)

Authentication		Encryption	
Protocol	**Algorithm**	**Protocol**	**Algorithm**
AH	MD5	ESP	DES
AH	MD5	ESP	3DES
AH	SHA	ESP	DES
AH	SHA	ESP	3DES
ESP	MD5	ESP	DES
ESP	MD5	ESP	3DES
ESP	SHA	ESP	DES
ESP	SHA	ESP	3DES

Encapsulation

In the process of applying either AH or ESP to an IP packet, the original IP packet is modified. Outbound packets are rebuilt with additional IPSec headers in a process known as encapsulation, while inbound packets are stripped of their IPSec headers in a process known as decapsulation. Before leaving a host, outbound packets are encapsulated using a cryptographic key that is known to both communicating hosts. Inbound packets are decapsulated on the receiving side using the same cryptographic key, thereby recovering the original datagram. If encryption is used, any packet that is intercepted on the IP network is unreadable to anyone without the encryption key. Any modifications to the IP packet while in transit are detected by authentication processing at the receiving host and discarded.

Transport mode and tunnel mode: The manner in which the original IP packet is modified depends on the encapsulation mode used. There are two encapsulation modes used by AH and ESP, transport and tunnel.

Transport mode encapsulation retains the original IP header. Therefore, when transport mode is used, the outer IP header reflects the original source and destination of the packet. Transport is most often used in a host-to-host scenario, where the data endpoints and the security endpoints are the same. A transport mode encapsulated datagram is routed, or transported, in the same manner as the original packet.

Figure 76 shows an IP packet that is encapsulated using AH in transport mode:

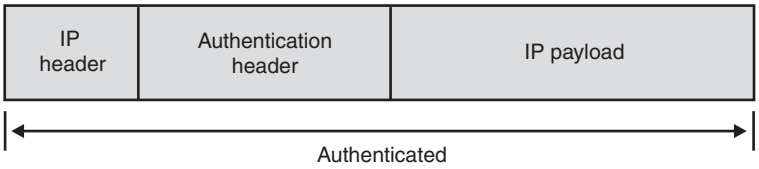


Figure 76. IP packet encapsulated using AH in transport mode

Figure 77 on page 844 shows an IP packet that is encapsulated using ESP in transport mode:

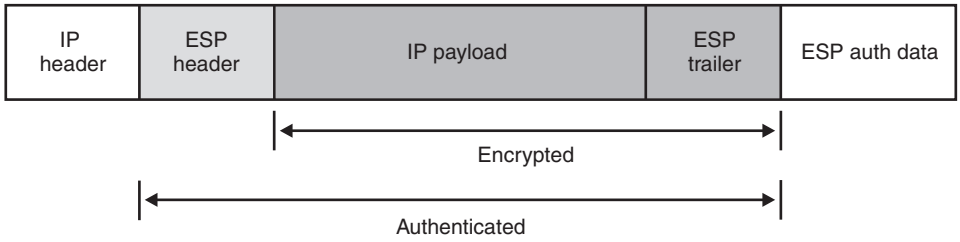


Figure 77. IP packet encapsulated using ESP in transport mode

Tunnel mode encapsulation builds a new IP header containing the source and destination address of the security endpoints. When tunnel mode is used, the outer IP header reflects the source and destination of the security endpoints, which might or might not be the same as the original source and destination IP address of the data connection. The choice of transport or tunnel mode depends on the structure of the network and relies heavily on logical connections between the endpoints. Tunnel mode is required if one of the IKE peers is a security gateway that is applying IPSec on behalf of another host or hosts. A datagram that is encapsulated in tunnel mode is routed, or tunneled, through the security gateways, with the possibility that the secure IPSec packet will not flow through the same network path as the original datagram.

Figure 78 shows an IP packet that is encapsulated using AH in tunnel mode:

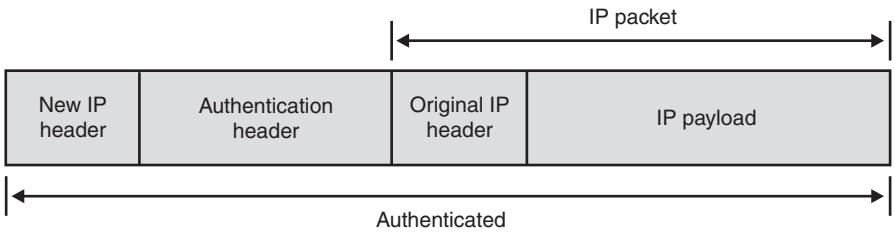


Figure 78. IP packet encapsulated using AH in tunnel mode

Figure 79 shows an IP packet that is encapsulated using ESP in tunnel mode:

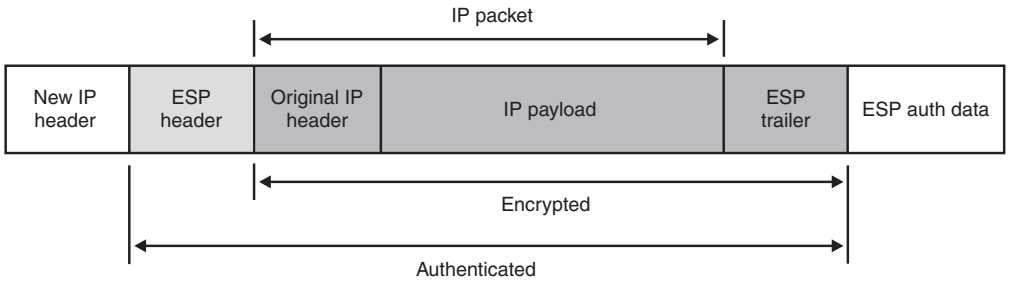


Figure 79. IP packet encapsulated using ESP in tunnel mode

Do not confuse tunnel mode encapsulation with IKE tunnel or IPSec tunnel. In this context, tunnel refers only to the method by which IPSec packets are constructed, while IKE and IPSec tunnels are conceptually defined as secure logical connections between hosts. IPSec tunnels can use transport mode or tunnel mode encapsulation.

For a dynamic tunnel, the choice of encapsulation mode is configured using the IpDataOffer statement in an IP security policy configuration file. For a manual

tunnel, the choice of IPSec protocol is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpDataOffer statement and the IpManVpnAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Guideline: In host-to-host scenarios, transport mode is commonly used because it does not incur the overhead of having to build an additional IP header. However, tunnel mode is equally valid.

Rule: If you are using sysplex-wide security associations, a dynamic security association cannot use a subnet or range that encompasses a DVIPA address. For takeover to work consistently, security associations must be negotiated for single IP addresses only. If two DVIPAs that are covered by the same security association are subsequently taken over by two different backup stacks, the coverage of the security association is ambiguous because it is then linked to two DVIPAs on two different stacks.

UDP encapsulation of IPSec ESP packets

When building an ESP packet, it can be further encapsulated by placing a UDP header in front of the ESP header. This is known as UDP encapsulation. UDP encapsulation is used to allow IPSec traffic to successfully traverse a NAT device. For more information on NAT traversal (NATT), see "IPSec and Network Address Translation (NAT) devices" on page 853.

There are two modes of UDP encapsulation:

- UDP-Encapsulated-Transport mode
- UDP-Encapsulated-Tunnel mode

As shown in Figure 80, UDP-Encapsulated-Transport mode inserts a UDP header in between the IP header and the ESP header of a normal transport mode ESP packet.

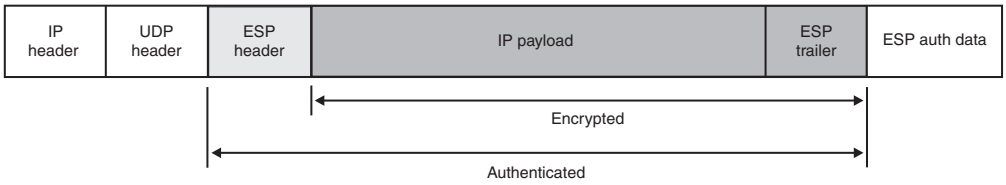


Figure 80. UDP-Encapsulated-Transport mode

As shown in Figure 81, UDP-Encapsulated-Tunnel mode inserts a UDP header in between the new IP header and the ESP header of a normal tunnel mode ESP packet.

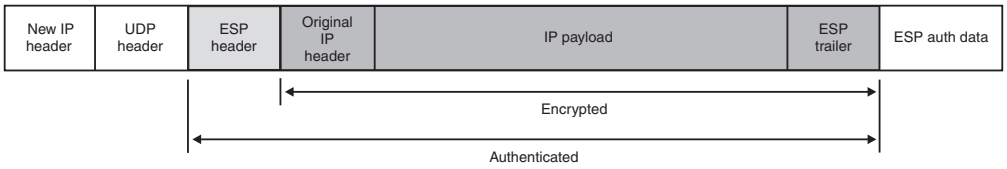


Figure 81. UDP-Encapsulated-Tunnel mode

When an IPSec UDP-encapsulated packet is built, the source and destination port values in the UDP header are set to the IKE port value of 4500.

Configure the choice of transport or tunnel mode using the IpDataOffer statement in the IP security policy configuration file. For more details about the IpDataOffer statement, refer to *z/OS Communications Server: IP Configuration Reference*.

The decision to use a UDP-encapsulated mode is not configured, but instead inferred, when a NAT is detected between two IKE daemons.

IPSec and symmetric key management

At the center of encryption and authentication is the notion of a cryptographic key. Security endpoints use keys to encrypt and decrypt data. The IPSec protocols create security association keys that are directional. As shown in Figure 82, the key that is used to encrypt outbound data on one host is used to decrypt the same data on the remote host, while the key that is used to encrypt data on the remote host is used to decrypt data on the local host.

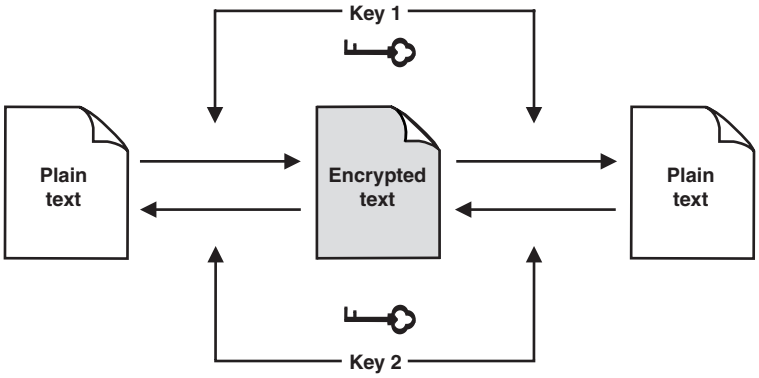


Figure 82. Symmetric encryption

This type of encryption is known as symmetric, because it requires that both hosts use the same keys on the same data.

Manual key management

IPSec keys and values can be configured manually. Manual management of keys was the only non-proprietary option for implementing IPSec before the standardization of the IKE protocols. In a manual IPSec configuration, the keys that are used to encrypt data, and the Security Parameter Index (SPI) values that are used to uniquely identify a security association, are determined by the administrator and configured beforehand on both hosts. However, the steps that you must take to manually generate encryption keys can become quite cumbersome, as shown in the following examples:

- A single IPSec connection requires two keys, because encryption keys are unidirectional (one for inbound traffic, one for outbound traffic).
- A single IPSec connection can require up to four SPI values, depending on the type of IPSec protection required.
- The keys and SPI values must be individually installed on both the local and remote system.
- To help ensure that the keys are not compromised, they need to be changed periodically and updated on both hosts. Manual IPSec has no key refresh capabilities unless the security associations are deactivated, reconfigured with the new key, and then reactivated.
- During the refresh maintenance period, IPSec functionality is unavailable, disrupting any IP traffic requiring IPSec protection.

- Because of the disruptive nature of key refresh with manual IPSec, key lifetimes are frequently defined with much larger values, thus increasing the security exposure.
- These steps must be carried out for each remote host that communicates with IPSec, with different keys for each of them.

In a small installation with minimal security concerns, this manual process can work quite well, and might actually be preferable. However, given the large number of communicating hosts in a typical network, this manual process quickly becomes unwieldy, error prone, maintenance intensive, and not scalable. Nevertheless, z/OS IP security does support manual IPSec configuration for compatibility with systems that use it. If an IP filter rule specifies a manual ipsec action, the corresponding action is consulted to determine all aspects of the encryption and authentication policy for that data, including the identities of the communicating hosts, the keys that are used for encryption and authentication of outbound and inbound packets, and the SPI values. Manual IPSec protection is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpManVpnAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Dynamic Key Management - IKE and IPSec negotiations

The primary role of the IKE daemon in an IP security environment is the automatic management of cryptographic keys. Dynamic key management, as provided by the IKE daemon, removes much of the administrative burden that is associated with the creation, distribution, and maintenance of cryptographic keys. IKE provides the following services:

- Host authentication (ensuring that both hosts are certain of the other's identity)
- The negotiation of a security association as follows:
  - Agreeing on the type of traffic to be protected
  - Agreeing on the authentication and encryption algorithms to be used
  - Generating cryptographic keys
- Nondisruptive periodic refresh of keys
- The deletion of security associations whose lifetimes have expired

IKE operates at the application layer. IKE negotiations are communicated between two IKE peers by a series of UDP messages. Ports 500 and 4500 are used by the IKE daemon. The negotiation proceeds in phases. In each phase, IKE negotiates a security association with a remote host. During the initial phase, IKE negotiates a phase 1 (or ISAKMP) security association, which establishes a secure channel over which the IKE peers communicate. The phase 2 (or IPSec) security association is negotiated to provide data authentication and encryption for subsequent IP traffic. The distinction between a phase 1 security association and a phase 2 security association is what the security associations protect:

- Phase 1 security associations are used to protect IKE messages that are exchanged between two IKE peers, or security endpoints.
- Phase 2 security associations are used to protect IP traffic, as specified by the security policy for a specific type of traffic, between two data endpoints.

Every security association that the IKE daemon negotiates contains information about the type of traffic it is to protect, the IP addresses of the two endpoints, the type of encryption or authentication that is provided, the keys that are used to protect data, how often the keys are refreshed, and an identifier called a Security Parameter Index (SPI) that is used to uniquely identify the security association.

An IPSec configuration contains separate policies governing each phase. Although many of the same attributes apply for phase 1 and phase 2 negotiations and keys, there are two major differences:

- The phase 1 security association can specify only a single IP address for the security endpoints, while the phase 2 security association can specify a contiguous range or subnet as the data endpoint.
- The phase 1 security association must specify an encryption method, while encryption is optional for the phase 2 security association. An authentication method must be specified for both the phase 1 and phase 2 security association.

The exact specifications for each phase 1 security association and each phase 2 security association are configured in the IpSecConfig file. The specific IP security policy statements that apply to the phase 1 and phase 2 specifications are the KeyExchangeOffer statement (phase 1) and the IpDataOffer statement (phase 2). For more details about the KeyExchangeOffer statement and the IpDataOffer statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Because the IKE protocols deal with initializing keys, they must be capable of running over links where no security is assumed to exist. IKE addresses the problem of secure key distribution by automatically deriving the keying material using a Diffie-Hellman exchange during the phase 1 IKE negotiation. This automatic creation and distribution of the key during phase 1 eliminates the need to manually distribute the session key between remote sites. Besides the obvious administrative advantage of IKE, the manual method of key distribution is prone to key compromise.

In addition, IKE non-disruptively refreshes the session keys based on the security policy of the installation. IKE specifies that this can be based on time (lifetime) and bytes transmitted (life size). IKE provides a property called perfect forward secrecy (PFS), and if PFS is used, each phase 2 key is derived independently through a separate Diffie-Hellman exchange. With PFS, if a single key is compromised, the integrity of subsequently generated keys is not affected.

Phase 1

Before IKE can negotiate the security parameters and generate the keys that are used to protect data between the two hosts, it must have a way of protecting the negotiation itself. The phase 1 IKE negotiation provides this protection by performing two tasks:

- Authenticating the IKE peer

  Peer authentication is performed either by the pre-shared key or RSA signature method. For details of peer authentication, see "Peer authentication" on page 849.
- Generating cryptographic keys

  A Diffie-Hellman exchange is performed to create a shared secret between the two IKE peers. This shared secret is then utilized in the generation of keying material. Keys to encrypt and authenticate messages sent during phase 2 are produced from this keying material. Cryptographic keys utilized by phase 2 security associations are generated from this keying material. The creation of the Diffie-Hellman shared secret is secure, but computationally expensive.

The phase 1 security association contains the following:

- The key that is used to encrypt IKE messages
- The key that is used to authenticate IKE messages
- Keying material used to generate keys produced during phase 2

- The security endpoints (single IP addresses)
- The type of protection that is required (authentication and encryption)
- How often the keys should be renewed
- A Security Parameter Index (SPI) value, which is used together with the remote security endpoint IP address to uniquely identify the security association
- The Diffie-Hellman group, which is an attribute of the public key cryptography algorithm

Because the tasks of authentication and master key generation are so resource intensive, a phase 1 security association is usually refreshed less often than a phase 2 security association.

Peer authentication: Peer authentication is a critical part of a phase 1 negotiation. Before two hosts can participate in the negotiation of a security association, each must be authorized to negotiate with the other. IKE does not allow negotiation with a host that cannot be properly identified. IP addresses are not a guarantee of identity (an IP packet can be spoofed), and the IP address from an inbound packet alone is insufficient to prove the identity of a remote host. Therefore, the IKE daemon needs a more reliable method for determining the remote host's identity. This proof of identity is first presented during the phase 1 negotiation.

z/OS IP security implements two methods of host authentication as follows:

- RSA signature
- Pre-shared key

Each of these authentication methods provides a way for hosts to verify the identity of the other, and while the pre-shared key mechanism is easier to configure, the RSA signature method is more versatile, secure, and scalable. The choice of authentication method is configured for each IPSec connection, using the KeyExchangeOffer statement in an IP security policy configuration file. For more details about the KeyExchangeOffer statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Peer authentication is not the same as data authentication. Data authentication uses IPSec to authenticate an IP packet after an IPSec security association has been negotiated by two IKE daemons. Peer authentication is used during an IKE phase 1 negotiation to identify two IKE peers to each other before the establishment of any phase 2 security associations.

Guideline: As a matter of security, pre-shared keys should not be shared among multiple remote IKE peers. If the pre-shared key method of authentication is used, each remote host should have its own unique key and KeyExchangeRule specific to that host. If multiple remote hosts are identified by the same KeyExchangeRule, the RSA signature method of peer authentication should be used.

Identity information: Multiple identities can be assigned down to the level of individual IP addresses for each IKE negotiation, or for ease of configuration, a single identity can be assigned to represent the IKE daemon for the entire system. With either method, the identity that is assigned must be an identity that is known to the remote IKE peer. Similarly, the local server must know the identity of any remote IKE peer with which it is to negotiate. The identity can be one of the following types:

- X500dn (the host's X.500 distinguished name)
- IpAddr (an IPv4 address)

- Fqdn (a fully qualified domain name)
- UserAtFqdn (an e-mail address)

If the RSA signature method of peer authentication is used, the local identity must be in an X.509 digital certificate on the IKE daemon's key ring. If the pre-shared key method of peer authentication is used, any of the identity types can be used to identify this IKE peer. Because digital certificates are not used in the pre-shared key method, the only requirement for the use of an identity type is that it be known to both hosts.

Identity information is configured using the LocalSecurityEndpoint and RemoteSecurityEndpoint statements in an IP security policy configuration file. For more details about the LocalSecurityEndpoint and RemoteSecurityEndpoint statements, refer to *z/OS Communications Server: IP Configuration Reference*.

RSA signature: RSA signature is an authentication method that uses X.509 digital certificates. An X.509 digital certificate contains information that was verified by a certificate authority to uniquely identify a host. IKE peers exchange certificates during the phase 1 negotiation to identify each other. Of the information that is typically included in an X.509 digital certificate, there are four fields that are relevant to an IKE exchange:

Subject Name
A certificate's subject name is the unique name by which the host is known. The subject name is used to extract the host's X.500 distinguished name (X.500dn). IKE can use the X.500dn as an identifier for the remote host, or use one of the certificate's Subject Alternative Names.

Subject Alternative Name
An optional field, X.509 extensions, can contain a Subject Alternative Name field. Although the X.500dn is the most specific way to identify a certificate, the Subject Alternative Name can be used as well, often acting as a simple alias to identify the certificate. If the certificate includes optional X.509v3 extensions, the certificate can contain any or all of the following:

- IpAddr - an IPv4 address
- Fqdn - a fully qualified domain name
- UserAtFqdn - an e-mail address

Subject Public Key Info
The public key is used for encrypting information that can be decrypted only with the certificate's private key. Likewise, information that is encrypted using the certificate's private key can be decrypted only with the public key. Using this scheme, information from the owner can be encrypted, digitally signed, and authenticated. IKE uses the public key to assist in verifying a host's identity.

Issuer Name
Certificate signing is a method by which a certificate is verified by a trusted third party. A signed certificate contains the subject name and key identifier of the certificate authority that signed it, known as the issuer.

Because there are many commercial and private certificate authorities, it is possible for a host to own multiple certificates that are signed by different certificate authorities. Depending on a site's policy, only certain certificate authorities might be trusted. Therefore, a z/OS IP security host might request that a peer use only certificates that are signed by a specific trusted certificate authority. Two

configuration parameters refer to certificate authorities, the SupportedCertAuth parameter of the iked.conf file, and the CaLabel parameter of the RemoteSecurityEndpoint statement. The SupportedCertAuth parameter names a specific certificate authority that is recognized by the IKE daemon. The CaLabel parameter of the RemoteSecurityEndpoint statement indicates that a remote peer should use only certificates that are signed by a certificate authority in the list of CaLabels.

Pre-shared key: The pre-shared key method of authentication enables a remote host to authenticate itself by providing a secret key, which is known to both hosts. This key is pre-configured by the administrator, and is used along with the Diffie-Hellman shared secret to derive cryptographic keys used to protect and authenticate data that flows during the phase 1 negotiation. The pre-shared key is a shared secret between the two IKE peers, and any host that does not know the shared key cannot enter into negotiation. IKE maintains a list of all the remote hosts that are authorized to negotiate. This list contains the identity of the remote host and the pre-shared key known to that host.

Negotiation modes for phase 1: The phase 1 negotiation that takes place between two IKE peers happens in one of two modes, Main mode or Aggressive mode.

Main mode is more secure because it encrypts the identities of the two hosts that are contained in the IKE messages, but somewhat slower because more message exchanges are required. Main mode requires a total of six messages, three from the initiator and three from the responder.

Aggressive mode is faster, in that fewer messages are exchanged. Aggressive mode requires only three messages, two from the initiator and one from the responder. However, the identity of the two hosts is not protected in Aggressive mode. An IKE implementation is not required to support Aggressive mode.

The choice of mode depends in part on the security needs of the installation. If it is important that the identities of either host are not to be in clear text on the network, use Main mode. Otherwise, if both hosts support it, Aggressive mode can be used for faster, more efficient key exchange. The choice of negotiation mode is configured using the KeyExchangeAction statement in an IP security policy configuration file. For more details about the KeyExchangeAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Phase 2

The purpose of phase 2 negotiation is to establish a set of parameters known as a security association, which is used to protect specific types of IP traffic. The phase 2 security association contains the keys that are used to encrypt and decrypt IPSec packets on the host, authenticate IPSec packets on the host, or both. The phase 2 security association is negotiated for a specific set of data endpoints for a specific type of traffic, and contains the following:

- The keys that are used to encrypt, if encryption is being used
- The keys that are used to authenticate
- The data endpoints, either a single IP address or range of IP addresses
- The protocol of the traffic to be protected, either a single protocol or all protocols
- The ports of the traffic to be protected, either a single port or all ports
- The IPSec protocol that is used to protect the data: AH, ESP, or both
- The type of authentication algorithm to be used
- The type of encryption algorithm to be used, if encryption is being used

- How to build the IPSec packets (tunnel, transport, UDP-Encapsulated-Tunnel, or UDP-Encapsulated-Transport)
- How often the keys should be refreshed
- A security parameter index (SPI) value, used together with the remote security endpoint IP address to uniquely identify the security association
- The Diffie-Hellman group for perfect forward secrecy (PFS)

A phase 2 negotiation can begin only after the completion of a corresponding phase 1 security association. The encryption methods that are agreed upon during the phase 1 negotiation are used to protect the data that is exchanged during the phase 2 negotiation. For instance, if the KeyExchangeRule between two security endpoints specified SHA1 and 3DES, the IKE data that is exchanged during the phase 2 negotiation is authenticated using HMAC-SHA and encrypted using 3DES encryption. Although both phase 1 and phase 2 security associations might use the same authentication and encryption methods, this is not required. For instance, a phase 1 security association can specify SHA1 authentication and 3DES encryption, while a phase 2 security association might use ESP with HMAC-MD5 authentication and DES encryption.

The keys that are generated during the phase 2 negotiation can be derived from the phase 1 master key to amortize the cost of the phase 1 key generation. As an alternative, you can configure the phase 2 negotiation to use perfect forward secrecy (PFS) for stronger security. Each has its advantage as follows:

- If PFS is used, phase 2 negotiation does not derive its keys from the master key, but instead generates new keying material using the Diffie-Hellman algorithm. Because it is independently derived, the resultant phase 2 key is more secure.
- If PFS is not used, phase 2 negotiation is completed much more quickly, but the resultant phase 2 key is less secure.

In either case, phase 2 does not incur the same degree of processing overhead that is involved in phase 1 negotiation with the remote IKE peer.

When phase 2 negotiation has completed, the security association is available for use by the stack. Now, any traffic that is mapped to this security association by an IP filter rule is IPSec-protected. Since each phase 2 security association corresponds to a single unique phase 1 security association, the identity of the remote peer is implicitly authenticated when the phase 2 security association is used.

Policy for phase 2 security associations is defined by referencing an IpDynVpnAction statement on an IpFilterRule statement. For more details on referencing an IpDynVpnAction statement on an IpFilterRule statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Only tunnel or transport mode encapsulation can be specified on the IpDataOffer statement. The decision to use UDP-Encapsulated-Transport mode or UDP-Encapsulated-Tunnel mode is made heuristically by the z/OS IKE daemon. If a NAT is detected in the process of creating a phase 1 security association, all phase 2 security associations negotiated under the protection of that phase 1 security association are negotiated utilizing UDP encapsulation. In this case, the z/OS IKE daemon internally converts all IpDataOffer statements containing ESP with tunnel mode encapsulation to UDP-Encapsulated-Tunnel mode, and IpDataOffer statements containing ESP with transport mode encapsulation to UDP-Encapsulated-Transport mode. Any IpDataOffer statements that are configured to use AH authentication are ignored, since the IPSec protocols do not allow for AH authentication when NAT is being used.

The phase 2 parameter values supported can differ between IPSec implementations. For example, z/OS provides configuration to negotiate a phase 2 security association for specific port and specific protocol values. Many implementations of IPSec support negotiating only a wide security association, covering all ports and protocols.

IPSec and Network Address Translation (NAT) devices

There are inherent incompatibilities between IPSec and Network Address Translation (NAT) functions. RFC 3715, *IPsec-Network Address Translation (NAT) Compatibility Requirements*, provides a description of the problems that arise when IPSec is used to protect traffic that traverses a NAT. One basic problem is that when IPSec security associations traverse a NAT, the NAT is unable to update IP addresses and checksums that are part of the encapsulated data (encrypted, authenticated, or both). RFCs 3947 and 3948 define mechanisms that enable specific uses of IPSec to traverse one or more NAT devices.

- RFC 3947, *Negotiation of NAT-Traversal in the IKE*, allows an IKE daemon to detect when one or more NATs are being traversed.
- RFC 3948, *UDP Encapsulation of IPsec ESP Packets*, defines two IPSec encapsulation modes, UDP-Encapsulated-Tunnel mode and UDP-Encapsulated-Transport mode. These modes facilitate the traversal of IPSec traffic through a NAT by encapsulating ESP packets within a UDP packet.

This support is known as NAT traversal (NATT).

UDP-Encapsulated-Tunnel mode and UDP-Encapsulated-Transport mode security associations are negotiated by IKE when NAT traversal is enabled and a NAT is detected. Manually configured security associations do not support UDP-Encapsulated-Tunnel mode or UDP-Encapsulated-Transport mode.

There are several reasons why Network Address Translation might be used. One reason is to economize on the use of public addresses within the internal network, using a public address only when data must be globally routed. A second reason is to hide internal IP addresses from network segments outside the internal IP address domain.

It should be noted that the NAT traversal support, as defined by the IETF, transmits internal IP addresses to its IPSec peer. These internal addresses are not exposed on the external network. However, the internal addresses are available for display at the remote security endpoint. If you are considering using this support, evaluate whether transmitting internal IP addresses to an IPSec peer is acceptable from a security policy perspective.

There are several terms used to describe various IP address mapping techniques. The term NAT describes a technique where a 1-to-1 address translation function is performed. An internal-external IP address mapping is maintained by the NAT device. When NAT is used, IP addresses are translated, but ports are unchanged. The mapping can be static or dynamic. For a static mapping, there is a definition in the NAT that always translates IP address x.x.x.x to IP address y.y.y.y. An outbound packet is not needed to establish the mapping. For a dynamic mapping, the NAT has a pool of IP addresses that are assigned as needed, so IP address x.x.x.x might be mapped to IP address y.y.y.y one time, and to IP address z.z.z.z at another time. The mapping is established when an outbound packet is processed.

The term network address port translation (NAPT) describes a technique where multiple internal IP addresses are translated into a single public IP address. As

part of this translation process, the TCP and UDP ports in the packets are translated. NAPT is sometimes referred to as port address translation (PAT).

The terms *in front of* and *behind* are used to convey which IP address a NAT is translating. When a NAT is said to be in front of a client, it means the client's address will be translated by the NAT. When a server is said to be behind a NAT, it means that server's address will be translated by the NAT.

Rule: z/OS Communications Server does not support negotiating security associations that traverse a NAPT.

NATT support level

z/OS supports NAT traversal as defined in RFCs 3947 and 3948. Platforms that have implemented their NAT traversal support using pre-RFC drafts might not inter-operate with implementations that are compliant with RFC 3947 and 3948. z/OS does provide limited support for the following pre-RFC implementations:

- draft-ietf-ipsec-nat-t-ike-02 (pre-RFC draft of RFC 3947), and draft-ietf-ipsec-udp-encaps-02 (pre-RFC draft of RFC 3948)
- draft-ietf-ipsec-nat-t-ike-03 (pre-RFC draft of RFC 3947), and draft-ietf-ipsec-udp-encaps-03 (pre-RFC draft of RFC 3948)

Dynamic structures used to map security associations

The negotiation of a phase 2 security association results in the dynamic creation of new filters used to map the security association to a specific type of traffic. These dynamic filters are then added to the filter table and remain as long as the security association is available for use. Additional structures are added to the filter table in the special case that the remote security endpoint is a gateway behind a NAT device.

Anchor filters and dynamic filters

Filters with ipsec actions are flagged as anchor filters. Anchor filters are neither permit or deny rules, but rather serve as place holders for dynamic filters in the ordered list of filter rules. A dynamic filter is an extension of an anchor filter and is created when a phase 2 security association is created. Each phase 2 security association that is negotiated is associated with two dynamic filters, an inbound filter and an outbound filter. When an IP packet matches an anchor filter rule, there is a secondary search for a matching dynamic filter rule. If one is not found, the packet is denied, unless the packet is an outbound packet and on-demand negotiations are allowed. In that case, an IKE negotiation ensues to create a security association and the matching dynamic filter. If a dynamic filter already exists, the action taken is to permit with IPSec processing applied. The dynamic filter rule indicates which security association should be used when applying IPSec processing, because there is a 1-to-1 correspondence between dynamic filter pairs (inbound and outbound) and phase 2 security associations. For a sample display of anchor and dynamic filters, see "Displaying active filters with the ipsec command" on page 958.

NATT anchor and NATT dynamic filters

When the remote peer is a security gateway behind a NAT, the dynamic anchor still serves as a place holder in the ordered list of filter rules. However, the dynamic filters that are created when a phase 2 security association is created are handled differently. When the phase 2 negotiation is successful, the dynamic filter pair that is created contains the 5-tuple information for which the security association was negotiated:

- The data endpoints, either a single IP address or range of IP addresses

- The protocol of the traffic to be protected, either a single protocol or all protocols
- The ports of the traffic to be protected, either a single port or all ports

For the case where the remote peer is a security gateway behind a NAT, this 5-tuple might not be unique for a security association. An additional structure, a NATT anchor, is generated to anchor dynamic filters that share the same 5-tuple information. The dynamic filter is then an extension to the NATT anchor and is flagged as a NATT dynamic. For a sample display of NATT anchor and NATT dynamic filters, see"Displaying active filters with the ipsec command" on page 958.

NAT resolution filters (NRF)

When the remote peer is a security gateway behind a NAT, a connection level filter structure is also created for TCP and UDP connections, the NAT resolution filter (NRF). Unlike the NATT anchor and NATT dynamic filters, which are created when the phase 2 security association is created, the NRF is created when the first inbound packet is received for the connection over the phase 2 security association. The NRF contains a single source and destination IP address, a single source and destination port value, and a single protocol. This connection-level filter is needed to determine the phase 2 security association that will be used for outbound data. For a sample display of NAT resolution filters, see "Displaying active filters with the ipsec command" on page 958.

Remote port translation

When the remote peer is a security gateway behind a NAT, the remote data endpoint of the client is represented by the security gateway's public IP address. From the z/OS server's perspective, multiple clients residing behind the security gateway are all identified by the IP address of their corresponding security gateway, not the client's private IP address. Consequently, if two clients behind the security gateway were to open a connection to the same service (FTP, for example) from the same ephemeral port (1024, for example), the two connections could not be uniquely identified. Traditionally, the combination of local and remote IP addresses and ports would be enough to distinguish the connections. Because in this instance, they would appear to be coming from the same IP address, they would look identical. To avoid such conflicts, Communications Server provides remote port translation for TCP and UDP connections when needed. Without remote port translation, the first inbound TCP connection request for a unique 5-tuple would succeed. However, a subsequent inbound connection request using the same 5-tuple would fail. Communications Server's remote port translation allows subsequent inbound connections using the same 5-tuple to succeed by translating the remote port value to a different ephemeral port.

Remote port translation does not need to be configured or enabled. It is built into Communications Server's NAT traversal support and is always in effect when the remote peer is a security gateway behind a NAT. When a remote port is translated, message EZ0827I, remote port translated, is logged. Remote port translation determines whether the source port in an inbound packet represents a duplicate port. If it does, an attempt is made to assign an unused ephemeral port value. If the inbound packet matches a configured filter rule that covers all ports, any unused ephemeral port value can be assigned. If the configured filter rule applies only to a range of ports, the remote port is translated only to an unused ephemeral port in that range. If an unused port cannot be found, the connection request fails. When a remote port value is translated, there is both an original remote port and translated remote port for a connection. Various commands, such as Netstat, enable you to view connection information including the port, or even to select display

data based on port. For connection data, if only one remote port is being provided, the translated port is displayed or used for a selection, if remote port translation has been done.

The **ipsec -o display** command provides a display of port mappings in effect. For a sample display of remote port translation, see "Displaying remote port translation with the ipsec command" on page 970.

Steps for preparing the z/OS system for IP security

Before you begin: You need to be familiar with the concepts of IP filter logging and IPSec protection. See "Overview of using IP security" on page 824. You also need to be familiar with the commands used to administer IP security. See "Commands used to administer IP security" on page 823.

Perform the following steps to prepare the z/OS system for IP security.

1. Develop a site policy to address the security needs of your installation.

   The following are some questions that you must consider before beginning to build a robust IP security policy:

   - Do you want a default-allow or default-deny policy?
   - If a default-deny policy is desired, what traffic is allowable as critical to the proper functioning of the system?
   - What hosts are allowed to send data to the secure host?
   - Do you understand the network topology, and where the communicating hosts are located within the network?
   - Are there Network Address Translation (NAT) devices in the network topology?
   - What type of traffic is allowable from those hosts?
   - What general network services that rely on well-known ports do you want to allow?
   - Will you forward any packets you receive that are not destined for you?
   - Who is allowed to configure the IP security policy?
   - Are you running with one TCP/IP stack, or more than one TCP/IP stack?
   - Will all TCP/IP stacks share the same policy definitions, or will they have unique differences?
   - Will traffic that is sent or received by the secure host traverse the Internet?
   - Do you want to participate in a VPN? (If so, IPSec is required.)
   - If IPSec is required:
     - What are the IPSec capabilities of the remote endpoints?
     - How do you want to authenticate participating hosts:
       - Pre-shared key?
       - RSA signature? (requires digital certificates)
     - What type of authentication is needed for the data that is involved?
     - Is encryption needed for the data that is involved, and how strong should the encryption algorithms be?
     - Will all participating hosts use the same level of data encryption and authentication, or do you need to define unique policies for individual hosts?

2. Identify the resources to be secured.

Create a worksheet for each TCP/IP stack you will enable for IP security. This information aids in determining which interfaces and connected networks are secured. You can later use the information from the worksheets to provide values for IP security policy configuration statements. Figure 83 includes a sample worksheet:

```
Host name of z/OS system _____________________________________

Complete for each TCP/IP stack on this host:
TCP/IP stack name _____________IPSECURITY-enabled (Y/N)______
Network interface(s)
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________


Identities, other than IPv4 address, by which the IKE daemon will be known
(e.g., X500dn, Fqdn, UserAtFqdn)
____________________
____________________
____________________
____________________
____________________
```

Figure 83. Sample worksheet for stack security

Security class is an optional designation for a network interface. You can assign network interfaces a security class (1-255) that is used to group interfaces with similar security requirements. This concept is an extension of the traditional notion of secure and nonsecure interfaces, to allow for more than two classes. The secure and nonsecure model can still apply if only two security classes are defined.

Complete a table of remote hosts and subnetworks with which this host needs to transfer data, as shown in Table 34. The remote hosts can optionally be grouped in a user-defined zone to simplify the number of IP filter rules that are required. For instance, you might want to define an internal, external, trusted, Internet, business partner, or other zone that has meaning to your site.

Rule: If the remote IKE peer resides on a security gateway to the remote host, the IP address of the remote host might not match the IP address of the remote IKE peer. If this is the case, tunnel-mode encapsulation of IPSec traffic is required.

Table 34. Table of remote hosts and subnetworks

Remote IP address or subnet	User-defined zone of remote hosts	Identity of remote IKE peer (IpAddr, X500dn, Fqdn, or UserAtFqdn)	IPv4 address of remote IKE peer	Allowed list of services (Telnet, FTP, Web, EE, ALL, or other)	Security action (permit, deny, ipsec)	IPSec security level, if applicable

3. Modify the default IP filter policy (optional).

   After the site policy is established, you can modify the default IP filter policy if necessary. For a description and examples of how to modify the default IP filter policy, see "Default IP filter policy and IP security policy" on page 830.

   Guideline: You should define IPSECRULE statements that permit access from at least one administrative machine. In the event that Policy Agent is unavailable, or the IP security configuration files contain errors, the active default policy would deny access to the IP security-enabled stack. Should this situation occur, if you have added the appropriate IPSECRULE statements to the default policy, you can still access the secure z/OS host and make the necessary administrative changes to correct the problem.

4. Set up the key required applications:

   - TCP/IP

     To enable IP security on a z/OS stack, make the following changes in the TCP/IP profile:

     – Add IPCONFIG IPSECURITY.

     – Reserve ports 500 and 4500 for the IKE daemon. The ports should be reserved for the user ID that the IKE daemon is running under. In this example, the IKE daemon is running under the IKED user ID:

       ```
       500   UDP        IKED
       4500  UDP        IKED
       ```

     For information on the IPCONFIG and PORT statements, refer to *z/OS Communications Server: IP Configuration Reference*.

   - Policy Agent

     For information on configuring the Policy Agent, see Chapter 14, "Policy-based networking," on page 717.

   - TRMD

     For information on configuring TRMD, see "TRMD" on page 815.

   - Syslogd

     For information on configuring the syslog daemon, see "Configuring the syslog daemon (syslogd)" on page 175.

   - IKED

     IKED can be started from a z/OS UNIX command line or as an MVS procedure. The iked.conf file controls the overall behavior of the IKE daemon, including:

     – The logging level of the IKE daemon

     – The logging level of the Policy Agent when performing IP security-related tasks on behalf of the IKE daemon

     – The name of the RACF key ring that is owned by the IKE daemon

     – Whether IKE messages are echoed to STDOUT when the daemon is started for the UNIX System services shell

     – How long to wait when attempting to connect to the Policy Agent

     – The list of certificate authorities that are acceptable for RSA signature negotiation

     Most of the configuration parameters have a default value and do not require modification.

     Rule: If the RSA signature method is to be used in any phase 1 IKE negotiation, the name of the IKE key ring must be included in the iked.conf file.

Tip: If the RSA signature method is used, including a list of supported certificate authorities enhances the performance of certificate searches.

Guideline: The logging levels of the IKE daemon (IkeSyslogLevel) and Policy Agent (PagentSyslogLevel) should not be changed from their default values for normal day-to-day operation. Higher logging levels can affect performance and should be used for temporary diagnostic purposes only. The default PagentSyslogLevel of 0 prevents the IKE daemon from logging diagnostic information about its interactions with Policy Agent. The default IkeSyslogLevel of 1 provides basic informational and error messages. The IkeSyslogLevel can be set to 0 to disable IKE syslog messages entirely. It can be set higher to identify the source of an error; for example, if you experience problems with security association negotiations due to a configuration error, you might enable IkeSyslogLevel 4 (debugging information for security association negotiations).

For detailed syntax and a description of the iked.conf file, and details on starting the IKE daemon as an MVS procedure, refer to *z/OS Communications Server: IP Configuration Reference*.

5. Define access controls for key required applications:
   - TRMD

     TRMD runs as an authorized program and requires RACF setup. TRMD must be able to run as a started task and have superuser authority. For sample RACF commands, refer to the EZARACF member of SEZAINST.
   - Policy Agent

     Policy Agent runs as an authorized program and requires RACF setup. Policy Agent must be able to run as a started task and have superuser authority. For sample RACF commands, refer to the EZARACF member of SEZAINST and "Step 1: General configuration" on page 741.
   - IKED

     For the steps to prepare for running the IKE daemon, see Appendix E, "Steps for preparing to run IP security," on page 1215.
6. Configure the IKE daemon to use digital certificates (optional).

   See Appendix E, "Steps for preparing to run IP security," on page 1215.
7. Configure additional encryption products (optional).
   - 3DES support

     For triple DES, the IP security level 3 feature, FMID JIP614K, is required.
   - ICSF

     There are several options available on z/OS to perform encryption in hardware. The ICSF product is required to support these various options. For a description and information regarding the configuration of these options, refer to *z/OS Cryptographic Services ICSF Administrator's Guide*.

     For the RACF commands to authorize ICSF, see Appendix E, "Steps for preparing to run IP security," on page 1215.
8. Create IP security policy configuration files.

   After the security needs of your installation are established, the next step is to create one or more IP security policy configuration files. There are two main configuration files in which IP security policy configuration is stored:

   IpSecConfig
   : Configured on a per-stack basis and contains only IP security policy configuration that applies to the stacks for which it is configured.

CommonIpSecConfig
Contains IP security policy configuration that applies to every stack on the system and can be used to hold shared definitions.

The IpSecConfig file can refer to various definitions in the CommonIpSecConfig file, and can override policy definitions in the CommonIpSecConfig file.

Although Policy Agent uses LDAP to store policy for some other policy types, IP security policy configuration is stored exclusively in human-readable text files, either on an HFS or in an MVS data set. These configuration files are then read by Policy Agent when it initializes and before being installed into the stack.

IP security policy configuration

This section describes the general steps for creating IP security policy configuration files for the most common configurations. Configuring a complete and specific IP security policy that meets the needs of any installation is beyond the scope of this text, but guidance for more advanced configurations is provided.

"Overview of configuring IP security policy" describes the CommonIpSecConfig file, the IpSecConfig file, and the general content, structure, and use of these files.

"Component policies of IP security policy configuration files" on page 864 describes the types of policies contained in IP security policy configuration files.

"Steps for configuring IP security policy" on page 887 describes the steps for manually creating IP security policy configuration files.

"Quick start using IP filtering and IPSec host-to-host" on page 874 describes a complete IP security policy allowing connections from a secure server to an administrative machine on an internal network, and represents the minimum configuration needed to provide IPSec protection with dynamic key management between two hosts. This section also describes the use of the **ipsec** command to display filters and security associations.

"Configuring specific security models" on page 889 provides more examples and describes the configuration needed for common security models.

Overview of configuring IP security policy

There are three options for IP security policy configuration for a system:

- Use a CommonIpSecConfig file that applies to all stacks on the system, enforcing a consistent policy. In this instance, an IpSecConfig file is not necessary, but the IpSecConfig statement must still be specified.
- Use a unique and separate IpSecConfig file for each stack on the system. In this instance, a CommonIpSecConfig file is not necessary.
- Use both a CommonIpSecConfig and an IpSecConfig file.
  - The CommonIpSecConfig file can be used as a common repository for frequently used definitions, which can be referenced by any IpSecConfig file.
  - The IpSecConfig file can contain unique statements that apply only to the stack for which it is configured, and can reference statements that are defined in the CommonIpSecConfig file.

Although not an error, note that when using the last approach, it is possible for duplicate statements to exist in the CommonIpSecConfig file and the IpSecConfig

file (for example, two IpFilterRule statements with the same name). In this case, the statement in the IpSecConfig file is honored. Statements in the stack-specific IpSecConfig file always take precedence over the CommonIpSecConfig file.

Structure of an IP security configuration file

The CommonIpSecConfig file and the IpSecConfig file have exactly the same structure. They are comprised of a number of statements that define items that are used to define policy, such as policies, rules, actions, groups, and objects. Statement names and attribute names are not case sensitive, though they appear in mixed case in this book for readability. Only user-defined names are case sensitive. For the complete syntax of all IP security policy statements, refer to *z/OS Communications Server: IP Configuration Reference*.

An IP security policy configuration statement has the following generic form:

```
StatementType     user-defined name
{
   Attribute1     value1
   Attribute2     value2
   .
   .
   .
}
```

Statements often contain other inline statements in a recursive form:

```
StatementType1     user-defined name
{
   Attribute1      value1
   StatementType2  optional user-defined name
   {
      Attribute1   value1
      Attribute2   value2
   }
   Attribute2      value2
}
```

There are three main sections in an IP security configuration file, identified by the following three statements:

- IpFilterPolicy
- KeyExchangePolicy
- LocalDynVpnPolicy

Additional statements that define rules, actions, groups, and objects are found both in the main body of the configuration file and within any of these other three policy blocks. A high-level view of an IP security configuration file follows. Although the statement blocks are shown in a specific order, the ordering is arbitrary.

```
IpFilterPolicy #(required)
{
  <local statements>
}

KeyExchangePolicy #(optional)
{
  <local statements>
}

LocalDynVpnPolicy #(optional)
{
```

```
  <local statements>
}

  <global statements>
```

Groups: Groups provide a method to combine related objects in a meaningful way into sets. The following IP security policy configuration statements can be used as groups:

- IpAddrGroup
- IpFilterGroup
- IpServiceGroup
- KeyExchangeGroup
- LocalDynVpnGroup

Reference statements: A reference statement can be recognized by the suffix Ref. References provide a convenient way to reuse definitions, eliminating the need to repeatedly specify things such as host addresses and common services. Nearly all statements in the IpSecConfig file can be referenced, and all action statements must be referenced. To be referenced, an IP security policy configuration statement must be given a user-defined name. Names of statements can be up to 32 characters. The following IP security policy configuration statements can be referenced, and therefore reused:

- IpAddr
- IpAddrSet
- IpDataOffer
- IpDynVpnAction
- IpFilterGroup
- IpFilterRule
- IpGenericFilterAction
- IpLocalStartAction
- IpManVpnAction
- IpService
- IpServiceGroup
- IpTimeCondition
- KeyExchangeAction
- KeyExchangeGroup
- KeyExchangeOffer
- KeyExchangeRule
- LocalDynVpnGroup
- LocalDynVpnRule
- LocalSecurityEndpoint
- RemoteSecurityEndpoint

Steps for configuring IP security policy using only a CommonIpSecConfig file

Perform the following steps to configure IP security policy using only a CommonIpSecConfig file.

1. In the main Policy Agent configuration file, include a line that identifies the CommonIpSecConfig file, as follows:

```
CommonIpSecConfig      /etc/common.ipsecpol
```

2. In the main Policy Agent configuration file, include a line with the TcpImage statement for each IP security stack to be configured:

```
TcpImage     /etc/TCPCS.image
TcpImage     /etc/TCPCS2.image
⋮
```

3. In each configuration file that was identified in the second field shown in step 2, include a line that contains IpSecConfig with no file name, as follows:

```
In /etc/TCPCS.image:
IpSecConfig

In /etc/TCPCS2.image:
IpSecConfig
```

All stacks on the z/OS system will adhere to the policy that is specified in the /etc/common.ipsecpol file.

Steps for configuring IP security policy using only a stack-specific IpSecConfig file

Perform the following steps to configure IP security policy using only a stack-specific IpSecConfig file.

1. In the main Policy Agent configuration file, include a line with the TcpImage statement for each stack to be configured, as follows:

```
TcpImage     /etc/TCPCS.image
TcpImage     /etc/TCPCS2.image
⋮
```

2. In each configuration file that was identified in the second field shown in step 1, include a line that identifies the stack-specific IpSecConfig file, as follows:

```
In /etc/TCPCS.image:
IpSecConfig     /etc/TCPCS.ipsecpol

In /etc/TCPCS2.image:
IpSecConfig     /etc/TCPCS2.ipsecpol
```

Each stack on the z/OS system will adhere to the policy that is specified by its unique policy file. Stack TCPCS uses the policy that is configured in /etc/TCPCS.ipsecpol, while stack TCPCS2 uses the policy that is configured in /etc/TCPCS2.ipsecpol.

Steps for configuring IP security policy configuration using both approaches

Perform the following steps to configure IP security policy configuration using both approaches.

1. In the main Policy Agent configuration file, include a line that identifies the CommonIpSecConfig file, as follows:

```
CommonIpSecConfig    /etc/common.ipsecpol
```

2. In the main Policy Agent configuration file, include a line with the TcpImage statement for each stack to be configured, as follows:

```
TcpImage   /etc/TCPCS.image
TcpImage   /etc/TCPCS2.image
⋮
```

3. In each configuration file that was identified in step 2 on page 863, include a line that identifies the stack-specific IpSecConfig file, as follows:

```
In /etc/TCPCS.image:
IpSecConfig     /etc/TCPCS.ipsecpol

In /etc/TCPCS2.image:
IpSecConfig     /etc/TCPCS2.ipsecpol
```

Any statements in the CommonIpSecConfig file are added to the policy for each stack when the policy is initialized. Either file, /etc/TCPCS.ipsecpol or /etc/TCPCS2.ipsecpol, can refer to statements in /etc/common.ipsecpol. In the case of duplicate names, any named statement in the stack-specific IpSecConfig file overrides a statement with the same name in the CommonIpSecConfig file.

Component policies of IP security policy configuration files

There are three types of policies in IP security policy configuration files:

- IP filter policy (IpFilterPolicy statement)
- Key exchange policy (KeyExchangePolicy statement)
- Local dynamic VPN policy (LocalDynVpnPolicy statement)

IP filter policy

An IP filter policy can stand alone to provide IP filtering and IPSec protection with manual key management. Used in conjunction with the two other policies, it is also required to provide IPSec protection with dynamic key management (IKE). Because filtering is crucial to secure traffic on a host, an IP security policy that contains no IpFilterPolicy statement block or an empty IpFilterPolicy statement block is considered an error, leaving the default policy that is provided by the stack in effect.

The IpFilterPolicy statement block consists of:

- A set of global configuration options
- An ordered list of IP filter rules (IpFilterRule statements)

The purpose of the global configuration options is to control global policy items, such as whether logging is active or whether on-demand security association negotiations are allowed, and so forth. These global options apply to all of the IP filter rules that are contained in the policy. Each IP filter rule, in turn, contains data endpoints, traffic descriptions, and actions. When a packet entering or leaving the system matches the data endpoints and traffic description in an IP filter rule, the associated action is taken. If the action is an ipsec action, additional action statements are coded that define the parameters of the IPSec security association.

Following is a sample IpFilterRule statement that allows Web traffic on an internal server, and a description of each line in the sample:

```
1  IpFilterRule             InternalNetWeb
2       {
3          IpSourceAddr          9.1.1.1
4          IpDestAddrSet         9.1.1.0/24
5          IpService
6          {
7             SourcePortRange        80
8             DestinationPortRange   1024 65535
9             Protocol               tcp
```

```
10           Direction              bidirectional InboundConnect
11           Routing                local
12           SecurityClass          0
13        }
14        IpGenericFilterActionRef  permit-nolog
15     }
```

Line	Description
1	The IpFilterRule keyword, followed by a required user-defined name for this rule.
2	An open brace ({) marks the start of an IpFilterRule statement block.
3	The source address of the rule. Outbound IP packets that match this rule must have 9.1.1.1 as the source address in the IP header.
4	The destination address of the rule. Outbound IP packets that match this rule must have an address in the range of 9.1.1.0 - 9.1.1.255 as the destination address in the IP header.
5	The IpService statement block describes the type of traffic that is allowed between the two data endpoints. The IpService block in this rule is inline to the rule, meaning that the entire definition of the IP service is included within the rule. Many policy statements, including the IpService statement, can be referenced rather than included inline.
6	An open brace ({) marks the start of the IpService statement block.
7	The range of source ports that is allowed for outbound packets. The value can be a single number, or a range of ports.
8	The range of destination ports that is allowed for outbound packets. The value can be a single number, or a range of ports.
9	The specific protocol that is allowed by this rule.
10	The direction specification for the IP packet. Bidirectional indicates that this rule allows outbound traffic from local address 9.1.1.1 on local port 80 to any address in subnet 9.1.1.0/24 using any ephemeral port (that is, 1024-65535), and inbound traffic from any address in subnet 9.1.1.0/24 using any ephemeral port to local address 9.1.1.1 on local port 80. Without the use of the bidirectional keyword, it would be necessary to create two filter rules, one for outbound traffic and one for inbound traffic. InboundConnect indicates that the rule will match inbound TCP connection attempts as well as bidirectional data on an established connection, but it will not match outbound TCP connection attempts.
11	The routing information for a packet matching this rule. For a packet to match this rule, the traffic must be local. In other words, this rule does not allow any packets that must be forwarded to another network node. Possible values are local, routed, or either.
12	The security class of the interface on which the packet must arrive or leave. The security class of an interface is defined using the SECCLASS parameter on the LINK statement or IPCONFIG DYNAMICXCF statement for that interface in the TCP/IP profile. Each interface on the system can be assigned a SECCLASS between 1 and 255 inclusive. If the SECCLASS parameter is not coded on the LINK statement or IPCONFIG DYNAMICXCF statement for an interface, the interface defaults to a

SECCLASS of 255. In the SecurityClass parameter above, the value 0 indicates that a packet that matches this rule is not restricted and can traverse any interface (1-255).

13 A close brace (}) marks the end of the IpService block.

14 The action that is taken on a packet that matches this rule. In this case, permit the packet and log all occurrences of a match. All IP filter rules must include an IpGenericFilterActionRef statement. The IpGenericFilterAction statement itself should be defined elsewhere, outside of the IpFilterRule block, either in the CommonIpSecConfig file or in the IpSecConfig file:

```
IpGenericFilterAction     permit-nolog
{
  IpFilterAction          permit
  IpFilterLogging         no
}
```

15 A close brace (}) marks the end of the IpFilterRule statement block.

Rule: Do not include policy action statements inline. They must be referenced.

Example 1: Permit rule allowing outbound FTP client connections from the local host (9.1.1.1) to a remote FTP server (9.1.1.2):

```
IpFilterRule              FTP-client
{
      IpSourceAddr           9.1.1.1
      IpDestAddr             9.1.1.2
      IpService
      {
        SourcePortRange      1024 65535
        DestinationPortRange 21
        Protocol             tcp
        Direction            bidirectional OutboundConnect
        Routing              local
        SecurityClass        0
      }
      IpService
      {
        SourcePortRange      1024 65535
        DestinationPortRange 20
        Protocol             tcp
        Direction            bidirectional InboundConnect
        Routing              local
        SecurityClass        0
      }
      IpGenericFilterActionRef  permit
}
```

Normal (non-passive mode) FTP requires that the FTP client be allowed to initiate outbound connections to port 21, and be able to receive inbound connections from port 20. The `IpGenericFilterAction permit` block must be defined elsewhere, in either the IpSecConfig file or the CommonIpSecConfig file:

```
IpGenericFilterAction     permit
{
  IpFilterAction          permit
}
```

Example 2: Deny rule that blocks all traffic from all private address spaces that is inbound to a public interface:

```
IpFilterRule              deny-private
{
      IpSourceAddrGroupRef   PrivateAddrs
```

```
      IpDestAddr             all
      IpService
      {
        SourcePortRange      0
        DestinationPortRange 0
        Protocol             all
        Direction            inbound
        Routing              either
        SecurityClass        0
      }
      IpGenericFilterActionRef  deny-log
}
```

The IpSourceAddrGroupRef parameter references an IP address group that is presumed to be defined elsewhere, in either the IpSecConfig file or the CommonIpSecConfig file:

```
IpAddrGroup   PrivateAddrs
{
   IpAddrSet  10.0.0.0/8
   IpAddrSet  172.16.0.0/12
   IpAddrSet  192.168.0.0-192.168.255.255
}
```

The IpGenericFilterActionRef parameter references an IpGenericFilterAction statement that is presumed to be defined elsewhere, in either the IpSecConfig file or the CommonIpSecConfig file:

```
IpGenericFilterAction     deny-log
{
  IpFilterAction          deny
  IpFilterLogging         yes
}
```

Example 3: An ipsec rule that requires IPSec protection for all traffic between the secure server and an administrative machine on the internal network:

```
IpFilterRule                Rule2Admin
{
     IpSourceAddrRef            InternalServerAddressA1
     IpDestAddrRef              AdminClient
     IpServiceRef               All-traffic-local
     IpGenericFilterActionRef   ipsec
     IpDynVpnActionRef          Silver-TransportMode
}
```

The use of multiple references in this example makes the IP filter rule easier to read. For each referenced object or action, there should be a corresponding definition elsewhere, in either the IpSecConfig file or the CommonIpSecConfig file:

```
IpAddr                InternalServerAddressA1
{
  Addr                9.1.1.1
}

IpAddr                AdminClient
{
  Addr                9.1.1.2
}

IpService             All-traffic-local
{
  Protocol            all
  Direction           bidirectional
  Routing             local
  SecurityClass       0
}
```

```
IpGenericFilterAction  ipsec
{
  IpFilterAction       ipsec
  IpFilterLogging      yes LogDeny
}

IpDynVpnAction         Silver-TransportMode
{
  Initiation           either
  Pfs                  None
  IpDataOfferRef       SHA-DES-Transport
}

IpDataOffer            SHA-DES-Transport
{
  HowToEncap           transport
  HowToEncrypt         DES
  HowToAuth            ESP HMAC_SHA
}
```

Ordering of IP filter rules: The previous examples show individual IP filter rules. A complete IP filter policy contains any number of IP filter rules, configured in much the same manner. It is important to remember that IP filter rules in an IP filter policy are searched in the order listed. Because it is possible for a packet to match more than one rule, a search for a matching filter rule stops after the first match is found, even if there are additional matches further down in the list. Use the **ipsec** command traffic test option (**ipsec -t**) as an aid in determining which IP filter rule an IP packet matches.

The command-line arguments to the **ipsec -t** command are a set of characteristics that describe a particular IP packet. The existing set of filter rules are searched for potential matches. Unlike normal filter processing, which stops the search after a match is found, the **ipsec -t** command displays all matching filter rules. Input to the **ipsec -t** command does not have to specify all possible filtering criteria from an IP packet. The output of the **ipsec -t** command must be inspected to determine which of the returned rules match for a given case.

For instance, an IP filter rule for ICMP can be configured for a specific type and code value, while the traffic test does not provide ICMP type and code as inputs. If more than one IP filter rule matches on the ICMP protocol, they are all displayed. You must determine, from among those listed, which rule applies for a specific IP packet.

For a complete description of the **ipsec** command, including the **ipsec -t** option, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Key exchange policy

A key exchange policy is required by IKE to provide dynamic key management. It contains the definitions about how the negotiation of keys are to be protected, and which hosts are allowed to negotiate. The absence of a key exchange policy is not considered an error, but without it, the IKE daemon is unable to provide dynamic key management.

A key exchange policy consists of an ordered list of key exchange rules. A key exchange rule consists of a set of security endpoints, and an action to be taken when the two security endpoints engage in a phase 1 IKE negotiation.

Optionally, a key exchange rule can contain a shared key known only to the two negotiating entities that are described in the rule. When an IKE negotiation is initiated, the current list of key exchange rules is searched for a match, based on four criteria:

- The identity of the local IKE peer, if known
- The identity of the remote IKE peer, if known
- The location (IPv4 address) of the local IKE peer, if needed to distinguish it or if local identity is not known
- The location (IPv4 address) of the remote IKE peer, if needed to distinguish it or if remote identity is not known

Following is a sample KeyExchangeRule block that allows an IKE negotiation between IKE daemons at 9.2.2.2 and 9.4.4.4. A description of each line in the sample follows the sample.

```
1   KeyExchangeRule             ZoneB_KeyExRule1
2   {
3      LocalSecurityEndpoint
4      {
5         Identity              IpAddr 9.2.2.2
6         Location              9.2.2.2
7      }
8      RemoteSecurityEndpoint
9      {
10        Identity              X500dn CN=ZoneB Cert,T=IKE ServerB,OU=endicott,O=ibm,C=US
11        Location              9.4.0.0/16
12        CaLabel               CA4endicott
13     }
14     KeyExchangeActionRef     Gold-RSA
15     SharedKey                Ascii TheEagleHasLanded
16 }
```

Line	Description
1	The KeyExchangeRule keyword, followed by a required user-defined name.
2	An open brace ({) marks the beginning of the KeyExchangeRule statement block.
3	The LocalSecurityEndpoint statement identifies a local security endpoint, or local IKE peer.
4	An open brace ({) marks the beginning of the LocalSecurityEndpoint statement block.
5	The identity of the local security endpoint that must match this rule. This can be one of four types: • IpAddr • X500dn • Fqdn • UserAtFqdn In the example, an IP address is used as the identity value.
6	The IPv4 address of the local IKE peer.
7	A close brace (}) marks the end of the LocalSecurityEndpoint statement block.
8	The RemoteSecurityEndpoint statement identifies a remote security endpoint, or remote IKE peer. The RemoteSecurityEndpoint statement can

also be used to define a related group of remote IKE peers by using wildcard values for identity and location.

9 An open brace ({) marks the beginning of the RemoteSecurityEndpoint statement block.

10 The identity of the remote security endpoint that must match this rule. This can be one of four types:

- IpAddr
- X500dn
- Fqdn
- UserAtFqdn

In the example, an X.500 distinguished name is used as the identity value.

11 The IPv4 subnetwork that defines a group of remote IKE peers.

12 Used only for RSA signature peer authentication. Specifies the certificate authority that is advertised to the remote security endpoint as an acceptable authority. The value for this parameter must be the label of a certificate authority that is defined in RACF. The CaLabel parameter can be specified multiple times.

13 A close brace (}) marks the end of the RemoteSecurityEndpoint statement.

14 A reference to a key exchange action that has been defined elsewhere, in either the IpSecConfig file or the CommonIpSecConfig file, as follows:

```
KeyExchangeAction            Gold-RSA
{
   HowToInitiate              main
   HowToRespond               main
   KeyExchangeOffer
   {
      HowToEncrypt               3DES
      HowToAuthMsgs              SHA1
      HowToAuthPeers             RsaSignature
   }
}
```

The KeyExchangeAction statement specifies the detailed parameters that govern a phase 1 negotiation between these two security endpoints, such as who can begin the negotiation and what type of encryption is used.

15 An optional shared key used only for pre-shared key host authentication.

16 A close brace (}) marks the end of the KeyExchangeRule statement block.

Example 1: Following is a key exchange rule for an Aggressive-mode phase 1 negotiation using pre-shared key authentication:

```
KeyExchangeRule              Admin_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  Admin_IKED
  KeyExchangeActionRef       Bronze-PSK
  SharedKey                  Ascii TheEagleHasLanded
}
```

This rule defines the parameters for the phase 1 IKE negotiation between two hosts that are identified by the security endpoints Internal_IKED and Admin_IKED (presumed to be defined elsewhere in the policy file). The specifics of the negotiation are covered by the Bronze-PSK action as follows:

```
KeyExchangeAction        Bronze-PSK
{
   HowToInitiate         Aggressive
   HowToRespond          Aggressive
   KeyExchangeOffer
   {
      HowToEncrypt       DES
      HowToAuthMsgs      SHA1
      HowToAuthPeers     PreSharedKey
   }
}
```

The optional SharedKey parameter is required only when the two hosts use the pre-shared key authentication method for the phase 1 negotiation.

Example 2: Following is a KeyExchangeRule statement for a Main-mode phase 1 negotiation using RSA signature authentication:

```
KeyExchangeRule              ZoneA_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  ZoneA_IKED
  KeyExchangeActionRef       Silver-RSA
}
```

The referenced objects are presumed to be defined elsewhere in the policy file. This rule defines the parameters for the phase 1 IKE negotiation between two hosts that are identified by the security endpoints Internal_IKED and ZoneA_IKED. The specifics of the negotiation are covered by the Silver-RSA action as follows:

```
KeyExchangeAction      Silver-RSA
{
   HowToInitiate       main
   HowToRespond        main
   KeyExchangeOffer
   {
      HowToEncrypt     DES
      HowToAuthMsgs    SHA1
      HowToAuthPeers   RsaSignature
   }
}
```

Ordering of key exchange rules: The previous two examples show individual key exchange rules. A complete key exchange policy contains any number of key exchange rules. Key exchange rules in a key exchange policy are searched in the order listed. In the process of an IKE negotiation, Policy Agent searches the list of active key exchange rules to locate a best match. It is possible for more than one key exchange rule to match a pending IKE negotiation. For this reason, the list of key exchange rules in the key exchange policy should be ordered from most specific to least specific in much the same way as the IP filter rules. If the key exchange policy contains key exchange rules with both unique and wildcard security endpoints, the most specific definitions should be placed higher in the list than the wildcard definitions.

For instance, there might be one key exchange rule governing a connection from an internal administrative machine, and another key exchange rule governing all other hosts on the internal network, as follows:

```
KeyExchangeRule              Admin_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  Admin_IKED
  KeyExchangeActionRef       Bronze-PSK
  SharedKey                  Ascii TheEagleHasLanded
```

```
}

KeyExchangeRule              ZoneA_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  ZoneA_IKED
  KeyExchangeActionRef       Silver-RSA
}
```

In this case, because the remote IKE peer that is defined by the remote security endpoint Admin_IKED matches both the Admin_KeyExRule1 and ZoneA_KeyExRule1 rules, the Admin_KeyExRule1 rule should be placed ahead of the ZoneA_KeyExRule1 rule in the key exchange policy as follows:

```
KeyExchangePolicy
{
  KeyExchangeRuleRef         Admin_KeyExRule1
  KeyExchangeRuleRef         ZoneA_KeyExRule1
  KeyExchangeRuleRef         ZoneB_KeyExRule1
  KeyExchangeRuleRef         ZoneC_KeyExRule1
}
```

Local dynamic VPN policy

A local dynamic VPN policy is required only if the IPSec negotiation is started through command-line activation using the **ipsec** command, or through automatic activation due to a local dynamic VPN policy update. IPSec negotiations can be initiated in one of four ways:

- On-demand

  The negotiation initiates when an outbound IP packet matches a filter rule with an ipsec action.
- Remotely

  The remote peer initiates the request, and the local IKE daemon responds.
- Command-line activation

  The **ipsec** command enables you to manually initiate an IPSec negotiation. (Not to be confused with manual tunnels, which do not use the IKE daemon at all.)
- Autoactivation

  The negotiation begins when either the stack or the IKE daemon initializes and both are active, or when the local dynamic VPN policy is updated.

A local dynamic VPN policy is required only in the last two cases.

A local dynamic VPN policy consists of an unordered list of local dynamic VPN rules. The negotiation for a phase 2 security association requires that the two communicating hosts agree on two data endpoints that the security association covers, the protocols the security association covers, and the ports that the security association covers. This information is then stored in the phase 2 security association, which is consulted each time relevant IPSec traffic needs to be encapsulated or decapsulated. The purpose of the local dynamic VPN rule is to define these requirements for each security association that is configured.

Following is a sample LocalDynVpnRule statement that defines the parameters for the negotiation of a phase 2 security association for TN3270 traffic between a server (9.1.1.1) and a client (9.4.4.100). A description of each line follows the sample.

```
1  LocalDynVpnRule      TelnetSA
2  {
3          LocalIP           9.1.1.1
4          RemoteIP          9.4.4.100
```

```
5          LocalDataPort     23
6          RemoteDataPort    0
7          Protocol          tcp
8          Autoactivate      yes
9  }
```

Line	Description
1	The LocalDynVpnRule keyword and user-defined name.
2	An open brace ({) marks the start of the LocalDynVpnRule statement block.
3	The local address of IP traffic that this security association is to protect. The address can be either a single address or a range of addresses. However, if the address is not a single address, this security association must be negotiated for tunnel mode.
4	The remote address of IP traffic that this security association is to protect. The address can be either a single address or a range of addresses. However, if the address is not a single address, this security association must be negotiated for tunnel mode.
5	The local ports for IP traffic that this security association is to protect. The port must be either a single port or all ports. A range of ports is not allowed. A value of 0 indicates all ports.
6	The remote ports for IP traffic that this security association is to protect. The port must be either a single port or all ports. A range of ports is not allowed. A value of 0 indicates all ports.
7	The protocol that this security association is to protect. This value can be numeric. It must define a single protocol or all protocols.
8	Indicates that this security association is to be activated when the stack and IKE daemon are active. No user intervention is required.
9	A close brace (}) marks the end of the LocalDynVpnRule statement block.

An on-demand security association does not require a local dynamic VPN rule definition. All of the parameters for the negotiation of an on-demand phase 2 security association can be inferred from one of two places, either the packet that began the on-demand activation or the filter rule on which the packet matched. The packet always provides a single value for address, port, and protocol. The filter rule, however, can allow for a range of values for IP address, port, or protocol. The granularity setting of the IpLocalStartAction statement determines whether the information is taken from the packet or from the matching filter rule. For more information regarding the IpLocalStartAction statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Security associations can be defined as wide or narrow with respect to IP addresses or ports and protocols. If a security association is wide with respect to IP address, the same security association is used to protect data between multiple endpoints. If a security association is wide with respect to ports and protocols, the same security association is used to protect multiple traffic types. Conversely, a narrowly defined security association can be used to protect specific data endpoints (based on IP address) or specific traffic types (based on commonly used ports and protocols for network services).

Example 1 - wide security association: The following rule allows any type of traffic to flow between PublicServerAddressA1 and SubnetC using the same security association. PublicServerAddressA1 and SubnetC can be defined in either

the IpSecConfig file or CommonIpSecConfig file. The AutoActivate parameter causes the IKE negotiation to initiate when the stack or IKE initializes.

```
LocalDynVpnRule       ZoneC_VPN-All-traffic
{
     LocalIpRef          PublicServerAddressA1
     RemoteIpSetRef      SubnetC
     LocalDataPort       0
     RemoteDataPort      0
     Protocol            all
     AutoActivate        yes
}

IpAddr                PublicServerAddressA1
{
  Addr                9.3.3.3
}

IpAddrSet             SubnetC
{
  Prefix              9.6.0.0/16
}
```

Example 2 - narrow security association: If narrow security associations are used for IPSec-protected FTP traffic, two VPN definitions are required, one for the data connection and one for the control connection. The following rules are from the server's perspective. The FTP client connecting from BranchOfficeAddressC1 to the PublicServerAddressA1 ports 20 and 21 uses the respective ZoneC FTP VPNs.

```
LocalDynVpnRule             ZoneC_VPN-FTP-Data
{
     LocalIpRef                PublicServerAddressA1
     RemoteIpRef               BranchOfficeAddressC1
     LocalDataPort             20
     RemoteDataPort            0
     Protocol                  tcp
}

LocalDynVpnRule             ZoneC_VPN-FTP-Control
{
     LocalIpRef                PublicServerAddressA1
     RemoteIpRef               BranchOfficeAddressC1
     LocalDataPort             21
     RemoteDataPort            0
     Protocol                  tcp
}

IpAddr                      PublicServerAddressA1
{
  Addr                      9.3.3.3
}

IpAddr                      BranchOfficeAddressC1
{
  Addr                      9.5.5.5
}
```

Quick start using IP filtering and IPSec host-to-host

The following sample shows a complete IP security policy allowing connections from a secure server (9.1.1.1) to an administrative machine (9.1.1.2) on an internal network. It represents the absolute minimum number of items that need to be configured for IKED to provide IPSec protection with dynamic key management between two hosts.

This IP security policy allows IKE negotiations in the clear (UDP, port 500 traffic), while authenticating and encrypting all other traffic using the ESP IPSec protocol. The policy relies almost exclusively on the z/OS IP security policy defaults, including MD5 and DES for the phase 1 ISAKMP security association and ESP/MD5 ESP/DES for the phase 2 IPSec security association. For a complete description of IP security policy configuration statements and their defaults, refer to *z/OS Communications Server: IP Configuration Reference*.

```
#------------------------------------------------------
# Quick-Start IP Security policy
#------------------------------------------------------
IpFilterPolicy
{
  PreDecap                    off
  FilterLogging               on
  AllowOnDemand               yes

  IpFilterRule                QuickStartRule1
  {
    IpSourceAddr              9.1.1.1
    IpDestAddr                9.1.1.2
    IpService
    {
      SourcePortRange         500
      DestinationPortRange    500
      Protocol                udp
      Direction               bidirectional
      Routing                 local
    }
    IpGenericFilterActionRef  permit
  }

  IpFilterRule                QuickStartRule2
  {
    IpSourceAddr              9.1.1.1
    IpDestAddr                9.1.1.2
    IpService
    {
      Direction               bidirectional
      Routing                 local
    }
    IpGenericFilterActionRef  ipsec
    IpDynVpnActionRef         TransportMode
  }
}

KeyExchangePolicy
{
  KeyExchangeRule             QuickStart_KeyExRule
  {
     LocalSecurityEndpoint
     {
       Identity               IpAddr 9.1.1.1
       Location               9.1.1.1
     }
     RemoteSecurityEndpoint
     {
       Identity               IpAddr 9.1.1.2
       Location               9.1.1.2
     }
    KeyExchangeActionRef      QuickStart_KeyExAction
    SharedKey                 Ascii TheEagleHasLanded
  }
}

#------------------------------------------------------
```

```
# Reusable actions
#-------------------------------------------------------
IpGenericFilterAction    permit
{
  IpFilterAction         permit
}

IpGenericFilterAction    ipsec
{
  IpFilterAction         ipsec
  IpFilterLogging        yes LogDeny
}

KeyExchangeAction        QuickStart_KeyExAction
{
  KeyExchangeOffer
  {
    HowToAuthPeers       PreSharedKey
  }
}

IpDynVpnAction           TransportMode
{
  IpDataOffer
  {
    HowToEncap           transport
  }
}
```

For all IKE negotiations, there must be a corresponding and consistent configuration on the remote host. In this case, if the remote system is running with z/OS IP security, the corresponding policy for the remote system can be generated merely by transposing all instances of local and remote IP addresses.

Displaying filters, rules, and actions

To display the filter rules for the quick start policy after they have been installed in the stack, enter the following UNIX System Services command:

```
ipsec -f display -r detail -c current

CS V1R7 ipsec  TCPIP Name: TCPCS  Fri Mar 18 12:00:03 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 6
Logging:  Yes             Predecap: No                DVIPSec:  Yes
NatKeepAlive:  20

FilterName:                   QuickStartRule1
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
Protocol:                     UDP(17)
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
```

```
|                       SourceAddress:               9.1.1.1
|                       SourceAddressPrefix:         n/a
|                       SourceAddressRange:          n/a
|                       SourceAddressGranularity:    Packet
|                       SourcePort:                  500
|                       SourcePortRange:             n/a
|                       SourcePortGranularity:       Rule
|                       DestAddress:                 9.1.1.2
|                       DestAddressPrefix:           n/a
|                       DestAddressRange:            n/a
|                       DestAddressGranularity:      Packet
|                       DestPort:                    500
|                       DestPortRange:               n/a
|                       DestPortGranularity:         Rule
|                       OrigRmtConnPort:             n/a
|                       RmtIDPayload:                n/a
|                       ********************************************************************************
|                       FilterName:                  QuickStartRule1
|                       FilterNameExtension:         2
|                       GroupName:                   n/a
|                       LocalStartActionName:        n/a
|                       VpnActionName:               n/a
|                       TunnelID:                    0x00
|                       Type:                        Generic
|                       State:                       Active
|                       Action:                      Permit
|                       Scope:                       Local
|                       Direction:                   Inbound
|                       OnDemand:                    n/a
|                       SecurityClass:               0
|                       Logging:                     None
|                       Protocol:                    UDP(17)
|                       ICMPType:                    n/a
|                       ICMPCode:                    n/a
|                       OSPFType:                    n/a
|                       TCPQualifier:                n/a
|                       ProtocolGranularity:         Rule
|                       SourceAddress:               9.1.1.2
|                       SourceAddressPrefix:         n/a
|                       SourceAddressRange:          n/a
|                       SourceAddressGranularity:    Packet
|                       SourcePort:                  500
|                       SourcePortRange:             n/a
|                       SourcePortGranularity:       Rule
|                       DestAddress:                 9.1.1.1
|                       DestAddressPrefix:           n/a
|                       DestAddressRange:            n/a
|                       DestAddressGranularity:      Packet
|                       DestPort:                    500
|                       DestPortRange:               n/a
|                       DestPortGranularity:         Rule
|                       OrigRmtConnPort:             n/a
|                       RmtIDPayload:                n/a
|                       ********************************************************************************
|                       FilterName:                  QuickStartRule2
|                       FilterNameExtension:         1
|                       GroupName:                   n/a
|                       LocalStartActionName:        n/a
|                       VpnActionName:               TransportMode
|                       TunnelID:                    Y0
|                       Type:                        Dynamic Anchor
|                       State:                       Active
|                       Action:                      Permit
|                       Scope:                       Local
|                       Direction:                   Outbound
|                       OnDemand:                    Yes
|                       SecurityClass:               0
```

```
|                       Logging:                    Deny
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              9.1.1.1
|                       SourceAddressPrefix:        n/a
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                9.1.1.2
|                       DestAddressPrefix:          n/a
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 QuickStartRule2
|                       FilterNameExtension:        2
|                       GroupName:                  n/a
|                       LocalStartActionName:       n/a
|                       VpnActionName:              TransportMode
|                       TunnelID:                   Y0
|                       Type:                       Dynamic Anchor
|                       State:                      Active
|                       Action:                     Permit
|                       Scope:                      Local
|                       Direction:                  Inbound
|                       OnDemand:                   Yes
|                       SecurityClass:              0
|                       Logging:                    Deny
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              9.1.1.2
|                       SourceAddressPrefix:        n/a
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                9.1.1.1
|                       DestAddressPrefix:          n/a
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 DenyAllRule_Generated__________Inbnd
|                       FilterNameExtension:        n/a
|                       GroupName:                  n/a
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
```

```
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Deny
|                       Scope:                      Both
|                       Direction:                  Inbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    None
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 DenyAllRule_Generated__________Outbnd
|                       FilterNameExtension:        n/a
|                       GroupName:                  n/a
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Deny
|                       Scope:                      Both
|                       Direction:                  Outbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    None
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
```

```
RmtIDPayload:                 n/a
*************************************************************************

6 entries selected
```

Each IP service in the example uses the bidirectional keyword. Therefore, two rules are created for each IP service, one outbound and one inbound. When the IP filter rules are expanded in this way, the specific filter rules are distinguished from each other by a unique numeric value in the FilterNameExtension field.

Note that the last two deny rules are not explicitly coded in the IpSecConfig file, but are added by the system in keeping with a default-deny policy.

For more information on displaying active filters with the **ipsec** command, see "Displaying active filters with the ipsec command" on page 958.

To view the quick start filter rules using the **pasearch** command, issue the following command:

```
pasearch -v f
```

```
TCP/IP pasearch CS V1R7                   TCP/IP Image:      TCPCS
  Date:                 03/18/2005        Time:  12:00:32
  IPSec Instance Id:    1111096508

policyRule:             QuickStartRule1
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               104               ForLoadDist:       False
  Priority:             4                 Sequence Actions:  Don't Care
  No. Policy Action:    1                 ConditionListType: CNF
  IpSecType:            policyIpFilter
  policyAction:         permit
   ActionType:          IpFilter GenericFilter
   Action Sequence:     0
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    05:00             To TimeOfDay UTC:  05:00
   TimeZone:            Local
  IpSec Condition Summary:                NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All
     ToAddr:            All
    Destination Address:
     FromAddr:          All
     ToAddr:            All
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
  Condition Work Level:      0
    Group Number:       1                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:           NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All
```

```
|                          ToAddr:           All
|                        Destination Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Service Condition:
|                          Protocol:         0
|                          Direction:        0
|                          RouteType:        0                SecurityClass:     0
|                      IpSec Condition Work:                   NegativeIndicator: Off
|                       IpFilter Condition:
|                        Source Address:
|                          FromAddr:         9.1.1.1
|                          ToAddr:           9.1.1.1
|                        Destination Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Service Condition:
|                          Protocol:         0
|                          Direction:        0
|                          RouteType:        0                SecurityClass:     0
|                      Condition Work Level:     1
|                        Group Number:       2                Cond Count:        2
|                        Ignore:             No
|                      IpSec Condition Work Summary:          NegativeIndicator: Off
|                       IpFilter Condition:
|                        Source Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Destination Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Service Condition:
|                          Protocol:         0
|                          Direction:        0
|                          RouteType:        0                SecurityClass:     0
|                      IpSec Condition Work:                   NegativeIndicator: Off
|                       IpFilter Condition:
|                        Source Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Destination Address:
|                          FromAddr:         9.1.1.2
|                          ToAddr:           9.1.1.2
|                        Service Condition:
|                          Protocol:         0
|                          Direction:        0
|                          RouteType:        0                SecurityClass:     0
|                      Condition Work Level:     2
|                        Group Number:       3                Cond Count:        2
|                        Ignore:             No
|                      IpSec Condition Work Summary:          NegativeIndicator: Off
|                       IpFilter Condition:
|                        Source Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Destination Address:
|                          FromAddr:         All
|                          ToAddr:           All
|                        Service Condition:
|                          Protocol:         0
|                          Direction:        0
|                          RouteType:        0                SecurityClass:     0
|                      IpSec Condition Work:                   NegativeIndicator: Off
|                       IpFilter Condition:
|                        Source Address:
|                          FromAddr:         All
|                          ToAddr:           All
```

```
|                          Destination Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Service Condition:
|                           Protocol:          UDP  (17)
|                            SrcPortFrom:      500               SrcPortTo:         500
|                            DestPortFrom:     500               DestPortTo:        500
|                           Direction:         Bidirectional
|                           RouteType:         Local             SecurityClass:     0
|
|                        IpFilter Action:      permit
|                          Version:            3                 Status:            Active
|                          Scope:              GenericFilter
|                          ipFilterAction:     Permit            IpFilterLogging:   No
|
|                      policyRule:             QuickStartRule2
|                        Rule Type:            IpFilter
|                        Version:              3                 Status:            Active
|                        Weight:               103               ForLoadDist:       False
|                        Priority:             3                 Sequence Actions:  Don't Care
|                        No. Policy Action:    2                 ConditionListType: CNF
|                        IpSecType:            policyIpFilter
|                        policyAction:         ipsec
|                         ActionType:          IpFilter GenericFilter
|                         Action Sequence:     0
|                        policyAction:         TransportMode
|                         ActionType:          IpFilter DynamicVpn
|                         Action Sequence:     0
|                        Time Periods:
|                         Day of Month Mask:
|                         First to Last:       11111111111111111111111111111111
|                         Last to First:       11111111111111111111111111111111
|                         Month of Yr Mask:    111111111111
|                         Day of Week Mask:    1111111  (Sunday - Saturday)
|                         Start Date Time:     None
|                         End Date Time:       None
|                         Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
|                         Fr TimeOfDay UTC:    05:00             To TimeOfDay UTC:  05:00
|                         TimeZone:            Local
|                        IpSec Condition Summary:                NegativeIndicator: Off
|                         IpFilter Condition:
|                          Source Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Destination Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Service Condition:
|                           Protocol:          0
|                           Direction:         0
|                           RouteType:         0                 SecurityClass:     0
|                        Condition Work Level:      0
|                          Group Number:       1                 Cond Count:        2
|                          Ignore:             No
|                        IpSec Condition Work Summary:           NegativeIndicator: Off
|                         IpFilter Condition:
|                          Source Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Destination Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Service Condition:
|                           Protocol:          0
|                           Direction:         0
|                           RouteType:         0                 SecurityClass:     0
|                        IpSec Condition Work:                   NegativeIndicator: Off
```

```
                        IpFilter Condition:
                         Source Address:
                          FromAddr:          9.1.1.1
                          ToAddr:            9.1.1.1
                         Destination Address:
                          FromAddr:          All
                          ToAddr:            All
                         Service Condition:
                          Protocol:          0
                          Direction:         0
                          RouteType:         0                 SecurityClass:     0
                       Condition Work Level:      1
                         Group Number:       2                 Cond Count:        2
                         Ignore:             No
                       IpSec Condition Work Summary:           NegativeIndicator: Off
                        IpFilter Condition:
                         Source Address:
                          FromAddr:          All
                          ToAddr:            All
                         Destination Address:
                          FromAddr:          All
                          ToAddr:            All
                         Service Condition:
                          Protocol:          0
                          Direction:         0
                          RouteType:         0                 SecurityClass:     0
                       IpSec Condition Work:                   NegativeIndicator: Off
                        IpFilter Condition:
                         Source Address:
                          FromAddr:          All
                          ToAddr:            All
                         Destination Address:
                          FromAddr:          9.1.1.2
                          ToAddr:            9.1.1.2
                         Service Condition:
                          Protocol:          0
                          Direction:         0
                          RouteType:         0                 SecurityClass:     0
                       Condition Work Level:      2
                         Group Number:       3                 Cond Count:        2
                         Ignore:             No
                       IpSec Condition Work Summary:           NegativeIndicator: Off
                        IpFilter Condition:
                         Source Address:
                          FromAddr:          All
                          ToAddr:            All
                         Destination Address:
                          FromAddr:          All
                          ToAddr:            All
                         Service Condition:
                          Protocol:          0
                          Direction:         0
                          RouteType:         0                 SecurityClass:     0
                       IpSec Condition Work:                   NegativeIndicator: Off
                        IpFilter Condition:
                         Source Address:
                          FromAddr:          All
                          ToAddr:            All
                         Destination Address:
                          FromAddr:          All
                          ToAddr:            All
                         Service Condition:
                          Protocol:          All
                          Direction:         Bidirectional
                          RouteType:         Local             SecurityClass:     0

                       IpFilter Action:      ipsec
```

```
|                          Version:            3                 Status:            Active
|                          Scope:              GenericFilter
|                          ipFilterAction:     IPSec             IpFilterLogging:   Yes Logdeny
|
|                        IpFilter Action:      TransportMode
|                          Version:            3                 Status:            Active
|                          Scope:              DynamicVpn
|                          Initiation:         Either            VpnLife:           1440
|                          Pfs:                None              IpDataOfferNum:    1
|                          IPDataOffer:        0
|                           HowToEncap:        Transport         HowToEncrypt:      DES
|                           HowToAuth:         ESP               HowToAuthAlgr:     HMAC_MD5
|                           RefLifeTmPropose:  240
|                           RefLifeTmAcptMin:  120               RefLifeTmAcptMax:  480
|                           RefLifeSzPropose:  None
|                           RefLifeSzAccept :  None
|
|                      policyRule:             DenyAllRule_Generated__________Inbnd
|                        Rule Type:            IpFilter
|                        Version:              3                 Status:            Active
|                        Weight:               102               ForLoadDist:       False
|                        Priority:             2                 Sequence Actions:  Don't Care
|                        No. Policy Action:    0
|                        IpSecType:            policyIpFilter
|                        Time Periods:
|                         Day of Month Mask:
|                         First to Last:       11111111111111111111111111111111
|                         Last to First:       11111111111111111111111111111111
|                         Month of Yr Mask:    111111111111
|                         Day of Week Mask:    1111111  (Sunday - Saturday)
|                         Start Date Time:     None
|                         End Date Time:       None
|                         Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
|                         Fr TimeOfDay UTC:    05:00             To TimeOfDay UTC:  05:00
|                         TimeZone:            Local
|                        IpSec Condition Summary:                NegativeIndicator: Off
|                         IpFilter Condition:
|                          Source Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Destination Address:
|                           FromAddr:          All
|                           ToAddr:            All
|                          Service Condition:
|                           Protocol:          All
|                           Direction:         Inbound
|                           RouteType:         Either            SecurityClass:     0
|
|                      policyRule:             DenyAllRule_Generated__________Outbnd
|                        Rule Type:            IpFilter
|                        Version:              3                 Status:            Active
|                        Weight:               101               ForLoadDist:       False
|                        Priority:             1                 Sequence Actions:  Don't Care
|                        No. Policy Action:    0
|                        IpSecType:            policyIpFilter
|                        Time Periods:
|                         Day of Month Mask:
|                         First to Last:       11111111111111111111111111111111
|                         Last to First:       11111111111111111111111111111111
|                         Month of Yr Mask:    111111111111
|                         Day of Week Mask:    1111111  (Sunday - Saturday)
|                         Start Date Time:     None
|                         End Date Time:       None
|                         Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
|                         Fr TimeOfDay UTC:    05:00             To TimeOfDay UTC:  05:00
|                         TimeZone:            Local
|                        IpSec Condition Summary:                NegativeIndicator: Off
```

```
   IpFilter Condition:
    Source Address:
     FromAddr:          All
     ToAddr:            All
    Destination Address:
     FromAddr:          All
     ToAddr:            All
    Service Condition:
     Protocol:          All
     Direction:         Outbound
     RouteType:         Either            SecurityClass:     0
```

For more information on displaying filter rules with the **pasearch** command, see "Displaying filter rules with the pasearch command" on page 974.

To display the key exchange rules and actions for the quick start IP security policy, issue the following command:

```
pasearch -v k

TCP/IP pasearch CS V1R7                   TCP/IP Image:      TCPCS
  Date:                 03/17/2005        Time:  16:37:06
  IPSec Instance Id:    1111094901

policyRule:             QuickStart_KeyExRule
  Rule Type:            KeyExchange
  Version:              3                 Status:            Active
  Weight:               101               ForLoadDist:       False
  Priority:             1                 Sequence Actions:  Don't Care
  No. Policy Action:    1
  IpSecType:            policyKeyExchange
  policyAction:         QuickStart_KeyExAction
   ActionType:          KeyExchange
   Action Sequence:     0
  Time Periods:
   Day of Month Mask:   00000000000000000000000000000000
   Month of Yr Mask:    000000000000
   Day of Week Mask:    0000000  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      00:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:                NegativeIndicator: Off
   KeyExchange Condition:
    LocalSecurityEndPoint:
     Location:          9.1.1.1
     Identity:
      IpAddr:
       FromAddr:        9.1.1.1
       ToAddr:          9.1.1.1
    RemoteSecurityEndPoint:
     Location:
      FromAddr:         9.1.1.2
      ToAddr:           9.1.1.2
     Identity:
      IpAddr:
       FromAddr:        9.1.1.2
       ToAddr:          9.1.1.2

  KeyExchange Action:   QuickStart_KeyExAction
    Version:            3                 Status:            Active
    HowToInitiate:      Main              HowToRespond:      Either
    AllowNat:           No
    KeyExchangeOffer:   0
     HowToEncrypt:      DES               HowToAuthMsgs:     MD5
     HowToAuthPeers:    PresharedKey      DHGroup:           Group1
```

```
     RefLifeTmPropose:  480
     RefLifeTmAcptMin:  240              RefLifeTmAcptMax:  1440
     RefLifeSzPropose:  None
     RefLifeSzAccept :  None
```

Activating the quick start security association

The quick start policy enables on-demand activation of a security association between the two endpoints. The security association is a wide security association allowing any type of traffic. Therefore, it can be activated by sending any type of traffic from the local host at 9.1.1.1 to the remote host at 9.1.1.2 as follows:

```
ping -s 9.1.1.1 9.1.1.2
CS V1R7: Pinging host 9.1.1.2
sendto(): EDC5111I Permission denied.
```

Because the security association does not exist, the initial ping attempt fails. After the negotiation for the security association has activated and the security association is established, a subsequent ping attempt succeeds as follows:

```
ping -s 9.1.1.1 9.1.1.2
CS V1R7: Pinging host 9.1.1.2
Ping #1 response took 0.001 seconds.
```

Displaying the quick start security associations

Use the **ipsec** command to display both the phase 1 and phase 2 security associations between 9.1.1.1 and 9.1.1.2. The following command displays the phase 1 security associations:

```
ipsec -k display -r detail

CS V1R7 ipsec  TCPIP Name: TCPCS  Thu Mar 17 16:37:54 2005
Primary:  IKE tunnel      Function: Display           Format:   Detail
Source:   IKED            Scope:    Current           TotAvail: n/a

TunnelID                      K1
KeyExchangeRuleName:          QuickStart_KeyExRule
KeyExchangeActionName:        QuickStart_KeyExAction
LocalEndPoint:                9.1.1.1
LocalIDType:                  IPV4
LocalID:                      9.1.1.1
RemoteEndPoint:               9.1.1.2
RemoteIDType:                 IPV4
RemoteID:                     9.1.1.2
ExchangeMode:                 Main
State:                        DONE
AuthenticationAlgorithm:      Hmac_Md5
EncryptionAlgorithm:          DES
DiffieHellmanGroup            1
AuthenticationMethod:         PresharedKey
InitiatorCookie:              0X4A0A99F121B8BB08
ResponderCookie:              0X663A1774B8A7544F
Lifesize:                     0K
CurrentByteCount:             216b
Lifetime:                     480m
LifetimeRefresh:              2005/03/17 22:19:07
LifetimeExpires:              2005/03/18 00:32:16
Role:                         Initiator
AssociatedDynamicTunnels:     1
NATTSupportLevel:             None
NATInFrntLclScEndPnt:         No
NATInFrntRmtScEndPnt:         No
zOSCanInitiateP1SA:           Yes
AllowNat:                     No
*************************************************************************

1 entries selected
```

In addition to information relating specifically to the phase 2 security association, use the **ipsec -y display** command to find the phase 1 that protects it. The ParentIKETunnelID field shows the associated phase 1, which is the same as the TunnelID from the previous **ipsec -k display** command.

```
ipsec -y display -r detail

CS V1R7 ipsec  TCPIP Name: TCPCS  Thu Mar 17 16:38:53 2005
Primary:  Dynamic tunnel  Function: Display           Format:   Detail
Source:   Stack           Scope:    Current           TotAvail: 1

TunnelID                      Y2
VpnActionName:                TransportMode
State:                        Active
LocalEndPoint:                9.1.1.1
RemoteEndPoint:               9.1.1.2
HowtoEncap:                   Transport
HowToAuth:                    ESP
 AuthAlgorithm:               Hmac_Md5
 AuthInboundSpi:              1335895732
 AuthOutboundSpi:             2732268544
HowToEncrypt:                 DES
 EncryptInboundSpi:           1335895732
 EncryptOutboundSpi:          2732268544
OutboundPackets:              1
OutboundBytes:                264
InboundPackets:               1
InboundBytes:                 264
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2005/03/17 19:22:40
LifetimeExpires:              2005/03/17 20:32:16
CurrentTime:                  2005/03/17 16:38:53
VPNLifeExpires:               2005/03/18 16:32:16
ParentIKETunnelID:            K1
LocalDynVpnRule:              n/a
NAT Traversal Topology:
  UdpEncapMode:               No
  LclNATDetected:             No
  RmtNATDetected:             No
  RmtIsGw:                    No
  RmtIsZOS:                   No
  zOSCanInitP2SA:             Yes
  SrcNATOARcvd:               n/a
  DstNATOARcvd:               n/a
*******************************************************************************

1 entries selected
```

Steps for configuring IP security policy

Perform the following steps to configure IP security policy.

1. Determine the number of zones to be protected. A zone typically equates to a subnetwork that is reachable by the host server, but can be any group of IP addresses that are conceptually related. The internal network can be defined in one or several zones, while any external network or group of networks can be placed in separate zones. Each zone should have meaning related to the site's security policy. If a zone maps to a physical interface, optionally assign a security class to all interfaces in that zone.
2. For each zone, determine what services are allowed and define an IpService statement for each desired service. Services are defined by the protocols and well-known ports that they use.

3. Determine the data endpoints to be protected. Typically, this is both a local and remote IP address, or subnetwork.
4. Determine what level of security is needed between each set of data endpoints. The level of security can be deny, permit, or ipsec.
5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged.
6. If IPSec is required between any two endpoints:
   a. Configure a KeyExchangePolicy statement that defines the parameters of the phase 1 negotiation:
      1) Determine the required type and strength of protection for the phase 1 security association.
      2) Decide what type of peer authentication will be used:
         - If RSA signature, set up RACF certificates and certificate authority information.
         - If pre-shared key, create a secret key that is known to both peers.
      3) Decide whether NAT traversal will be allowed. If the network topology contains one or more NAT devices that must be traversed by the phase 1 security association, NAT traversal should be allowed.
      4) Configure a KeyExchangeOffer statement.
      5) Determine negotiation mode, Main or Aggressive.
      6) Configure a KeyExchangeAction statement.
      7) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement.
      8) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action.
      9) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block.
   b. Configure an IpDynVpnAction statement that defines the control of the phase 2 negotiation:
      1) Determine the required type and strength of IPSec protection for the phase 2 security association.
      2) Determine whether tunnel or transport mode is required.
      3) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation.
      4) Determine which peer should be allowed to initiate the negotiation.
   c. Decide how the security association is to be activated:
      1) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.
      2) Configure an optional IpLocalStartAction statement, if the security association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host security association). Include a reference to the IpLocalStartAction statement in the IP filter rule.
      3) Either create an IpFilterRule statement that allows IPSec traffic (AH and ESP), or set the global PreDecap parameter of the IpFilterPolicy statement to off.

7. Define an IpFilterRule statement for each set of data endpoints. The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500.

   Rule: To allow IKE negotiations for security associations, IKE traffic (port 500, and optionally port 4500 for NAT traversal) must be permitted in the clear.

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include the IpFilterRule statements in the IpFilterPolicy block.
10. Include all configured statements in the IpSecConfig file. For information on the use of the IpSecConfig and CommonIpSecConfig files, see "Steps for configuring IP security policy using only a CommonIpSecConfig file" on page 862, "Steps for configuring IP security policy using only a stack-specific IpSecConfig file" on page 863, and "Steps for configuring IP security policy configuration using both approaches" on page 863.

For examples of the use of these steps, see "Configuring specific security models."

Configuring specific security models

Setting up IP security configuration files can be a complex task, as there are many powerful features, options, and controls. However, after the security needs of the business are identified, implementing an IP security policy becomes a matter of translating the requirements to a Policy Agent configuration file.

The choice of protection model primarily depends on the network topology. Although it is perfectly permissible to follow a single model when configuring IP security policy, the z/OS IP security function enables any number of models to be installed concurrently. Commonly, one set of rules governs internal network traffic, another protects traffic from connected networks, and a third provides security for traffic that is routed over the Internet. The following scenarios presume that you are configuring a secure server that is a multihomed host that is connected to an internal, an external, and a wide-area network that traverses the Internet. The configuration guidelines that are presented in the following sections are based on three business models:

- Trusted internal network (permit, deny)

  In the trusted internal network model, the server is protecting traffic that originates from hosts inside a privately controlled network. IP packets on the internal network are not generally subject to the stringent restrictions that are placed on traffic that is generated from outside the business. This model is usually more tolerant, given that users inside the company need access to internal network resources and services, such as the Web, FTP, and Telnet.

- Business partner (permit, deny, strong IPSec protection)

  The business partner model consists of two interconnected networks, with the server protecting traffic that originates from hosts outside the internal network. Typically, two separate networks are physically connected to the z/OS server. Because the traffic is not restricted to internal hosts, security is usually somewhat tighter than in the trusted internal network model. Each business partner has no physical control over the machine of the other business partner.

The services that are provided are determined by the needs of the business, but typically include many of the same services that are provided to the internal network, such as access to a Web server, FTP, and Telnet. Though many services might be allowed between business partners, the need for confidentiality and authentication of data is more stringent than in the trusted internal network model, because there is little to no control over the other network. IPSec is often specified to authenticate and optionally encrypt data that flows between the two networks.

- Branch office (permit, deny, strong tunnel-mode IPSec protection)

  The branch office model consists of two networks whose IP connectivity relies on the Internet. The server is protecting traffic that originates from hosts outside the internal network, which at some point is routed over the Internet. Because there is no control over any data that traverses the Internet, the need for security is greatest in this model. The services that are provided are based on business need, but typically include a subset of what is available internally. All traffic that traverses the Internet carrying vital information should be secured using some form of authentication and encryption.

Figure 84 shows a sample network for all three security models.

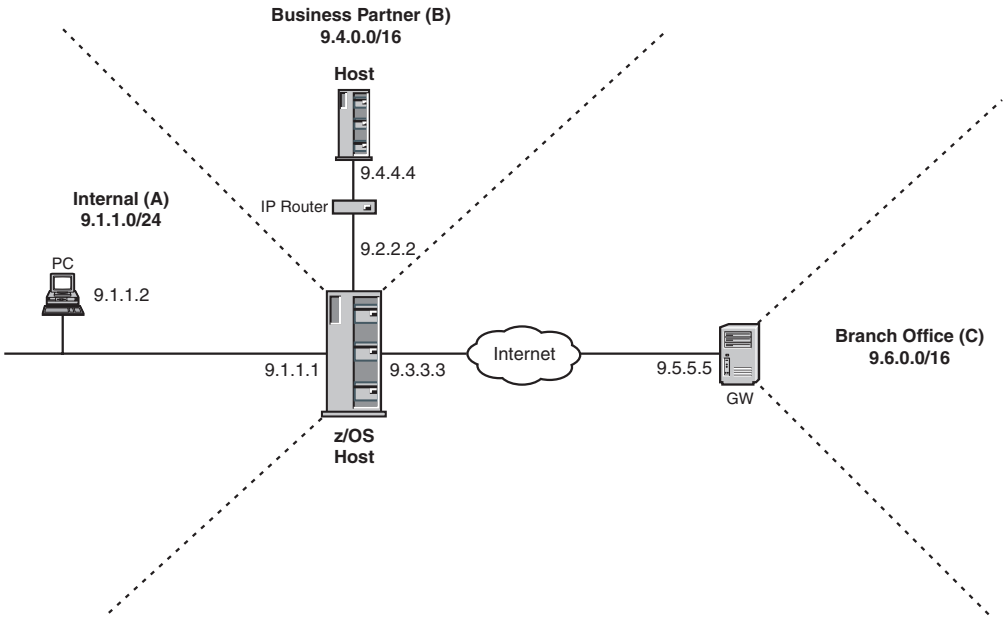


Figure 84. Security model network

The following subsections describe how to configure these models using the steps described in "Steps for configuring IP security policy" on page 887. The policy examples assume that a default-deny policy is in place. Any traffic not explicitly permitted is blocked.

Steps for configuring the trusted internal network model (simple IP filtering)

The following statements, concepts, and files are covered in the discussion of this model:

- IpFilterRule
- IpService
- IpGenericFilterAction

- References
- Groups
- IpSecConfig
- CommonIpSecConfig

Figure 85 shows the trusted internal network portion of the security model network.

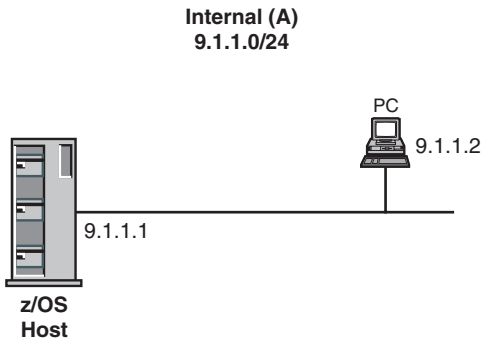


Figure 85. Trusted internal network model

For this example, assume that the following requirements must be met to control traffic on the internal network:

- Give FTP access to an administrator from a host inside the internal network (9.1.1.2) to the secure server (9.1.1.1). Log any traffic that matches this traffic pattern.
- Allow Web access (HTTP, port 80) to the secure server from any host on the internal network (9.1.1.0/24). Do not log any traffic that matches this traffic pattern.
- Deny all other traffic.

Perform the following steps to meet these requirements and configure the trusted internal network model.

1. Determine the number of zones to be protected.

   There is only one zone for this example, the internal network 9.1.1.0/24.

2. For each zone, determine what services are allowed and define an IpService block for each desired service. Optionally, assign a security class to all interfaces in each zone.

   There are two services stated in the example requirements, HTTP and FTP. The traffic is local to this host and, therefore, the routing is designated as local. No forwarding of these services is allowed.

   Because the entire internal network is defined in one zone, you can define a unique security class for the interface with address 9.1.1.1. For this example, the SECCLASS parameter of all internal network interfaces is assigned the arbitrary value of 1, which can be interpreted to mean a trusted network. If you specify the SecurityClass parameter in the IpService block, the related interface must be assigned the same value on the SECCLASS parameter of the LINK statement in the TCP/IP profile. In this example, the traffic is allowed only over an interface with a SECCLASS parameter value of 1, presumed to be the interface connected to the internal network.

```
IpService
{
  SourcePortRange       80
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional InboundConnect
  Routing               local
  SecurityClass         1
}
```

Because normal FTP uses two well-known ports, two services are required, one for the control connection and one for the data connection:

```
IpService
{
  SourcePortRange       21
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional InboundConnect
  Routing               local
  SecurityClass         1

}

IpService
{
  SourcePortRange       20
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional OutboundConnect
  Routing               local
  SecurityClass         1

}
```

The InboundConnect keyword is used for services that are not allowed to initiate a TCP connection. The OutboundConnect keyword is used for services that are not allowed to receive a TCP connection request. If neither keyword is specified, either side can initiate a TCP connection.

3. Determine the data endpoints to be protected.

   There are two sets of data endpoints to be protected in this example, representing the connection from the administrative machine, and all the other hosts on the subnetwork:

   ```
   Local  Address of secure server: 9.1.1.1
   Remote Address of administrative machine: 9.1.1.2

   Local  Address of secure server: 9.1.1.1
   Remote Address of all hosts on internal network: 9.1.1.0/24
   ```

4. Determine what level of security is needed between each set of data endpoints.

   In this example, only permit is required. Therefore, no IPSec information is needed. Because z/OS IP security policy implicitly provides a default-deny policy, all other traffic is denied.

5. Configure an IpGenericFilterAction statement for the level of security (permit, deny, ipsec) that is required, including whether the connection is logged.

   Because the example requirement is to permit two types of traffic with different logging requirements, two actions are needed as follows:

   ```
   IpGenericFilterAction    permit-log
   {
      IpFilterAction       permit
      IpFilterLogging       yes
   ```

```
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, configure a KeyExchangePolicy statement that defines the parameters of the phase 1 negotiation, configure an IpDynVpnAction statement that defines the control of the phase 2 negotiation, and decide how the security association is to be activated.

   IPSec is not required in this example. If there was sensitive data flowing through the internal network that needed to be confidential, IPSec could be specified to encrypt some IP packets, thereby effectively securing information that travels between two hosts on the internal network.

7. Define an IpFilterRule block for each set of data endpoints. Each rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500.

   In this example, the source address refers to an address on the secure host. The destination address refers to remote hosts. The IpService statements are the ones defined in step 2 on page 891. Note that the IpGenericFilterAction statement must reference a previously defined action.

```
IpFilterRule             AdminFTP
{
   IpSourceAddr          9.1.1.1
   IpDestAddr            9.1.1.2
   IpService
   {
      SourcePortRange       21
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional InboundConnect
      Routing               local
      SecurityClass         1

   }
   IpService
   {
      SourcePortRange       20
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional OutboundConnect
      Routing               local
      SecurityClass         1
   }
   IpGenericFilterActionRef  permit-log
}
IpFilterRule             InternalNetWeb
{
   IpSourceAddr          9.1.1.1
   IpDestAddrSet         9.1.1.0/24
   IpService
   {
      SourcePortRange       80
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional InboundConnect
```

```
      Routing               local
      SecurityClass         1
   }
   IpGenericFilterActionRef  permit-nolog
}

IpGenericFilterAction        permit-log
{
  IpFilterAction             permit
  IpFilterLogging            yes
}

IpGenericFilterAction        permit-nolog
{
  IpFilterAction             permit
  IpFilterLogging            no
}
```

Because IPSec is not required in this example, no filters for IKE traffic are needed.

8. Include the IpFilterRule statements in the IpFilterPolicy block.

   The IP filter rules and their relative placement within the IpFilterPolicy block should be from most specific to least specific. Because the AdminFTP rule controls traffic from a specific host, it should be placed before the InternalNetWeb rule. Note that to enable logging of the individual rules, filter logging must be enabled at the global level of the IP filter policy with the FilterLogging parameter.

```
IpFilterPolicy
{
   FilterLogging            on

   IpFilterRule             AdminFTP
   {
      IpSourceAddr          9.1.1.1
      IpDestAddr            9.1.1.2
      IpService
      {
         SourcePortRange       21
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpService
      {
         SourcePortRange       20
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional OutboundConnect
         Routing               local
         SecurityClass         1
      }
      IpGenericFilterActionRef  permit-log
   }
   IpFilterRule             InternalNetWeb
   {
      IpSourceAddr          9.1.1.1
      IpDestAddrSet         9.1.1.0/24
      IpService             WebServer
      {
         SourcePortRange       80
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
```

```
            Routing               local
            SecurityClass         1
         }
         IpGenericFilterActionRef  permit-nolog
      }
   }
```

9. Include all configured statements in the IpSecConfig file.

   The IpFilterPolicy statement and the IpGenericFilterAction statements are placed in the IpSecConfig file in no particular order, although the IpSecConfig file is easier to read if logically related items are placed close together. For further ease of reading and maintenance, the IpSecConfig file should be documented with comments, which begin with the number sign (#).

   The completed IpSecConfig file for the internal network with two filter rules follows:

```
# IP Security policy for Secure Server
#########################
# IpFilterPolicy block   #
#########################
IpFilterPolicy
{
   FilterLogging            on
   #Allow admin FTP; log traffic
   IpFilterRule             AdminFTP
   {
      IpSourceAddr          9.1.1.1
      IpDestAddr            9.1.1.2
      IpService
      {
         SourcePortRange       21
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpService
      {
         SourcePortRange       20
         DestinationPortRange  1024 65535
         Protocol              tcp
         Routing               local
         SecurityClass         1
         Direction             bidirectional OutboundConnect
      }
      IpGenericFilterActionRef  permit-log
   }
   #Allow LAN Web traffic; don't log
   IpFilterRule             InternalNetWeb
   {
      IpSourceAddr          9.1.1.1
      IpDestAddrSet         9.1.1.0/24
      IpService
      {
         SourcePortRange       80
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpGenericFilterActionRef  permit-nolog
   }
```

```
##########################
# Generic Filter Actions   #
##########################
IpGenericFilterAction    permit-log
{
   IpFilterAction        permit
   IpFilterLogging       yes
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction        permit
   IpFilterLogging       no
}
```

10. Define an IP filter group for each zone, and include the IP filter rules that belong to that zone.

    In step 9 on page 895, both IpFilterRule statements include a reference to statements defined outside of the IpFilterRule block, the IpGenericFilterAction statements. Some other information, such as IP addresses and services, is undoubtedly to be needed more than once. Changing these occurrences to reference objects eliminates repeated typing of the same information and adds clarity to the configuration file. To take advantage of references, the reusable statements must be given a name.

    - Single IP addresses are defined by the IpAddr statement, which contains one parameter, Addr:

      ```
      IpAddr         InternalNetServerAddress
      {
         Addr        9.1.1.1
      {

      IpAddr         InternalNetAdminAddress
      {
         Addr        9.1.1.2
      }
      ```

    - Ranges or subnetworks are defined by the IpAddrSet statement, which contains either a Range or Prefix attribute:

      ```
      IpAddrSet      InternalNet
      {
         Prefix      9.1.1.0/24
      }
      ```

    - To be referenced, each IpService statement needs a name:

      ```
      IpService      WebServer
      {
         SourcePortRange       80
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }

      IpService     FTPServer-Control
      {
         SourcePortRange       21
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpService     FTPServer-Data
      {
      ```

```
   SourcePortRange       20
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional OutboundConnect
   Routing               local
   SecurityClass         1
}
```

- FTP is composed of two individual services, and both can be condensed into an IpServiceGroup that references the two FTP services:

```
IpServiceGroup    FTPServer
{
   IpServiceRef   FTPServer-Control
   IpServiceRef   FTPServer-Data
}
```

- The IP filter rules can be grouped as well. Because the filter rules that apply to the internal network naturally relate to each other in the sense that they apply to the same security zone, they can be combined into an IpFilterGroup statement:

```
IpFilterGroup     InternalNetZoneA
{
   IpFilterRef    AdminFTP
   IpFilterRef    InternalNetWeb
}
```

  Notice that just as the list of IpFilterRule statements in the IpFilterPolicy block is ordered, the list of IpFilterRef statements in the IpFilterGroup block is also ordered. The InternalNetWeb rule applies to all of the IP addresses in the network, including the administrative machine. However, the AdminFTP rule is more specific because it applies only to a specific address within that network. The more specific rule is placed first in the list.

Now that all reusable statements have been identified and separately defined, they can be incorporated into any statement that requires that reusable statement type. The modified IpSecConfig file using references follows. Note that by adding names and organizing related statements, the purpose of the IpFilterPolicy statement is clarified.

```
# IP Security policy for Secure Server
##########################
# IpFilterPolicy block   #
##########################
IpFilterPolicy
{
   FilterLogging            on
   IpFilterGroupRef     InternalNetZoneA
}

##########################
# Security Zones         #
##########################
IpFilterGroup     InternalNetZoneA
{
   IpFilterRuleRef    AdminFTP
   IpFilterRuleRef    InternalNetWeb
}

##########################
# Filter rules           #
##########################
#Allow admin FTP; log traffic
IpFilterRule                AdminFTP
{
   IpSourceAddrRef          InternalNetServerAddress
```

```
   IpDestAddrRef             InternalNetAdminAddress
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  permit-log
}

#Allow LAN Web traffic; don't log
IpFilterRule                 InternalNetWeb
{
   IpSourceAddrRef           InternalNetServerAddress
   IpDestAddrSetRef          InternalNet
   IpServiceRef              WebServer
   IpGenericFilterActionRef  permit-nolog
}

######## All reusable reference statements defined below ########

############################
# Generic Filter Actions   #
############################
IpGenericFilterAction    permit-log
{
   IpFilterAction        permit
   IpFilterLogging       yes
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction        permit
   IpFilterLogging       no
}


########################
# Reusable Services    #
########################
IpService      WebServer
{
   SourcePortRange       80
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}

IpService     FTPServer-Control
{
   SourcePortRange       21
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}

IpService     FTPServer-Data
{
   SourcePortRange       20
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional OutboundConnect
   Routing               local
   SecurityClass         1
}

############################
# Reusable Service Groups  #
```

```
##########################
IpServiceGroup    FTPServer
{
   IpServiceRef   FTPServer-Control
   IpServiceRef   FTPServer-Data
}

##########################
# Reusable IP Addresses    #
##########################
IpAddr         InternalNetServerAddress
{
   Addr        9.1.1.1
}

IpAddr         InternalNetAdminAddress
{
   Addr        9.1.1.2
}

IpAddrSet      InternalNet
{
   Prefix      9.1.1.0/24
}
```

This IpSecConfig file gives FTP access to the administrator and Web access to everyone in the internal network. By relying heavily on the abstracted use of references, the policy is not only more self-explanatory, but changes to any referenced object are propagated to any statement that references it. So, if the IP address of the administrative machine or internal subnetwork changes, you merely have to make one change to an IpAddr or IpAddrSet statement, rather than modify a large number of instances in multiple rules.

Using a CommonIpSecConfig file for reusable statements: The IpSecConfig file should be tailored to the specific stack to which it belongs. However, as the policy files for IP security are being constructed, a large number of statements can be reused. Reusable statements can be placed in a common file, the CommonIpSecConfig file, which is available to all stacks. Statements in the CommonIpSecConfig file are read by all stacks on the system, providing a convenient way to store common definitions that they can all share. If you are operating in a sysplex, you can also place a CommonIpSecConfig file on shared DASD or in a shared zFS directory so that stacks in a multiple sysplex image have access to the same common configuration file.

Assuming that all statements that might be used later are placed in a CommonIpSecConfig file, the IpSecConfig file from step 10 now reads as follows:

```
# IP Security policy for Secure Server
#########################
# IpFilterPolicy block   #
#########################
IpFilterPolicy
{
   FilterLogging            on
   IpFilterGroupRef     InternalNetZoneA
}

#########################
# Security Zones         #
#########################
IpFilterGroup     InternalNetZoneA
{
   IpFilterRuleRef    AdminFTP
   IpFilterRuleRef    InternalNetWeb
```

```
}

########################
# Filter rules         #
########################
#Allow admin FTP; log traffic
IpFilterRule                AdminFTP
{
   IpSourceAddrRef          InternalNetServerAddress
   IpDestAddrRef            InternalNetAdminAddress
   IpServiceGroupRef        FTPServer
   IpGenericFilterActionRef permit-log
}

#Allow LAN Web traffic; don't log
IpFilterRule                InternalNetWeb
{
   IpSourceAddrRef          InternalNetServerAddress
   IpDestAddrSetRef         InternalNet
   IpServiceRef             WebServer
   IpGenericFilterActionRef permit-nolog
}
```

This IpSecConfig file references the following reusable statements:

- InternalNetServerAddress
- InternalNetAdminAddress
- InternalNet
- FTPServer
- FTPServer-Control
- FTPServer-Data
- permit-log
- permit-nolog
- WebServer

IpFilterRule statements can also be placed in the CommonIpSecConfig file, because some IP filter rules apply to all addresses. If certain IpFilterRule statements are to apply globally to all stacks on the system, they can go into the CommonIpSecConfig file. Use a value of `all` for the IpSourceAddr and IpDestAddr attributes. For instance, if all stacks on a z/OS system need rules permitting dynamic routing traffic (OSPF or RIP, for example), the statements that define this type of traffic can be placed in the CommonIpSecConfig file and referenced in the stack-specific file:

```
IpFilterRule                  AllowOmprouteLocalNolog
{
  IpSourceAddr                all
  IpDestAddr                  all
  IpServiceGroupRef           Omproute-local
  IpGenericFilterActionRef    Permit-nolog
}

IpServiceGroup                Omproute-local
{
  IpServiceRef                OSPF-local
  IpServiceRef                RIP-local
}

IpService                     OSPF-local
{
  Protocol                    OSPF
  Direction                   bidirectional
```

```
  Routing                       local
}

IpService                       RIP-local
{
  Protocol                      UDP
  SourcePortRange               520
  DestinationPortRange          520
  Direction                     bidirectional
  Routing                       local
}
```

With these definitions in the CommonIpSecConfig file, any stack needing global permission to send and receive routing information merely needs to include the following statement in the IpFilterPolicy block of its specific IpSecConfig file:

```
IpFilterRuleRef  AllowOmprouteLocalNolog
```

Steps for configuring the business partner model (host-to-host with IPSec)

The following statements and concepts are covered in the discussion of this model:

- Dynamic host-to-host IKE negotiations
- Key exchange rules
- Local and remote security endpoints
- Use of wildcards in Location and Identity
- IpLocalStartAction
- Granularity
- RSA signature peer authentication
- Certificates and certificate authorities
- On-demand activation
- Remote activation

Figure 86 shows the business partner portion of the security model network.

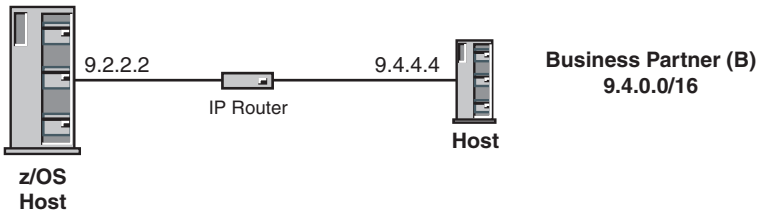


Figure 86. Business partner model

The business partner model is similar to that of the internal network, but an increased need for security means that a greater number of connections rely on data authentication and encryption. Although some services are allowed in the clear, sensitive data needs IPSec protection. To apply IPSec to a connection, the traffic that flows over that connection must match an IP filter rule that has an ipsec action.

For this example, assume you must meet the following requirements to allow network communications from a business partner in an untrusted zone B over a connected network (9.4.0.0/16) to a public IP address (9.2.2.2) on this host:

- Allow IKE traffic from untrusted zone B to this host.

- Allow secure FTP traffic (using TLS/SSL) from untrusted zone B to a secure FTP server running on this host.

  Secure FTP has its own security mechanism. Although IPSec can be used together with TLS/SSL, using both adds processing expense. For this example, secure FTP is allowed without IPSec protection.
- Allow Enterprise Extender (EE) traffic from untrusted zone B to an EE service running on this host using a dynamic IPSec tunnel with strong encryption and authentication.

  Because there is no encryption mechanism used in this example for EE, IPSec provides the secure service.
- Allow FTP traffic from untrusted zone B to an FTP server running on this host using a dynamic IPSec tunnel with strong AH authentication.
- The dynamic IPSec tunnel for EE comes up when outbound EE traffic is detected (on-demand activation).
- A dynamic IPSec tunnel for normal FTP control activates for each remote host that initiates an FTP connection (remote activation).
- A dynamic IPSec tunnel for normal FTP data activates for each outbound data connection to a remote host (local activation).
- Peers authenticate themselves using the RSA signature method.

Perform the following steps to meet these requirements and configure the business partner model.

1. For each zone, determine what services are allowed:
   - IKE traffic
   - Normal FTP traffic
   - Secure FTP traffic
   - Enterprise Extender traffic
2. Define an IpService statement for each desired service.
   - IKE uses UDP, port 500 for message exchanges:

     ```
     IpService                   IKE-local
     {
       SourcePortRange           500
       DestinationPortRange      500
       Protocol                  UDP
       Direction                 bidirectional
       Routing                   local
       SecurityClass             0
     }
     ```
   - For normal FTP traffic, allow inbound connections but do not allow outbound connection requests, other than data. Two services are required, one for the control connection and one for the data connection:

     ```
     IpService                   FTPServer-Control
     {
       SourcePortRange           21
       DestinationPortRange      1024 65535
       Protocol                  tcp
       Direction                 bidirectional InboundConnect
       Routing                   local
       SecurityClass             0
     }

     IpService                   FTPServer-Data
     {
       SourcePortRange           20
       DestinationPortRange      1024 65535
     ```

```
  Protocol                  tcp
  Direction                 bidirectional OutboundConnect
  Routing                   local
  SecurityClass             0
}
```

- For secure FTP traffic, allow inbound connections but do not allow outbound connection requests, other than data. Two services are required, one for the control connection and one for the data connection.

```
IpService                   SecureFTPServer-Control
{
  SourcePortRange           990
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   SecureFTPServer-Data
{
  SourcePortRange           989
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional OutboundConnect
  Routing                   local
  SecurityClass             0
}
```

- For Enterprise Extender traffic, allow both inbound and outbound traffic:

```
IpService                   Enterprise-Extender
{
  SourcePortRange           12000 12004
  DestinationPortRange      12000 12004
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

IP services can be grouped together for convenience, flexibility, and reuse of configuration. Any IP filter rule that includes an IP service group will be expanded to include all of the services from that group. The following IP service groups include the IP services that define the FTPServer and the SecureFTPServer:

```
IpServiceGroup FTPServer
{
   IpServiceRef FTPServer-Control
   IpServiceRef FTPServer-Data
}
IpServiceGroup SecureFTPServer
{
   IpServiceRef Secure-FTPServer-Control
   IpServiceRef Secure-FTPServer-Data
}
```

3. Determine the data endpoints to be protected. In this example, the local data endpoint is the public address of the server, and the remote data endpoint is the entire subnetwork in zone B:

```
Local public IP address:  9.2.2.2
Remote subnet: 9.4.0.0/16
```

4. Determine what level of security is needed between each set of data endpoints:

```
IKE traffic - permit
Secure FTP traffic - permit
EE traffic - IPSec encryption and authentication
FTP traffic - IPSec authentication
```

5. Configure an IpGenericFilterAction statement for the level of security that is required. The requirements stated that IKE and secure FTP should be allowed, and that EE and normal FTP traffic required IPSec protection. Notice that the parameters of the IpGenericFilterAction statement include ipsec as well as permit and deny. Because no logging requirements were specified, two actions are needed, permit and ipsec:

```
IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, do the following:
   a. Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:
      1) Determine the required type and strength of protection for the phase 1 security association. Phase 1 security associations must use both authentication and encryption. Because strong encryption was specified for phase 2, use strong encryption for phase 1:

         ```
         Authentication, SHA1 algorithm
         Encryption, 3DES algorithm
         DHGroup, Diffe-Hellman Group2
         ```

         Tip: Of the two choices, Group2 provides greater protection than Group1.

      2) Decide what type of peer authentication is used:

         ```
         RSA signature
         ```

         Because the remote data endpoint represents multiple hosts in this example, RSA signature is used. RSA signature authentication provides greater flexibility, scalability, and security than pre-shared key authentication. Because there are potentially multiple remote peers in the business partner's subnet, RSA signature is a reasonable choice.

         The use of RSA signature requires setup of RACF certificates and certificate authority information. Both IKE peers need access to a key ring with at least one X.509 digital certificate to identify itself. To set up the IKE daemon for certificates, see Appendix E, "Steps for preparing to run IP security," on page 1215. The site should decide what certificate authorities are recognized. A certificate authority can be an outside commercial entity, or it can be defined locally in RACF. For information about installing certificate authorities in RACF, refer to *z/OS Security Server RACF Security Administrator's Guide*.

         For the purposes of this example, a certificate authority with label CA4BusinessPartner is used, which is presumed to have been defined as a certificate authority in the RACF database. The label CA4BusinessPartner should be added to the iked.conf file as a recognized certificate authority as follows:

```
SupportedCerthAuth CA4BusinessPartner
```

Guideline: For more efficient processing of certificates, you should code SupportedCertAuth for each acceptable certificate authority that will be used to sign certificates for remote IKE peers.

3) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation, including type of peer authentication, strength of encryption, and how often the phase 1 keys are refreshed. Because KeyExchangeOffer statements are reusable, give it a descriptive name as follows:

```
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

4) Decide whether NAT traversal will be allowed.

   In this example, the following parameter is used:

```
AllowNat No
```

5) Determine the negotiation mode. Main mode is used for phase 1, providing added security by encrypting the identities of the two IKE peers during the phase 1 negotiation.

6) Configure a KeyExchangeAction statement that defines the control information for the phase 1 negotiation. The key exchange action determines the mode of the phase 1 negotiation and the parameters that are included in the KeyExchangeOffer statement. Although multiple KeyExchangeOffer statements are acceptable, only one is required. Both peers must agree on the parameters. Use the KeyExchangeOffer statement that was configured in step 6a3:

```
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate main
   HowToRespond  main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}
```

7) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement. In this example, the local security endpoint is the local host. An identity and IP address of the local IKE peer is required as follows:

```
LocalSecurityEndpoint  Public_IKED
{
   Identity      IpAddr  9.2.2.2
   Location      9.2.2.2
}
```

   Similarly, the same information is needed for remote hosts. Notice that in the RemoteSecurityEndpoint statement, an IP address range is specified for both identity and location. Although it is possible to configure a RemoteSecurityEndpoint statement for every host in the remote subnetwork, unless they have unique key exchange requirements, it is not necessary. Wildcards can be used for the Identity and Location parameters. Any of the identity types described earlier can potentially contain wildcards to include a group of remote hosts with similar identities.

```
RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      IpAddr   9.4.0.0/16
   Location      9.4.0.0/16
   CaLabel       CA4BusinessPartner
}
```

The use of an IPv4 address on the Identity parameter for the Local_IKED security endpoint requires that the X.509 digital certificate for the local IKE daemon include the IPv4 address in the Subject Alternative Name field of the certificate. The use of an IPv4 subnet on the Identity parameter for the ZoneB_IKED security endpoint requires that an IPv4 address within that subnet (9.4.0.0/16) appear in the Subject Alternative Name field of the X.509 digital certificate for the remote IKE daemon.

The inclusion of the CaLabel parameter in the ZoneB_IKED security endpoint emphasizes the fact that the local host requests that the remote host use only certificates that are signed by the certificate authority that is identified by the label CA4BusinessPartner.

Rule: For RSA signature mode authentication, the identity of a security endpoint must be contained in its X.509 digital certificate, either in the Subject Name field or the Subject Alternative Name field.

For detailed specifications on the use of wildcards in the RemoteSecurityEndpoint statement, refer to *z/OS Communications Server: IP Configuration Reference*.

8) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action:

```
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}
```

9) Include the key exchange rule in the KeyExchangePolicy statement:

```
KeyExchangePolicy
{
   KeyExchangeRuleRef    ZoneB_KeyExRule1
}
```

b. Configure an IpDynVpnAction statement that defines the parameters of the phase 2 negotiation as follows:

1) Determine the required type and strength of IPSec protection for the phase 2 security association.

In this example, there is a unique requirement for each traffic type. Therefore, there are two types of IPSec protection for each type of traffic as follows:

```
EE traffic - strong encryption: ESP, 3DES
             strong ESP authentication: Hmac SHA
Normal FTP traffic - strong AH authentication, Hmac SHA
```

This information eventually translates into two IpDataOffer statements.

2) Determine whether tunnel or transport mode is required.

Tunnel mode is required only if either endpoint of the security association is a secure gateway. The choice is optional in a host-to-host configuration, but transport mode is typically used.

3) Configure IpDataOffer statements that define the parameters of the phase 2 negotiation.

- Authenticated offer for FTP:

```
IpDataOffer TRAN-AHSHA-NOENCR
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    AH HMAC_SHA
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

- Encrypted and authenticated offer for EE:

```
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    ESP HMAC SHA
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

4) Determine which peer is allowed to initiate the negotiation.

   Because the EE VPN activates when outbound EE traffic is detected, you need to be able to start the negotiation locally. The remote host initiates the negotiation of the security association for the FTP control connection, so local initiation is not required. The FTP data connection is initiated from the local host, so to activate the security association, local initiation is required.

5) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation.

   The IpDynVpnAction statement can control which host is allowed to initiate the negotiation. You must allow the IPSec VPN for the FTP control connection to be initiated by the remote peer, and the data connection by the local host. EE should be allowed to initiate an IKE negotiation locally.

   - Authenticated VPN action for FTP:

```
IpDynVpnAction FTP-vpnaction
{
   Initiation  either
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-AHSHA-NOENCR
}
```

   - Encrypted and authenticated VPN action for EE traffic:

```
IpDynVpnAction EE-vpnaction
{
   Initiation  localonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-3DES
}
```

c. Decide how the security association is to be activated as follows:

1) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.

   Neither the EE nor FTP VPNs require a LocalDynVpnPolicy statement, because neither is command-line activated or autoactivated.

2) Configure an optional IpLocalStartAction statement, if the security association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host security association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

   Before a phase 2 negotiation can initiate, the IKE daemon needs to know the IP addresses, ports, and protocols that the security association covers. In most cases, these can be inferred from the filter rule, or from the packet that started the on-demand activation. An IpLocalStartAction statement explicitly defines where these parameters are obtained. The granularity setting determines whether the information comes from the matching filter rule, or from the packet. By explicitly specifying packet, you can guarantee that a new security association is created for each connection request.

```
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranularity       packet
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}
```

   In this example, specifying `packet` is not required. IKE detects that the filter rule has a range of ports specified and resorts to packet granularity instead of the default of `rule`. Specifying `packet` is required, however, if the EE and FTP rules had all ports specified. In that case, one security association is negotiated for all ports, not a single port. To avoid confusion, it is better to specify your intention.

3) Set the global PreDecap parameter of the IpFilterPolicy statement to `off`, or create an IpFilterRule statement that allows IPSec traffic (AH and ESP).

   When the system begins to process encapsulated traffic, the encapsulated packets are subject to filtering. The encrypted and authenticated packets are denied unless they are explicitly allowed. There are two options to configure this that are subtly different. The first is less restrictive, allowing AH and ESP packets from anywhere and not subjecting them to filtering until after the packets have had their IPSec headers removed. The second method is to add an IpFilterRule statement that explicitly permits the IPSec-encapsulated traffic. This option adds an additional layer of security, in that you can control exactly who is allowed to send encrypted traffic. This protects system resources by preventing the processing of IPSec packets unnecessarily, such as might happen in the case of a malicious attack of IPSec packets. Although more secure, this solution is less efficient because it requires additional processing resources to filter all IPSec-encapsulated traffic.

   For the first option, in the main IpFilterPolicy block, code `PreDecap off` as follows:

```
IpFilterPolicy
{
   PreDecap off
```

```
...
}
```

For the second option, configure a filter rule in the IpFilterPolicy block that explicitly allows AH and ESP traffic as follows:

```
IpFilterRule           Allow-IPSec-traffic
{
   IpSourceAddr        9.2.2.2
   IpDestAddrSet       9.4.0.0/16
   IpService                   AH-traffic
   {
      Protocol                 AH
      Direction                bidirectional
      Routing                  local
      SecurityClass            0
   }
   IpService                   ESP-traffic
   {
      Protocol                 ESP
      Direction                bidirectional
      Routing                  local
      SecurityClass            0
   }
   IpGenericFilterRef    permit-nolog
}
```

Tip: In most cases, unless the extra security is deemed a necessity, use the `PreDecap off` global option to allow IPSec packets to flow with minimum overhead.

7. Define an IpFilterRule statement for each set of data endpoints.

   The secure host and subnetwork B represent the data endpoints. The IpService statements were defined in step 2 on page 902. Place the IpDynVpnAction statements that were created in step 6 on page 904 with the appropriate rule.

```
IpFilterRule              ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}


IpFilterRule              FTPServer-ZoneB
{
   IpSourceAddr              9.2.2.2
   IpDestAddrSet             9.4.0.0/16
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            FTP-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
IpFilterRule             EE-ZoneB
{
   IpSourceAddr          9.2.2.2
   IpDestAddrSet         9.4.0.0/16
   IpService
   {
      SourcePortRange       12000 12004
      DestinationPortRange  12000 12004
      Protocol              udp
      Direction             bidirectional
      Routing               local
```

```
      SecurityClass         0
   }
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
```

Because both IKE and Secure FTP need to be permitted without IPSec protection, an IpFilterRule statement for both is not needed. The two services can be combined into one rule.

Tip: Any services that share the same data endpoints and the same security requirements can be placed together in one IpFilterRule statement.

8. Include the IpFilterRule statements in the IpFilterPolicy block.

   The IpFilterRule statement allowing IKE traffic should always be at the top of the list. The rest of the IpFilterRule statements are disjointed, and their relative placement within the IpFilterPolicy is irrelevant.

   Tip: If it is known that one particular type of traffic is the most frequent, placing that rule near the top of the list results in faster filter lookups.

9. Define an IP filter group for each zone and include the IpFilterRule statements that belong to that zone. Although the creation of reference objects and groups is not mandatory, they provide for ease of maintenance as the IP security policy grows more complex.

   A completely configured policy, including all objects and their references, is as follows:

```
# IpFilterPolicy for secure public server

IpFilterPolicy
{
   PreDecap             off
   IpFilterGroupRef     ZoneB
}

KeyExchangePolicy
{
   KeyExchangeRuleRef   ZoneB_KeyExRule1
}

###### All reusable statements follow #######
IpFilterGroup           ZoneB
{
   IpFilterRuleRef      ZoneB-Permitted-traffic
   IpFilterRuleRef      FTPServer-ZoneB  #IPSec-protected
   IpFilterRuleRef      EE-ZoneB         #IPSec-protected
}

####################################
# IpFilterRules                    #
#   defines:                       #
#      data endpoints              #
#      Allowed services            #
#      Actions (permit, deny, ipsec) #
####################################
IpFilterRule            ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}

IpFilterRule           EE-ZoneB
```

```
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

IpFilterRule           FTPServer-ZoneB
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         FTP-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

######################
# Local Start Actions #
######################
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranularity       packet
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}

###################
# IpService groups #
###################
IpServiceGroup            FTPServer
{
   IpServiceRef           FTPServer-Control
   IpServiceRef           FTPServer-Data
}

IpServiceGroup            SecureFTPServer
{
   IpServiceRef           SecureFTPServer-Control
   IpServiceRef           SecureFTPServer-Data
}

################################
# Services provided by this host #
################################

IpService                    IKE-local
{
  SourcePortRange            500
  DestinationPortRange       500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Control
{
  SourcePortRange            990
  DestinationPortRange       1024 65535
```

```
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   SecureFTPServer-Data
{
  SourcePortRange           989
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional OutboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   Enterprise-Extender
{
  SourcePortRange           12000 12004
  DestinationPortRange      12000 12004
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}

IpService                   FTPServer-Control
{
  SourcePortRange           21
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   FTPServer-Data
{
  SourcePortRange           20
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional OutboundConnect
  Routing                   local
  SecurityClass             0
}

######################
# Security Endpoints #
######################
LocalSecurityEndpoint  Public_IKED
{
   Identity      IpAddr  9.2.2.2
   Location      9.2.2.2
}

RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      IpAddr   9.4.0.0/16
   Location      9.4.0.0/16
   CaLabel       CA4BusinessPartner
}

#########################
# Generic filter actions #
#########################

IpGenericFilterAction    permit-nolog
```

```
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}

################################
# Key Exchange offers          #
#   defines:                   #
#     Authentication type      #
#     Encryption type          #
#     Peer authentication method #
#     Refresh limits           #
################################
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

################################
# Key Exchange Actions         #
#  defines:                    #
#    Negotiation mode          #
#    List of Key exchange offers #
################################
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate main
   HowToRespond  main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}

#################################
# KeyExchangeRules              #
#   defines:                    #
#      A pair of security endpoints  #
#      permitted in IKE negotiations #
#################################
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}

############################
# Data Offers              #
#   defines:               #
#      Encapsulation mode  #
#      Authentication type #
#      Encryption type     #
#      Refresh limits      #
############################
### Authenticated offer ###
IpDataOffer TRAN-AHSHA-NOENCR
```

```
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    AH HMAC_SHA
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

### Encrypted offer ###
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    DoNot
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

############################
# Dynamic VPN Actions      #
#   defines:               #
#     Initiation role      #
#     Pfs group            #
#     Lifetime of connection #
#     List of Data offers  #
############################
IpDynVpnAction FTP-vpnaction
{
   Initiation  either
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-AHSHA-NOENCR
}

IpDynVpnAction EE-vpnaction
{
   Initiation  localonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-3DES
}

###############
# IP addresses #
###############

IpAddr    PublicServerAddress
{
   Addr   9.2.2.2
}

IpAddrSet ZoneB-subnet
{
   Prefix 9.4.0.0/16
}
```

10. Include all configured statements in the IpSecConfig file.

Steps for configuring the business partner with NAT model (host-to-host with IPSec)

The following statements and concepts are covered in the discussion of this model:

- AllowNat and NatKeepAliveInterval parameters on the KeyExchangePolicy and KeyExchangeAction statements
- IKE traffic on UDP port 4500, in addition to port 500
- NAT implications for host-to-host dynamic IKE negotiations:
  - Local and remote data endpoints
  - Local and remote security endpoints
  - IKE initiator and responder roles
  - Restriction on HowToAuth protocol (AH not supported)
- Using wildcards for location and identity
- RSA signature peer authentication
- Certificates and certificate authorities
- CaLabel
- SupportedCertAuth

The previous business partner model assumed a network topology with both business partners using public IP addresses in their internal networks. Often one or both businesses have an internal network utilizing IETF-defined private IP addresses (10.0.0.0/8, 172.16.0.0./12, and 192.168.0.0/16). Private IP addresses cannot be routed outside an internal network. Network Address Translation (NAT) is used to create a mapping of private addresses to public addresses and perform the necessary translation as packets traverse the NAT device.

When IPSec security associations traverse a NAT, there are problems because the NAT is unable to update IP addresses and checksums that are part of the encapsulated data (encrypted, authenticated, or both). The IETF has defined a solution known as NAT traversal (NATT) that allows IPSec security associations to successfully traverse a NAT device.

Figure 87 shows the business partner with NAT topology when the business partner model topology has been modified to include private addressing in each business' private network with a NAT device in front of each private network.

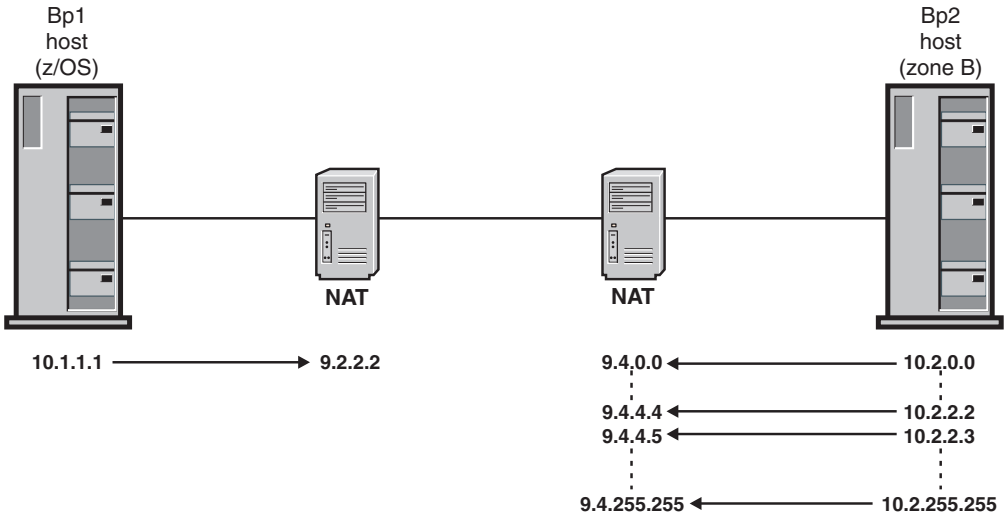


Figure 87. Business partner with NAT model

The steps in this section will describe the configuration considerations and requirements when the NATT solution is implemented to traverse NAT devices in a host-to-host environment. The business partner with NAT model has the same basic security requirements as the business partner model. Configuration statements added or changed for the business partner with NAT model are shown in **bold**. The example describes the policy for business partner 1 (BP1).

For this example, assume you must meet the following requirements to allow network communications from a business partner (BP2) in an untrusted zone B behind a NAT over a connected network (9.4.0.0/16) to a server on this host that is behind a NAT:

- IKE traffic from untrusted zone B that is behind a NAT is allowed to this host that is behind a NAT.
- Secure FTP traffic (using TLS/SSL) from untrusted zone B is allowed to a secure FTP server running on this host.
- Enterprise Extender (EE) traffic from untrusted zone B is allowed to an EE service running on this host using a dynamic IPSec tunnel with strong encryption and authentication.
- FTP traffic from untrusted zone B is allowed to an FTP server running on this host using a dynamic IPSec tunnel with strong authentication.
- The dynamic IPSec tunnel for EE is activated when outbound EE traffic is detected (on-demand activation).
- A dynamic IPSec tunnel for normal FTP control activates for each remote host that initiates an FTP connection (remote activation).
- A dynamic IPSec tunnel for normal FTP data activates for each remote host that initiates an FTP data connection (remote activation).
- Peers authenticate themselves using the RSA signature method.

Perform the following steps to meet these requirements and configure the business partner with NAT model.

1. For each zone, determine what services are allowed:
   - IKE traffic
   - Normal FTP traffic
   - Secure FTP traffic
   - Enterprise Extender traffic
2. Define an IpService statement for each desired service.
   - IKE traffic

     When NAT traversal is allowed, IKE uses UDP port 500 and port 4500 for message exchanges. NAT traversal is controlled by the AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements.

```
IpService                   IKE-local-500
{
  SourcePortRange           500
  DestinationPortRange      500
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}

IpService                   IKE-local-4500
{
  SourcePortRange           4500
```

```
  DestinationPortRange      4500
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

- Normal FTP traffic

  When a NAT is being traversed, FTP clients typically need to use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server, allowing inbound connections but not outbound connection requests. For more information on active and passive mode FTP, see "FTP considerations when a NAT is being traversed" on page 951.

  Two services are required, one for the control connection and one for the data connection. The range of server ports specified for the data connection reflects the port range specified for PASSIVEDATAPORTS in the server's FTP.DATA file. For more information on PASSIVEDATAPORTS, refer to *z/OS Communications Server: IP Configuration Reference*.

  The following definitions show the services required for the FTP server for EPSV mode:

```
IpService                   FTPServer-Control
{
  SourcePortRange           21
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   FTPServer-Data-Passive
{
  SourcePortRange           50000 50200
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}
```

- Secure FTP traffic

  Again, when a NAT is being traversed, FTP clients typically need to use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server, allowing inbound connections but not outbound connection requests. Two services are required, one for the control connection and one for the data connection.

```
IpService                   SecureFTPServer-Control
{
  SourcePortRange           990
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   SecureFTPServer-Data-Passive
{
  SourcePortRange           50201 50400
  DestinationPortRange      1024 65535
  Protocol                  tcp
```

```
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}
```

- Enterprise Extender traffic

  Allow both inbound and outbound connections.

```
IpService                   Enterprise-Extender
{
  SourcePortRange           12000 12004
  DestinationPortRange      12000 12004
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

3. Determine the data endpoints that are to be protected.

   In this case, the local data endpoint is the private address of the server, local IP address 10.1.1.1.

   The remote data endpoint is the business partner's internal network. The z/OS implementation does not require you to code private addresses for the remote endpoint; the network address translated public addresses should be configured as the remote data endpoint. In this example, the private addresses in the BP2 internal network (10.2.0.0/16) are translated into the public address range 9.4.0.0/16. Thus, the remote subnetwork is 9.4.0.0/16.

4. Determine what level of security is needed between each set of data endpoints.

```
IKE traffic - permit
Secure FTP traffic - permit
EE traffic - IPSec encryption and authentication
FTP traffic - IPSec authentication
```

5. Configure an IpGenericFilterAction statement for the level of security that is required. The requirements stated that IKE and secure FTP should be allowed, and that EE and normal FTP traffic required IPSec protection. Because no logging requirements were specified, two actions are needed, permit and ipsec.

```
IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, do the following:

   a. Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:

      1) Determine the required type and strength of protection for the phase 1 security association. Phase 1 security associations must use both authentication and encryption. Because strong encryption was specified for phase 2, use strong encryption for phase 1:

```
Authentication, SHA1 algorithm
Encryption, 3DES algorithm
DHGroup, Diffe-Hellman Group2
```

Tip: Of the two choices, Group2 provides greater protection than Group1.

2) Decide what type of peer authentication is used:

```
RSA signature
```

Because the remote data endpoint represents multiple hosts in this example, RSA signature is used. RSA signature authentication provides greater flexibility, scalability, and security than pre-shared key authentication. Because there are potentially multiple remote peers in the business partner's subnet, RSA signature is a reasonable choice.

The use of RSA signature requires setup of RACF certificates and certificate authority information. Both IKE peers need access to a key ring with at least one X.509 digital certificate to identify itself. To set up the IKE daemon for certificates, see Appendix E, "Steps for preparing to run IP security," on page 1215. The site should decide what certificate authorities are recognized. A certificate authority can be an outside commercial entity, or it can be defined locally in RACF. For information about installing certificate authorities in RACF, refer to *z/OS Security Server RACF Security Administrator's Guide*.

For the purposes of this example, a certificate authority with label CA4BusinessPartner is used, which is presumed to have been defined as a certificate authority in the RACF database. The label CA4BusinessPartner should be added to the iked.conf file as a recognized certificate authority as follows:

```
SupportedCerthAuth CA4BusinessPartner
```

Guideline: For more efficient processing of certificates, you should code SupportedCertAuth for each acceptable certificate authority that will be used to sign certificates for remote IKE peers.

3) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation, including type of peer authentication, strength of encryption, and how often the phase 1 keys are refreshed. Because KeyExchangeOffer statements are reusable, give it a descriptive name as follows:

```
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RSAsignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

4) Decide whether NAT traversal will be allowed.

The AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements allows IKE to advertise NAT traversal support. When a phase 1 security association is being negotiated, if both IKE peers support NAT traversal, several IKE payloads are exchanged that allow the peers to determine if one or more NAT devices are being traversed. If a NAT is being traversed, the IKE peers negotiate a UDP-Encapsulated-Transport or UDP-Encapsulated-Tunnel mode phase 2 security association to allow IPSec traffic to traverse the NATs successfully. If a NAT is not being traversed, a standard transport or tunnel mode phase 2 security association is negotiated.

The AllowNat parameter can be specified on the overall KeyExchangePolicy statement or for a specific KeyExchangeAction statement. If AllowNat is No, IKE does not advertise its support for NAT traversal. You might want to disable NAT traversal for interoperability purposes (certain remote resources might not be able to tolerate NAT traversal protocols) or security concerns (for example, NAT traversal exposes private internal addresses to the IKE peer). If NAT traversal is disabled and there is a NAT device in the path, ipsec processing might fail to negotiate the security association or fail to send data over the security association.

For this model, allow NAT traversal support to be advertised for all phase 1 security associations by coding the following on the KeyExchangePolicy statement:

```
AllowNat Yes
```

A NatKeepAliveInterval parameter is also provided on the KeyExchangePolicy statement. A NAT keep-alive timer is maintained to ensure that NAT mappings do not expire. If z/OS is behind a NAT, a NAT keep-alive timer is started with the interval specified on the NatKeepAliveInterval parameter. If z/OS is behind a NAT that is using static mappings that will not expire, the NatKeepAliveInterval parameter should be set to 0. It is not necessary to run a NAT keep-alive timer in this case.

For this model, since static mapping is being used for the NAT in front of BP1, the NatKeepAliveInterval is set to 0:

```
NatKeepAliveInterval   0
```

5) Determine the negotiation mode. Main mode is used for phase 1 in this example, providing added security by encrypting the identities of the two IKE peers during the phase 1 negotiation.
6) Configure a KeyExchangeAction statement that defines the control information for the phase 1 negotiation. The key exchange action determines the mode of the phase 1 negotiation and the parameters that are included in the KeyExchangeOffer statement. Although multiple KeyExchangeOffer statements are acceptable, only one is required. Both peers must agree on the parameters. Use the KeyExchangeOffer statement that was configured previously:

   ```
   KeyExchangeAction Main-RSA-SHA1-3DES-DH2
   {
      HowToInitiate main
      HowToRespond  main
      KeyExchangeOfferRef RSA-SHA1-3DES-DH2
   }
   ```

7) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement. In this example, the local security endpoint is the local host. An identity and IP address of the local IKE peer is required as follows:

   ```
   LocalSecurityEndpoint  Local_IKED
   {
      Identity      Fqdn  Server.BP1.example.com
      Location      10.1.1.1
   }
   ```

   The value specified for Identity on the LocalSecurityEndpoint statement is sent to the remote peer to identify this endpoint. The remote peer uses this value to determine if a security association can

be negotiated and which policy should be used. Identity can be specified in different formats, such as an IP address or fully qualified domain name.

The business partner model specified an IP address for Identity, and an IP address could be used in this model. However, BP1's private IP address 10.1.1.1 is not meaningful in the BP2 configuration, so one of the other Identity types was chosen. The fully qualified domain name (Fqdn) identifies the local security endpoint.

An identity and IP address are also needed for the remote hosts. Notice that in the following RemoteSecurityEndpoint statement, the Identity and Location attributes use a wildcard to include a group of remote hosts with similar identities. It is possible to wildcard the values because the key exchange requirements are the same for the hosts included in the specified range.

The Identity attribute is specified as `Fqdn` (fully qualified domain name), and the Location attribute is the range of public IP addresses used by the BP2 internal network. BP2 internal private addresses are not coded because they have no meaning for BP1.

```
RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      Fqdn   *.BP2.example.com
   Location      9.4.0.0/16
   CaLabel       CA4BusinessPartner
}
```

The use of a fully-qualified domain name on the Identity parameter for the Local_IKED security endpoint requires that the X.509 digital certificate for the local IKE daemon include the fully-qualified domain name in the Subject Alternative Name field of the certificate. The use of a wild-carded domain name on the Identity parameter for the ZoneB_IKED security endpoint requires that a fully-qualified domain name ending with .BP2.example.com appear in the Subject Alternative Name field of the X.509 digital certificate for the remote IKE daemon.

The inclusion of the CaLabel parameter in the ZoneB_IKED security endpoint emphasizes the fact that the local host requests that the remote host use only certificates that are signed by the certificate authority that is identified by the label CA4BusinessPartner.

Rule: For RSA signature mode authentication, the identity of a security endpoint must be contained in its X.509 digital certificate, either in the Subject Name field or the Subject Alternative Name field.

For detailed specifications on the use of wildcards in the RemoteSecurityEndpoint statement, refer to *z/OS Communications Server: IP Configuration Reference*.

8) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action:

```
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}
```

9) Include the key exchange rule in the KeyExchangePolicy statement block. Also include the AllowNat parameter:

```
KeyExchangePolicy
{
   AllowNat              Yes
   KeyExchangeRuleRef    ZoneB_KeyExRule1
}
```

b. Configure an IpDynVpnAction statement that defines the parameters of the phase 2 negotiation as follows:

   1) Determine the required type and strength of IPSec protection for the phase 2 security association.

      In this example, there is a unique requirement for each traffic type. Therefore, there are two types of IPSec protection for each type of traffic as follows:

      ```
      EE traffic - strong encryption: ESP, 3DES
                   strong ESP authentication: Hmac SHA
      Normal FTP traffic - strong ESP authentication, Hmac SHA
      ```

      The business partner model used AH authentication for normal FTP traffic, but the NATT solution defined by the IETF is based on the ESP protocol. The AH protocol is not supported for NATT, and ESP authentication is used in this model.

      This information eventually translates into two IpDataOffer statements.

   2) Determine whether tunnel or transport mode is required.

      Tunnel mode is required only if either endpoint of the security association is a secure gateway. The choice is optional in a host-to-host configuration, but transport mode is typically used.

   3) Configure IpDataOffer statements that define the parameters of the phase 2 negotiation.

      - Authenticated offer for FTP:

        ```
        IpDataOffer TRAN-ESPSHA-NOENCR
        {
           HowToEncap Transport
           HowToEncrypt DoNot
           HowToAuth    ESP HMAC_SHA
           RefreshLifetimeProposed 240
           RefreshLifetimeAccepted 120 480
           RefreshLifesizeProposed  none
           RefreshLifesizeAccepted  none
        }
        ```

      - Encrypted and authenticated offer for EE:

        ```
        IpDataOffer TRAN-ESPSHA-3DES
        {
           HowToEncap Transport
           HowToEncrypt 3DES
           HowToAuth    ESP HMAC_SHA
           RefreshLifetimeProposed 240
           RefreshLifetimeAccepted 120 480
           RefreshLifesizeProposed  none
           RefreshLifesizeAccepted  none
        }
        ```

   4) Determine which peer is allowed to initiate the negotiation.

      When an IKE peer is behind a NAT device, there are implications regarding which peer can initiate. In a host-to-host configuration, IKE should only initiate to a peer whose IP address is unambiguous. If the peer is not behind a NAT, or if the peer's NAT mapping is static, the address is unambiguous. In this model, assume that static mapping is being used for both BP1 and BP2.

Because the EE VPN activates when outbound EE traffic is detected, you need to be able to start the negotiation locally. For PASV mode or EPSV mode FTP, both the FTP control connection and the FTP data connection initiate the negotiation of the security association remotely, so local initiation is not required.

5) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation.

   The IpDynVpnAction statement can control which host is allowed to initiate the negotiation. Both the FTP control connection and data connection will be initiated remotely. EE should be allowed to initiate an IKE negotiation locally.

   - Authenticated VPN action for FTP:

```
IpDynVpnAction FTP-vpnaction
{
   Initiation  remoteonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-NOENCR
}
```

   - Encrypted and authenticated VPN action for EE traffic:

```
IpDynVpnAction EE-vpnaction
{
   Initiation  localonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-3DES
}
```

c. Decide how the security association is to be activated as follows:

1) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.

   Neither the EE nor FTP VPNs require a LocalDynVpnPolicy statement, because neither is command-line activated or autoactivated.

2) Configure an optional IpLocalStartAction statement, if the security association is to be activated locally (that is, on-demand, command-line, or autoactivated) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host security association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

   Before a phase 2 negotiation can initiate, the IKE daemon needs to know the IP addresses, ports, and protocols that the security association covers. In most cases, these can be inferred from the filter rule, or from the packet that started the on-demand activation. An IpLocalStartAction statement explicitly defines where these parameters are obtained. The granularity setting determines whether the information comes from the matching filter rule, or from the packet. By explicitly specifying packet, you can guarantee that a new security association is created for each connection request.

```
IpLocalStartAction          ZoneB-Start-Action
{
   AllowOnDemand            yes
   LocalPortGranularity     packet
   RemotePortGranularity    packet
   ProtocolGranularity      packet
   LocalIpGranularity       packet
```

```
   RemoteIpGranulatiry       packet
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}
```

In this example, specifying `packet` for LocalPortGranularity and RemotePortGranularity is not required. IKE detects that the filter rule has a range of ports specified and resorts to packet granularity instead of the default of `rule`. Specifying `packet` is required, however, if the EE rules had all ports specified. In that case, one security association is negotiated for all ports, not a single port. To avoid confusion, it is better to specify your intention.

3) Set the global PreDecap parameter of the IpFilterPolicy statement to `off`, or create an IpFilterRule statement that allows IPSec traffic (AH and ESP).

   In this example, set `PreDecap off` in the IpFilterPolicy statement

7. Define an IpFilterRule statement for each set of data endpoints.

   The secure host and zone B represent the data endpoints. The IpService statements were defined previously. Place the IpDynVpnAction statements that were created previously with the appropriate rule. Notice that the parameters of the IpGenericFilterAction statement include ipsec as well as permit and deny. Because no logging requirements were specified, two actions are needed, permit and ipsec.

```
IpFilterRule              ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}


IpFilterRule                 FTPServer-ZoneB
{
IpSourceAddr              10.1.1.1
IPDestAddrSet             9.4.0.0/16
IpServiceRef              FTPServer-Control
   IpServiceRef              FTPServer-Data-Passive
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            FTP-vpnaction
}

IpFilterRule             EE-ZoneB
{
IpSourceAddr              10.1.1.1
IPDestAddrSet             9.4.0.0/16
IpServiceRef          Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
```

   Because IKE port 500, IKE port 4500, and secure FTP without IPSec protection are permitted, a separate IpFilterRule statement for each is not needed. The services can be combined into one IpFilterRule statement.

   Tip: Any services that share the same data endpoints and the same security requirements can be placed together in one IpFilterRule statement.

8. Include the IpFilterRule statements in the IpFilterPolicy block.

9. Define an IP filter group for each zone and include the IpFilterRule statements that belong to that zone. Although the creation of reference objects and groups is not mandatory, they provide for ease of maintenance as the IP security policy grows more complex.
10. Include all configured statements in the IpSecConfig file.

A completely configured policy, including all objects and their references, is as follows:

```
# IpFilterPolicy for secure public server

IpFilterPolicy
{
   PreDecap             off
   IpFilterGroupRef     ZoneB
}

KeyExchangePolicy
{
   AllowNat             Yes
   NatKeepAliveInterval 0
   KeyExchangeRuleRef   ZoneB_KeyExRule1
}

###### All re-usable statements follow #######
IpFilterGroup           ZoneB
{
   IpFilterRuleRef      ZoneB-Permitted-traffic
   IpFilterRuleRef      FTPServer-ZoneB  #IPSec-protected
   IpFilterRuleRef      EE-ZoneB         #IPSec-protected
}

####################################
# IpFilterRules                    #
#   defines:                       #
#      data endpoints              #
#      Allowed services            #
#      Actions (permit, deny, ipsec) #
####################################
IpFilterRule            ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}

IpFilterRule           EE-ZoneB
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

IpFilterRule           FTPServer-ZoneB
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
```

```
   IpDynVpnActionRef         FTP-vpnaction
}

######################
# Local Start Actions #
######################
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranulatiry       packet
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}

###################
# IpService groups #
###################
IpServiceGroup            FTPServer
{
   IpServiceRef           FTPServer-Control
   IpServiceRef           FTPServer-Data-Passive
}

IpServiceGroup            SecureFTPServer
{
   IpServiceRef           SecureFTPServer-Control
   IpServiceRef           SecureFTPServer-Data-Passive
}

#################################
# Services provided by this host #
#################################

IpService                    IKE-local-500
{
  SourcePortRange            500
  DestinationPortRange       500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    IKE-local-4500
{
  SourcePortRange            4500
  DestinationPortRange       4500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}


IpService                    SecureFTPServer-Control
{
  SourcePortRange            990
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}
```

```
IpService                    SecureFTPServer-Data-Passive
{
  SourcePortRange            50201 50400
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    Enterprise-Extender
{
  SourcePortRange            12000 12004
  DestinationPortRange       12000 12004
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    FTPServer-Control
{
  SourcePortRange            21
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                   FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }

######################
# Security Endpoints #
######################
LocalSecurityEndpoint  Local_IKED
{
   Identity      Fqdn  Server.BP1.example.com
   Location      10.1.1.1
   CaLabel       CA4BusinessPartner
}

RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      Fqdn   *.BP2.example.com
   Location      9.4.0.0/16
}

#########################
# Generic filter actions #
#########################

IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
```

```
                    {
                       IpFilterAction       ipsec
                       IpFilterLogging       no
                    }

                    ################################
                    # Key Exchange offers          #
                    #   defines:                   #
                    #     Authentication type      #
                    #     Encryption type          #
                    #     Peer authentication method #
                    #     Refresh limits           #
                    ################################
                    KeyExchangeOffer RSA-SHA1-3DES-DH2
                    {
                       HowToEncrypt    3DES
                       HowToAuthMsgs   SHA1
                       HowToAuthPeers  Rsa Signature
                       DHGroup         Group2
                       RefreshLifetimeProposed  480
                       RefreshLifetimeAccepted  240 1440
                       RefreshLifesizeProposed  none
                       RefreshLifesizeAccepted  none
                    }

                    ################################
                    # Key Exchange Actions         #
                    #  defines:                    #
                    #    Negotiation mode          #
                    #    List of Key exchange offers #
                    ################################
                    KeyExchangeAction Main-RSA-SHA1-3DES-DH2
                    {
                       HowToInitiate main
                       HowToRespond  main
                       KeyExchangeOfferRef RSA-SHA1-3DES-DH2
                    }

                    ###################################
                    # KeyExchangeRules                #
                    #   defines:                      #
                    #      A pair of security endpoints  #
                    #      permitted in IKE negotiations #
                    ###################################
                    KeyExchangeRule             ZoneB_KeyExRule1
                    {
                       LocalSecurityEndpointRef  Local_IKED
                       RemoteSecurityEndpointRef ZoneB_IKED
                       KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
                    }

                    ############################
                    # Data Offers              #
                    #   defines:               #
                    #      Encapsulation mode  #
                    #      Authentication type #
                    #      Encryption type     #
                    #      Refresh limits      #
                    ############################
                    ### Authenticated offer ###
                    IpDataOffer TRAN-ESPSHA-NOENCR
                    {
                       HowToEncap Transport
                       HowToEncrypt DoNot
                       HowToAuth    ESP HMAC_SHA
                       RefreshLifetimeProposed 240
                       RefreshLifetimeAccepted 120 480
```

```
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

### Encrypted Authenticated offer ###
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    ESP HMAC_SHA
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

############################
# Dynamic VPN Actions      #
#   defines:               #
#     Initiation role      #
#     Pfs group            #
#     Lifetime of connection #
#     List of Data offers  #
############################
IpDynVpnAction FTP-vpnaction
{
   Initiation  remoteonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-NOENCR
}

IpDynVpnAction EE-vpnaction
{
   Initiation  localonly
   Pfs         group2
   VpnLife     1440
   IpDataOfferRef TRAN-ESPSHA-3DES
}

###############
# IP addresses #
###############

IpAddr    PrivateServerAddress
{
   Addr   10.1.1.1
}

IpAddrSet ZoneB-subnet
{
   Prefix 9.4.0.0/16
}
```

Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)

The following topics are covered in the discussion of this model:

- Host-to-gateway
- Gateway-to-gateway
- Autoactivation
- Command-line activation
- Using wildcards for location and identity
- Pre-shared key peer authentication

Figure 88 shows the branch office portion of the security model network.

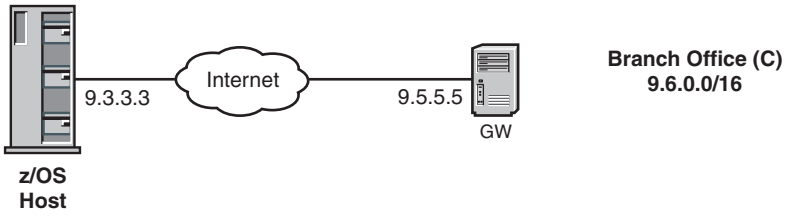


Figure 88. Branch office model

Transport-mode IPSec is used exclusively for transporting encrypted data directly between two hosts. Tunnel-mode IPSec encapsulation must be used if one of the security endpoints is a security gateway routing traffic for any number of hosts. In the branch office model, there can be multiple hosts behind a security gateway that is providing the security mechanism for all hosts that reside behind the security gateway. The base assumption is that the local secure server represents a single data endpoint and a single security endpoint, whereas the remote subnetwork represents multiple data endpoints which share a common security endpoint, namely the security gateway host. Traffic from the local server to the branch office gateway is IPSec protected, while traffic from the branch office gateway to the hosts behind the security gateway need not be encrypted or even filtered.

As in the previous models, a complete IP security policy defines the traffic that is allowed between the local server and the zone representing the branch office. However, the difference lies primarily in where the remote security endpoint is situated. In the previous examples, all of the IPSec protection was provided on a host-to-host basis. Each set of communicating endpoints had a single dynamic VPN that represented a secure channel of communication between two hosts, local and remote. In contrast, any scenario that involves an IPSec security gateway can require that one VPN carry traffic for multiple hosts. This difference is highlighted in this branch office example.

For this example, assume the following requirements to allow network communications from zone C, a branch office network (9.6.0.0/16), to a public IP address (9.3.3.3) on this host. The hosts on the branch office network connect to the Internet through the public branch office gateway server (9.5.5.5).

- Allow IKE traffic from branch office zone C to this host.
- Allow EE traffic from branch office zone C to an EE service running on this host, using a dynamic VPN with strong authentication and encryption. The VPN should be up when the stack initializes.
- Allow normal FTP traffic from branch office zone C to an FTP server running on this host, using a dynamic VPN with strong authentication and encryption. The VPN is command-line activated. Only one VPN should be established to carry all FTP traffic. There will not be one VPN per connection, but rather one security association will be negotiated for all of the remote FTP clients.

  Restriction: Negotiating a single security association for multiple remote clients is possible only when the remote security endpoint is acting as a secure gateway.
- Peers authenticate themselves using the pre-shared key method.

Perform the following steps to meet these requirements and configure part 1 of the branch office model.

1. Determine the number of zones to be protected.

The branch office represents one zone, zone C.

2. For each zone, determine what services are allowed and define an IpService block for each desired service. Services are defined by their protocols and the well-known ports they use.

   The definitions that describe FTP traffic can be combined in an IP service group as follows:

```
IpServiceGroup               FTPServer
{
  IpService                  FTPServer-Control
  {
    SourcePortRange            21
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }
  IpService                  FTPServer-Data
  {
    SourcePortRange            20
    Protocol                   tcp
    Direction                  bidirectional OutboundConnect
    Routing                    local
    SecurityClass              0
  }
  IpService                  FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }
}
```

   The traffic pattern for Enterprise Extender can be defined in one IpService block as follows:

```
IpService                    Enterprise-Extender
{
  SourcePortRange            12000 12004
  DestinationPortRange       12000 12004
  Protocol                   udp
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}
```

3. Determine the data endpoints to be protected. Typically, this is both a local and remote IP address or subnetwork.

   In this example, for the local host's public IP address, 9.3.3.3, define the following:

```
IpAddr                       PublicServerAddressA1
{
  Addr                       9.3.3.3
}
```

   For the remote subnet, 9.6.0.0/16, the following is defined:

```
IpAddrSet                    SubnetC
{
  Prefix                     9.6.0.0/16
}
```

   Because it is necessary to permit IKE traffic between the local public server and the remote branch office gateway, the IP address of the remote gateway must also be defined as follows:

```
IpAddr                      BranchOfficeGateway
{
  Addr                      9.5.5.5
}
```

4. Determine what level of security is needed between each set of data endpoints.

   In this example, strong authentication and encryption is needed for both EE and FTP traffic, so ESP authentication and ESP encryption are used.

5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged, as follows:

```
IpGenericFilterAction       ipsec
{
  IpFilterAction            ipsec
}
```

6. If IPSec is required between any two endpoints, do the following:

   a. Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:

      1) Determine the required type and strength of protection for the phase 1 security association.

         In this example, SHA1 authentication is used because it is more secure than MD5, and 3DES encryption is used because it is more secure than DES.

      2) Decide what type of peer authentication to use.

         In this example, pre-shared key authentication is specified in the requirements. Typically, the RSA signature method is preferable, given its numerous advantages, but for the purposes of example the pre-shared key method is used here. Because there is only one remote IKE peer in the branch office scenario (the remote security gateway), and because it is relatively simple to configure, pre-shared key authentication is a reasonable choice.

      3) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation as follows:

```
KeyExchangeOffer            SHA1-3DES-PSK
{
  HowToEncrypt              3DES
  HowToAuthMsgs             SHA1
  HowToAuthPeers            PresharedKey
}
```

      4) Decide whether NAT traversal will be allowed.

         In this example, the following parameter is used:

```
AllowNat No
```

      5) Determine the negotiation mode, Main or Aggressive.

         Because security is a priority in the branch office model, the more secure Main mode is used for the phase 1 negotiation.

      6) Configure a KeyExchangeAction statement as follows:

```
KeyExchangeAction           Gold-PSK
{
  HowToInitiate             main
  HowToRespond              main
  KeyExchangeOfferRef       SHA1-3DES-PSK
}
```

7) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement as follows:

```
LocalSecurityEndpoint        Public_IKED
{
  Identity                   IpAddr 9.3.3.3
  LocationRef                PublicServerAddressA1
}

RemoteSecurityEndpoint       ZoneC_IKED
{
  Identity                   Fqdn gateway.BO.example.com
  LocationRef                BranchOfficeGateway
}
```

8) Configure a KeyExchangeRule statement that includes the two endpoints, the key exchange action, and the pre-shared key as follows:

```
KeyExchangeRule              ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef   Public_IKED
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
  PresharedKey               abracadabra
}
```

9) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block, as follows:

```
KeyExchangePolicy
{
    KeyExchangeRuleRef         ZoneC_KeyExRule1
}
```

b. Configure an IpDynVpnAction statement defining the control of the phase 2 negotiation, as follows:

1) Determine the required type and strength of IPSec protection for the phase 2 security association.

   In this example, authentication is ESP HMAC_SHA and encryption is ESP 3DES.

2) Determine whether tunnel or transport mode is required.

   Tunnel mode is required when one of the security endpoints is a security gateway.

3) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation, as follows:

```
IpDataOffer                  SHA-3DES-Tunnel
{
  HowToEncap                 tunnel
  HowToEncrypt               3DES
  HowToAuth                  ESP HMAC_SHA
}
```

4) Determine which peer is allowed to initiate the negotiation.

   Because the EE VPN is up when the stack starts, you need to be able to start the negotiation locally (that is, autoactivate it). The FTP VPN is command-line activated, so you need to be able to start the negotiation locally.

5) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation, as follows:

```
IpDynVpnAction              Gold-TunnelMode
{
  Initiation                either
  Pfs                       Group2
  IpDataOfferRef            SHA-3DES-Tunnel
}
```

c. Decide how the security association is activated.

1) Optionally, configure a local dynamic VPN policy, if command-line activated or autoactivated.

A LocalDynVpnRule statement is required for each security association because none are on-demand activated, five for the EE traffic and two for the FTP traffic, as follows:

```
LocalDynVpnRule             ZoneC_VPN-EE1
{
   LocalIpRef                PublicServerAddressA1
   RemoteIpSetRef            SubnetC
   LocalDataPort             12000
   RemoteDataPort            12000
   Protocol                  UDP
   AutoActivate              Yes
}

LocalDynVpnRule             ZoneC_VPN-EE2
{
   LocalIpRef                PublicServerAddressA1
   RemoteIpSetRef            SubnetC
   LocalDataPort             12001
   RemoteDataPort            12001
   Protocol                  UDP
   AutoActivate              Yes
}

LocalDynVpnRule             ZoneC_VPN-EE3
{
   LocalIpRef                PublicServerAddressA1
   RemoteIpSetRef            SubnetC
   LocalDataPort             12002
   RemoteDataPort            12002
   Protocol                  UDP
   AutoActivate              Yes
}

LocalDynVpnRule             ZoneC_VPN-EE4
{
   LocalIpRef                PublicServerAddressA1
   RemoteIpSetRef            SubnetC
   LocalDataPort             12003
   RemoteDataPort            12003
   Protocol                  UDP
   AutoActivate              Yes
}

LocalDynVpnRule             ZoneC_VPN-EE5
{
   LocalIpRef                PublicServerAddressA1
   RemoteIpSetRef            SubnetC
   LocalDataPort             12004
   RemoteDataPort            12004
   Protocol                  UDP
   AutoActivate              Yes
}

LocalDynVpnRule             ZoneC_VPN-FTP-Data
{
   LocalIpRef                PublicServerAddressA1
```

```
   RemoteIpSetRef             SubnetC
   LocalDataPort              20
   RemoteDataPort             0
   Protocol                   TCP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-FTP-Control
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              21
   RemoteDataPort             0
   Protocol                   TCP
   AutoActivate               Yes
}
```

2) Configure an optional IpLocalStartAction statement, if the security association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host security association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

   If the local secure server initiates an IKE negotiation, it must be able to identify a remote IKE peer. However, in this case, the remote data endpoint is not a single host, but multiple endpoints in a subnetwork, ZoneC. The IpLocalStartAction statement is used to identify the remote IKE peer, and is required if the local IKE daemon initiates an ipsec connection with a remote security gateway.

```
IpLocalStartAction             StartZoneC
{
   RemoteSecurityEndpointRef   ZoneC_IKED
}
```

3) Create an IpFilterRule statement that allows IPSec traffic (AH and ESP), or set the global IpFilterPolicy statement parameter PreDecap to `off`.

   In this example, `PreDecap off` is used in the IpFilterPolicy statement.

7. Define an IpFilterRule statement for each set of data endpoints. The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500. See the IpAddr and IpAddrSet statements configured in step 3 on page 931.

```
IpFilterRule                 Rule1C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrRef              BranchOfficeGateway
   IpServiceRef               IKE
   IpGenericFilterActionRef   permit
}

IpFilterRule                 Rule2C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrSetRef           SubnetC
   IpServiceRef               Enterprise-Extender
   IpServiceGroupRef          FTPServer
```

```
        IpGenericFilterActionRef   ipsec
        IpDynVpnActionRef          Gold-TunnelMode
        IpLocalStartActionRef      StartZoneC
    }
```

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include all IpFilterGroup references and, optionally, any additional IpFilterRule statements as needed, in the IpFilterPolicy block.
10. Include all configured statements in the IpSecConfig file.

Following is a complete IP security policy for traffic from the local secure server to zone C, assuming that the reusable statements have been included in the CommonIpSecConfig policy file:

```
#------------------------------------------------------
# Filter Policy for Secure Server
#------------------------------------------------------
IpFilterPolicy
{
  PreDecap                  off
  FilterLogging             on
  AllowOnDemand             no
  IpFilterGroupRef          ZoneC
}

#------------------------------------------------------
# KeyExchange Policy for Secure Server
#------------------------------------------------------
KeyExchangePolicy
{
  KeyExchangeRuleRef        ZoneC_KeyExRule1
}

#------------------------------------------------------
# LocalDynVpn Policy for Secure Server
#------------------------------------------------------
LocalDynVpnPolicy
{
  LocalDynVpnGroupRef       ZoneC_BranchOfficeVPNs
}

#######################################################
#               Connectivity Profile
#             Secure Server To Zone C
#
#   Server to Trusted Branch Office Network
#
#######################################################
IpFilterGroup               ZoneC
{
   #-----------------------------------------------------
   # Permitted Zone C traffic:
   #    Allow IKE traffic from the gateway IKE Server
   #    for branch office to this host.
   #
   #    IKE  (UDP port 500) - IKE negotiations
   #-----------------------------------------------------

   IpFilterRule               Rule1C
   {
     IpSourceAddrRef          PublicServerAddressA1
     IpDestAddrRef            BranchOfficeGateway
     IpServiceRef             IKE
     IpGenericFilterActionRef permit
```

```
   }

   #-------------------------------------------------------
   # IPSec-protected Zone C traffic:
   #
   # Enterprise Extender (ports 12000-12004)
   #    FTP Server - SubnetC to PublicServerAddressA
   #-------------------------------------------------------

   IpFilterRule                 Rule2C
   {
     IpSourceAddrRef            PublicServerAddressA1
     IpDestAddrSetRef           SubnetC
     IpServiceRef               Enterprise-Extender
     IpServiceGroupRef          FTPServer
     IpGenericFilterActionRef   ipsec
     IpDynVpnActionRef          Gold-TunnelMode
     IpLocalStartActionRef      StartZoneC
   }
}

IpLocalStartAction             StartZoneC
{
   AllowOnDemand               yes
   RemoteSecurityEndpointRef   ZoneC_IKED
}

KeyExchangeRule              ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef   Public_IKED
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
}

#-------------------------------------------------------
# Zone C LocalDynVpnRules
#
# Setup SAs for EE traffic from branch office zone C to
# EE (UDP ports 12000-12004).
#-------------------------------------------------------

LocalDynVpnGroup             ZoneC_BranchOfficeVPNs
{
   LocalDynVpnRule              ZoneC_VPN-EE1
   {
     LocalIpRef                 PublicServerAddressA1
     RemoteIpSetRef             SubnetC
     LocalDataPort              12000
     RemoteDataPort             12000
     Protocol                   UDP
     AutoActivate               Yes
   }

   LocalDynVpnRule              ZoneC_VPN-EE2
   {
     LocalIpRef                 PublicServerAddressA1
     RemoteIpSetRef             SubnetC
     LocalDataPort              12001
     RemoteDataPort             12001
     Protocol                   UDP
     AutoActivate               Yes
   }

   LocalDynVpnRule              ZoneC_VPN-EE3
   {
     LocalIpRef                 PublicServerAddressA1
     RemoteIpSetRef             SubnetC
```

```
      LocalDataPort             12002
      RemoteDataPort            12002
      Protocol                  UDP
      AutoActivate              Yes
    }

    LocalDynVpnRule             ZoneC_VPN-EE4
    {
      LocalIpRef                PublicServerAddressA1
      RemoteIpSetRef            SubnetC
      LocalDataPort             12003
      RemoteDataPort            12003
      Protocol                  UDP
      AutoActivate              Yes
    }

    LocalDynVpnRule             ZoneC_VPN-EE5
    {
      LocalIpRef                PublicServerAddressA1
      RemoteIpSetRef            SubnetC
      LocalDataPort             12004
      RemoteDataPort            12004
      Protocol                  UDP
      AutoActivate              Yes
    }

#-------------------------------------------------------
# Setup SAs for FTP traffic from branch office zone C
# to an FTP server running on this host using a dynamic
# vpn (TCP port 20, 21).
#-------------------------------------------------------

    LocalDynVpnRule             ZoneC_VPN-FTP-Data
    {
      LocalIpRef                PublicServerAddressA1
      RemoteIpSetRef            SubnetC
      LocalDataPort             20
      RemoteDataPort            0
      Protocol                  TCP
      AutoActivate              Yes
    }

    LocalDynVpnRule             ZoneC_VPN-FTP-Control
    {
      LocalIpRef                PublicServerAddressA1
      RemoteIpSetRef            SubnetC
      LocalDataPort             21
      RemoteDataPort            0
      Protocol                  TCP
      AutoActivate              Yes
    }
}
```

Steps for configuring the branch office with NAT model (host-to-gateway with IPSec)

The following NAT implications for host-to-gateway dynamic IKE negotiations are covered in the discussion of this model:

- Local and remote data endpoints
- Local and remote security endpoints
- IKE initiator and responder roles
- Restriction on HowToAuth protocol (AH not supported)

The previous branch office model assumed a network topology with both the host and the security gateway, as well as the hosts behind the security gateway, using public IP addresses. Often one or both security endpoints are behind a NAT utilizing a private IP address.

Modifying the branch office model topology to include a NAT in front of the security gateway, the branch office with NAT topology becomes as shown in Figure 89:

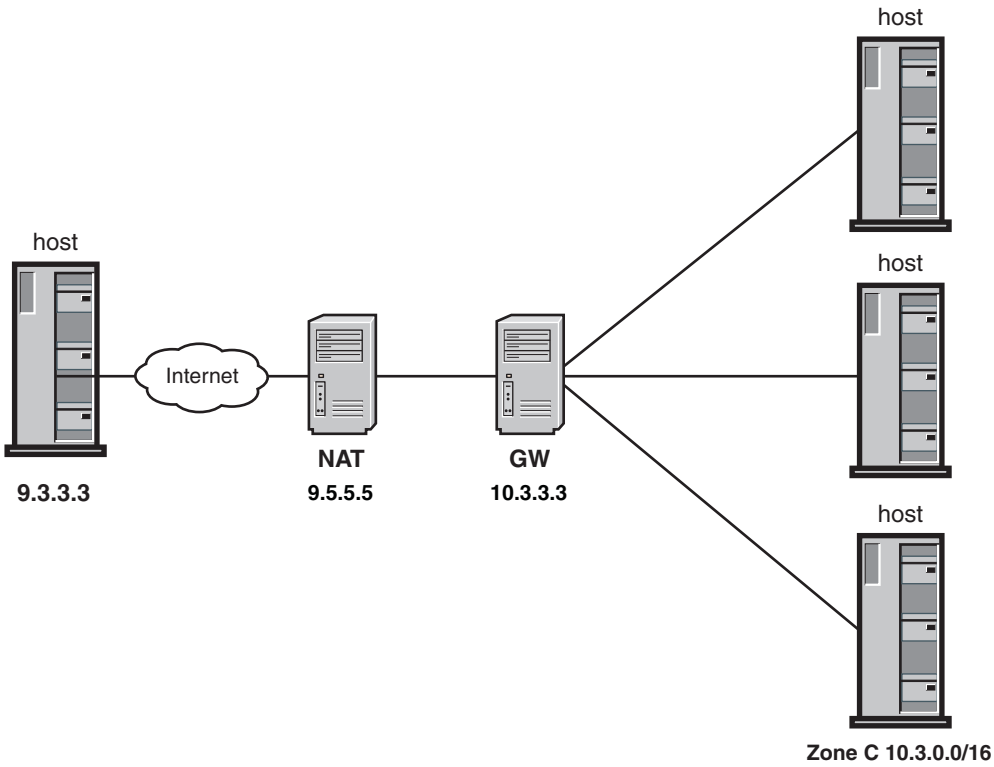


Figure 89. Branch office with NAT model

This example will describe the configuration considerations and requirements when the NATT solution is implemented to traverse NAT devices in a host-to-security gateway environment. The branch office with NAT model has the same basic security requirements as the branch office model. Configuration statements added or changed for the branch office with NAT model are shown in **bold**. The example describes the policy for host 9.3.3.3.

For this example, assume the following requirements to enable network communications from zone C, a branch office network using private IP addresses (10.3.0.0/16), to a public IP address (9.3.3.3) on this host. The hosts on the branch office network connect to the Internet through the branch office gateway server, which is behind a NAT. In this model, the NAT has a static mapping of the security gateway's private address 10.3.3.3 to the public address 9.5.5.5.

- Allow IKE traffic from the branch office zone C security gateway to this host.
- Allow EE traffic from branch office zone C to an EE service running on this host, using a dynamic VPN with strong authentication and encryption. Only one host behind the security gateway (branch office zone C) will be able to send EE traffic.

Guideline: In most cases, EE hosts should not be located behind a security gateway that is behind a NAT. Instead, a host-to-host security association should be negotiated for each EE host.

For the previous branch office scenario, the VPN was brought up when the z/OS stack initialized by setting the AutoActivate parameter to Yes on the LocalDynVpnRule statement. In this scenario, the hosts behind the gateway do not have public addresses that can be configured in the policy. Therefore, initiation from host 9.3.3.3 to the security gateway 9.5.5.5 becomes ambiguous because the IP address of the remote data endpoint is unknown. z/OS does not allow initiation of a UDP-Encapsulated-Tunnel mode security association to a security gateway. The security associations between the security gateway 9.5.5.5 and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as responder.

- Allow normal FTP traffic from branch office zone C to an FTP server running on this host, using a dynamic VPN with strong authentication and encryption. For the previous branch office scenario, the VPN was activated by an administrator from the z/OS UNIX command line. Again, z/OS does not allow initiation of a UDP-Encapsulated-Tunnel mode security association to a security gateway. The security associations between the security gateway and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as the responder.
- Peers authenticate themselves using the pre-shared key method.

Perform the following steps to meet these requirements and configure the branch office with NAT model.

1. Determine the number of zones to be protected.

   The branch office represents one zone, zone C.

2. For each zone, determine what services are allowed and define an IpService block for each desired service. Services are defined by their protocols and the well-known ports that they use.

   The definitions that describe FTP traffic can be combined in an IP service group. Typically, FTP clients use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server when the client is behind a NAT. For more information on active and passive mode FTP, see "FTP considerations when a NAT is being traversed" on page 951.

   The range of server ports specified for the data connection reflects the port range specified for PASSIVEDATAPORTS in the server's FTP.DATA file. For more information on PASSIVEDATAPORTS, refer to *z/OS Communications Server: IP Configuration Reference*.

   The following definitions show the services required for the server for PASV or EPSV:

```
IpServiceGroup              FTPServer
{
  IpService                 FTPServer-Control
  {
    SourcePortRange           21
    Protocol                  tcp
    Direction                 bidirectional InboundConnect
    Routing                   local
    SecurityClass             0
  }
  IpService                 FTPServer-Data-Passive
  {
    SourcePortRange           50000 50200
    Protocol                  tcp
    Direction                 bidirectional InboundConnect
```

```
    Routing                   local
    SecurityClass             0
  }
}
```

The traffic pattern for Enterprise Extender can be defined in one IpService block as follows:

```
IpService                   Enterprise-Extender
{
  SourcePortRange           12000 12004
  DestinationPortRange      12000 12004
  Protocol                  udp
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

3. Determine the data endpoints to be protected.

   In this example, for the local host's public IP address, 9.3.3.3, define the following:

   ```
   IpAddr                     PublicServerAddressA1
   {
     Addr                     9.3.3.3
   }
   ```

   In this case, the remote data endpoints are in the branch office's internal network. The z/OS NATT implementation does not require the coding of private addresses for the remote endpoints. Instead, the security gateway's public address is treated as the remote data endpoint. The NAT is using a static mapping for the security gateway, so the public address of the gateway is specified. If the NAT was using dynamic mappings, the range of public IP addresses to which the security gateway could be mapped would need to be included in this definition.

   ```
   IpAddr                     BranchOfficeGateway
   {
     Addr                     9.5.5.5
   }
   ```

   This IP address is also needed to permit IKE traffic between the local public server and the remote branch office gateway.

4. Determine what level of security is needed between each set of data endpoints.

   In this example, strong authentication and encryption is needed for both EE and FTP traffic, so ESP authentication and ESP encryption are used. ESP must be used when NAT traversal support is being used.

5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged, as follows:

   ```
   IpGenericFilterAction      ipsec
   {
     IpFilterAction           ipsec
   }
   ```

6. If IPSec is required between any two endpoints, do the following:

   a. Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:

      1) Determine the required type and strength of protection for the phase 1 security association.

In this example, SHA1 authentication is used because it is more secure than MD5, and 3DES encryption is used because it is more secure than DES.

2) Decide what type of peer authentication to use.

   In this example, pre-shared key authentication is specified in the requirements. Typically, the RSA signature method is preferable, given its numerous advantages, but for the purposes of example the pre-shared key method is used here. Because there is only one remote IKE peer in the branch office scenario (the remote security gateway), and because it is relatively simple to configure, pre-shared key authentication is a reasonable choice.

3) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation as follows:

```
KeyExchangeOffer            SHA1-3DES-PSK
{
  HowToEncrypt              3DES
  HowToAuthMsgs             SHA1
  HowToAuthPeers            PresharedKey
}
```

4) Decide whether NAT traversal will be allowed.

   The AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements enables IKED to advertise NAT traversal support. For this model, allow NAT traversal support to be advertised for all phase 1 security associations by specifying Yes on the AllowNat parameter on the KeyExchangePolicy statement:

```
AllowNat Yes
```

   For this model, allow the NatKeepAliveInterval parameter to default to 20 seconds. When z/OS is behind a NAT, a NAT keep-alive timer is started, with the interval specified in NatKeepAliveInterval parameter on the KeyExchangePolicy statement. Because z/OS is not behind a NAT in this model, a NAT keep-alive is not kept regardless of the value specified or defaulted for NatKeepAliveInterval.

5) Determine the negotiation mode, Main or Aggressive.

   Because security is a priority in the branch office with NAT model, the more secure Main mode is used for the phase 1 negotiation.

6) Configure a KeyExchangeAction statement as follows:

```
KeyExchangeAction           Gold-PSK
{
  HowToInitiate             main
  HowToRespond              main
  KeyExchangeOfferRef       SHA1-3DES-PSK
}
```

7) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement as follows:

```
LocalSecurityEndpoint       Public_IKED
{
  Identity                  IpAddr 9.3.3.3
  LocationRef               PublicServerAddressA1
}

RemoteSecurityEndpoint      ZoneC_IKED
{
  Identity                  Fqdn gateway.BO.example.com
  LocationRef               BranchOfficeGateway
}
```

8) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action, as follows:

```
KeyExchangeRule              ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef   Public_IKED
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
  PresharedKey               abracadabra
}
```

9) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block, as follows:

```
KeyExchangePolicy
{
    AllowNat                 Yes
    KeyExchangeRuleRef       ZoneC_KeyExRule1
}
```

b. Configure an IpDynVpnAction statement defining the control of the phase 2 negotiation, as follows:

1) Determine the required type and strength of IPSec protection for the phase 2 security association.

   In this example, authentication is ESP HMAC_SHA and encryption is ESP 3DES.

2) Determine whether tunnel or transport mode is required.

   Tunnel mode is required when one of the security endpoints is a security gateway.

3) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation, as follows:

```
IpDataOffer                  SHA-3DES-Tunnel
{
  HowToEncap                 tunnel
  HowToEncrypt               3DES
  HowToAuth                  ESP HMAC_SHA
}
```

4) Determine which peer is allowed to initiate the negotiation.

   In this scenario, the hosts behind the gateway do not have public addresses that can be configured in the policy. Therefore, initiation from host 9.3.3.3 to the security gateway 9.5.5.5 becomes ambiguous because the IP address of the remote data endpoint is unknown. z/OS does not allow initiation of a UDP-Encapsulated-Tunnel mode security association to a security gateway. The security associations between the security gateway 9.5.5.5 and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as responder.

5) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation, as follows:

```
IpDynVpnAction               Gold-TunnelMode
{
  Initiation                 remoteonly
  Pfs                        Group2
  IpDataOfferRef             SHA-3DES-Tunnel
}
```

c. Decide how the security association is activated.

1) Only remote initiation of the security association is allowed, as specified for the Initiation parameter on the Gold-TunnelMode IpDynVpnAction statement. No LocalDynVpnRule or IpLocalStartAction statements are needed.

2) Create an IpFilterRule statement that allows IPSec traffic (ESP), or set the global IpFilterPolicy statement parameter PreDecap to `off`.

   In this example, `PreDecap off` is used in the IpFilterPolicy statement.

7. Define an IpFilterRule statement for each set of data endpoints. The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500 and 4500). See the IpAddr and IpAddrSet statements configured previously.

```
IpFilterRule                Rule1C
{
   IpSourceAddrRef           PublicServerAddressA1
   IpDestAddrRef             BranchOfficeGateway
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpGenericFilterActionRef  permit
}

IpFilterRule                Rule2C
{
   IpSourceAddrRef           PublicServerAddressA1
   IpDestAddrRef             BranchOfficeGateway
   IpServiceRef              Enterprise-Extender
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec
   IpDynVpnActionRef         Gold-TunnelMode
}
```

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include all IpFilterGroup references and, optionally, any additional IpFilterRule statements as needed, in the IpFilterPolicy block.
10. Include all configured statements in the IpSecConfig file.

The following is the complete IP security policy for traffic from the local secure server to zone C, assuming that the reusable statements have been included in the CommonIpSecConfig policy file:

```
#------------------------------------------------------
# Filter Policy for Secure Server
#------------------------------------------------------
IpFilterPolicy
{
  PreDecap                  off
  FilterLogging             on
  AllowOnDemand             no
  IpFilterGroupRef          ZoneC
}

#------------------------------------------------------
# KeyExchange Policy for Secure Server
#------------------------------------------------------
KeyExchangePolicy
{
  AllowNat                  Yes
  KeyExchangeRuleRef        ZoneC_KeyExRule1
}

######################################################
#               Connectivity Profile
#             Secure Server To Zone C
#
```

```
#   Server to Trusted Branch Office Network
#
#####################################################
IpFilterGroup               ZoneC
{
   #-------------------------------------------------------
   # Permitted Zone C traffic:
   #    Allow IKE traffic from the gateway IKE Server
   #    for branch office to this host.
   #
   #    IKE  (UDP port 500/4500) - IKE negotiations
   #-------------------------------------------------------

   IpFilterRule                Rule1C
   {
     IpSourceAddrRef           PublicServerAddressA1
     IpDestAddrRef             BranchOfficeGateway
     IpServiceGroupRef         IKE
     IpGenericFilterActionRef  permit
   }

   #-------------------------------------------------------
   # IPSec-protected Zone C traffic:
   #
   #    Enterprise Extender (ports 12000-12004)
   #    FTP Server - SubnetC to PublicServerAddressA
   #-------------------------------------------------------

   IpFilterRule                Rule2C
   {
     IpSourceAddrRef           PublicServerAddressA1
     IpDestAddrRef             BranchOfficeGateway
     IpServiceRef              Enterprise-Extender
     IpServiceGroupRef         FTPServer
     IpGenericFilterActionRef  ipsec
     IpDynVpnActionRef         Gold-TunnelMode
   }
}


KeyExchangeRule             ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef  Public_IKED
  RemoteSecurityEndpointRef ZoneC_IKED
  KeyExchangeActionRef      Gold-PSK
  PresharedKey              abracadabra
}
```

Steps for configuring the branch office model: Part 2 (gateway-to-gateway with IPSec)

It is not likely that the z/OS system will function strictly as a router or a firewall at the network perimeter, but it is possible to configure the z/OS system to provide the IPSec functionality that many secure gateway devices provide. This section includes instructions on how to configure a scenario in which the z/OS system is routing network traffic from inside the internal network. This functionality is similar to the functionality that is provided by the branch office gateway in the previous example. Here, there are multiple data endpoints on both the local side and the remote side, but only one pair of security endpoints, one local and one remote.

In this example, assume that the local z/OS system is acting as a secure gateway for hosts on an internal network A, and tunneling the IPSec-protected traffic to a remote secure gateway for subnetwork C. The following summarizes the requirements for this example:

- Permit IKE negotiations between the two security gateways, the secure local host and the secure remote gateway for subnetwork C.
- Permit traffic from the internal network to the internal interface on the secure local host.
- Add IPSec protection to any traffic that flows between the two secure gateways.

In this scenario, the z/OS system is a secure forwarding agent for the internal hosts, rather than a data endpoint. Traffic from the internal hosts that is destined for the remote network first comes to the secure local gateway in the clear. Before it is sent out to the remote network, it is IPSec encapsulated. The process is reversed for traffic that comes from the remote network. Traffic that comes from the remote network to the local secure gateway is IPSec decapsulated on the local secure host and forwarded to the internal host in the clear.

Perform the following steps to meet the above requirements and configure part 2 of the branch office model.

1. Permit IKE negotiations between the two secure gateways.

   Similar to the previous examples, UDP port 500 traffic must be allowed for IKE negotiations.

```
IpFilterRule                  Rule1AtoC
{
     IpSourceAddrRef            PublicServerAddressA1
     IpDestAddrRef              BranchOfficeGateway
     IpServiceRef               IKE
     IpGenericFilterActionRef   permit
}
```

2. Permit traffic from the internal network to the internal interface on the secure host.

```
IpFilterRule                  Rule2AtoC
{
     IpSourceAddrSetRef         SubnetC
     IpDestAddrSetRef           InternalNetworkA
     IpServiceGroupRef          All-traffic-routed
     IpGenericFilterActionRef   permit
}
```

   The bidirectional keyword on the IpService statement creates two filter rules, one inbound and one outbound. Expansion of this IpFilterRule statement is shown in Table 35.

Table 35. Expanded filter rule for internal traffic

Source	Destination	Routing	Direction	Action
SubnetC	InternalNetworkA	Routed	Outbound	permit
InternalNetworkA	SubnetC	Routed	Inbound	permit

   As required, traffic that enters the secure server from InternalNetworkA that is destined for SubnetC is permitted by the secure host as an inbound routed packet. Traffic that leaves the secure server from SubnetC destined for InternalNetworkA is permitted by the secure host as an outbound routed packet.

3. Add IPSec protection to any traffic that flows between the two secure gateways.

```
IpFilterRule                Rule3AtoC
{
     IpSourceAddrSetRef         InternalNetworkA
```

```
    IpDestAddrSetRef           SubnetC
    IpServiceGroupRef          All-traffic-routed
    IpGenericFilterActionRef   ipsec-log
}
```

Expansion of this rule is shown in Table 36.

Table 36. Expanded filter rule for remote traffic

Source	Destination	Routing	Direction	Action
InternalNetworkA	SubnetC	Routed	Outbound	ipsec
SubnetC	InternalNetworkA	Routed	Inbound	ipsec

Traffic that leaves the secure server from InternalNetworkA that is destined for SubnetC is permitted with ipsec. Traffic that enters the secure server from SubnetC that is destined for InternalNetworkA is permitted with ipsec.

Additional topologies

There are alternative VPN solutions that might be suitable for particular networks. Cascaded tunnels and nested tunnels are two common ways of building VPN connections. Although the configuration for these VPN solutions is not explicitly covered here, they are presented as alternatives to single end-to-end IPSec connections. Both configurations are supported by z/OS IP security.

Cascaded tunnels: If there are multiple hops from data endpoint to data endpoint, there might be security associations between any two hosts along the path. For example, the data might be authenticated from a host to the secure gateway, then encrypted for transportation over the Internet, then possibly authenticated and encrypted from the second secure gateway to the host on the other side, as shown in Figure 90:

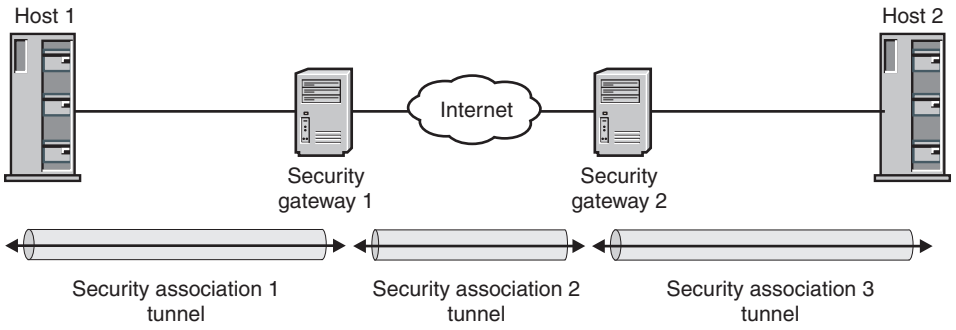


Figure 90. Cascaded tunnels

Nested tunnels: In a nested environment, data is encapsulated multiple times over multiple hops. If the local and remote hosts are both behind a secure gateway, there could be a tunnel-mode security association that carries the traffic from one secure gateway to the other. Meanwhile, a transport-mode security association could carry the traffic from one host to the other, end-to-end. In this case, the transport-mode security association is nested in the tunnel-mode security association. Data from the local host is encapsulated once when leaving the local machine, encapsulated again at the secure gateway, decapsulated at the other secure gateway, and decapsulated one final time at the remote host, leaving only the original packet.

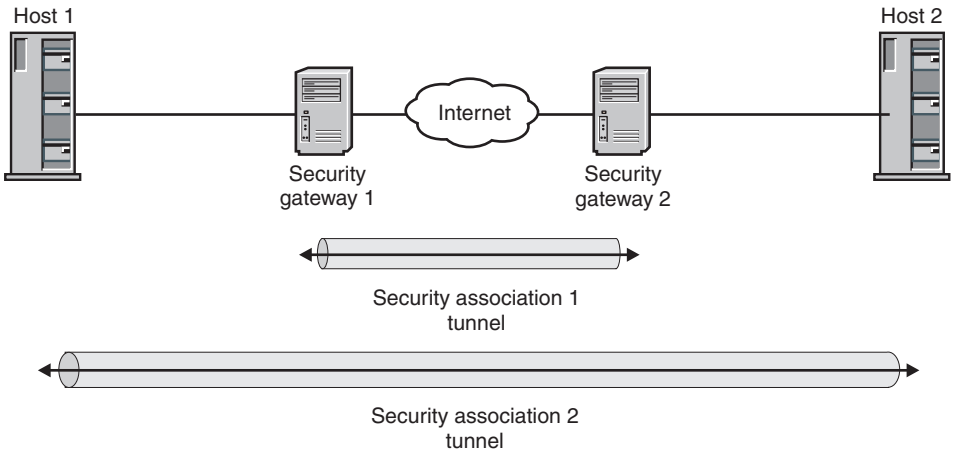


Figure 91. Nested tunnels

Configuration scenarios supported for NAT traversal

Communications Server can act as a host security association endpoint for UDP-encapsulated mode security associations that are negotiated to enable IPSec traffic to traverse a NAT. The business partner with NAT model describes Communications Server's host-to-host support. The branch office with NAT model describes Communications Server's host-to-security gateway support.

Rule: Communications Server does not support acting as a security gateway endpoint for UDP-Encapsulated-Tunnel mode security associations. This is different than the Communications Server support provided for tunnel mode security associations, where Communications Server can act as a security gateway, although Communications Server is not typically deployed in this manner.

Rule: Communications Server can negotiate a UDP-encapsulated mode security association that traverses a NAT device, but not a NAPT device. A NAPT device translates multiple internal IP addresses to a single public address, and translates the TCP or UDP port to make the connection unique.

The following figures show z/OS configuration support for UDP-encapsulated security associations. A security association is negotiated by two IKE peers, with one initiating the negotiation and the other acting in responder mode. The location of the NAT, and the NAT's functionality, affect which IKE peer can initiate the security association. A traditional dynamic NAT implementation requires outbound traffic to be sent first to create an address mapping, before inbound traffic can be accepted. When an IKE responder is behind a NAT, the NAT's address mapping must be static, allowing inbound traffic for the address to be received prior to outbound traffic being sent.

Host-to-host scenario 1 — z/OS-to-z/OS: Figure 92 on page 949 shows a NAT in front of both z/OS hosts. A configuration with a NAT in front of only one of the z/OS hosts is supported as well. If there is a NAT device in front of the responder, the NAT's address mapping must be static, allowing inbound traffic for the address to be received prior to outbound traffic being sent. A traditional dynamic NAT implementation requires outbound traffic to be sent first to create an address mapping, before inbound traffic can be accepted.

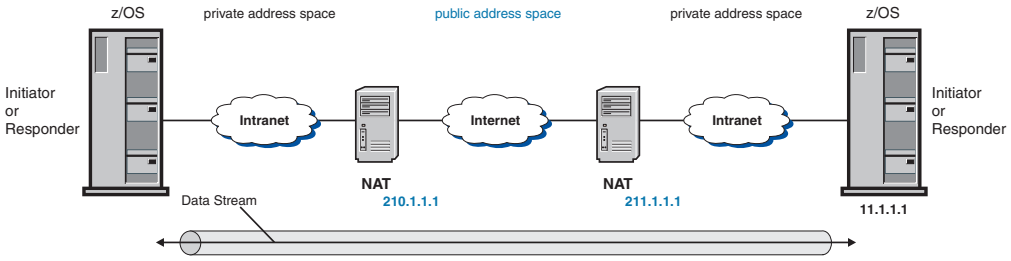


Figure 92. z/OS host to z/OS host, double NAT

Either UDP-Encapsulated-Transport mode or UDP-Encapsulated-Tunnel mode can be negotiated in a z/OS host-to-z/OS host configuration.

Host-to-host scenario 2 — z/OS-to-non-z/OS: Figure 93 shows a NAT in front of the z/OS host and the non-z/OS host. A configuration with a NAT in front of only one of the hosts is supported as well. If there is a NAT device in front of the responder, the NAT's address mapping must be static, allowing inbound traffic for the address to be received prior to outbound traffic being sent. A traditional dynamic NAT implementation requires outbound traffic to be sent first to create an address mapping, before inbound traffic can be accepted.

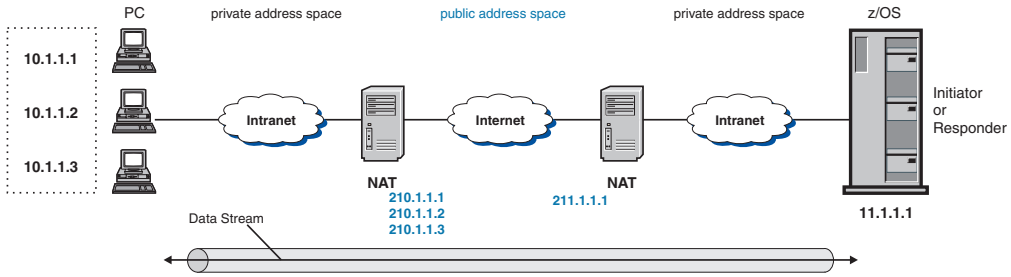


Figure 93. z/OS host to non-z/OS host, double NAT

Either UDP-Encapsulated-Transport mode or UDP-Encapsulated-Tunnel mode can be negotiated in a z/OS host-to-non-z/OS host configuration.

Interoperability Considerations: z/OS is typically used to provide a server function. The client initiates the phase 2 security association and data, with z/OS acting in the role of IKE responder and data responder. z/OS provides robust NAT traversal responder support, allowing it to interoperate with a variety of clients.

z/OS can also act as the initiator of the phase 2 security association and data. Potential incompatibilities exist in the following cases, depending on the support of the non-z/OS peer:

- Phase 2 security associations that protect specific ports, protocols, or both. When initiating such a phase 2 security association, z/OS represents the data being protected by the security association with the following:
  - The local IP address as it appears in the home list. This could be a private IP address if z/OS is behind a NAT.
  - The client's public IP address. If the client is behind a NAT, this would be the client's public address.
  - The specific port and protocol values.

  The Phase 2 security association negotiation should succeed if the non-z/OS peer supports receiving this specification.

- Tunnel mode phase 2 security associations that protect all ports and protocols. When initiating such a phase 2 security association, z/OS does not explicitly include the IP addresses of the data being protected in the negotiation of the security association. This allows the non-z/OS peer to view the protected data in terms of the IP addresses that it understands, the public address of the remote endpoint and the private address of the local endpoint. The phase 2 security association negotiation is expected to be successful.

  If data is initiated from z/OS over the security association, the data packet contains:

  – The local IP address as it appears in the home list. This could be a private IP address if z/OS is behind a NAT.
  – The client's public IP address. If the client is behind a NAT, this would be the client's public address.

  If the non-z/OS peer supports receiving a packet with these IP addresses, the data flow should be successful. Once z/OS receives data packets from the non-z/OS peer, z/OS will send packets containing the IP addresses used by the peer.

Host-to-security gateway scenario: The z/OS host is limited to acting in responder mode in a host-to-security gateway configuration when a NAT is traversed. The phase 2 security association negotiation must be initiated by the security gateway. Data must be initiated by the client behind the security gateway.

Figure 94 shows both the security gateway and the host behind a NAT. z/OS also supports acting in responder mode when only one endpoint is behind a NAT, either the security gateway or the z/OS host.

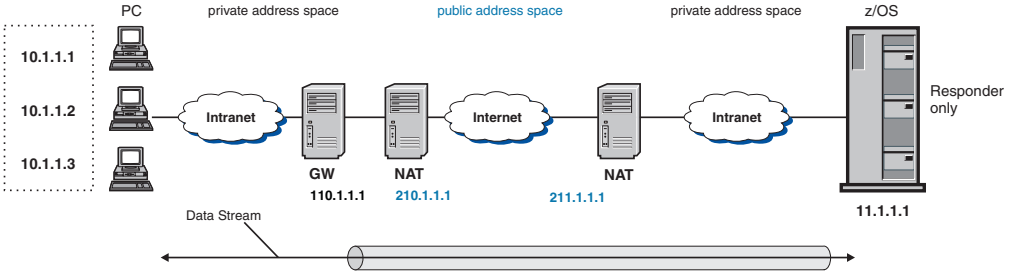


Figure 94. z/OS in a host-to-security gateway configuration

Figure 94 shows the clients and the security gateway as separate devices. However, whenever a security association is negotiated to protect something other than a single IP address (for example, a range of IP addresses), that IKE daemon negotiating the security association is acting as a security gateway.

When there is a NAT device in front of the z/OS host (acting as responder), the NAT's address mapping must be static, allowing inbound traffic for the address to be received prior to outbound traffic being sent. A traditional dynamic NAT implementation requires outbound traffic to be sent first to create an address mapping, before inbound traffic can be accepted.

When the security gateway is behind a NAT, the individual hosts behind the NAT cannot be distinguished. If only one NAT address is available, all security associations negotiated between the security gateway (GW) and z/OS are negotiated using the NAT address, and have the same security characteristics. If

multiple security characteristics are required to protect the traffic behind the security gateway, more NAT addresses are needed so that z/OS can locate different policies based on the NAT address.

FTP considerations when a NAT is being traversed: FTP requires both a control connection and a data connection. For active mode FTP, the client initiates the control connection and provides IP address and port information for the server to initiate the data connection.

There are some cases where the network configuration does not allow the server to initiate the data connection. For example:

- If the client is behind a NAT, the client provides its private IP address to the server. The server is unable to use the private IP address to set up the data connection.
- If the client is behind a firewall that is configured to block incoming connection requests from the outside network, the inbound data connection request is dropped by the firewall.

For FTP to work in these cases, passive mode (PASV) FTP must be used. For passive mode FTP, the client initiates the control connection. In response to the control connection request, the server provides the IP address and port information for establishing the data connection to the client. The client is then able to initiate the data connection. Since the data connection request is initiated by the client behind the NAT or firewall, the active mode problem is resolved.

These cases are situations where the FTP client is behind a NAT or firewall. An additional consideration arises when the FTP server is behind a NAT. Passive mode (PASV) requires the server to provide the IP address and port to the client for the data connection initiation. When the FTP server is behind a NAT, the IP address provided is the private address, not the public address. This address does not have meaning to the client, which causes the data connection initiation to fail.

Extended passive (EPSV) mode resolves this problem by eliminating the IP address from the data that the server provides to the client. With EPSV, the server provides the port for the data connection, and the client connects to the same IP address that was used for the control connection.

For more information on FTP, see Chapter 11, “Transferring files using FTP,” on page 541. For a sample FTP.DATA data set, or information on the LOCSIte subcommand, refer to *z/OS Communications Server: IP User's Guide and Commands*.

Additional configuration concerns for NAT traversal: Following are some additional configuration concerns for NAT traversal:

- When using NAT traversal, z/OS views its own address as the one configured in the z/OS home list. If the z/OS host is behind a NAT, this address is a private address. Otherwise, it is the public address of the z/OS host. z/OS views its IPSec peer and the remote IP connection endpoint as a public IP address. If a NAT is in front of the IPSec peer, the z/OS host perceives the IPSec peer and connection endpoint addresses to be that of the NAT. The z/OS implementation does not use private addresses in its configuration to describe the remote IKE peer or remote IP connection endpoint.
- RFC 3947 states that pre-shared keys cannot be used with Main mode, unless group shared keys for all those behind the NAT are deployed. The use of group pre-shared keys is considered a security risk. You should not configure

pre-shared keys in Main mode when multiple remote peers reside behind a NAT and the peers do not map to unique RemoteSecurityEndpoint specifications.

- If a NAT address is coming out of a dynamic NAT pool, addresses assigned to a host from the pool can be assigned any of the pooled addresses. In this case, there are additional considerations. When z/OS is the responder, the IPSec policy intended for a host must be broad enough to cover the entire range of IP addresses in the dynamic NAT pool. When z/OS is the initiator and the responder is behind a NAT, the identity of the target host can be ambiguous. A z/OS should not be configured to initiate to an ambiguous target host.
- When a remote security endpoint resides behind a NAT, its identity must be unique. During a phase 1 negotiation, the remote security endpoint sends its identity in an ID payload. The IKE daemon can manage multiple remote security endpoints using the same ID when those endpoints are not behind a NAT. However, when a remote security endpoint is behind a NAT, it must use a unique ISAKMP identity.
- When the remote security endpoint is a security gateway behind a NAT, ICMP traffic over the security association has limited support. Non-TCP, non-UDP, and non-ICMP traffic is not supported.
- z/OS allows traffic for a TCP connection to continue as long as the integrity of the connection can be verified. Two cases where z/OS can no longer verify the integrity of the connection are:
  - Adding or removing IPSec protection for a TCP connection

    When a TCP connection traverses a NAT, the TCP connection must be restarted after a filter policy change that causes the connection's traffic to change from IPSec-protected traffic to clear text, or from clear text to IPSec-protected traffic.
  - NAT IP address remapping

    If the peer's IP address is remapped by a NAT due to a timeout or reboot of the NAT device, the TCP connection must be restarted.

Configuring the IKE daemon

This section describes considerations and steps for configuring the IKE daemon.

The IKE daemon's purpose is to manage dynamic IPSec tunnels. The IKE daemon is not involved in the actual filtering, encapsulation, or decapsulation of packets. The IKE daemon is not required for the configuration or use of IP filters when no IpDynVpnAction statements are utilized.

Multiple TCP/IP stacks

A one-to-many relationship can exist between an instance of the IKE daemon and stacks configured with IPCONFIG IPSECURITY. A single instance of the IKE daemon can service all stacks configured with IPCONFIG IPSECURITY on a single z/OS image. Only one instance of the IKE daemon can run on a single z/OS image.

The IKE daemon can run concurrently with the z/OS Integrated Security Services Firewall Technologies isakmpd on a single z/OS image. The IKE daemon services only stacks configured with IPCONFIG IPSECURITY. The z/OS Integrated Security Services Firewall Technologies isakmpd daemon services only stacks configured with IPCONFIG FIREWALL.

Run-time environment

The IKE daemon is a z/OS UNIX application, and it requires the Hierarchical File System (HFS) or a compatible file system such as zFS. The IKE daemon can be started from an MVS started procedure, from the z/OS shell, with the AUTOLOG statement in the TCP/IP profile, or by using the COMMNDxx member of PARMLIB. The IKE daemon must be started by a RACF-authorized user ID, and it must reside in an APF-authorized library.

The IKE daemon uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and STDOUT are used for major events such as initialization, termination, and error conditions. Syslogd is used for logging events related to dynamic IPSec tunnel management. CTRACE is used for detailed tracing and debugging.

The IKE daemon uses a standard message catalog. The message catalog must be in the HFS. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

Language Environment run-time considerations

When starting the IKE daemon from a started or cataloged procedure, you should usually start the IKE daemon directly from the SEZALOAD data set using PGM=IKED. However, there is a situation where you might want to start the IKE daemon using BPXBATCH.

When the IKE daemon is started using PGM=IKED, the STDENV DD card, if used, is passed directly to the IKE daemon program. Language Environment does not get access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS= environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM= parameter and the options must be specified before any IKE daemon options. However, the PARM= statement allows a maximum of 100 characters. If the desired Language Environment run-time options plus IKE daemon parameters exceeds 100 characters, consider using BPXBATCH to start the IKE daemon. When PGM=BPXBATCH is used, the Language Environment variable _CEE_RUNOPTS can be included on the STDENV DD card to specify run-time options in excess of 100 characters long.

IKE daemon configuration source information

The IKE daemon obtains configuration information from two sources.

- IKE daemon configuration file

  The IKE daemon configuration file contains the IkeConfig statement. Parameters on the IkeConfig statement are global IKE daemon operational parameters. For details on the IKE daemon configuration file and the IkeConfig statement, refer to *z/OS Communications Server: IP Configuration Reference*.

  The IKE daemon configuration file is read when the IKE daemon initializes and can be reread dynamically. For more information, see "Controlling the IKE daemon" on page 957.

- Policy Agent

  IP security policy includes dynamic IPSec tunnel configuration information required by the IKE daemon. The IKE daemon obtains IP security policy from the Policy Agent. The IKE daemon obtains IP security policy when connecting to the Policy Agent and whenever the Policy Agent informs the IKE daemon of a

change in IP security policy. The IKE daemon connects only to stacks configured with IPCONFIG IPSECURITY, if there is an IP security policy defined for a stack. For details on configuring IP security policy, see "Configuring specific security models" on page 889.

Policy Agent considerations

The IKE daemon cannot perform management of dynamic IPSec tunnels until it has obtained IP security policy from the Policy Agent. While the Policy Agent is running, it can inform the IKE daemon of dynamic changes to IP security policy. Once the IKE daemon has obtained an IP security policy, the Policy Agent can be stopped without impacting the IKE daemon. However, any changes to the IP security policy will not be detected until the Policy Agent is restarted, nor will IKE detect newly activated or reactivated stacks. The IKE daemon reconnects to the Policy Agent when the Policy Agent is restarted.

Steps for configuring the IKE daemon

Perform the following steps to configure the IKE daemon:

1. Create the IKE daemon configuration file.

   A sample configuration file is provided in /usr/lpp/tcpip/samples/IBM/EZAIKCFG.

   Following is the search order used by the IKE daemon to locate the configuration data set or file:

   a. If the environment variable IKED_FILE has been defined, the IKE daemon uses the value as the name of an MVS data set or HFS file to access the configuration data.
   b. /etc/security/iked.conf

2. Set the _BPX_JOBNAME environment variable (optional).

   When starting the IKE daemon from the z/OS shell, the environment variable _BPX_JOBNAME should be set. This enables a specific job name to be used when reserving ports for the IKE daemon. This name can also be used with the STOP or MODIFY console commands.

   For more information on _BPX_JOBNAME, refer to *z/OS UNIX System Services Planning*

3. Reserve the ports (optional).

   Update the PORT statement in PROFILE.TCPIP to reserve ports 500 and 4500 for the IKE daemon. Add the name of the member containing the IKE daemon cataloged procedure or the name as set using _BPX_JOBNAME:

   ```
   PORT
                  500 UDP IKED
                  4500 UDP IKED
   ```

4. Update the IKE daemon cataloged procedure.

   If the IKE daemon is to be started by a procedure, create the cataloged procedure by copying the sample in SEZAINST(IKED) to your system or recognized PROCLIB. Specify IKE daemon parameters and change the data set names to suit your local configuration. Following is a copy of the sample:

```
//IKED     PROC
//*
//* IBM Communications Server for z/OS
//* SMP/E distribution name: EZBIKPRC
//*
//* 5694-A01 (C) Copyright IBM Corp. 2005
//* Licensed Materials - Property of IBM
//* "Restricted Materials of IBM"
//* Status = CSV1R7
//*
//*
//IKED     EXEC PGM=IKED,REGION=0K,TIME=NOLIMIT,
//       PARM='ENVAR("_CEE_ENVFILE=DD:STDENV")/'
//*
//* Provide environment variables to run with the desired
//* configuration.  As an example, the data set or file specified by
//* STDENV could contain:
//*
//*   IKED_FILE=/etc/security/iked.conf2
//*   IKED_CTRACE_MEMBER=CTIIKE01
//*
//* For information on the above environment variables, refer to the
//* IP Configuration Reference.
//*
//STDENV   DD DUMMY
//* Sample MVS data set containing environment variables:
//*STDENV   DD DSN=TCPIP.IKED.ENV(IKED),DISP=SHR
//* Sample HFS file containing environment variables:
//*STDENV   DD PATH='/etc/security/iked.env',PATHOPTS=(ORDONLY)
//*
//* Output written to stdout and stderr goes to the data set or
//* file specified with SYSPRINT or SYSOUT, respectively.
//SYSPRINT DD SYSOUT=*
//SYSOUT   DD SYSOUT=*
```

Figure 95. IKE cataloged procedure

5. Authorize the IKE daemon to the external security manager.

   See "Step 2: Authorizing the IKE daemon to the external security manager" on page 1215.

6. Configure and start syslogd.

   The IKE daemon uses the local4 facility when writing messages to syslogd. For performance purposes, it is recommended that syslogd utilize zFS as its underlying file system. For more information on syslogd, see "Configuring the syslog daemon (syslogd)" on page 175.

7. Update the IKE daemon environment variables (optional).

   The following environment variables are used by the IKE daemon and can be tailored to a particular installation:

   IKED_CTRACE_MEMBER
   The IKED_CTRACE_MEMBER variable is used by the IKE daemon to locate a parmlib member for IKE daemon CTRACE customization. For more information on the TCP/IP services component trace for the IKE daemon, refer to *z/OS Communications Server: IP Diagnosis Guide*.

   IKED_FILE
   The IKED_FILE variable is used by the IKE daemon in the search order for the IKE daemon configuration file. For details on the search order used for locating this configuration file, see step 1 on page 954.

8. Setup the IKE daemon for RSA signature mode authentication (optional).

See “Step 4: Setting up the IKE server for RSA signature mode authentication (optional)” on page 1217.

Starting the IKE daemon

After the necessary external security manager authorization has been defined (see “Step 2: Authorizing the IKE daemon to the external security manager” on page 1215), the IKE daemon can be started from an MVS procedure, from the z/OS shell, or using the AUTOLOG statement.

- You can start the IKE daemon procedure from the MVS operator console. A sample start procedure is provided in SEZAINST(IKED).
- You can start the IKE daemon from the z/OS shell by starting OMVS and then issuing the iked command.
- You can use the AUTOLOG statement to start the IKE daemon automatically during TCP/IP initialization by inserting the name of the IKE daemon start procedure into the AUTOLOG statement in the PROFILE.TCPIP data set:

  ```
  AUTOLOG
    IKED
  ENDAUTOLOG
  ```

  Tip: When implementing multiple stacks enabled for IP security, adding an AUTOLOG statement for the IKE daemon might not be optimal. If the IKE daemon is listed in an AUTOLOG statement of a stack’s profile, the IKE daemon is cancelled if it is already running when that stack starts. In a multiple IP security stack environment, this could disrupt traffic on other IP security stacks. Use another method to automate starting the IKE daemon when the system is IPLed, such as using the COMMNDxx member of PARMLIB. For more information about the use and configuration of the COMMNDxx member of PARMLIB, refer to *z/OS MVS Initialization and Tuning Reference*.

When running from an MVS procedure, the environment variables can be set using the STDENV DD statement in the IKE daemon procedure. For information concerning the environment variables used by IKE daemon, see step 7 on page 955 in “Steps for configuring the IKE daemon” on page 954.

Stopping the IKE daemon

The IKE daemon can be stopped as follows:

- From MVS, issue:

  ```
  STOP procname
  ```

  If the IKE daemon was started from a cataloged procedure, *procname* is the member name of that procedure. If the IKE daemon was started from the z/OS shell and the environment variable _BPX_JOBNAME was set, *procname* is the same as _BPX_JOBNAME. If the IKE daemon was started from the z/OS shell and _BPX_JOBNAME was not set, *procname* is *useridX*, where X is the sequence number set by the system. To determine the sequence number, from the SDSF LOG window on TSO, issue:

  ```
  /d omvs,u=userid
  ```

  This command shows the programs running under this user ID. For more information on _BPX_JOBNAME, refer to *z/OS UNIX System Services Planning*.
- From a superuser ID in the z/OS shell, issue the kill command to the process ID (PID) associated with the IKE daemon. The IKE daemon PID is recorded in /var/ike/iked.pid.

Controlling the IKE daemon

You can control the IKE daemon from the operator's console using the MODIFY command. MODIFY commands are available to perform the following functions:

- Rereading the configuration file

  The MODIFY *procname*,REFRESH command is used to reread the IKE daemon configuration file. Not all IkeConfig statement parameters can be updated using this command. For information on which parameters can be dynamically changed, refer to the parameter descriptions for the IkeConfig statement of the IKE daemon configuration file in the *z/OS Communications Server: IP Configuration Reference*.
- Displaying the configuration file parameters

  The MODIFY *procname*,DISPLAY command is used to display configuration values currently being used by the IKE daemon.

For more information on the MODIFY command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Verifying policy installation

This section describes the console messages and commands that are used to verify policy installation.

Console messages

After the IP security policy has been configured, start the TCP/IP stacks, Policy Agent, and IKED. A series of console messages is issued if the installation of IP security policy was successful.

The following console messages indicate that Policy Agent and IKE have completed initialization:

```
EZZ8432I PAGENT INITIALIZATION COMPLETE
EZD1046I IKE INITIALIZATION COMPLETE
```

The following console messages indicate that the processing of IP security policy is complete:

```
EZZ8771I PAGENT CONFIG POLICY PROCESSING COMPLETE FOR TCPCS : IPSEC
EZD1068I IKE POLICY UPDATED FOR STACK TCPCS
```

If there were errors in the configuration files, Policy Agent issues the following message to the console:

```
EZZ8438I PAGENT POLICY DEFINITIONS CONTAIN ERRORS FOR TCPCS : IPSEC
```

Look at the Policy Agent log to find and correct the error.

Displaying TCP/IP configuration

To display whether the TCP/IP stack is configured with IPCONFIG IPSECURITY, issue the netstat -f command and look for the following field:

```
IpSecurity: Yes
```

To display whether the TCP/IP stack is configured for sysplex-wide security associations, issue the **ipsec -f display** command. The DVIPSec field in the header of the command display shows whether or not the DVIPSEC keyword has been coded in the TCP/IP profile:

```
ipsec -f display | head -n 6

CS V1R7 ipsec  TCPIP Name: TCPCS  Tue Mar 22 12:26:49 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 130
Logging:  Yes             Predecap: No                DVIPSec:  No
NatKeepAlive:  20
```

Displaying active filters with the ipsec command

Use the **ipsec -f display** command to display active filter rules, configured filter rules from IP security policy configuration files, and the default IP filter rules from the TCP/IP profile. The scope on the command, as indicated by the -c option, determines which source is queried:

-c policy
Shows IP filters as configured in the IP security policy configuration files.

-c profile
Shows default IP filters as configured in the TCP/IP profile.

-c current
Shows active IP filters in the stack. The active filters that are shown can be the default IP filters as defined in the TCP/IP profile, or IP filters as configured in the IP security policy configuration files, depending on which policy is active at the time the command is issued. The output of the display indicates the source of the current active filters.

The output of the command can be quite voluminous, so you might want to redirect the output of the display to a file.

The information in the report header of the report output indicates how many filters are active, and also indicates the source of the filters, whether from the default IP filter policy or the IP security policy from the Policy Agent.

```
ipsec -f display

CS V1R7 ipsec  TCPIP Name: TCPCS  Thu Mar 17 16:44:43 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Profile   Scope:    Current           TotAvail: 4
Logging:  Yes             Predecap: No                DVIPSec:  Yes
NatKeepAlive:  20
```

If the source field shows `Stack Policy`, the IP security policy is installed and active.

If the source field shows `Stack Profile`, the IP security policy is either not installed or the **ipsec -f default** command was issued. Either issue the **ipsec -f reload** command, or correct the IP security policy configuration.

Filter displays can be abbreviated to include only specific named rules. To view a named filter rule, use the -n option as follows:

```
ipsec -f display -n Rule2Admin

CS V1R7 ipsec  TCPIP Name: TCPCS  Tue Mar 22 12:30:01 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 104
Logging:  Yes             Predecap: No                DVIPSec:  No
NatKeepAlive:  20

FilterName:                   Rule2Admin
FilterNameExtension:          1
```

```
                    GroupName:                  Admin
                    LocalStartActionName:       n/a
                    VpnActionName:              Silver-TransportMode
                    TunnelID:                   Y0
                    Type:                       Dynamic Anchor
                    State:                      Active
                    Action:                     Permit
                    Scope:                      Local
                    Direction:                  Outbound
                    OnDemand:                   No
                    SecurityClass:              0
                    Logging:                    Deny
                    Protocol:                   All
                    ICMPType:                   n/a
                    ICMPCode:                   n/a
                    OSPFType:                   n/a
                    TCPQualifier:               n/a
                    ProtocolGranularity:        Rule
                    SourceAddress:              9.1.1.1
                    SourceAddressPrefix:        n/a
                    SourceAddressRange:         n/a
                    SourceAddressGranularity:   Packet
                    SourcePort:                 All
                    SourcePortRange:            n/a
                    SourcePortGranularity:      Rule
                    DestAddress:                9.1.1.2
                    DestAddressPrefix:          n/a
                    DestAddressRange:           n/a
                    DestAddressGranularity:     Packet
                    DestPort:                   All
                    DestPortRange:              n/a
                    DestPortGranularity:        Rule
                    OrigRmtConnPort:            n/a
                    RmtIDPayload:               n/a
                    ***************************************************************************
                    FilterName:                 Rule2Admin
                    FilterNameExtension:        2
                    GroupName:                  Admin
                    LocalStartActionName:       n/a
                    VpnActionName:              Silver-TransportMode
                    TunnelID:                   Y0
                    Type:                       Dynamic Anchor
                    State:                      Active
                    Action:                     Permit
                    Scope:                      Local
                    Direction:                  Inbound
                    OnDemand:                   No
                    SecurityClass:              0
                    Logging:                    Deny
                    Protocol:                   All
                    ICMPType:                   n/a
                    ICMPCode:                   n/a
                    OSPFType:                   n/a
                    TCPQualifier:               n/a
                    ProtocolGranularity:        Rule
                    SourceAddress:              9.1.1.2
                    SourceAddressPrefix:        n/a
                    SourceAddressRange:         n/a
                    SourceAddressGranularity:   Packet
                    SourcePort:                 All
                    SourcePortRange:            n/a
                    SourcePortGranularity:      Rule
                    DestAddress:                9.1.1.1
                    DestAddressPrefix:          n/a
                    DestAddressRange:           n/a
                    DestAddressGranularity:     Packet
                    DestPort:                   All
```

```
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*************************************************************************

2 entries selected
```

Anchor filters and dynamic filters

After a security association is negotiated, the **ipsec -f display** command shows the addition of two dynamic filters that were created when the security association was created, corresponding to the inbound and outbound anchor filters. Dynamic filters are placed ahead of the anchor filters in the filter table, so dynamic filters are searched first when IP filtering is performed. In the following sample output, note that two dynamic filters have been added to the filter table subsequent to the activation of a phase 2 security association. The Type field indicates whether the filter is a dynamic anchor filter or a dynamic filter:

```
ipsec -f dis -n Rule2Admin

CS V1R7 ipsec  TCPIP Name: TCPCS  Tue Mar 22 12:31:07 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 106
Logging:  Yes             Predecap: No                DVIPSec:  No
NatKeepAlive:  20

FilterName:                   Rule2Admin
FilterNameExtension:          1
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y22
Type:                         Dynamic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.1
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.1.1.2
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*************************************************************************
FilterName:                   Rule2Admin
FilterNameExtension:          1
GroupName:                    Admin
```

```
                    LocalStartActionName:         n/a
                    VpnActionName:                Silver-TransportMode
                    TunnelID:                     Y0
                    Type:                         Dynamic Anchor
                    State:                        Active
                    Action:                       Permit
                    Scope:                        Local
                    Direction:                    Outbound
                    OnDemand:                     No
                    SecurityClass:                0
                    Logging:                      Deny
                    Protocol:                     All
                    ICMPType:                     n/a
                    ICMPCode:                     n/a
                    OSPFType:                     n/a
                    TCPQualifier:                 n/a
                    ProtocolGranularity:          Rule
                    SourceAddress:                9.1.1.1
                    SourceAddressPrefix:          n/a
                    SourceAddressRange:           n/a
                    SourceAddressGranularity:     Packet
                    SourcePort:                   All
                    SourcePortRange:              n/a
                    SourcePortGranularity:        Rule
                    DestAddress:                  9.1.1.2
                    DestAddressPrefix:            n/a
                    DestAddressRange:             n/a
                    DestAddressGranularity:       Packet
                    DestPort:                     All
                    DestPortRange:                n/a
                    DestPortGranularity:          Rule
                    OrigRmtConnPort:              n/a
                    RmtIDPayload:                 n/a
                    ************************************************************************
                    FilterName:                   Rule2Admin
                    FilterNameExtension:          2
                    GroupName:                    Admin
                    LocalStartActionName:         n/a
                    VpnActionName:                Silver-TransportMode
                    TunnelID:                     Y22
                    Type:                         Dynamic
                    State:                        Active
                    Action:                       Permit
                    Scope:                        Local
                    Direction:                    Inbound
                    OnDemand:                     No
                    SecurityClass:                0
                    Logging:                      Deny
                    Protocol:                     All
                    ICMPType:                     n/a
                    ICMPCode:                     n/a
                    OSPFType:                     n/a
                    TCPQualifier:                 n/a
                    ProtocolGranularity:          Rule
                    SourceAddress:                9.1.1.2
                    SourceAddressPrefix:          n/a
                    SourceAddressRange:           n/a
                    SourceAddressGranularity:     Packet
                    SourcePort:                   All
                    SourcePortRange:              n/a
                    SourcePortGranularity:        Rule
                    DestAddress:                  9.1.1.1
                    DestAddressPrefix:            n/a
                    DestAddressRange:             n/a
                    DestAddressGranularity:       Packet
                    DestPort:                     All
                    DestPortRange:                n/a
```

```
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
********************************************************************
FilterName:                   Rule2Admin
FilterNameExtension:          2
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.2
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.1.1.1
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
********************************************************************

4 entries selected
```

NATT anchor and NATT dynamic filters

Using the **ipsec -f** command after the activation of two phase 2 security associations in the branch office with NAT model, the filter structure looks like the following:

```
CS V1R7 ipsec  TCPIP Name: TCPCS  Wed Mar 30 16:36:23 2005
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 34
Logging:  Yes             Predecap: Yes               DVIPSec:  Yes
NatKeepAlive:  20

FilterName:                   Rule2C
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NATT Dynamic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
```

```
|                       SecurityClass:               0
|                       Logging:                     All
|                       Protocol:                    All
|                       ICMPType:                    n/a
|                       ICMPCode:                    n/a
|                       OSPFType:                    n/a
|                       TCPQualifier:                n/a
|                       ProtocolGranularity:         Rule
|                       SourceAddress:               9.3.3.3
|                       SourceAddressPrefix:         n/a
|                       SourceAddressRange:          n/a
|                       SourceAddressGranularity:    Packet
|                       SourcePort:                  All
|                       SourcePortRange:             n/a
|                       SourcePortGranularity:       Rule
|                       DestAddress:                 9.5.5.5
|                       DestAddressPrefix:           n/a
|                       DestAddressRange:            n/a
|                       DestAddressGranularity:      Packet
|                       DestPort:                    All
|                       DestPortRange:               n/a
|                       DestPortGranularity:         Rule
|                       OrigRmtConnPort:             n/a
|                       RmtIDPayload:                10.3.1.1
|                       ****************************************************************************
|                       FilterName:                  Rule2C
|                       FilterNameExtension:         1
|                       GroupName:                   n/a
|                       LocalStartActionName:        StartZoneC
|                       VpnActionName:               Gold-TunnelMode
|                       TunnelID:                    Y3
|                       Type:                        NATT Dynamic
|                       State:                       Active
|                       Action:                      Permit
|                       Scope:                       Local
|                       Direction:                   Outbound
|                       OnDemand:                    No
|                       SecurityClass:               0
|                       Logging:                     All
|                       Protocol:                    All
|                       ICMPType:                    n/a
|                       ICMPCode:                    n/a
|                       OSPFType:                    n/a
|                       TCPQualifier:                n/a
|                       ProtocolGranularity:         Rule
|                       SourceAddress:               9.3.3.3
|                       SourceAddressPrefix:         n/a
|                       SourceAddressRange:          n/a
|                       SourceAddressGranularity:    Packet
|                       SourcePort:                  All
|                       SourcePortRange:             n/a
|                       SourcePortGranularity:       Rule
|                       DestAddress:                 9.5.5.5
|                       DestAddressPrefix:           n/a
|                       DestAddressRange:            n/a
|                       DestAddressGranularity:      Packet
|                       DestPort:                    All
|                       DestPortRange:               n/a
|                       DestPortGranularity:         Rule
|                       OrigRmtConnPort:             n/a
|                       RmtIDPayload:                10.3.2.2
|                       ****************************************************************************
|                       FilterName:                  Rule2C
|                       FilterNameExtension:         1
|                       GroupName:                   n/a
|                       LocalStartActionName:        StartZoneC
|                       VpnActionName:               Gold-TunnelMode
```

```
|                       TunnelID:                   Y0
|                       Type:                       NATT Anchor
|                       State:                      Active
|                       Action:                     Permit
|                       Scope:                      Local
|                       Direction:                  Outbound
|                       OnDemand:                   No
|                       SecurityClass:              0
|                       Logging:                    All
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              9.3.3.3
|                       SourceAddressPrefix:        n/a
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                9.5.5.5
|                       DestAddressPrefix:          n/a
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 Rule2C
|                       FilterNameExtension:        1
|                       GroupName:                  n/a
|                       LocalStartActionName:       StartZoneC
|                       VpnActionName:              Gold-TunnelMode
|                       TunnelID:                   Y0
|                       Type:                       Dynamic Anchor
|                       State:                      Active
|                       Action:                     Permit
|                       Scope:                      Local
|                       Direction:                  Outbound
|                       OnDemand:                   No
|                       SecurityClass:              0
|                       Logging:                    All
|                       Protocol:                   All
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               n/a
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              9.3.3.3
|                       SourceAddressPrefix:        n/a
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 All
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                9.6.0.0
|                       DestAddressPrefix:          16
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
```

```
RmtIDPayload:                 n/a
*******************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NATT Dynamic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 10.3.1.1
*******************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y3
Type:                         NATT Dynamic
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
Protocol:                     All
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
```

```
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 10.3.2.2
|                       ********************************************************************
|                       FilterName:                   Rule2C
|                       FilterNameExtension:          2
|                       GroupName:                    n/a
|                       LocalStartActionName:         StartZoneC
|                       VpnActionName:                Gold-TunnelMode
|                       TunnelID:                     Y0
|                       Type:                         NATT Anchor
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     No
|                       SecurityClass:                0
|                       Logging:                      All
|                       Protocol:                     All
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.5.5.5
|                       SourceAddressPrefix:          n/a
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   All
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.3.3.3
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ********************************************************************
|                       FilterName:                   Rule2C
|                       FilterNameExtension:          2
|                       GroupName:                    n/a
|                       LocalStartActionName:         StartZoneC
|                       VpnActionName:                Gold-TunnelMode
|                       TunnelID:                     Y0
|                       Type:                         Dynamic Anchor
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     No
|                       SecurityClass:                0
|                       Logging:                      All
|                       Protocol:                     All
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 n/a
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.6.0.0
```

```
SourceAddressPrefix:          16
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
*************************************************************************

8 entries selected
```

The inbound dynamic anchor filter protects TCP traffic from source address 9.6.0.0/16, source port any, to destination address 9.3.3.3, destination port 21. The inbound NATT anchor filter protects TCP traffic from source address 9.5.5.5, source port any, to destination address 9.3.3.3, destination port 21. The two inbound NATT dynamic filters also protect TCP traffic from source address 9.5.5.5, source port any, to destination address 9.3.3.3, destination port 21. However, the two NATT dynamic filters were negotiated for separate clients behind the security gateway. You can see that the first inbound NATT dynamic is for a host behind the security gateway using internal address 10.3.1.1 (value in the RmtIDpayload field).

The second inbound NATT dynamic is for a host behind the security gateway using internal address 10.3.2.2. The internal address of the data endpoint is what makes each NATT dynamic unique.

NAT resolution filters

Use the -h option on the **ipsec -f** command to display any NRFs associated with the displayed filters. After two clients behind the security gateway have connected to host 9.3.3.3 using FTP, the NRFs could look like the following (The display has been truncated to include only the NRFs):

```
FilterName:                   Rule2C
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NRF
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
Protocol:                     TCP(6)
ICMPType:                     n/a
ICMPCode:                     n/a
OSPFType:                     n/a
TCPQualifier:                 None
ProtocolGranularity:          Rule
SourceAddress:                9.3.3.3
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   21
```

```
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.5.5.5
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     34732
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              34732
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule2C
|                       FilterNameExtension:          1
|                       GroupName:                    n/a
|                       LocalStartActionName:         StartZoneC
|                       VpnActionName:                Gold-TunnelMode
|                       TunnelID:                     Y3
|                       Type:                         NRF
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Outbound
|                       OnDemand:                     No
|                       SecurityClass:                0
|                       Logging:                      All
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 None
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.3.3.3
|                       SourceAddressPrefix:          n/a
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   21
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.5.5.5
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     65535
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              34732
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule2C
|                       FilterNameExtension:          2
|                       GroupName:                    n/a
|                       LocalStartActionName:         StartZoneC
|                       VpnActionName:                Gold-TunnelMode
|                       TunnelID:                     Y2
|                       Type:                         NRF
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     No
|                       SecurityClass:                0
|                       Logging:                      All
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
```

```
TCPQualifier:                None
ProtocolGranularity:         Rule
SourceAddress:               9.5.5.5
SourceAddressPrefix:         n/a
SourceAddressRange:          n/a
SourceAddressGranularity:    Packet
SourcePort:                  34732
SourcePortRange:             n/a
SourcePortGranularity:       Rule
DestAddress:                 9.3.3.3
DestAddressPrefix:           n/a
DestAddressRange:            n/a
DestAddressGranularity:      Packet
DestPort:                    21
DestPortRange:               n/a
DestPortGranularity:         Rule
OrigRmtConnPort:             34732
RmtIDPayload:                n/a
*******************************************************************************
FilterName:                  Rule2C
FilterNameExtension:         2
GroupName:                   n/a
LocalStartActionName:        StartZoneC
VpnActionName:               Gold-TunnelMode
TunnelID:                    Y3
Type:                        NRF
State:                       Active
Action:                      Permit
Scope:                       Local
Direction:                   Inbound
OnDemand:                    No
SecurityClass:               0
Logging:                     All
Protocol:                    TCP(6)
ICMPType:                    n/a
ICMPCode:                    n/a
OSPFType:                    n/a
TCPQualifier:                None
ProtocolGranularity:         Rule
SourceAddress:               9.5.5.5
SourceAddressPrefix:         n/a
SourceAddressRange:          n/a
SourceAddressGranularity:    Packet
SourcePort:                  65535
SourcePortRange:             n/a
SourcePortGranularity:       Rule
DestAddress:                 9.3.3.3
DestAddressPrefix:           n/a
DestAddressRange:            n/a
DestAddressGranularity:      Packet
DestPort:                    21
DestPortRange:               n/a
DestPortGranularity:         Rule
OrigRmtConnPort:             34732
RmtIDPayload:                n/a
*******************************************************************************
```

There are two NRF inbound/outbound entry pairs associated with the NATT anchor. In this example, two clients behind the security gateway have an FTP connection with host 9.3.3.3. The first outbound NRF entry is for:

```
source address 9.3.3.3, source port 21
destination address 9.5.5.5, destination port 34732
protocol TCP
```

The destination port is shown in the DestPort field. This value can be a translated value. The OrigRmtConnPort field indicates the original remote connection port, prior to remote port translation by Communications Server. In this example, the first outbound NRF shows that DestPort and OrigRmtConnPort are both 34732. For more information, see "Remote port translation" on page 855.

The second outbound NRF entry is for:

```
source address 9.3.3.3, source port 21
destination address 9.5.5.5, destination port 65535
protocol TCP
```

The original remote connection port (OrigRmtConnPort) is 34732. Because the values in DestPort and OrigRmtConnPort do not match, you can tell that the value was translated by Communications Server's remote port translation function. For more information, see "Remote port translation" on page 855.

The TunnelID field provides information on which phase 2 security association the traffic will be sent over. In this example, the phase two security associations are identified by the labels Y2 and Y3 respectively.

Displaying remote port translation with the ipsec command

As seen in the previous example, the remote data endpoint is represented by the security gateway's public IP address (9.5.5.5), not the client's IP address (10.3.1.1 or 10.3.2.2). Using the **ipsec -o display** command after the activation of two FTP connections in the branch office with NAT model, the port mappings look like the following:

```
CS V1R7 ipsec  TCPIP Name: TCPCS  Wed Mar 30 16:37:59 2005
Primary:  NATT Port Trans Function: Display            Format:   Detail
Source:   Stack           Scope:    Current            TotAvail: 2

RmtIpAddress:         9.5.5.5
Protocol:             TCP(6)
TransRmtConnPort:     34732
OrigRmtConnPort:      34732
RmtInnerIpAddress:    10.3.1.1
*************************************************************************
RmtIpAddress:         9.5.5.5
Protocol:             TCP(6)
TransRmtConnPort:     65535
OrigRmtConnPort:      34732
RmtInnerIpAddress:    10.3.2.2
*************************************************************************

2 entries selected
```

In both entries, you can see that the remote IP address (RmtIpAddress) value is 9.5.5.5, the IP address of the branch office gateway, the protocol is TCP (6), and the original remote connection port (OrigRmtConnPort) is 34732. The first entry shows that the translated remote connection port (TransRmtConnPort) is also 34732. The remote inner IP address contains the private address of the client behind the security gateway that initiated the connection, 10.3.1.1. The second entry shows that the remote connection port was translated to a value of 65535 (TransRmtConnPort), and that the client initiating the connection is using private IP address 10.3.2.2.

Table 37 on page 971 details several places where one or both of the remote port values are displayed or used for a selection.

Table 37. Original and translated port values

Function	How remote port values are used	Which remote port, original or translated?
Netstat displays of connection data, such as Netstat ALL/-A, Netstat ALLConn/-a, and Netstat COnn/-c	The Netstat command has many options to display connection information, including the remote port value. In some cases, the Netstat command takes a remote port value as a selector.	Translated remote port
Netstat display of the VIPA connection routing table (netstat VCRT/-V)	This command displays the remote port value in the sport or Source field, depending on the flavor of the report generated, and allows you to select based on port.	Translated remote port
Packet trace	Packet trace displays packet data as it was received or sent. If the packet is authenticated but not encrypted, the port is visible in the packet trace data.	Original remote port
IPSecurity syslog messages: • `EZD0814I packet permitted` • `EZD0815I packet denied by policy` • `EZD0821I packet denied, no tunnel` • `EZD0822I packet denied, tunnel inactive` • `EZD0832I packet denied by NAT traversal processing`	For an inbound packet, the sport field in these messages contains a remote port value. For an outbound packet, the dport field in these messages contains a remote port value. These messages also have an origport field.	The sport and dport fields contain the translated remote port, and the origport field contains the original remote port.
Dynamic anchor, displayed with the **ipsec -f** command and the **ipsec -t** command	The dynamic anchor that is configured in the Policy Agent configuration file can specify the remote port as a single port, a range of ports, or all ports. This specification of the remote port controls the range of ports that the original port can be translated to. Both the original port and the translated port for a connection will fit the range of ports coded.	The configured remote port value is displayed. A packet's original port is used to match on this rule. Both the original port and translated port are included in the remote port value displayed.
NATT anchor, displayed with the **ipsec -f** command and the **ipsec -t** command	The NATT anchor, which is created as a result of the security association negotiation, contains a specific remote port or all ports.	Original remote port if specific port displayed

Table 37. Original and translated port values (continued)

Function	How remote port values are used	Which remote port, original or translated?
NATT dynamic, displayed with the **ipsec -f** command and the **ipsec -t** command	The NATT dynamic, which is created as a result of the security association negotiation, contains a specific remote port or all ports.	Original remote port if specific port displayed
NAT resolution filter (NRF), displayed with the **ipsec -f** command with the -h option	The NAT resolution filter is a connection level filter, and contains both the translated remote port and the original remote port.	Translated remote port and original remote port. A packet's translated port is used to match on this rule.
ipsec -t command	The **ipsec** traffic test command allows a remote port value to be specified as a filter selection criteria.	Original remote port

Displaying security associations with the ipsec command

Displaying IKE tunnel information with the ipsec command

Use the **ipsec -k display** command to display IKE tunnel information.

```
ipsec -k display

CS V1R7 ipsec  TCPIP Name: TCPCS  Wed Mar 30 17:20:44 2005
Primary:  IKE tunnel      Function: Display           Format:   Detail
Source:   IKED            Scope:    Current           TotAvail: n/a

TunnelID                      K1
KeyExchangeRuleName:          ZoneC_KeyExRule1
KeyExchangeActionName:        Gold-PSK
LocalEndPoint:                9.3.3.3
LocalIDType:                  IPV4
LocalID:                      9.3.3.3
RemoteEndPoint:               10.3.1.1
RemoteIDType:                 USERFQDN
RemoteID:                     gateway.poughkeepsie.ibm.com
ExchangeMode:                 Main
State:                        DONE
AuthenticationAlgorithm:      Hmac_Sha
EncryptionAlgorithm:          3DES
DiffieHellmanGroup            2
AuthenticationMethod:         PresharedKey
InitiatorCookie:              0XD08B69783E9F9F24
ResponderCookie:              0X2980CD07CB84787B
Lifesize:                     0K
CurrentByteCount:             1216b
Lifetime:                     240m
LifetimeRefresh:              2005/03/30 19:52:25
LifetimeExpires:              2005/03/30 20:32:27
Role:                         Responder
AssociatedDynamicTunnels:     2
NATTSupportLevel:             RFC
NATInFrntLclScEndPnt:         No
NATInFrntRmtScEndPnt:         Yes
zOSCanInitiateP1SA:           Yes
```

```
AllowNat:                       No
*************************************************************************

1 entries selected
```

The setting of the AllowNat field indicates whether or not NAT traversal support was advertised to the IKE peer. If AllowNat is Yes, the negotiation might or might not have detected a NAT. If the NATInFrntLclScEndPnt field is Yes, a NAT device was detected in front of the local security endpoint. If the NATInFrntRmtScEndPt field is Yes, a NAT device was detected in front of the remote security endpoint.

Displaying IPSec tunnel information with the ipsec command

Use the **ipsec -y display** command to display IKE tunnel information.

```
ipsec -y display -a Y39

TunnelID                        Y39
VpnActionName:                  TransportMode
State:                          Active
LocalEndPoint:                  9.2.2.2
RemoteEndPoint:                 9.4.4.4
HowtoEncap:                     Transport
HowToAuth:                      ESP
 AuthAlgorithm:                 Hmac_Md5
 AuthInboundSpi:                3574499869
 AuthOutboundSpi:               279454707
HowToEncrypt:                   DES
 EncryptInboundSpi:             3574499869
 EncryptOutboundSpi:            279454707
OutboundPackets:                6
OutboundBytes:                  218
InboundPackets:                 4
InboundBytes:                   136
Lifesize:                       0K
LifesizeRefresh:                0K
CurrentByteCount:               0b
LifetimeRefresh:                2005/04/01 15:05:44
LifetimeExpires:                2005/04/01 16:14:08
CurrentTime:                    2005/04/01 12:29:25
VPNLifeExpires:                 2005/04/02 12:14:08
ParentIKETunnelID:              K9
LocalDynVpnRule:                n/a
NAT Traversal Topology:
  UdpEncapMode:                 Yes
  LclNATDetected:               No
  RmtNATDetected:               Yes
  RmtIsGw:                      No
  RmtIsZOS:                     Yes
  zOSCanInitP2SA:               Yes
  SrcNATOARcvd:                 10.2.2.2
  DstNATOARcvd:                 9.2.2.2
```

The NAT Traversal Topology fields show additional information when a NAT was detected in the path between the IKE peers. The setting of the UdpEncapMode field indicates whether a UDP-encapsulated mode security association has or has not been negotiated. If NAT Traversal is supported by both IKE peers and one or more NATs are detected, UdpEncapMode is set to Yes. The RmtNATDetected field is Yes if a NAT is detected in front of the remote peer. The RmtIsGW field is Yes if the remote peer is acting as a security gateway.

Tip: Use the -b option of the **ipsec -y display** command to show the ports and protocols of the dynamic filter that are associated with the phase 2 security

association. The following excerpt from the **ipsec -y display** using the -b option indicates a Telnet connection from a remote host:

```
AssociatedFiltProtocol:     TCP(6)
AssociatedFiltSrcPort:      23
AssociatedFiltDestPort:     0
```

Displaying filter rules with the pasearch command

The configured IP filter rules and associated actions can also be viewed from the perspective of the Policy Agent. The **pasearch** command provides a way to view all Policy Agent configuration, of which IP security is a subset. In contrast to the information that is provided by the **ipsec** command, the detailed information that is provided by the **pasearch** command does not reflect the stack's active use of the IP security policy, but offers a relatively static view of configured IP security values that were generated from the IpSecConfig file and CommonIpSecConfig file. For more information on displaying policy based networking information, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Verifying filter action

To quickly determine which filter rule applies to a specific type of traffic, use the **ipsec** traffic test command (**ipsec -t**). This command returns all of the rules in the current filter table that match the given traffic type.

For example, to test which filter rule matches an incoming FTP connection request from remote IP address 9.1.1.2 to local IP address 9.1.1.1, issue the following command. The input values represent the remote address, local address, protocol, remote port, local port, direction, and security class of the packet.:

```
ipsec -t  9.1.1.2 9.1.1.1 tcp 0 21 in 0

CS V1R7 ipsec  TCPIP Name: TCPCS  Fri Mar 18 12:13:44 2005
Primary:  IP Traffic Test Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    n/a               TotAvail: 8
TestData: 9.1.1.2  9.1.1.1  tcp 0 21 in 0

FilterName:                  Rule2Admin
FilterNameExtension:         2
GroupName:                   Admin
LocalStartActionName:        n/a
VpnActionName:               Silver-TransportMode
TunnelID:                    Y0
Type:                        Dynamic Anchor
State:                       Active
Action:                      Permit
Scope:                       Local
Direction:                   Inbound
OnDemand:                    No
SecurityClass:               0
Logging:                     Deny
Protocol:                    All
ICMPType:                    n/a
ICMPCode:                    n/a
OSPFType:                    n/a
TCPQualifier:                n/a
ProtocolGranularity:         Rule
SourceAddress:               9.1.1.2
SourceAddressPrefix:         n/a
SourceAddressRange:          n/a
SourceAddressGranularity:    Packet
SourcePort:                  All
SourcePortRange:             n/a
SourcePortGranularity:       Rule
DestAddress:                 9.1.1.1
```

```
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule1A
|                       FilterNameExtension:          15
|                       GroupName:                    ZoneA
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 Connect Inbound
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.1.1.0
|                       SourceAddressPrefix:          24
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   All
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.1.1.1
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     21
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule1A
|                       FilterNameExtension:          18
|                       GroupName:                    ZoneA
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 Connect Outbound
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.1.1.0
|                       SourceAddressPrefix:          24
```

```
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   21
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.1.1.1
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule1A
|                       FilterNameExtension:          19
|                       GroupName:                    ZoneA
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
|                       ICMPCode:                     n/a
|                       OSPFType:                     n/a
|                       TCPQualifier:                 Connect Inbound
|                       ProtocolGranularity:          Rule
|                       SourceAddress:                9.1.1.0
|                       SourceAddressPrefix:          24
|                       SourceAddressRange:           n/a
|                       SourceAddressGranularity:     Packet
|                       SourcePort:                   20
|                       SourcePortRange:              n/a
|                       SourcePortGranularity:        Rule
|                       DestAddress:                  9.1.1.1
|                       DestAddressPrefix:            n/a
|                       DestAddressRange:             n/a
|                       DestAddressGranularity:       Packet
|                       DestPort:                     All
|                       DestPortRange:                n/a
|                       DestPortGranularity:          Rule
|                       OrigRmtConnPort:              n/a
|                       RmtIDPayload:                 n/a
|                       ********************************************************************************
|                       FilterName:                   Rule1A
|                       FilterNameExtension:          20
|                       GroupName:                    ZoneA
|                       LocalStartActionName:         n/a
|                       VpnActionName:                n/a
|                       TunnelID:                     0x00
|                       Type:                         Generic
|                       State:                        Active
|                       Action:                       Permit
|                       Scope:                        Local
|                       Direction:                    Inbound
|                       OnDemand:                     n/a
|                       SecurityClass:                0
|                       Logging:                      None
|                       Protocol:                     TCP(6)
|                       ICMPType:                     n/a
```

```
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               Connect Outbound
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              9.1.1.0
|                       SourceAddressPrefix:        24
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 50000
|                       SourcePortRange:            50200
|                       SourcePortGranularity:      Rule
|                       DestAddress:                9.1.1.1
|                       DestAddressPrefix:          n/a
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 Rule1All-Permit
|                       FilterNameExtension:        6
|                       GroupName:                  ZoneAll
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Permit
|                       Scope:                      Local
|                       Direction:                  Inbound
|                       OnDemand:                   n/a
|                       SecurityClass:              0
|                       Logging:                    None
|                       Protocol:                   TCP(6)
|                       ICMPType:                   n/a
|                       ICMPCode:                   n/a
|                       OSPFType:                   n/a
|                       TCPQualifier:               Connect Outbound
|                       ProtocolGranularity:        Rule
|                       SourceAddress:              0.0.0.0
|                       SourceAddressPrefix:        0
|                       SourceAddressRange:         n/a
|                       SourceAddressGranularity:   Packet
|                       SourcePort:                 53
|                       SourcePortRange:            n/a
|                       SourcePortGranularity:      Rule
|                       DestAddress:                0.0.0.0
|                       DestAddressPrefix:          0
|                       DestAddressRange:           n/a
|                       DestAddressGranularity:     Packet
|                       DestPort:                   All
|                       DestPortRange:              n/a
|                       DestPortGranularity:        Rule
|                       OrigRmtConnPort:            n/a
|                       RmtIDPayload:               n/a
|                       ************************************************************************
|                       FilterName:                 Rule2All-Deny
|                       FilterNameExtension:        2
|                       GroupName:                  ZoneAll
|                       LocalStartActionName:       n/a
|                       VpnActionName:              n/a
|                       TunnelID:                   0x00
|                       Type:                       Generic
|                       State:                      Active
|                       Action:                     Deny
|                       Scope:                      Both
```

```
                    Direction:                  Inbound
                    OnDemand:                   n/a
                    SecurityClass:              0
                    Logging:                    All
                    Protocol:                   All
                    ICMPType:                   n/a
                    ICMPCode:                   n/a
                    OSPFType:                   n/a
                    TCPQualifier:               n/a
                    ProtocolGranularity:        Rule
                    SourceAddress:              0.0.0.0
                    SourceAddressPrefix:        0
                    SourceAddressRange:         n/a
                    SourceAddressGranularity:   Packet
                    SourcePort:                 All
                    SourcePortRange:            n/a
                    SourcePortGranularity:      Rule
                    DestAddress:                0.0.0.0
                    DestAddressPrefix:          0
                    DestAddressRange:           n/a
                    DestAddressGranularity:     Packet
                    DestPort:                   All
                    DestPortRange:              n/a
                    DestPortGranularity:        Rule
                    OrigRmtConnPort:            n/a
                    RmtIDPayload:               n/a
                    ************************************************************************
                    FilterName:                 DenyAllRule_Generated__________Inbnd
                    FilterNameExtension:        n/a
                    GroupName:                  n/a
                    LocalStartActionName:       n/a
                    VpnActionName:              n/a
                    TunnelID:                   0x00
                    Type:                       Generic
                    State:                      Active
                    Action:                     Deny
                    Scope:                      Both
                    Direction:                  Inbound
                    OnDemand:                   n/a
                    SecurityClass:              0
                    Logging:                    None
                    Protocol:                   All
                    ICMPType:                   n/a
                    ICMPCode:                   n/a
                    OSPFType:                   n/a
                    TCPQualifier:               n/a
                    ProtocolGranularity:        Rule
                    SourceAddress:              0.0.0.0
                    SourceAddressPrefix:        0
                    SourceAddressRange:         n/a
                    SourceAddressGranularity:   Packet
                    SourcePort:                 All
                    SourcePortRange:            n/a
                    SourcePortGranularity:      Rule
                    DestAddress:                0.0.0.0
                    DestAddressPrefix:          0
                    DestAddressRange:           n/a
                    DestAddressGranularity:     Packet
                    DestPort:                   All
                    DestPortRange:              n/a
                    DestPortGranularity:        Rule
                    OrigRmtConnPort:            n/a
                    RmtIDPayload:               n/a
                    ************************************************************************

                    8 entries selected
```

An incoming FTP connection request matches all of the rules shown in the example. The first rule that is returned does not always match a specific packet, depending on how much detail you provide as the input to the **ipsec -t** command. However, the Rule2Admin rule is the best match in this case, so the search for a matching filter ends there. The matching rule in this case is an ipsec rule, as indicated by the designation `Dynamic Anchor`. Therefore, IPSec processing is applied to this packet.

Tip: When using the **ipsec -t** command, provide as much detailed input as possible. The more detailed the input to the command, the more narrow the results of the search will be.

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Security associations

This section describes the following:

- Activating a security association
- Verifying the activation of a security association
- Verifying the use of an active security association
- Refreshing security associations
- Deactivating security associations

Activating a security association

Negotiations can be initiated in one of four ways:

- Remote activation

  When a remote IKE peer initiates a negotiation with the local IKE daemon, no action is required. If the IP security policy has been configured correctly and is consistent with the policy of the remote IKE peer, a security association is established. No operator message is issued when a remote activation has occurred, but the syslog does contain a record of all IKE activity. The **ipsec -y display** command can also be used to view all of the active security associations.

- On-demand activation

  An on-demand security association is activated when some outbound traffic matches an ipsec rule that allows on-demand activation. The `ondemand` field of the filter display indicates whether or not on-demand activation is allowed for that rule.

- Automatic activation

  The local IKE daemon initiates a negotiation for an autoactivated security association when it connects to the TCP/IP stack. IKE also initiates a negotiation for an autoactivated security association when the **ipsec -f reload** command is issued, changing the active filter rule set from default IP filter rules to Policy Agent filter rules. No operator message is issued when an autoactivation has occurred, but the syslog does contain a record of all IKE activity. The **ipsec -y display** command can also be used to view all of the active security associations.

- Command-line activation

  The **ipsec** command can be used as follows to activate a security association that has been defined by a LocalDynVpnRule statement:

```
ipsec -y activate -l ZoneC_VPN-EE1

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 12:25:50 2004
```

```
Primary:  Dynamic tunnel  Function: Activate

Selection Data                                  Status
ZoneC_VPN-EE1                                   Activating
```

The output of the command indicates the status of the activation.

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Verifying the activation of a security association

After a security association has been activated, it can be displayed with the **ipsec -y display** command. To view all active security associations, issue the following command:

```
ipsec -y display
```

You can use the **ipsec** command to view all of the active phase 1 security associations, or limit the report to a single phase 1 security association by using the -a option.

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Verifying the use of an active security association

If filter logging is enabled for the selected filter rule, the log indicates whether a packet has been permitted with IPSec processing applied. Among the information available for a typical filter log entry are the rule name, the action, and the tunnel ID:

```
Jun 30 16:46:14 MVS175 TRMD.TCPCS[30]: EZD0814I Packet permitted: 06/30/2004
16:46:11.71 filter rule= Rule2Admin ext= 1 sipaddr= 10.1.1.175 dipaddr= 10.1.1.172
proto= tcp(6) sport= 1032 dport= 1035 Interface= 192.168.175.172 ( 0 ) secclass= 255
dest= local len= 52 vpnaction= Silver-TransportMode tunnelID= Y2
```

The **ipsec -y display** command also outputs a field with the number of bytes of traffic that have been protected by a particular security association.

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Refreshing security associations

When a security association is refreshed, the encryption keys are changed. Refreshing a security association periodically prevents the keys from being compromised by an outside party. Phase 1 ISAKMP security associations and phase 2 IPSec security associations are refreshed automatically, based on the lifetime or life size that was negotiated between the two IKE peers.

Tip: For phase 1, these parameters are specified in the KeyExchangeOffer statement. For phase 2, these parameters are specified in the IpDataOffer statement.

You can also refresh security associations from the z/OS UNIX command line, but this should only be necessary in exceptional conditions because the IKE daemon is normally responsible for refreshing the keys at configured intervals. Exceptional conditions might include the compromise of a key or the failure to receive an informational IKE message from a remote host.

Phase 1

Each phase 1 security association is identified by a tunnel ID, a number with a prefix of *K*. To manually refresh a phase 1 ISAKMP security association, issue the **ipsec -k display** command to find the tunnel ID. Then issue the **ipsec -k refresh** command for that ID as follows:

```
ipsec -k refresh -a K1

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:33:45 2004
Primary:  IKE tunnel      Function: Refresh


Tunnel ID    Status
K1           Refreshing
```

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Phase 2

Each phase 2 security association is identified by a tunnel ID, a number with a prefix of *Y*. To manually refresh a phase 2 IPSec security association, issue the **ipsec -y display** command to find the tunnel ID. Then issue the **ipsec -y refresh** command for that ID as follows:

```
ipsec -y refresh -a Y2

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:40:16 2004
Primary:  Dynamic tunnel  Function: Refresh

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           ZoneC_VPN-EE1                                    Refreshing
```

The phase 2 IPSec security association can also be identified by the local dynamic VPN rule with which it is associated, if one exists, as follows:

```
ipsec -y refresh -l ZoneC_VPN-EE1

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:42:29 2004
Primary:  Dynamic tunnel  Function: Refresh

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           ZoneC_VPN-EE1                                    Refreshing
```

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Deactivating security associations

When a security association is deleted, all of the information that is stored in the security association is deleted from the TCP/IP stack and from IKED, along with the dynamic filters that were created when the security association was created. After deletion, the security association is no longer available for use. Traffic that was protected by the old security association is denied until a new security association is subsequently activated.

When a parent phase 1 security association is deactivated, all of the associated phase 2 security associations are deleted as well. Be careful when deleting phase 1 security associations, because all traffic that uses the security association and its associated phase 2 security associations are dropped until new security associations can be negotiated.

- To delete a phase 1 security association and all of the phase 2 security associations it is protecting, issue the following command:

```
ipsec -k deactivate -a K1

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:56:29 2004
Primary:  IKE tunnel      Function: Deactivate

Tunnel ID    Status
K1           Deactivating
```

- To delete all phase 1 security associations and all phase 2 security associations, use the `-a all` option as follows:

```
ipsec -k deactivate -a all

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:59:45 2004
Primary:  IKE tunnel      Function: Deactivate

All IKE tunnels Deactivating
```

- To delete a phase 2 security association, issue the following command:

```
ipsec -y deactivate -a Y2

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:53:33 2004
Primary:  Dynamic tunnel  Function: Deactivate

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           n/a                                              Deactivating
```

  The `n/a` in the LocalDynVpnRuleName field indicates that no local dynamic VPN rule name is associated with this security association. The security association was either remotely activated or was activated on-demand.

- To delete all phase 2 security associations, use the `-a all` option as follows:

```
ipsec -y deactivate -a all

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 14:58:07 2004
Primary:  Dynamic tunnel  Function: Deactivate

All dynamic tunnels Deactivating
```

For detailed information about the use of the **ipsec** command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Modifying active IP security policy

This section describes the effects of changes to security-related files and of issuing the **ipsec -f default** command.

IP security policy files

The CommonIpSecConfig and IpSecConfig files for a TCP/IP stack with IPSECURITY defined can be modified while the stack is active as follows:

- Policy Agent configuration, including IP security, is updated at each configured refresh interval that is specified on the TcpImage statement. The default is 30 minutes.
- If Policy Agent was started with the -i option and the configuration files are stored in an HFS file, changes to any of the configuration files are detected and updated dynamically without user intervention.
- Policy Agent configuration, including IP security, can be updated at any time by issuing the MODIFY PAGENT,REFRESH command or the MODIFY PAGENT,UPDATE command from the console.

Policy Agent image configuration files

For an active TCP/IP stack with IPSECURITY defined, the following conditions apply:

- Changing the file name that is identified by an existing IpSecConfig statement causes Policy Agent to update and install the IP security policy as defined in the new file. If the new IP security policy file contains errors, the policy is not updated and the existing IP security policy remains in effect.
- Removing an existing IpSecConfig statement from a Policy Agent image configuration file activates the default IP filter policy.

Policy Agent main configuration file

For an active TCP/IP stack with IPSECURITY defined, removing an existing TcpImage statement from the Policy Agent configuration file has the following effects on IP security policy:

- Existing IP filters remain.
- Existing security associations remain.
- Traffic continues to flow in the same way it did before the TcpImage statement was removed, including IPSec-protected traffic.
- New security associations cannot be activated.
- Existing security associations cannot be refreshed and are deleted when the refresh period expires.

If the intent of removing the TcpImage statement is to remove IP filters from the stack, an alternative is to modify the IP security policy to install a filter rule that permits all inbound and outbound traffic. Also, before restarting the stack, the IPSECURITY parameter should be removed from the IPCONFIG statement of the relevant stack.

Active security associations and the ipsec -f default command

Any active security associations that were negotiated for IPSec-protected traffic are not deleted when the **ipsec -f default** command is issued. However, they are deleted if, while the default policy is in effect, any associated IP filter rules from the IP filter policy are deleted or modified in such a way that the filter rule no longer encompasses the scope of the security association. In that case, the security association will be deleted when the IP security policy is reloaded.

For example, security associations are not deleted by the following sequence of actions:

1. The **ipsec -f default** command is issued.

   Security associations remain active in the stack and in IKE, though unavailable for use.
2. No modification is made to the IP filter policy in the IP security configuration files.

   Security associations remain active in the stack and in IKE, though unavailable for use.
3. The **ipsec -f reload** command is issued.

   Security associations remain active in the stack and in IKE, and are available for use.

Security associations are deleted by the following sequence of actions:

1. The **ipsec -f default** command is issued.

Security associations remain active in the stack and in IKE.

2. The IpFilterRule statement that is associated with an active security association is deleted.
3. The IP security policy is updated by issuing the MODIFY PAGENT,REFRESH command from the console.

   Existing security associations are deleted.
4. The **ipsec -f reload** command is issued.

   Security associations have been deleted.

In either case, security associations are never available for use when the default IP filter policy is in effect.

Considerations for sysplex-wide security associations

z/OS IP security supports sysplex-wide security associations (SWSA). In a sysplex environment, SWSA provides for distribution of IPSec security associations to the target stacks of distributed DVIPAs. To enable this support, you must code IPCONFIG IPSECURITY in the TCP/IP profile, as well as DVIPSEC in the IPSEC block.

In a sysplex environment, the IKE daemon can detect movement of a DVIPA to or from an IP security or Firewall Technologies stack. To re-establish the security associations of a DVIPA, the DVIPSEC option must be specified in the TCP/IP profile of both the stack that the DVIPA is being moved from and the stack detecting the movement. For details of this movement, see "Sysplex-wide security associations" on page 328. For information on the IPCONFIG and IPSEC statements that need to be added to the TCP/IP profile to configure this support, refer to *z/OS Communications Server: IP Configuration Reference*.

When a DVIPA is moved from one IP security stack in a sysplex to another IP security stack, and both stacks have the DVIPSEC option specified, an attempt is made to automatically re-establish security associations on the backup stack. The IKE daemon on the system that is assuming control of the DVIPA attempts to renegotiate new security associations to replace the ones that were on the system that previously owned the DVIPA. If these attempts fail due to configuration errors or connectivity errors, manual intervention might be required. Phase 1 security association or phase 2 security association negotiations that were in progress at the time of the DVIPA movement are lost. However, if these negotiations were for a refresh, a new negotiation is started in the process of assuming control of the DVIPA.

In NATT configurations where IKE can act only as the responder, the IKE daemon does not attempt to renegotiate a new security association for a UDP-encapsulated mode security association. Sysplex distribution is possible in these configurations, but the recovery of the security associations when the DVIPA moves is not supported. For configurations where IKE is limited to only a responder role in negotiating a UDP-encapsulated mode security association, as well as interoperability considerations, see "Configuration scenarios supported for NAT traversal" on page 948.

When IKE is limited to only a responder role, the security association must be reestablished by peer initiation. The interoperability considerations for establishing an initial phase 2 security association are relevant to the renegotiation of the phase 2 security association due to the movement of a DVIPA. For example, a host-to-host UDP-Encapsulated-Tunnel mode security association protecting

specific protocols or ports that was initially initiated from a non-z/OS client, might not be able to be renegotiated when the z/OS system assuming control of the DVIPA initiates the negotiation. If this is the case, the security association must be reestablished by peer initiation.

When a DVIPA is moved from one IP security stack in a sysplex to another IP security stack, and one or both stacks do not have the DVIPSEC option specified, the security associations that are associated with that DVIPA must be re-established by issuing the **ipsec** command, on-demand activation, or by a peer initiation.

Guidelines:

- If a DVIPA is manually deleted and that address has no backup, the IKE daemon might not be able to terminate the tunnels in which the DVIPA is a security endpoint. To avoid this problem, use the **ipsec** command to deactivate the DVIPA's IKE tunnels before manually deleting the DVIPA.
- This support does not address the dynamic relocation of static filter rules and VPN policy definitions to the target system in the sysplex. It is up to you to ensure that the necessary filter rules and IP security policy definitions exist on all participating systems in the sysplex. If the necessary filter rules and IP security policy definitions do not exist, the IKE daemon might not be able to re-establish all security associations. For a description of the SWSA function, see "Sysplex-wide security associations" on page 328.

Rule: Because the renegotiation of a security association after a DVIPA move requires the sysplex stack to initiate an IKE negotiation, the sysplex stack must be allowed to initiate. You must code the Initiation attribute on the IpDynVpnAction statement as `localonly` or `either`.

Shadow security associations

When an IP security stack is the target of a DVIPA, it receives a copy (shadow) of any active security associations for the DVIPA. To display the shadow security associations, use the following command:

```
ipsec -y display -s

ZCS V1R7 ipsec  TCPIP Name: TCPCS  Wed Jun 30 16:24:04 2004
Primary:  Dynamic tunnel  Function: display (shadows)  Format:   Detail
Source:   Stack           Scope:    Current            TotAvail: 1

TunnelID                      Y17
VpnActionName:                E-N-TRAN-AHMD5-DES
State:                        Shadow
LocalEndPoint:                10.172.1.20
RemoteEndPoint:               10.1.1.175
HowtoEncap:                   Transport
HowToAuth:                    AH
 AuthAlgorithm:               Hmac_Md5
 AuthInboundSpi:              215887401
 AuthOutboundSpi:             3906679633
HowToEncrypt:                 DES
 EncryptInboundSpi:           1044458494
 EncryptOutboundSpi:          1914793660
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2004/06/30 19:23:57
LifetimeExpires:              2004/06/30 20:23:57
```

```
CurrentTime:                2004/06/30 16:24:04
VPNLifeExpires:             2004/07/01 16:23:57
ParentIKETunnelID:          K16
LocalDynVpnRule:            n/a
```

Interoperability with z/OS Integrated Security Services Firewall Technologies

z/OS IP security is compatible with z/OS Integrated Security Services Firewall Technologies in the following ways:

- The FWKERN daemon and the IKE daemon can be run on the same z/OS image.
- You can run a firewall stack and an IP security stack on the same z/OS image.
- A particular z/OS TCP/IP stack can be configured for IPSECURITY or FIREWALL, but not both.
- An IP security stack can negotiate a security association with a firewall stack.
- If running in a sysplex:
  - An IP security stack can be the target of a distributed DVIPA from a firewall stack.
  - A firewall stack can be the target of a distributed DVIPA from an IP security stack.

z/OS IP security supports negotiating UDP-encapsulated mode security associations to allow IPSec traffic to traverse a NAT. z/OS Integrated Security Services Firewall Technologies does not support UDP-encapsulated mode security associations. Therefore:

- A firewall stack is unable to negotiate a UDP-encapsulated mode security association with an IP security stack.
- If running in a sysplex:
  - For UDP-encapsulated ESP traffic, a firewall stack cannot be the target of a distributed DVIPA from an IP security stack.
  - With UDP-encapsulated ESP traffic, a firewall stack should not be configured as the backup for an IP security stack.

Sample IP security policy files

A sample stack-specific policy is located in /usr/lpp/samples/pagent_IPSec.conf.

A sample common policy is located in /usr/lpp/samples/pagent_CommonIPSec.conf.

Chapter 18. Application Transparent Transport Layer Security (AT-TLS) data protection

The Transport Layer Security (TLS) protocol defined in RFC 2246 provides communications privacy over the Internet. The protocol enables client and server applications to communicate in a way that is designed to prevent eavesdropping, tampering, and message forgery. To implement TLS protocols, typically, applications must be modified to incorporate a TLS toolkit. Modifying applications requires significant development overhead, ongoing maintenance for each application, and application-specific knowledge of the parameters needed to implement TLS for that application.

Application Transparent Transport Layer Security (AT-TLS) consolidates TLS implementation in one location, reducing or eliminating application development overhead, maintenance, and parameter specification. AT-TLS is based on z/OS System SSL, and transparently implements these protocols in the TCP layer of the stack. As shown in Figure 96, most applications do not need any awareness of the security negotiations and encryption done by TCP/IP on its behalf. However, you might want some applications to be aware of AT-TLS or have control over the security functions being performed by TCP/IP. For example, if the application is a server requesting client authentication, you might want the application to get the partner certificate or the user ID associated with the partner certificate, or the application might negotiate in cleartext with its partner to decide whether a secure session is necessary. If both agree to a secure session, the application needs to tell AT-TLS to set up a secure session. The SIOCTTLSCTL ioctl provides the interface for the application to query or control AT-TLS.

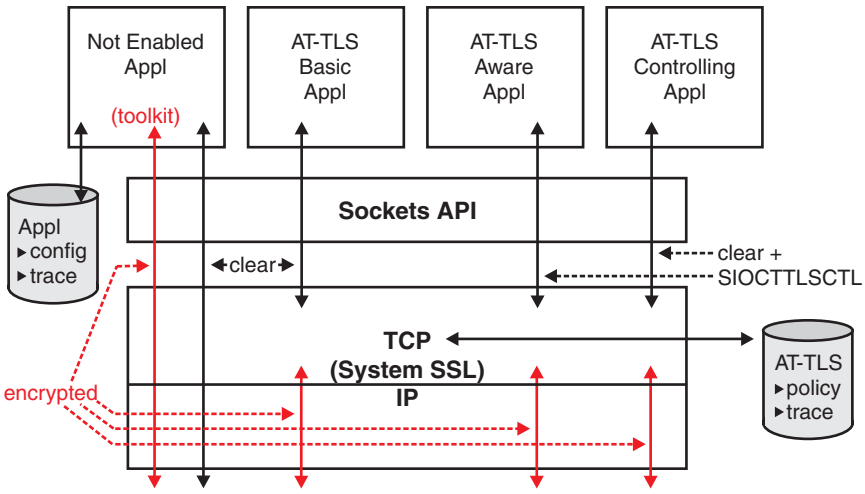


Figure 96. Application Transparent TLS

In all cases, an application using a socket enabled with AT-TLS continues to send and receive text data in the clear while encrypted data flows over the network. This allows the use of TLS with applications that cannot be modified or that cannot incorporate one of the available tool kits. The partner application must also support TLS protocols, either by using AT-TLS or an available TLS toolkit.

AT-TLS configuration in PROFILE.TCPIP

AT-TLS support is controlled by the TTLS or NOTTLS parameter on the TCPCONFIG statement in PROFILE.TCPIP. AT-TLS is enabled by specifying TTLS. The information required to negotiate secure connections is provided to the stack by AT-TLS policies configured in Policy Agent. When AT-TLS is enabled and a newly established connection is first used, the TCP layer of the stack searches for a matching AT-TLS policy installed from the Policy Agent. If no policy is found, the connection is made without AT-TLS involvement.

TCP/IP stack initialization access control

A TCP/IP stack initializes before Policy Agent installs configured policies into the stack. This leaves a window of time where connections that should be covered by AT-TLS are clear text connections. The RACF resource EZB.INITSTACK.*sysname.tcpname* in the SERVAUTH class is used to block stack access, except for the user IDs permitted to the resource. While in this initialization window, any socket request from an unauthorized application receives the same errno (EAGAIN with JrTcpNotActive) received prior to the stack coming up.

Tip: The TN3270E server running inside the stack is always considered authorized. If you want to prevent Telnet connections during this window, you can delay the start of the TN3270E server until after AT-TLS policy is installed. Alternatively, you can define a NACUSERID for the TN3270E port that is not permitted to the INITSTACK resource profile.

Checking is done only if the TCP/IP profile activates AT-TLS. If there is no profile in the SERVAUTH class covering this resource name, all socket requests fail, including those from Policy Agent. Checking ceases the first time that the Policy Agent indicates AT-TLS policy is complete, or if a TCP/IP profile change deactivates AT-TLS.

When the limited access window begins, non-scrollable message EZZ4248E is written to the system console stating that TCP/IP is waiting for Policy Agent to install AT-TLS policies. The message is released when the restriction ends.

You must permit a limited set of administrative applications to the profile to ensure full initialization of the stack. If Policy Agent is dependent on other applications in your environment, they must also be permitted. You can permit other applications that do not require AT-TLS and that you want to start prior to general applications. At a minimum, the following applications must be permitted to the profile:

- Policy Agent
- OMPROUTE
- SNMP subagent

For examples of the security product commands needed to create this resource profile name and grant users access to it, refer to member EZARACF in sample data set SEZAINST.

Options for configuring AT-TLS security

AT-TLS is configured using a set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the AT-TLS policy for each TCP/IP stack. In a complex environment, this file can become large. For this reason, there are two alternatives for creating the Policy Agent files.

Option 1: Manual configuration

You can manually create the AT-TLS policy configuration files by coding all the required statements in an HFS file or MVS data set. There are a large number of powerful configuration options provided by AT-TLS policy statements that permit advanced users to carefully fine-tune AT-TLS policy on a per-stack basis. This chapter describes the procedure for creating an AT-TLS policy by manually creating and editing the configuration files. For details on the AT-TLS policy statements, refer to *z/OS Communications Server: IP Configuration Reference*.

Option 2: Use the z/OS Network Security Configuration Assistant

IBM provides a configuration GUI that you can use to generate the Policy Agent files. The z/OS Network Security Configuration Assistant is a standalone application that runs under the Windows operating system and requires no network connectivity or setup. You can download the GUI from the Communications Server family downloadable tools Web page. Through a series of wizards and online help panels, you can use the GUI to create AT-TLS configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the GUI, there are three types of reusable objects:

- Traffic Descriptors that define the local application, by describing the TCP traffic with ports or identifying the application using its jobname.
- Security Levels that define the level of security, such as the encryption level.
- Requirement Maps that map Traffic Descriptors to Security Levels. A single Requirement Map should contain a complete set of security requirements that will govern the level of security for multiple IP traffic types.

For each TCP/IP stack, you create a set of Connectivity Rules that indicate the data endpoints and indicate which Requirement Map will govern security between the data endpoints.

The configuration GUI comes with a number of IBM-supplied Traffic Descriptors, Security Levels, and Requirement Maps that are easily applied, or you can use the IBM-supplied definitions as the basis for your own set of reusable objects.

The GUI can dramatically reduce the amount of time that is required to create AT-TLS policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security, use of the GUI is encouraged to ensure that you have a consistent and easily manageable interface for implementing AT-TLS security.

If you plan to configure Policy Agent manually, you cannot take advantage of the GUI. This chapter primarily targets those who plan to use option 1, manual configuration. However, if you are using the GUI, reading this chapter will help you understand security concepts and the relationship between Policy Agent and AT-TLS function.

AT-TLS policy configuration

AT-TLS policy is provided to the stack by the Policy Agent. The Policy Agent main configuration file contains a TcpImage statement for each stack that is to receive policy, and can optionally contain a CommonTTLSConfig statement that identifies a shared AT-TLS policy file. The TcpImage statement identifies the z/OS UNIX file or MVS data set that contains policy for that stack. This policy file can contain a TTLSConfig statement to identify the z/OS UNIX file or MVS data set that contains the AT-TLS policy. The TTLSConfig statement is required for each stack that is to receive AT-TLS policy. If both a TTLSConfig statement and a CommonTTLSConfig statement are defined, the specified CommonTTLSConfig file is processed before the TTLSConfig policy file specified for that stack.

Within the AT-TLS policy file, AT-TLS rules define a set of conditions that are compared to connections when policy is mapped during connect, or at the first select for readable or writable, poll for readable or writable, send, receive, or SIOCTTLSCTL ioctl. If a rule match is found, AT-TLS transparently provides TLS protocol control for the connection based on the security attributes specified in the actions associated with the rule.

AT-TLS rules

A TTLSRule statement consists of a set of conditions that are compared against the connection being checked. When a match is found, policy lookup stops and the connection is assigned the actions associated with the rule. The rule conditions are:

- LocalAddr - Local IP address or addresses
- RemoteAddr - Remote IP address or addresses
- LocalPortRange - Local port or ports
- RemotePortRange - Remote port or ports
- Jobname - Job name of the owning application or wildcard job name
- Userid - User ID of the owning process or wildcard user ID
- Direction - Inbound if applied to a passive socket (established by accept), Outbound if applied to an active socket (established by connect), or Both

Direction and at least one other condition must be specified. Other rule considerations include:

- If a condition is not specified, that condition is not considered when comparing the rule and the connection for a match.
- Multiple values can be specified for the IP address and port conditions, either directly in the condition or as a referenced group.
- IPv6 addresses are valid in all environments.

Each TTLSRule statement can also have a priority. Priority values can be integers in the range 1 to 255, with 255 being the highest priority. When assigning priorities, you should skip some values to allow for future rule insertion between existing rules. Policy Agent orders rules in alphabetical order within priority.

Tip: If connections can map to more than one rule, always use priority and leave priority space between rules.

AT-TLS actions

A TTLSRule statement can reference up to three actions:

- The TTLSGroupActionRef parameter includes the name of a globally defined TTLSGroupAction statement
- The TTLSEnvironmentActionRef parameter includes the name of a globally defined TTLSEnvironmentAction statement
- The TTLSConnectionActionRef parameter includes the name of a globally defined TTLSConnectionAction statement

The TTLSGroupActionRef parameter is required, and the TTLSGroupAction statement must specify TTLSEnabled ON or TTLSEnabled OFF. If TTLSEnabled OFF is specified, no additional specifications are needed. Otherwise, the AT-TLS environment action is required and the AT-TLS connection action is optional. Each action represents a scope of control.

When an AT-TLS action statement is deleted or replaced, it is considered stale. New connections will not map to a stale action. Connections that mapped to an action that later becomes stale continue to use the resources associated with the stale action until the connection closes.

AT-TLS group action

This action defines whether AT-TLS is enabled, allows trace settings, and provides the ability to set unique Language Environment variables for the Language Environment process that will be started for the action. Many TTLSRule statements can reference the same TTLSGroupAction statement. In a simple implementation of AT-TLS, you need to specify only TTLSEnabled ON on the TTLSGroupAction statement.

The AT-TLS group action represents a single Language Environment process and enclave, and initializes one instance of the System SSL DLL. Global attributes are owned by this action. Each AT-TLS group has a main task, a logging task, and a dynamic pool of pthreads that handle System SSL secure environment and secure connection management. When a stale group has no remaining connections, the pthreads, tasks, and Language Environment process are removed.

Guideline: Use as few AT-TLS group actions as necessary.

AT-TLS environment action

This action requires a key ring name (either RACF or gskkyman format), and the handshake role (client or server) this half of the connection will assume. Cipher suites and trace settings can also be set. There are several advanced parameters available if needed, but in a simple implementation of AT-TLS, you need to specify only a key ring and the handshake role.

The AT-TLS environment action is used to create System SSL environments. A key ring and SSL Session ID (SID) cache are examples of attributes owned by a System SSL environment. The AT-TLS environment action initializes a System SSL environment within the Language Environment process that was created to represent an AT-TLS group action. Several System SSL environments can exist within a single AT-TLS group. The same TTLSEnvironmentAction statement can be used to create similar System SSL environments in the same or different groups. AT-TLS dynamically creates instances of System SSL environments as needed. AT-TLS deletes System SSL environments when they have had no connections using them for a period of ten to twenty minutes. If the TTLSEnvironmentAction statement used to create a System SSL environment becomes stale, AT-TLS deletes the environment when it has no remaining connections. Connections associated with the same server application or same client user ID can share a System SSL environment. Connections that share an existing System SSL environment avoid

the processing required to initialize an environment, such as opening a key ring. Connections between the same partners can also reuse recent session information in the SID cache, allowing them to use the SSL short handshake that requires less processing. System SSL connection resources are released when the connection closes.

AT-TLS connection action

The AT-TLS connection action represents attributes at the connection level. This action is optional, and is needed only when a subset of connections within an AT-TLS environment must have different parameters. Handshake role, security version, cipher suites, and tracing are examples of attributes that can be changed at the connection level. In a simple implementation of AT-TLS, you do not need to specify this action.

System SSL connections are initialized within a System SSL environment. Use the AT-TLS connection action to override attributes specified at the SSL environment layer. System SSL connection resources are released when the connection closes.

Getting started with AT-TLS

Assume you have a TCP client and server application pair running on z/OS platforms. This application handles sensitive data, and you want this application to be used only with the TLS protocol. The server application runs under the job name of XYZSRV, and creates a passive TCP socket bound to IP address INADDR_ANY and port 5000. The client application runs as a command, issued by TSO or z/OS UNIX interactive users, and connects to port 5000.

To complete AT-TLS security setup for this sample environment, you need to create both server and client key rings. The server key ring needs to contain a server certificate, and any certificates used to sign it. The server needs access to the private keys of the server certificate. The client key ring needs the root certificate used to sign the server certificates. For a TLS/SSL primer and some step-by-step examples, see Appendix B, "TLS/SSL security," on page 1167. For more information on managing key rings and certificates with RACF and the RACDCERT command, refer to *z/OS Security Server RACF Security Administrator's Guide*. For detailed information on managing key rings and certificates with gskkyman, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*.

Configuring the server system

On each z/OS system where you run the server application, see Table 38 for the tasks needed to configure the server.

Table 38. AT-TLS configuration for the server system

Task	Specification
Create key ring	Create server key ring with server certificate and necessary certificate authority certificates.
Create Policy Agent files	1. Create a Policy Agent main configuration file containing a TcpImage statement for the server stack. 2. Create a Policy Agent image configuration file for the server stack.
Add AT-TLS configuration	Add a TTLSConfig statement to the Policy Agent image configuration file, identifying the TTLSConfig policy file location: `TTLSConfig` *`serverpath`*

Table 38. AT-TLS configuration for the server system (continued)

Task	Specification

Add statements to the TTLSConfig policy file	Add the AT-TLS policy statements to the *serverpath* file:
	`TTLSRule XYZServerRule`
	`{`
	`  LocalPortRange 5000`
	`  JobName XYZSRV`
	`  Direction Inbound`
	`  TTLSGroupActionRef XYZGroup`
	`  TTLSEnvironmentActionRef XYZServerEnvironment`
	`}`
	`TTLSGroupAction XYZGroup`
	`{`
	`  TTLSEnabled On`
	`}`
	`TTLSEnvironmentAction XYZServerEnvironment`
	`{`
	`  TTLSKeyRingParms`
	`  {`
	`  Keyring server_key_ring`
	`  }`
	`  HandshakeRole SERVER`
	`  Trace 7`
	`}`

Set up InitStack access control	1. Define the EZB.INITSTACK.*sysname.tcpname* profile for each AT-TLS stack. 2. Permit administrative applications to use the stack before AT-TLS is initialized. For examples of the security product commands needed to create this resource profile name and grant users access to it, refer to member EZARACF in sample data set SEZAINST.
Enable AT-TLS	Set TCPCONFIG TTLS in PROFILE.TCPIP.

Configuring the client systems

On each z/OS system where you run the client application, see Table 39 for the tasks needed to configure the client.

Table 39. AT-TLS configuration for the client system

Task	Specification
Create key ring	Create a client key ring for each client with necessary certificate authority certificates. If using client authentication, also attach each client's certificate to their key ring. **Tip:** To simplify AT-TLS policy, use the same RACF key ring name for every client. SystemSSL will qualify the key ring name with the current UserID when accessed.
Create Policy Agent files	1. Create a Policy Agent main configuration file containing a TcpImage statement for the client stack. 2. Create a Policy Agent image configuration file for the client stack.
Add AT-TLS configuration	Add a TTLSConfig statement to the Policy Agent image configuration file, identifying the TTLSConfig policy file location: `TTLSConfig clientpath`

Table 39. AT-TLS configuration for the client system (continued)

Task	Specification

Add statements to the TTLSConfig policy file	Add the AT-TLS policy statements to the *clientpath* file:
	`TTLSRule XYZClientRule`
	`{`
	`  RemotePortRange 5000`
	`  Direction Outbound`
	`  TTLSGroupActionRef XYZGroup`
	`  TTLSEnvironmentActionRef XYZClientEnvironment`
	`}`
	`TTLSGroupAction XYZGroup`
	`{`
	`  TTLSEnabled On`
	`}`
	`TTLSEnvironmentAction XYZClientEnvironment`
	`{`
	`  TTLSKeyRingParms`
	`  {`
	`  Keyring client_key_ring`
	`  }`
	`  HandshakeRole CLIENT`
	`  Trace 7`
	`}`

Set up InitStack access control	1. Define the EZB.INITSTACK.*sysname*.*tcpname* profile for each AT-TLS stack. 2. Permit administrative applications to use the stack before AT-TLS is initialized. For examples of the security product commands needed to create this resource profile name and grant users access to it, refer to member EZARACF in sample data set SEZAINST.
Enable AT-TLS	Set TCPCONFIG TTLS in PROFILE.TCPIP.

Steps for starting AT-TLS and verifying its operation

You are now ready to start the sample AT-TLS environment and verify its operation.

Before you begin:

1. Perform the tasks in Table 38 on page 992 and Table 39 on page 993.
2. Review your syslogd configuration to verify that messages written by Policy Agent and TCP/IP stacks are saved in the desired files. AT-TLS syslogd messages are written to the daemon facility by default.
3. Start syslogd.

Perform the following steps to start AT-TLS and verify its operation:

1. Start the TCP/IP stacks.
2. Start the administrative applications required for successful Policy Agent execution, such as OMPROUTE, LDAP, and the name server.
3. Start Policy Agent on all participating systems and verify that there were no policy errors in processing the policy files.
4. Verify that the participating TCP/IP stacks have received AT-TLS policy and released console message EZZ4248E.

5. Start server application and verify that it starts without errors.
6. Start client applications. Review the AT-TLS trace messages in the syslogd output on both the client and server systems. Verify that connections are mapping to the intended policy and no handshake errors occur. The info messages `EZD1281I TTLS Map` and `EZD1283I TTLS Initial Handshake` show the policy used and result of TLS handshake negotiation. The error message `EZD1286I TTLS Error` shows any failures.

For information on common AT-TLS startup errors, refer to *z/OS Communications Server: IP Diagnosis Guide*.

Application compatibility with AT-TLS

Most applications can use AT-TLS. However, some applications should not be configured to use AT-TLS. Any application that is already configured to use SSL or TLS protocols should not use AT-TLS. Use of AT-TLS would result in encrypting data that is already encrypted. The receiving partner would not be able to decipher the doubly-encrypted data. If the application can be configured to use clear text, AT-TLS can then provide support.

TCP/IP applications that already support the TLS protocol include:

- TN3270E Telnet server
- FTP server
- FTP client
- Sendmail
- DCAS (Express Logon server)
- Bind 9.*x*, which uses open SSL

AT-TLS does not support Web servers using the Fast Response Cache Accelerator (FRCA) support in TCP/IP. AT-TLS ignores policy for connections using FRCA. The sockets are treated as if they did not match any AT-TLS rules.

AT-TLS does not support applications that use the Pascal sockets API. AT-TLS ignores all Pascal sockets. The sockets are treated as if they did not match any AT-TLS rules. TCP/IP applications that use the Pascal API include:

- TSO TN3270E Telnet client
- SMTP server
- LPD server

Policy considerations

Policy is a powerful tool for configuring and managing your applications that use AT-TLS. Before configuring the AT-TLS policy, review the general syntax rules in the Policy Agent and policy applications chapter of *z/OS Communications Server: IP Configuration Reference*.

Reusable objects

With several of the AT-TLS rule conditions and action parameters, you can reference named objects. If you are going to use the same definition in several rules or actions, it is easier to create a single named object and refer to it, rather than repeating the definition. This also makes changing these definitions easier and more accurate. For example, the IpAddrGroup statement can be used to identify

groups of IP addresses that are used in several rules. You might find it useful to define TTLSRule statements that reference an IP address group with a name, such as LocalHost. In each stack's AT-TLS policy file, you would define that IP address group with all of the addresses local to that stack. That single reference can then be used throughout the policy to easily represent all local IP addresses, without re-coding the local addresses in each rule condition.

CommonTTLSConfig file

This file should contain all of the policy that is common to multiple stacks. When the policy agent encounters a TTLSConfig statement inside the policy file for a given stack, the contents of the CommonTTLSConfig file are logically added before the contents of the TTLSConfig file. If the policy file for a given stack does not have a TTLSConfig statement, the CommonTTLSConfig file is not processed for that stack. Rules and actions in the CommonTTLSConfig file can reference objects, such as a local IpAddrGroup statement, that are defined in the TTLSConfig file. Rules and actions in the TTLSConfig file can also reference objects that are defined in the CommonTTLSConfig file.

If Policy Agent encounters multiple objects of the same type and name, the last occurrence is the one that is used. You can take advantage of this by defining an object that is used on many stacks in the CommonTTLSConfig file, and then overriding it in the TTLSConfig file of a specific stack.

Exempting specific connections from AT-TLS

In some cases, you might want to have the majority of users of an application using AT-TLS, but then exempt a small group of users. You might find that it is simpler to define a broad AT-TLS policy for the application, and then define a higher priority rule for the exempt users. The simplest way to do this is to define a TTLSGroupAction statement with the TTLSEnabled parameter set to OFF. The exempt user rule would then reference this action. When the TTLSGroupAction statement specifies the TTLSEnabled parameter as OFF, the TTLSEnvironmentActionRef parameter on the corresponding rule is not required.

Action refresh

When Policy Agent is stopped and restarted, or when policy files are changed, policy objects that are currently in use might be deleted or replaced. When an AT-TLS action is deleted or replaced, connections using the old object continue processing without change. Connections that search AT-TLS policy after the change use the new action objects. A change to an AT-TLS group action causes a new Language Environment process to be created, along with new SSL environments for each user or application environment associated with that group. A change to an AT-TLS environment action causes new SSL environments to be initialized within each group with which that environment is associated. System SSL reopens key rings and certificates, and creates an empty session ID cache when it initializes an SSL environment.

There are cases when a change is made that is not reflected by a change in the action. For example, the default certificate in a key ring might change. The key ring name has not changed, but there is a need to open a new environment. Simply refreshing policy will not refresh the AT-TLS environment action in AT-TLS, because no values within the action have changed. To force a refresh in AT-TLS, some parameter must be changed. The EnvironmentUserInstance parameter can be used for this purpose. Incrementing the instance number forces a refresh of AT-TLS without changing any of the security parameters. Similarly,

changes to the contents of the environment file named in a group action will not be applied until the group action is changed. The GroupUserInstance parameter can be used to force an AT-TLS refresh of the group, creating a new Language Environment process using the new environment file contents.

Common setup considerations

Handshake role

Every secure connection must have one end using the HandshakeRole Client and the other end using the HandShakeRole Server or ServerWithClientAuth. In the SSL and TLS protocols, the client side sends a ClientHello handshake record and the server side responds with a ServerHello handshake record. If both sides send ClientHello records, the handshake will fail. If both sides wait for a ClientHello, the handshake will time out.

The role played in handshake is independent of whether the application's usage is that of client, server, or peer. It is also independent of which application does the listen() and accept() and which does the connect(). The primary consideration for deciding which role an application should play is certificate management.

- An application designated as HandshakeRole Server must have a user or site certificate on its key ring. It must have access to the private keys of its user or site certificate. It might also need other certificates on its key ring, required for authentication.

  The partner application designated as HandshakeRole Client must have a key ring that contains the root certificates required to authenticate the server's certificate. The client does not need access to the private keys of any certificate. Clients can share this key ring.
- An application designated as HandshakeRole ServerWithClientAuth might need additional root certificates on its key ring to authenticate client certificates. The server decides whether a client must present a user certificate and what constitutes an acceptable certificate.

  The partner application designated as HandshakeRole Client decides whether to present a user certificate when challenged. To present a user certificate, it must have a private key ring with that user certificate. It must have access to the private keys of the user certificate. The client key ring must also contain the root certificates required to authenticate the server's certificate.

Key ring specification

AT-TLS configuration can be simplified by using key rings with a common name stored in RACF. The SAF key ring name is specified as *userid/keyring*. The current user ID is used if only the key ring name is specified. A value that begins with a forward slash is interpreted as the path name of a gskkyman key ring.

Protocol versions

SSLv3 and TLSv1 provide stronger security than SSLv2. Support for SSLv2 is off by default in AT-TLS policy. You should turn it on only if you must support older applications that do not support the newer protocols.

Cipher suite specification

The set of SSL protocol cipher specifications to be allowed for the secure session can be set. You should not include any that you do not want to allow. Order is important. System SSL selects ciphers according to the server's order of usage

preference. The first cipher in the server's list that is also in the client's list is selected. Other implementations might work differently.

AT-TLS does not pass any cipher suites to System SSL by default. For the list of cipher suites supported and the default order used if none is specified, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*.

Trace options

The Trace value is interpreted by AT-TLS as a bit map. Each of the options is assigned a value that is a power of 2. You should add together the values of each option that you want to activate.

The default Trace value is 2, which provides error messages to syslogd. While you are deploying a new policy, you might find it beneficial to specify a Trace value of 6 or 7. This provides connection info messages, in addition to error messages in syslogd. The info messages provide positive feedback that connections are mapping to the intended policy.

Trace options event (8), flow (16), and data (32) are intended primarily for diagnosing problems. Trace values larger than 7 can cause a large number of trace records to be dropped instead of being sent to syslogd.

Tip: Use a TTLSConnectionAction with a higher Trace value to diagnose problems in a production environment. You can temporarily define a high priority TTLSRule with conditions that cover only a small number of problem connections. This temporary rule can reference the same TTLSGroupAction and TTLSEnvironmentAction that your production rule references, and a TTLSConnectionAction with the Trace level you want for diagnosis.

Action user instance

Use these fields to track policy changes. You can also use them to force a policy change to be recognized when no other fields require change, such as when a new certificate has been installed in a key ring.

Handshake timer

Certain configuration or application protocol mismatches can lead to stalled connections. The TLS handshake protocol always expects the end of the connection configured as HandshakeRole Client to send the first message. One connection stall scenario results when both ends of the connection are configured as HandshakeRole Server and wait for a ClientHello record from the other end. Another connection stall scenario results when the end of the connection that normally sends the first application data is configured as HandshakeRole Server, but the partner application is not configured to use a secure connection. The HandshakeRole Server end is waiting for a ClientHello, and the nonsecure end is waiting for application data.

You can use the HandshakeTimer action parameter to control the time that AT-TLS should wait during TLS handshake negotiation before resetting the connection. AT-TLS times two different handshake intervals, handshake start and handshake completion. The handshake start interval is intended to detect configuration problems that result in neither partner sending data. The handshake completion interval is intended to detect problems that might stall a handshake in one of the TLS protocol implementations.

The handshake start interval begins when AT-TLS is ready to begin a TLS handshake, and ends when the hello handshake record is received from the partner. On the initiating or active side of the connection, the handshake start interval used is five times the specified HandshakeTimer value, because it includes:

- The network time for the ClientHello record to reach the partner if HandshakeRole is Client.
- The time a connection spends on the partner's listen backlog.
- The time before the partner causes the connection to be mapped.
- The time spent on the partner AT-TLS work queue.
- The time spent by the partner initializing a new System SSL environment, if necessary.
- The network time for the partner's ServerHello or ClientHello record to be returned.

On the listening or passive side of the connection, the handshake start interval used is the specified HandshakeTimer value, because it includes only the network time for one or both hello records. Handshake start interval timeouts result in AT-TLS return code 5004 and a connection reset.

The handshake completion interval begins when the hello handshake record is received from the partner, and ends when the System SSL gsk_secure_connection_init() service returns to AT-TLS. The handshake completion interval used is the specified HandshakeTimer value on either active or passive connections. Handshake completion interval timeouts result in AT-TLS return code 5005 and a connection reset.

The HandshakeTimer action parameter has a default value of 10 seconds. If you determine that you are getting handshake time-outs that are caused by network delays or application workload rather than configuration or application errors, you should increase the value. On the other hand, if you determine that handshakes normally complete much faster in your environment and you would like to detect the occasional incorrectly configured partner more quickly, you can decrease the value.

Client authentication

When the HandshakeRole parameter is set to ServerWithClientAuth, a certificate request is sent to the client during the handshake. The client can send its certificate to the server, which can then validate the certificate. The level of validation done by the server is controlled by the setting of the ClientAuthType parameter. The following chart summarizes the differences between the parameter settings:

Table 40. ClientAuthType parameter settings

ClientAuthType	Client certificate required or optional	Certificate validation
PassThru	Optional	None
Full	Optional	Certificate is validated against keyring, if provided
Required	Required	Certificate is validated against keyring
SAFCheck	Required	Certificate is validated against keyring and must be associated with a user ID in the security product

If the certificate fails validation, the secure connection does not initialize successfully. The default setting for ClientAuthType is Required. For applications that do not issue AT-TLS ioctl calls to obtain the certificate or user ID of the client, the Required setting ensures that any client that connects provides a valid client certificate. The ClientAuthType setting of PassThru should only be used for applications that will get the certificate from AT-TLS using the SIOCTTLSCTL ioctl call and implement their own security methods to validate the certificate. The ClientAuthType setting of Full can be used for applications that want to validate the client certificate if provided, but do not require a client certificate to establish a secure connection. The ClientAuthType setting of SAFCheck provides an additional level of security when using client authentication by requiring AT-TLS to derive a user ID associated with the certificate. An application can then get the user ID from AT-TLS using the SIOCTTLSCTL ioctl after the secure connection has been established. The user ID can be used to either verify the user ID presented by the client during the application's protocol flows or eliminate the need for a user ID to be sent by the client.

AT-TLS access control considerations

Access to key rings and certificates is verified by System SSL when SSL environments are initialized. Access to certificate private keys is verified by ICSF when asymetric encryption services are requested that require the private keys. AT-TLS invokes System SSL services that cause these access control checks to occur on tasks created in the TCP/IP address space. TCP/IP replicates the security environment of the user running the application that owns the socket at the time AT-TLS policy is mapped, before invoking these System SSL services.

Several common application models were considered to determine the most appropriate time for replicating the security environment. Replication occurs when AT-TLS policy is mapped. Policy mapping occurs during processing of the first occurrence of connect, a SIOCTTLSCTL IOCTL, select for socket readable or writable, poll for socket readable or writable, or call that sends or receives data over the socket. This defers security environment replication for applications such as INETD until after the accept(), fork(), setuid(), and exec() sequence of services has established the server application process.

In the CICS socket environment, transaction security environments are not visible to AT-TLS support. The CICS job and all of its transactions appear to the stack as a single server application with a single z/OS UNIX process ID running in the security environment of the CICS job. All AT-TLS policy lookups, System SSL key ring authorization checks, and ICSF private key authorization checks are processed using the identity of the CICS job. Connections established, whether active or passive, can perform TLS handshake processing as either the client or server. All of the connections established by a single CICS job are able to share the session ID cache in the SSL environment. The CICS job should use a private key ring with a server certificate. The key ring used must contain the chain of root certificates needed to validate the server certificate it presents to the client. If the server requires client authentication, it must also have any other root certificates necessary to validate client certificates presented on its key ring.

Application model considerations

AT-TLS support provides for several typical socket application models. Socket applications with significantly different models might not benefit from AT-TLS.

Client application model

As shown in Figure 97, this type of application runs entirely within the security environment of a single user. Many users might use the application, but each usage is independent, runs in a separate process, and should not share System SSL information with other processes. All socket calls for each usage of the application are made from the same z/OS UNIX process. Most connections are active connections initiated with the connect() service by this application to a server. Some client applications, such as Web browsers, repeatedly connect to servers at the same or different IP addresses.

Some client applications, such as FTP or REXEC, support a second parallel connection with the server. This is often a passive connection, established by binding to an ephemeral port and listening for a single connection back from the server's IP address. For a description of an alternative method of mapping policy for these special cases, see "Secondary connection application model" on page 1005.

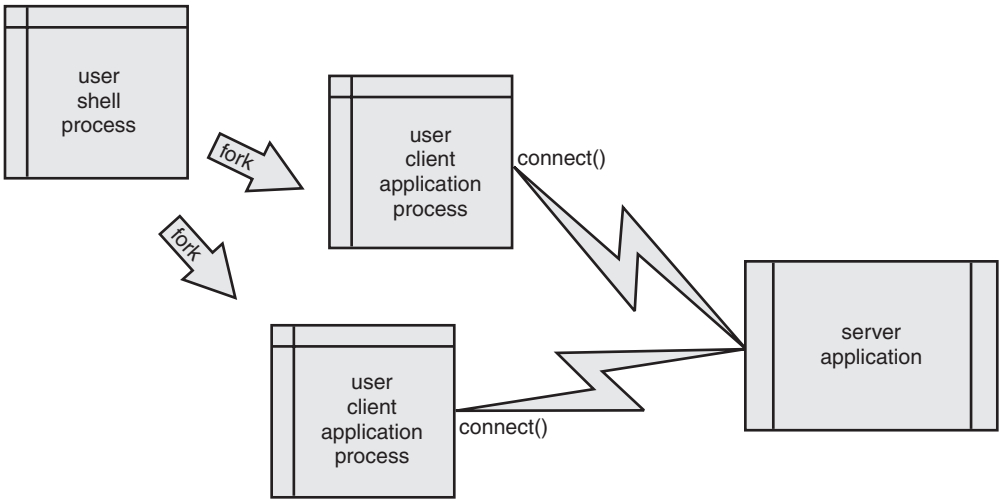


Figure 97. Client application model

All connections established by this application, whether active or passive, perform TLS handshake processing as a client. All of the connections established by a single process under the same user ID should be able to share the session ID cache in the SSL environment. If the server does not require client authentication, this client can use a shared key ring containing only the root certificates needed to validate the server's certificate. If the server does require client authentication, each user ID needs to own a key ring that contains that user's certificate, in addition to the necessary root certificates. RACF key ring names are qualified with the owning user IDand are specified as *userid/keyring*. If the user ID is not explicitly specified on the AT-TLS key ring parameter, the current user ID is used. Therefore, key ring names can be the same across different user IDs . If the same key ring name is used by all client user IDs, a single TTLSEnvironmentAction keyring parameter can represent all clients required to support client authentication. AT-TLS policy administration is simplified if the individual key rings all have the same key ring name.

Server application model

As shown in Figure 98 on page 1002, this type of application runs entirely within a single z/OS UNIX process. Connections can be either passive connections returned

from a listening socket by the accept() service, or active connections initiated with the connect() service. Communication partners can be client applications or peer servers. Connections can be processed by subtasks or pthreads within the server process. The initial read or write of data on the connection is done under the primary security environment of the server process. Some server applications allow a client to log in with a user ID on the server system and can place this client-specified identity on the subtask or pthread used to access resources on behalf of the client. The user ID associated with the server is used for AT-TLS purposes, regardless of this ability to change to the client-specified identity.

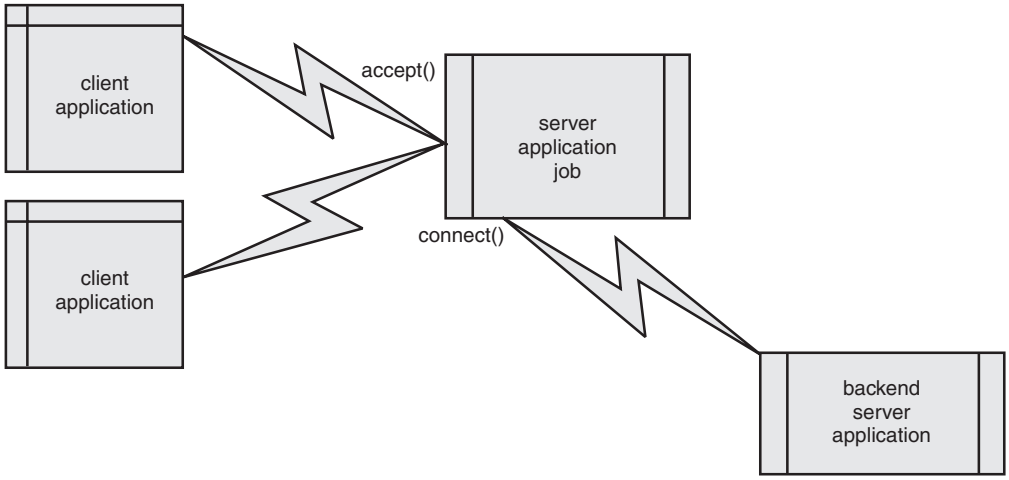


Figure 98. Server application model

Connections established by this application, whether active or passive, can perform TLS handshake processing as either a client or server. All of the connections established by a single process, under the same user ID and performing the handshake as a client or as a server, should be able to share a session ID cache in the SSL environment. The server process should use a private key ring with a server certificate. The key ring used must contain the chain of root certificates needed to validate the server certificate it presents to the client. If the server requires client authentication, it must also have any other root certificates necessary to validate the client certificates presented on its key ring.

Forked server application model

As shown in Figure 99 on page 1003, this type of application is forked by a daemon application, such as INETD, to handle a single passive connection to a socket that the daemon is listening on. The daemon invokes the bind(), listen(), and accept() services on the parent socket. It then forks a new process to handle each child connection, optionally changes the new process to a different identity, and optionally execs a configured server application. The server application reads and writes data on the child connection.

Some server applications support a second parallel connection with the client. This is often an active connection, established by connecting back to an ephemeral port opened by the client at the client's IP address. For a description of an alternative method of mapping policy for these special cases, see "Secondary connection application model" on page 1005.

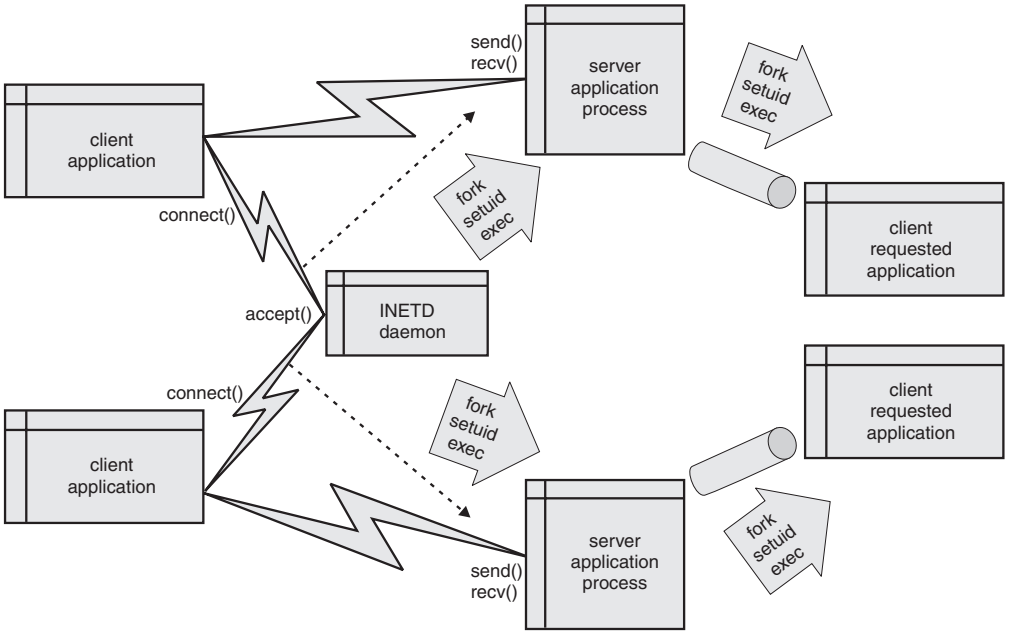


Figure 99. Forked server application model

Many server applications allow the client to log in with a user ID on the server system. Some server applications change the security environment of the server process to the client-specified identity prior to accessing resources on behalf of the client. Other server applications set up another communications path (pseudo terminal, named pipe, UNIX domain socket, and so on) and then fork an additional client process. The server changes the security environment in this client process to the client-specified identity, and then execs a client-specified program or login shell.

The initial child connection transferred by the daemon to the server process, and any additional connections established by the server process, perform TLS handshaking as the server, optionally requiring client authentication. Each forked server process runs under the same user ID. All of the server processes forked to handle child connections of the same parent connection should be able to share the session ID cache in the SSL environment. The server processes can use a shared key ring with a site certificate or a private key ring with a server certificate. The key ring used must contain the chain of root certificates needed to validate the site or server certificate it presents to the client. If the server requires client authentication, it must also have any other root certificates necessary to validate client certificates presented on its key ring.

Any client process forked by the server process is treated as a new application by AT-TLS. New connections established will not share an SSL environment with the server process that created them.

CICS transaction model

TCP/IP CICS socket support provides a CICS resource manager that invokes z/OS UNIX socket services. It also provides a CICS Listener transaction that accepts passive connections from a listening socket. Each listening socket has a configured transaction that is launched to process a single established connection.

For more information on configuring TCP/IP CICS socket support, refer to *z/OS Communications Server: IP CICS Sockets Guide*.

Advanced application considerations

Some applications need to be aware of when a secure connection is being used or examine the certificate presented by the partner. Other applications need to control if and when the TLS handshake occurs. These applications typically support both TLS and non-TLS connections over the same port. They define an application protocol for negotiating whether to use TLS and when to begin. In both cases, these applications need to be aware that TLS is being used on the connection. However, you might not want to, or might not be able to, use any SSL toolkits in the application. AT-TLS support provides the SIOCTTLSCTL ioctl commands that can be used in these situations.

Some applications establish a second connection using ephemeral ports or after the server has changed to a client supplied identity. These secondary connections need to be associated with the policy and security environment used on the primary connection. AT-TLS provides special support for these applications.

AT-TLS aware application considerations

Applications that need to examine the partner's certificate can issue the SIOCTTLSCTL IOCTL with request type TTLS_RETURN_CERTIFICATE to get the certificate at any time during a secure connection. Applications that are running under a policy with the HandshakeRole parameter set to CLIENT receive the server's certificate. Applications that are running under a policy with the HandshakeRole parameter set to ServerWithClientAuth receive the client's certificate if provided.

Applications configured as HandshakeRole ServerWithClientAuth that need to examine or use the user ID associated with the certificate in SAF can issue the SIOCTTLSCTL ioctl with request type TTLS_QUERY_ONLY or TTLS_RETURN_CERTIFICATE. If a partner certificate is available on the secure connection, AT-TLS uses a RACF service to extract the associated user ID. If no client certificate is available, or no user ID has been associated, the ioctl returns zero as the associated user ID length.

AT-TLS controlling application considerations

Applications that need to control AT-TLS behavior, using the SIOCTTLSCTL IOCTL with the TTLS_INIT_CONNECTION, TTLS_RESET_SESSION, or TTLS_RESET_CIPHER request flags, must have the ApplicationControlled parameter set to ON in their TTLSEnvironmentAdvancedParms or TTLSConnectionAdvancedParms statement. This causes AT-TLS to postpone the TLS handshake. After the connection is established, the application can issue the SIOCTTLSCTL IOCTL to get the current AT-TLS connection status and determine whether or not AT-TLS support is available on this connection. When the application is ready for AT-TLS to perform the TLS handshake, it issues the SIOCTTLSCTL IOCTL with request type TTLS_INIT_CONNECTION. The SIOCTTLSCTL IOCTL initiates an AT-TLS policy lookup, if one has not yet been done, and assigns a rule and actions to the connection if a match is found. For more IOCTL information, refer to *z/OS Communications Server: IP Programmer's Guide and Reference*.

Some application protocols provide a way for the client and server programs to negotiate whether to use the TLS or SSL protocol to protect data on the connection. This cleartext negotiation typically occurs very early in the connection. When both partners agree to use TLS, they initiate the handshake. These applications can take advantage of AT-TLS. The policy must indicate ApplicationControlled ON. After

the connection is established, the application can use the SIOCTTLSCTL IOCTL to determine whether AT-TLS support is configured, policy covers the connection, and the policy specifies ApplicationControlled ON. The application can then send and receive cleartext data to negotiate the use of TLS. When both partners agree, they each must initiate a secure connection. The AT-TLS application can use the SIOCTTLSCTL ioctl with TTLS_INIT_CONNECTION to initialize the secure connection. AT-TLS performs the initial handshake and provides encryption and decryption services for the application. The application can simply send and receive cleartext over the socket as it would if it were not a secure connection.

The application can also use the SIOCTTLSCTL IOCTL to reset the cipher or the session. Resetting the cipher causes AT-TLS to initiate another handshake. If the session is reset first, the subsequent handshake is a full handshake. If the session has not been reset, a short handshake is attempted for the subsequent handshake. The partner application must agree to the short handshake.

Secondary connection application model

Some applications create two connections between the client and server programs. These applications typically have a single primary connection that is bound to a well-known port on the server side. After exchanging some information, a second connection is established. This second connection often uses dynamically assigned ports on both ends. Examples of this behavior include the stderr connection in the rsh, rexec, and rlogin family of applications, as well as the firewall-friendly FTP data connection. It is often not possible to define a set of policy rule conditions to correctly map these secondary connections on the client side or when the server forks a new process, with a dynamic job name, for each connected client.

In other cases, this second connection is established after the server has changed to a client-supplied identity. Mapping this second connection to AT-TLS policy as a new and independent connection would force the use of a different System SSL environment. The client-supplied identity would need to have access to the certificate private keys. Normal FTP data connections are an example of this behavior.

AT-TLS provides an alternate method of mapping policy for these secondary connections. This alternate method causes the secondary connection to share the System SSL environment and security environment of the associated primary connection.

To activate the alternate policy mapping method, define a policy rule using conditions that will map the primary connection. In this policy, specify the SecondaryMap parameter with a value of ON. When this policy is mapped to a primary connection, an entry is made in an internal table. Future connections do a normal policy lookup, and then look in the internal table for an entry with the same process ID and pair of IP addresses. If a matching entry is found and the new connection has no mapped policy, or has a mapped policy with a lower priority than the matching entry, the new connection is marked as a secondary connection and uses the same policy and user ID as the primary connection.

You should use this alternate policy mapping method only for client applications and server applications that have a single primary connection. Careful consideration should be given before using it for non-forking server applications that accept multiple primary connections, such as MVRSHD (TCP/IP's combined rsh and rexec server for the TSO environment). The alternative method of policy mapping always associates secondary connections with the most recent primary

connection mapped by this process. When the process establishes multiple primary connections, the alternate mapping method is not able to reliably associate secondary connections with the correct primary connection. You should not use this alternate policy mapping method when the primary connections can map to different policies based on client IP address or multiple server listening port numbers. You should use normal policy mapping with a job name condition for the secondary connections of non-forking servers.

Advanced policy parameters

Advanced policy parameters are as follows:

SSLv2, SSLv3, and TLSv1
The SSLv2 protocol is supported but is not as secure as the SSLv3 or TLSv1 protocols. The AT-TLS default is to enable the SSLv3 and TLSv1 protocols but to disable the SSLv2 protocol. You should turn on SSLv2 only if you must support connections with older applications that do not support the more secure protocols.

ApplicationControlled
Some application protocols provide a way for the client and server programs to negotiate whether to use the TLS or SSL protocol to protect data on the connection. These applications can be modified to take advantage of AT-TLS. The policy must indicate ApplicationControlled ON. For more information, see "AT-TLS controlling application considerations" on page 1004.

HandshakeTimeout
This parameter controls the amount of time AT-TLS will wait during the SSL handshake. Timeouts can be caused by configuration mismatches or long delays in initial connection processing by applications. For more information, see "Handshake timer" on page 998.

ResetCipherTimer
The SSLv3 and TLSv1 protocols allow the encryption key to be renegotiated during a secure connection. This can provide a higher level of security for long running connections. The AT-TLS default is to not reset the cipher. You can specify a time interval to cause AT-TLS to request a reset of the cipher in the range of 1 to 1440 seconds. The cipher reset will be requested when the timer expires and the next application read or write completes. The time interval is restarted when the cipher has been changed. Both ends of the secure connection must agree to perform a re-handshake to renegotiate the cipher. The HandshakeRole Client end must initiate the re-handshake. The HandshakeRole Server can send an alert to the client requesting a re-handshake. The client is free to ignore or postpone the request. The server is free to refuse a re-handshake request sent by the client.

CertificateLabel
SystemSSL normally uses the certificate designated

as the key ring's default. Specify this parameter to explicitly identify the name of the certificate you want to use.

ClientAuthType
For policies that specify HandshakeRole ServerWithClientAuth, this parameter specifies what actions AT-TLS should take to authenticate the client certificate. For more information, see "Client authentication" on page 999.

SecondaryMap
This parameter is used only for applications that create secondary connections and meet certain criteria. For more information see "Secondary connection application model" on page 1005.

EnvFile
This parameter specifies the name of a file containing environment variable assignments that modify the behavior of z/OS Language Environment or System SSL. This parameter should normally be used only under the direction of IBM Service.

CtraceClearText
Applications that implement SSL or TLS can control whether non-encrypted application data is included in diagnostic traces. Lower layers have access to only encrypted data. When using AT-TLS, the TCP, PFS, and SOCKAPI layers have access to non-encrypted data. The AT-TLS default is to suppress this data in CTRACE records generated by these layers to protect the application's users. If you need to see this data in these records to diagnose a problem, you can set CtraceClearText ON.

SyslogFacility
This parameter specifies the name of the syslogd facility to which an AT-TLS group should write its messages. The AT-TLS default is to write syslogd messages to the daemon facility. Other TCP/IP functions, such as the SNMP TCP/IP subagent, also specify the daemon facility name when writing records to syslogd. The job name and syslog facility name are the same. Filters cannot be used to direct the records to different output files. If you want AT-TLS records to go to a different output file, you can change the syslog facility name in the TTLSGroupAction statement to direct the messages from that group to the Auth facility instead. You can then set up filtering based on the jobname and facility in the syslogd configuration file to direct AT-TLS records to a different output file.

GSK_LDAP_SERVER, GSK_LDAP_SERVER_PORT, GSK_LDAP_USER, GSK_LDAP_USER_PW, GSK_CRL_CACHE_TIMEOUT
Applications using HandshakeRole ServerWithClientAuth can optionally use a Certificate Revocation List (CRL) service. This service is provided by an LDAP server. This set of parameters is used to configure System SSL so that

it can contact the CRL service. Connections used by System SSL to contact the CRL service should not fall under an enabled AT-TLS policy because these connections can be made before AT-TLS policy has been installed.

GSK_SYSPLEX_SIDCACHE
Applications can use AT-TLS with Sysplex Distributor. Setting this parameter ON instructs System SSL to make the SID cache available across the sysplex. The server applications on different system images can share session information so that clients can request session reuse on subsequent connections.

GSK_V2_SIDCACHE_SIZE, GSK_V2_SESSION_TIMEOUT, GSK_V3_SIDCACHE_SIZE, GSK_V3_SESSION_TIMEOUT
Use these parameters to configure the System SSL SID cache in your application environment. The size parameters control how many different partners session information is cached for. The timeout parameters control how long the session information should be kept in the cache. When session information is found in the cache, connections can use the SSL short handshake that requires less processing.

Chapter 19. z/OS Load Balancing Advisor

The z/OS Load Balancing Advisor communicates with external load balancers and one or more Load Balancing Agents. The main function of the Load Balancing Advisor is to provide external TCP/IP load balancing solutions, such as the Cisco Content Switching Module (CSM), with recommendations on which TCP/IP applications and target z/OS systems within a z/OS sysplex are best equipped to handle new TCP/IP workload requests. These recommendations can then be used by the load balancer to determine how to route new requests to the target applications and systems (that is, how many requests should be routed to each target). The recommendations provided by the Advisor are dynamic, and can change as the conditions of the target systems and applications change. The recommendations include several key components:

- State of the target application and system

  This includes an indication of whether the target application and target system is currently active. This enables the load balancer to exclude systems that are not active or do not have the desired application running.
- z/OS Workload Management (WLM) system-wide recommendations

  WLM recommendations provide a relative measure of a target system's ability to handle new workload, as compared to other target systems in the sysplex. The WLM recommendations are derived using several measures, including each system's available CPU capacity, or capacity that can be displaced by higher importance workloads. The latter is important for scenarios where systems might be 100% utilized, but some might be running work that has a lower importance (as defined by the WLM policy) and can therefore be displaced by higher importance workload.
- z/OS WLM server-specific recommendations

  These recommendations are similar to the WLM system-wide recommendations, but also contain an indication of how well individual server applications are doing compared to the WLM policy goals that have been specified for that workload. These recommendations can be very useful in helping the load balancers avoid selecting application servers that are experiencing performance problems (that is, not meeting the specified WLM policy goals).
- Application server health from a TCP/IP perspective

  TCP/IP statistics for target applications are monitored to determine whether specific server applications are encountering problems keeping up with the current workload. For example, is a target TCP server application keeping up with TCP connection requests? Or are requests being rejected because the backlog queue is full? In scenarios where this occurs, the recommendations passed back to the load balancers are adjusted appropriately, so that the load balancer can direct fewer connections to any application that is experiencing these problems. These recommendations are provided for both UDP and TCP server applications. These recommendations are referred to as Communication Server weights in this chapter.

System overview

Figure 100 on page 1010 illustrates the relationship between the load balancer, a z/OS Load Balancing Advisor, and Load Balancing Agents.

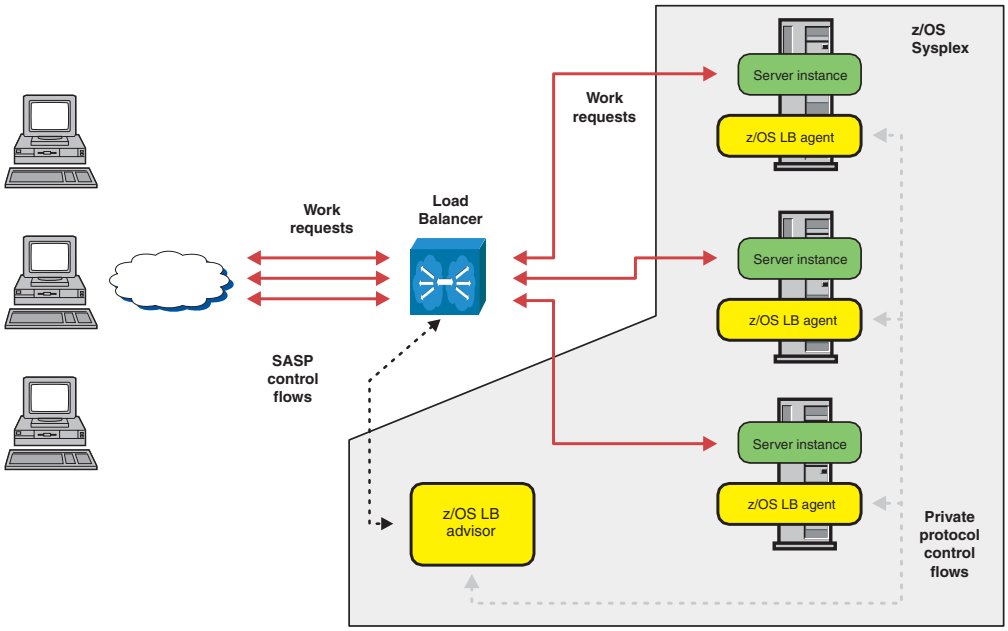


Figure 100. z/OS Load Balancing Advisor

The load balancer is configured with a list of systems and applications that it will balance. The load balancer tells the Load Balancing Advisor about the applications by specifying an IP address, port, and protocol, or about the systems by specifying an IP address. The Advisor is configured with a list of authorized load balancers and a list of Load Balancing Agents with which it can gather data, and with a poll interval at which the Agents update the Advisor's data. Each Agent gathers data on its own z/OS system about the TCP/IP stacks and applications running on that system. The Agent is configured with the information it needs to contact the Advisor. The Advisor consolidates the data from all its Agents, and returns the data to the load balancer to advise the load balancer about the status of the systems and applications.

Steps for configuring the z/OS Load Balancing Advisor

Before you begin: You must meet the following requirements:

- You must have at least one external load balancer that supports the Server/Application State Protocol (SASP). This load balancer must have IP connectivity to each z/OS system in the sysplex that is to participate in this load balancing methodology.
- All pre-V1R7 z/OS systems within the sysplex that you want to participate in workload balancing using this method require a new function APAR, as shown in Table 41.

Table 41. New function APARs needed for pre-V1R7 releases

Release level	APAR needed
V1R4	PQ90032
V1R5 and V1R6	PQ96293

Perform the following steps to configure the z/OS Load Balancing Advisor and one or more Load Balancing Agents. Where special considerations exist for

multiple TCP/IP stack environments, references are made to "Steps for configuring the z/OS Load Balancing Advisor in multiple TCP/IP stack (CINET) environments" on page 1027.

1. Evaluate TCP/IP workloads to be load balanced and select a load balancing solution. (optional)
2. Decide who will have authority to start the Advisor. (optional)
3. Decide who will have authority to start the Agents. (optional)
4. Authorize the Agents to use WLM services
5. Optionally configure the Advisor and Agents to automatically restart in case of application or system failure. (optional)
6. Configure and start syslogd.
7. Configure one Advisor per sysplex.
8. Configure one Agent per z/OS system in the sysplex.
9. Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on. (optional)
10. Start the TCP/IP stacks that the Advisor and the Agents will use.
11. Start the target applications that will be the targets of load balancing.
12. Customize WLM policies for the Advisor and Agents. (optional)
13. Start one Agent on each sysplex system that you want to participate in this method of workload balancing.
14. Start the one instance of the Advisor in the sysplex.
15. Configure the external load balancers.
16. Start the load balancers.
17. Verify that the Advisor system is functioning properly. (optional)

Step 1: Evaluate TCP/IP workloads to be load balanced and select a load balancing solution (optional)

The first steps involve identifying the TCP/IP applications that you want to load balance, the systems that these applications will be running on, and ensuring that these applications can exploit load balancing. When this is done, evaluate the load balancing techniques that best meet your requirements. You might want to use a combination of workload balancing solutions.

There are various technology choices for performing IP load balancing in a sysplex environment. For a discussion of some of these choices, see "Workload balancing" on page 386.

Step 2: Decide who will have authority to start the Advisor (optional)

Explicit authority should be granted to all users that can start the Advisor, to prevent unauthorized users from starting it. If you do not grant explicit authority, any user able to issue the START command can start the Advisor.

Tip: You might want to combine this step with the next two steps, "Step 3: Decide who will have authority to start the Agents (optional)," and "Step 4: Authorize the Agents to use WLM services" on page 1013, since these steps use similar commands.

Steps for granting authority to start the Advisor

Perform the following steps to grant authority to start the Advisor:

1. Ensure that the OPERCMDS class is active and RACLISTed, and RACLIST processing is enabled:

   ```
   SETROPTS CLASSACT(OPERCMDS)
   SETROPTS RACLIST (OPERCMDS)
   ```

2. Define the following OPERCMDS class profile using a security product like RACF:

   ```
   RDEFINE  OPERCMDS (MVS.SERVMGR.LBADV) UACC(NONE)
   ```

3. Grant the Advisor access to the OPERCMDS class:

   ```
   PERMIT   MVS.SERVMGR.LBADV CLASS(OPERCMDS) ACCESS(CONTROL) -
                    ID(userid)
   ```

4. Refresh the OPERCMDS class:

   ```
   SETROPTS RACLIST(OPERCMDS) REFRESH
   ```

5. Refer to the EZARACF sample in SEZAINST for specific instructions.

   All commands that you can issue against the Advisor are MODIFY commands, with the exception of the STOP command used to stop the Advisor. Therefore, you might also want to limit which users are able to issue MODIFY and STOP commands.

Step 3: Decide who will have authority to start the Agents (optional)

Explicit authority should be granted to all users that can start the Agents, to prevent unauthorized users from starting them. If you do not grant explicit authority, any user able to issue the START command can start the Agents.

Steps for granting authority to start the Agents

Perform the following steps to grant authority to start the Agents:

1. Ensure that the OPERCMDS class is active and RACLISTed, and RACLIST processing is enabled. If you have already done this for the Advisor, you can skip this step.

   ```
   SETROPTS CLASSACT(OPERCMDS)
   SETROPTS RACLIST (OPERCMDS)
   ```

2. Define the following OPERCMDS class profile using a security product like RACF:

   ```
   RDEFINE  OPERCMDS (MVS.SERVMGR.LBAGENT) UACC(NONE)
   ```

3. Grant the Agents access to the OPERCMDS class:

   ```
   PERMIT   MVS.SERVMGR.LBAGENT CLASS(OPERCMDS) ACCESS(CONTROL) -
                    ID(userid)
   ```

4. Refresh the OPERCMDS class:

   ```
   SETROPTS RACLIST(OPERCMDS) REFRESH
   ```

5. Refer to the EZARACF sample in SEZAINST for specific instructions.

All commands that you can issue against the Agents are MODIFY commands, with the exception of the STOP command used to stop the Agents. Therefore, you might also want to limit which users are able to issue MODIFY and STOP commands.

Step 4: Authorize the Agents to use WLM services

You might want or need to define the BPX.WLMSERVER resource profile to your security product and grant the Agents access to it. If you are using RACF and already have the resource profile defined and the FACILITY class is enabled, permit the Agents to the resource profile. If you are using a security product other than RACF that by default denies access to the resource profile, grant the Agents access to the resource profile. If you do not already have the resource profile defined and you are using RACF, consult the documentation of other programs and products that require WLM services and coordinate any potential changes with these programs and products.

Steps for defining the resource profile with RACF

Perform the following steps for RACF if you choose to define the resource profile:

1. Ensure that the FACILITY class is active and RACLISTed, and RACLIST processing is enabled:

   ```
   SETROPTS CLASSACT(FACILITY)
   SETROPTS RACLIST (FACILITY)
   ```

2. Define the following FACILITY class profile:

   ```
   RDEFINE  FACILITY (BPX.WLMSERVER) UACC(NONE)
   ```

3. Grant the Agent access to the FACILITY class:

   ```
   PERMIT   BPX.WLMSERVER CLASS(FACILITY) ACCESS(READ) -
                          ID(userid)
   ```

4. Refresh the FACILITY class:

   ```
   SETROPTS RACLIST(FACILITY) REFRESH
   ```

For more information, see the EZARACF sample in SEZAINST.

Step 5: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)

Although this step is optional, performing it will provide high availability to your target applications. In the event that an Agent fails, the Advisor would indicate that it has no information for any applications running on that system. As a result, target applications on the failing system would cease to receive new workload requests, in most cases, until the Agent is restarted. Automatically restarting the Agent on the same system would minimize this perceived outage. This can be accomplished using automation software or by defining an automatic restart manager (ARM) policy. For more information on defining ARM policies, refer to *z/OS MVS Setting Up a Sysplex*.

The Agent registers with ARM using the following values:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=LBAGENT
TERMTYPE=ELEMTERM
```

This indicates that if the Agent fails on this system, it should be restarted on this system only.

If the Advisor or its underlying system were to fail, the load balancer might continue to distribute workload requests according to the last set of information received from the Advisor, it might resort to preconfigured weights, or it might even stop distributing new work requests to the cluster. (The behavior depends upon the load balancer implementation; consult the load balancer documentation for details.) Therefore, it is important that the Advisor be restarted as soon as possible when a failure occurs, so that it can begin communicating with the load balancer and workload request distribution can resume normally. This restart capability should cover scenarios where the Advisor itself fails, and where the system that the Advisor is running on fails. The Advisor can run on any system in the sysplex and thus can be restarted on any system in the sysplex, as long as it is configured to use dynamic VIPAs and dynamic routing is in effect. The Advisor registers with ARM using the following values:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=LBADV
TERMTYPE=ALLTERM
```

This indicates that the Advisor should be restarted only on the same system in cases where the Advisor itself fails, and also restarted on a different system if the system fails. Using an ARM policy, you can indicate which systems are eligible for running the Advisor in the case of system failures. You also need to ensure that the specified backup systems have all the necessary configuration in place to enable the Advisor to be restarted there.

Some special considerations exist for scenarios where ARM is used and the TCP/IP stack address space terminates, as the result of a failure or of a planned operation. When the TCP/IP stack becomes unavailable, the Advisor also terminates, as it can no longer establish any TCP/IP communications. An ARM restart of the Advisor will likely fail, as the TCP/IP protocol stack will not be available when the restarts occur. You can handle these scenarios in the following ways:

- Planned outages of the TCP/IP stack

  Manually start the Advisor on another system, as soon as the Advisor terminates on the system where TCP/IP is stopped.
- Unplanned outages of the TCP/IP stack

  Ensure that an ARM policy (or other automation) is in place to quickly restart the TCP/IP stack on the same system. The Advisor also needs to be quickly restarted on the same system. This can be done by using an automation software package, or by using the TCP/IP profile AUTOLOG statement.

The AUTOLOG statement also has some important considerations:

- You should place the Advisor in the AUTOLOG statement list to ensure that it is started when TCP/IP is started on that system. However, you should specify the NOAUTOLOG parameter on the PORT reservation statements for the Advisor ports in the TCP/IP profile. This prevents TCP/IP from monitoring and attempting to restart the Advisor, as that could interfere with your automation logic or the ARM policy that you have put in place.
- The AUTOLOG function works best on systems where a single TCP/IP stack is active (INET environment). For CINET considerations, see "Step 5: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)" on page 1028.

Requirement: The Load Balancing Advisor and Agents do not run using a system key. Therefore, if you are using ARM registration, the started task IDs need to be permitted with UPDATE authority to the associated IXCARM.SYSTCPIP.*elemname* FACILITY class profile within the SAF product on your system.

Restrictions:

- If using AUTOLOG for the Agent, code the NOAUTOLOG parameter on the PORT reservation statement for the Agent port in the TCP/IP profile. This prevents the Agent from automatically being cancelled and restarted because the Agent does not listen on the port.
- If the Advisor is using IPv6 for the load balancer connections or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to those z/OS releases that support IPv6 DVIPAs. For information on IPv6 DVIPA support, refer to *z/OS Communications Server: New Function Summary*.

Step 6: Configure and start syslogd

The Advisor and Agent write most log messages and trace data to the syslog daemon (syslogd). A limited number of messages are written to the MVS console, but these are unaffected by syslogd configuration. For the Advisor and Agent to be able to write their log messages and trace data to syslogd, syslogd must be properly configured and started before the Advisor and Agent are started.

Because it is likely that you will be running an Agent on the same system as the Advisor, for better readability, you might want to configure syslogd to place Advisor and Agent log output in separate files. For further information, see "Configuring the syslog daemon (syslogd)" on page 175.

Tip: As more data is logged by the Advisor and Agent, performance of the Advisor and Agent can be adversely affected. The amount of data that is logged by the Advisor and Agent is determined by the debug_level statement. Backing the syslogd output file with a zFS file system instead of an HFS file system can minimize performance impacts caused by logging.

Step 7: Configure one Advisor per sysplex

There can be only one Advisor active in the sysplex at any given time. The Advisor reads configuration data from one file, which might exist as an HFS file or zFS file, a PDS or PDSE member, or a sequential data set. If you plan on allowing the Advisor to move within the sysplex in the case of failure, you probably want the Advisor configuration file or data set to exist on shared DASD, to make it accessible to all systems in the sysplex if necessary. The Advisor configuration file is specified on the CONFIG DD statement in the Advisor start procedure. A sample start procedure is provided in EZBLBADV in SEZAINST.

The Advisor configuration file serves three basic purposes:

- Defines the listening sockets for the load balancers and Agents
- Provides an access control list for specifying which load balancers and Agents can connect to the Advisor
- Customizes some optional parameters

A sample Advisor configuration file is provided in EZBLBADC in SEZAINST.

Define listening sockets/ports (required)

The Advisor maintains at least two, and up to three, listening sockets/ports, one for Agents to connect to and up to two for load balancers to connect to. There are separate IPv4 and IPv6 listening sockets for load balancers. If your TCP/IP stack is not IPv6 enabled, you will not be able to use the IPv6 listening socket.

The Advisor and Agent statements define addresses and ports on the local system and on remote systems. At times, it can be difficult to remember which statements refer to local sockets and which statements refer to remote addresses and ports.

Tip: Any statement containing the word *connection* refers to a local socket, and any statement containing the word *id* refers to a remote address and possibly a port.

Specify the local IPv4 address and port that the Advisor listens on for IPv4 load balancer connections with the lb_connection_v4 configuration statement. The default port for communications with load balancers is 3860.

The lb_connection_v6 statement does the equivalent for IPv6 that lb_connection_v4 does for IPv4. You can specify either or both of these statements. For CINET considerations regarding stack termination, see "Step 7: Configure one Advisor per sysplex" on page 1028.

Guideline: To enable movement of the Advisor to another system in the sysplex or to another TCP/IP stack on the same system in the event of failure of the Advisor or its underlying system, use a dynamic VIPA (DVIPA) for the address specified on the lb_connection_v4 and lb_connection_v6 statements. Furthermore, make this DVIPA a unique application-instance DVIPA (defined through VIPARANGE) rather than a multiple application-instance DVIPA (defined through VIPADEFINE), to enable movement of the Advisor if the Advisor itself failed. For CINET considerations with DVIPAs, see "Step 7: Configure one Advisor per sysplex" on page 1028.

Restriction: If the Advisor is using IPv6 for the load balancer connections, or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to those z/OS releases that support IPv6 DVIPAs. For information on IPv6 DVIPA support, refer to *z/OS Communications Server: New Function Summary*.

Specify the local port that the Advisor listens on for Agent connections with the agent_connection_port statement. If the Advisor's TCP/IP stack is IPv6 enabled, the Advisor opens a listening socket for Agents on the IPv6 unspecified address (::) on the port specified by this statement. This enables Agents to connect to the Advisor using either IPv4 or IPv6, and by using any address on the Advisor's system. If the TCP/IP stack is not IPv6 enabled, the Advisor opens a listening socket on the IPv4 unspecified address, 0.0.0.0. This enables Agents to connect to the Advisor using any IPv4 address on the Advisor's system.

Guideline: The port number used on the agent_connection_port statement should not be the same as the port used on the lb_connection_v4 statement or the lb_connection_v6 statement.

Rules:

- The port number specified on the agent_connection_port statement must match the port number specified on the Agents' advisor_id statement.
- If at least one IPv4 address is specified in the lb_id_list statement, the lb_connection_v4 statement must be specified. Similarly, if at least one IPv6 address is specified in the lb_id_list statement, the lb_connection_v6 statement must be specified.

Define the access control list (required)

The Advisor maintains loose security of which load balancers and which Agents are allowed to connect to it by maintaining an access control list. The access

control list specifies the remote IP address of the connecting load balancers and the remote IP address and port of the Agents that are allowed to connect.

Rule: Specify only complete IP addresses in access control lists; subnetworks, IP prefixes, or other types of wildcards are not allowed.

Specify the list of load balancers that are allowed to connect to the Advisor in the lb_id_list statement. Specify the list of Agents that are allowed to connect in the agent_id_list statement.

Rules:

- The addresses in the agent_id_list statement must match the addresses in the Agents' host_connection statement. For the purposes of high availability, the addresses specified in the Advisor's agent_id_list statement and the Agents' host_connection statement should be static VIPAs, to tolerate individual link outages on the hosts.
- If the lb_connection_v4 statement is specified, at least one IPv4 address must be specified on the lb_id_list statement. Similarly, if the lb_connection_v6 statement is specified, at least one IPv6 address must be specified on the lb_id_list statement.

Restriction: There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

Customizing optional statements

Customize the optional statements in the Advisor's configuration file.

The update_interval statement controls how often each Agent updates the Advisor with data, and depending upon load balancer implementation and configuration, can control how often load balancers are updated with information from the Advisor. The default value is 60 seconds. At each update interval, each Agent refreshes the Advisor with the status of each registered member for which the Agent is responsible. This information includes the status of the target application (active or not), whether it is an application member, the operator quiesce state of the member, and various weights that measure the target system and application's ability to handle additional workload requests. The lower the update interval, the more up-to-date the load balancer's data will be with respect to the target's availability and capability to handle additional workload requests. Of course, the lower the update interval, the higher the network traffic and CPU overhead is.

Depending upon load balancer implementation and configuration, the update_interval statement might also determine how often the load balancer is updated with data from the Advisor. If the load balancer supports the SASP push flag, and it has been specified in the load balancer, the Advisor sends data to the load balancer at least every update interval. Regardless of what the update interval is set to, if this push flag is supported and configured in the load balancer, certain events might cause the Advisor to update the load balancer with information before the update interval timer expires. These events include the starting or stopping of a target application, or the addition or deletion of a member's IP address on the Agent host.

Therefore, the update interval is a key factor in determining the latency period between when changes occur on the target application system, and when the load balancer is informed of them. Each Agent updates the Advisor with new information every update interval. The Advisor, in turn, updates the load balancer with changes in weights every update interval, if the load balancer supports the

push flag. In addition, if the push flag is supported and configured by the load balancer, the Advisor updates the load balancer with any change in the target's availability status as soon as it discovers such a change from the Agent, instead of waiting for the update interval to expire. Therefore, when the load balancer supports and configures the push flag, the maximum amount of latency expected between a change in a member's weight and when the load balancer is informed of it is twice the update interval (that is, one update interval for the Agent to report it to the Advisor and one update interval for the Advisor to report it to the load balancer). However, on the average, it should take one update interval for a change in the target application weight to reach the load balancer.

Use the optional wlm statement to specify the default type of WLM recommendation to be used for all groups. There are two choices for this, the basewlm and serverwlm values. If you do not specify this statement, the default is basewlm. If you want a specific group of applications to use a type of WLM recommendation other than the default, you can override the default WLM recommendation type for that group on a port number basis using the port_list statement. The WLM recommendation is a key component used in determining the net weight assigned to a member.

The type of WLM recommendation represented by the basewlm value indicates the overall displaceable capacity of the system where the application represented by the member resides, relative to the other systems in the sysplex. This is referred to as a WLM system weight recommendation. The serverwlm value represents a different type of WLM recommendation, in that it reflects how well an individual server application is performing from a WLM perspective (based on the WLM policy). This type of recommendation is a server-specific recommendation and is referred to as a server-specific WLM recommendation. Server-specific WLM recommendations are composed of two key elements:

- The amount of displaceable capacity available on the target system at the importance level of the application.
- How well the application is performing compared to the WLM goals for that application workload.

System members (port and protocol are zero) always use WLM system weight recommendations. Application members can use either type.

It is important that you choose the type of WLM recommendation that is best suited to each group of applications. Some types of applications are better suited to using WLM system weight recommendations rather than server-specific WLM recommendations. For most applications, server-specific WLM recommendations provide a more accurate way to distribute workload to their servers. However, when a server acts as an access point to applications that run in other address spaces (and therefore in a different service class), WLM system weight recommendations might be the preferred distribution method. The Sysplex Distributor function can also use server-specific WLM recommendations or WLM system weight recommendations. For examples of some applications that might be better represented by WLM system weight recommendations, see "Sysplex Distributor" on page 388.

The optional port_list statement enables you to override or specify parameters for members on a port number basis. The wlm parameter of the port_list statement enables you to override the value defined (or defaulted) on the wlm statement, for all members that use the port number specified. The actual WLM recommendation

type used is still dependent upon the value specified and the z/OS level of the Agents owning the members of the group.

When selecting the type of WLM recommendation to use for a given group, it is important to consider the following requirements:

- All members in a group must specify the same type of WLM recommendation (using the wlm or port_list statement).
- To use server-specific recommendations, no Agent reporting on behalf of a member of a group can be at a release level prior to z/OS V1R7.

For any groups where these requirements are not met, the Advisor uses WLM system weight recommendations, and a warning message is written to syslogd. The main rationale behind this is that WLM system weight recommendations and server-specific WLM recommendations cannot be directly compared to one another.

The Advisor can detect dynamically whether or not these requirements are being met. For example, if all owning Agents of a group, except one, support server-specific WLM recommendations, and the application on that one system is brought down, the WLM recommendation type would change dynamically from WLM system weight recommendations to server-specific WLM recommendations, provided the Advisor was configured to request server-specific WLM recommendations for that group. Similarly, if that same application is brought back up, the WLM recommendation type would dynamically switch back to WLM system weight recommendations. A similar circumstance would arise if the member owned by the Agent that does not support server-specific WLM recommendations was quiesced by the z/OS operator or the load balancer administrator.

The optional debug_level statement determines how much trace data is captured in the Advisor's log file.

Restriction: In most cases, you should not customize the debug_level statement, unless directed to do so by an IBM Service representative. Adding additional types of trace data can cause the amount of data captured to become voluminous. Reducing the amount of trace data from the default might make diagnosing a problem more difficult.

For more details on the Advisor configuration file statements, refer to *z/OS Communications Server: IP Configuration Reference*.

Step 8: Configure one Agent per z/OS system in the sysplex

There can be only one Agent active per z/OS system in the sysplex at any point in time. The Agent reads configuration data from one file, which can exist as an HFS or zFS file, a PDS or PDSE member, or a sequential data set. The Agent configuration file is specified on the CONFIG DD statement in the Agent start procedure. A sample start procedure is provided in EZBLBAGE in SEZAINST.

The Agent configuration file serves three basic purposes:

- Defines the IP address and port that the Agent binds to for communication with the Advisor
- Identifies the location (IP address and port) of the Advisor
- Customizes optional parameters

A sample Agent configuration file is provided in EZBLBAGC in SEZAINST.

The Advisor and Agent statements define addresses and ports on the local system and on remote systems. At times, it can be difficult to remember which statements refer to local sockets and which statements refer to remote addresses and ports.

Tip: Any statement containing the word *connection* refers to a local socket, and any statement containing the word *id* refers to a remote address and possibly port.

Defining the IP address and port to bind to for communications with the Advisor (required)

Specify the local IP address and port that the Agent binds to for communications with the Advisor on the host_connection statement. This is used as part of the Advisor's access control enforcement.

Rules:

- The IP address and port on the host_connection statement must be included on the agent_id_list statement in the Advisor's configuration file. The IP address on the host_connection statement can be an IPv6 address, if the Agent's system is running an IPv6-enabled TCP/IP stack and the Advisor has an IPv6-enabled TCP/IP stack available.
- If an IPv4 address is specified on the host_connection statement, an IPv4 address must be specified on the advisor_id statement. Similarly, if an IPv6 address is specified on the host_connection statement, an IPv6 address must be specified on the advisor_id statement.

Guidelines:

- For simplicity and consistency, you might want to specify the same port on the host_connection statement for each Agent, and reserve the same port for the Agent on each TCP/IP stack that an Agent will run on. For more information about port reservation, see "Step 9: Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on (optional)" on page 1021.
- The address in the host_connection statement (and therefore, also in the Advisor's agent_id_list statement) should be a static VIPA. For CINET considerations regarding the host_connection statement, see "Step 8: Configure one Agent per z/OS system in the sysplex" on page 1029.

Also see "Step 7: Configure one Advisor per sysplex" on page 1028 for CINET considerations regarding unique application-instance DVIPAs and stack affinity.

Restriction: If the Advisor is using IPv6 for the load balancer connections or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to those z/OS releases that support IPv6 DVIPAs. For information on IPv6 DVIPA support, refer to *z/OS Communications Server: New Function Summary*.

Identifying the location of the Advisor (required)

Specify the location of the Advisor on the advisor_id statement. This statement contains an IP address and port that the Advisor uses to listen for connections from Agents in the sysplex.

Guideline: Use a dynamic VIPA for this address, to enable the Advisor to be moved within the sysplex in the event of a failure of the Advisor or the Advisor's underlying system.

Rules:

- The port specified on the advisor_id statement must match the port specified on the Advisor's agent_connection_port statement.

- If an IPv4 address is specified on the advisor_id statement, an IPv4 address must be specified on the host_connection statement. Similarly, if an IPv6 address is specified on the advisor_id statement, an IPv6 address must be specified on the host_connection statement.

Customizing optional statements

Similar to the Advisor, the optional debug_level statement determines how much trace data is captured in the Agent's log file.

Restriction: In most cases, you should not customize this statement, unless directed to do so by an IBM Service representative. Adding additional types of trace data can cause the amount of data captured to become voluminous. Reducing the amount of trace data from the default might make diagnosing a problem more difficult.

For more details on the Agent configuration file statements, refer to *z/OS Communications Server: IP Configuration Reference*.

Step 9: Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on (optional)

There are several things to do to customize the TCP/IP profile to accommodate the Advisor and Agents.

- In the appropriate TCP/IP profiles that they will use, including the TCP/IP stacks that they could potentially move to, reserve the ports that the Advisor and Agents will use. All ports for the Advisor and Agent utilize the TCP protocol, and thus should be reserved for TCP. The Advisor has at least two ports, and potentially three ports, to reserve, including the ports specified on the following statements:
  - lb_connection_v4
  - lb_connection_v6
  - agent_connection_port

  For CINET considerations regarding the lb_connection_v4 and lb_connection_v6 statements, see "Step 9: Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on (optional)" on page 1029.
- If you use dynamic VIPAs for the Advisor as recommended, you need to configure the appropriate TCP/IP profiles in the sysplex for the DVIPA definition and usage. The preferred definitions would include VIPADEFINE with MOVEABLE IMMEDIATE, or VIPARANGE with MOVEABLE NONDISRUPTIVE. For more specific information, see "Using dynamic VIPAs (DVIPAs)" on page 305.

  Restriction: Do not mix Sysplex Distributor functions with these DVIPAs.
- If you are currently using the SHAREPORT or SHAREPORTWLM parameters on the TCP/IP profile PORT statement to enable multiple TCP applications to share the same port, some additional considerations might apply to your configuration. For example, if the TCP applications sharing the same port are also members of groups that are being reported to external load balancers with SASP, it is important to ensure that consistent criteria are used by the various load balancing components.

  When using the z/OS Load Balancing Advisor, all instances of a TCP application sharing the same port on a target system are reported to external load balancers using a single member entry, and therefore, a single recommendation. This recommendation reflects the average net weight calculated for all the servers sharing the same port on a target system, and is

based on the type of WLM recommendation configured on the Advisor. When the TCP connection requests reach a target TCP/IP stack and multiple applications are sharing the same port, the connections are then load balanced by TCP/IP across the multiple application server instances. How this load balancing is performed depends on whether the SHAREPORT or SHAREPORTWLM parameter is specified on the PORT statement. For more details on the PORT statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Guideline: If server-specific WLM recommendations are configured within the Advisor for a given group that contains servers that share the same port on a given system or TCP/IP stack, the SHAREPORTWLM parameter should also be specified on the PORT statement in the TCP/IP profile for these servers. This enables both the external load balancers and the internal TCP/IP load balancer to operate with the same type of recommendations when load balancing work requests to these servers. Similarly, if WLM system weight recommendations were configured in the Advisor for a group, the SHAREPORT parameter would probably be more appropriate.

Step 10: Start the TCP/IP stacks that the Advisor and the Agents will use

The TCP/IP stacks that the Advisor will use must be started prior to starting the Advisor. An Agent can be started before the TCP/IP stack it uses is started. If the TCP/IP stack that an Agent uses terminates, the Agent remains active and reestablishes communication with the TCP/IP stack once it becomes active again. For CINET considerations regarding Agent recovery after stack failure, see "Step 10: Start the TCP/IP stacks that the Advisor and the Agents will use" on page 1030.

Step 11: Start the target applications that will be the targets of load balancing

No modifications are necessary to these applications, their configurations, or start procedures, unless the load balancer is using dispatch mode for packet forwarding. For more information on dispatch mode, see "Step 15: Configure the external load balancers" on page 1023.

Step 12: Customize WLM policies for the Advisor and Agents (optional)

It is important that the Advisor and Agents receive an adequate amount of system resources to properly balance workloads. Part of this task involves making the Advisor and Agent run non-swappable. In addition, WLM can control the amount of system resources allocated to the Advisor and Agents.

Guideline: The Advisor and Agents should be assigned to the WLM SYSSTC service class to receive the proper dispatching priority. For more information on categorizing work into service classes, refer to *z/OS MVS Planning: Workload Management*.

Step 13: Start one Agent on each sysplex system you want to participate in this method of workload balancing

It does not matter whether the Agents are started before the Advisor, or whether the Advisor is started before the Agents. If the Advisor is started after the Agents are started, the Agent periodically attempts to connect to the Advisor. Only one

Agent can be started per z/OS system. Agents must be started from a start procedure as a started program (EXEC PGM=). They cannot be started under BPXBATCH. The IBM-supplied program properties table has entries to make the Agents run non-swappable. You should not override this entry to make the Agents run swappable.

Step 14: Start the one instance of the Advisor in the sysplex

As Agents connect to the Advisor, MVS console messages appear on the Advisor's MVS console and on the Agents' MVS consoles. Verify that each Agent you expect to connect to the Advisor has connected. You can also use the NETSTAT,CONN command on the Advisor's TCP/IP stack to see which Agents are currently connected. The Advisor must be started from a start procedure as a started program (EXEC PGM=). It cannot be started under BPXBATCH. The IBM-supplied program properties table has entries to make the Advisor run non-swappable. You should not override this entry to make the Advisor run swappable.

Step 15: Configure the external load balancers

Configure the load balancers with the location (IP address and port) of the Advisor. For maximum availability, this address should be defined as a DVIPA.

Restrictions:

- If the Advisor is using IPv6 for the load balancer connections or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to those z/OS releases that support IPv6 DVIPAs. For information on IPv6 DVIPA support, refer to *z/OS Communications Server: New Function Summary*.
- There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

You might be able to customize features of the load balancer's communication with the Advisor. The SASP protocol defines two features that the load balancer implementation might or might not allow to be configured. One determines whether the load balancer polls the Advisor for updated data, or whether updated data is pushed to the load balancer. The other determines whether only members that have updated data should be sent to the load balancer, or whether all members should be sent to the load balancer regardless of whether their data has changed or not. To determine whether these features can be customized, and how to perform the customization if available, consult your load balancer's documentation. If the load balancer is capable and configured to request that the Advisor push updated information to the load balancer, the Advisor will update the load balancer at least every update interval. If the load balancer is capable and configured to poll the Advisor for updated information, the Advisor will recommend to the load balancer that it poll every update interval. However, the load balancer can choose to disregard this guideline. Consult the load balancer documentation for the expected behavior in these circumstances.

You might want to consider having redundant load balancers configured alike for availability reasons. If so, you need to be aware of the load balancer's unique load balancer identifier (LB UID), sometimes referred to as the UID or UUID, which uniquely identifies a load balancer. Duplicate LB UIDs are not allowed and connection attempts to the Advisor from a load balancer using the same LB UID as an existing connection will force the existing connection to be broken and replaced by the new connection. Redundantly configured load balancers either need to have unique LB UIDs, if you want them to serve as hot standbys that are connected simultaneously along with the load balancer they are backing up, or if they are

configured with the same LB UID, they must remain unconnected from the Advisor until the original load balancer fails.

Some load balancers might be capable of using either dispatch or directed mode when forwarding packets to their destinations. External load balancers typically use a cluster IP address to represent the set of applications being load balanced. Client applications use this cluster IP address as the destination IP address for their requests. When a load balancer uses dispatch mode, the destination IP addresses for incoming IP packets is not changed. Instead, the load balancer forwards the packet to a target z/OS system by using the MAC address of a network adapter on that system. The receiving z/OS system inspects the destination IP address of the packet, and if it matches one of the IP addresses in its HOME list, accepts the packet. As a result, with dispatch mode, all target z/OS systems must have the load balancer's cluster IP address defined in their HOME list. However, it is important that these addresses are not advertised externally through dynamic routing protocols. One way to accomplish this is by defining these IP addresses as loopback addresses on z/OS.

With directed mode, the load balancer converts the destination IP address (that is, the cluster IP address) to an IP address owned by the target z/OS system, using technologies such as Network Address Translation (NAT). When IP packets for these connections are sent back to clients, the load balancer converts the source IP address (that is, the target z/OS system's IP address) back to the cluster IP address that the application had used on its request.

While dispatch mode eliminates the need for performing NAT, it does have some special considerations. For example, in Figure 104 on page 1042, both SYSA and SYSB have the same server, FTPD, bound to the same port number using INADDR_ANY. The packet will have the cluster IP address (both SYSA and SYSB are in the same cluster), so the load balancer will use the MAC address to decide to send the packet to SYSA or SYSB, and TCP/IP will then route the packet to the server.

Restrictions: When using dispatch mode, for the load balancer to function correctly, there are the following limitations:

- An OSA cannot be shared among LPARs.
- All target applications must bind to the IP address specified by INADDR_ANY and the cluster IP address must be defined to the stack; however, this must be done so that the address is not advertised (such as a loopback address).

If these restrictions are not met, load balancing will not be optimal because some servers will not get work routed to them.

With directed mode, either the destination IP address (server NAT) is modified in the packet itself, or both the destination and source IP addresses (server NAT and client NAT) are modified in the packet. The packet must return through the same load balancer that will recognize the changes and do the reverse mapping, so a packet can flow from the original destination to the original source.

Configure each load balancer with the members that represent the individual target application instances, or system members that generically represent a system in the sysplex, or both. Members that can share the same type of workload are defined under the same group. For example, TN3270 servers are defined under one group, and HTTP servers in another. Application members are defined by specifying an IP address, a nonzero port, and a nonzero protocol. System members are defined by specifying an IP address, and specifying the port and protocol to be zero. Members

that have only a port of zero or only a protocol of zero (that is, one but not both are zero) are not considered valid members and will not receive any data from the Advisor. The IP addresses of the members must represent valid, reachable addresses within the sysplex that are unique to a specific sysplex system. This excludes such addresses as the loopback addresses, and other non-advertised addresses.

Tip: The Advisor does not check for improperly configured members. After the entire z/OS Load Balancing Advisor system is operational, display all members registered by each load balancer and verify each member you expect to be available is flagged as available. Screen any unavailable members for coding errors in the member, such as incorrect IP addresses, ports, or protocols.

Guideline: For availability reasons, the IP addresses configured for each member should be VIPA addresses (static or dynamic). If the IP address of a physical interface fails and a member specifies that IP address, the Advisor still indicates that the member is available, as alternate routing paths to that member might exist. However, if no alternate routing paths exist, workload requests cannot be delivered to the target system. By using static or dynamic VIPAs in members, the chance of an alternate route being available when a physical interface fails is greatly increased, as long as at least one physical interface is still available.

Restrictions:

- All IP addresses configured in members belonging to the same group must exist within the same sysplex.
- All members belonging to the same group must be of the same type. That is, all members must be application members or all must be system members.
- Certain classes of IP addresses must not be coded for members in the load balancer. This includes the following classes of addresses:
  - Distributed DVIPAs (the address specified on a VIPADISTRIBUTE statement). Defining members with these addresses would combine two load balancing methodologies for the same workload, wasting system resources.
  - Deprecated IPv6 addresses. These are flagged as such in a NETSTAT HOME display. It is probably safest to not code any autoconfigured IPv6 addresses within members.
  - Addresses that are not unique within the sysplex.
  - Addresses that are not reachable from the load balancer, including:
    - Loopback addresses.
    - Unavailable IPv6 addresses. These might be marked as unavailable if duplicate address detection is in progress, has failed, or the interface ID is unknown. These addresses are displayed in a NETSTAT HOME display, including the reason they are marked unavailable.

Step 16: Start the load balancers

When a load balancer has connected, messages appear on the Advisor's MVS console. You can also use the Advisor's MODIFY *procname*,DISPLAY,LB command to see which load balancers are connected to the Advisor. For details on the Advisor's MODIFY command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Restriction: There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

Step 17: Verify that the Advisor system is functioning properly (optional)

View the MVS console of the Advisor and Agent systems after they are started to verify that the applications started correctly and are still running. If there are any failure messages, refer to the appropriate message description for the proper corrective action. View the syslogd files of the Advisor and Agent systems to see whether any error or warning messages were issued.

Use the following commands to verify that the Advisor system is functioning properly:

- Verify that the Advisor is started and connected to the expected load balancers by issuing the Advisor's MODIFY *procname*,DISPLAY,LB command. Verify that each load balancer is displayed. Take note of the LB INDEX displayed for each load balancer. This identifier is needed to display details of the load balancer, including its registered groups and members.

  Tip: Individual load balancers are identified in displays by their load balancer index (LB Index), which is generated by the Advisor when the load balancer first connects. To display the details of a particular load balancer, first obtain the LB Index by displaying the list of all load balancers using the Advisor's MODIFY *procname*,DISPLAY,LB command. When you have the LB Index, display the details of a particular load balancer using the appropriate LB Index on the INDEX parameter as follows:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=lb_index
  ```

- Verify that each load balancer configured and registered the proper groups and members with the Advisor by issuing the following command:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=assigned_lb_index
  ```

  This display should show all groups and members defined to the load balancer.

- Verify that each Agent has started properly and is communicating with the Advisor. On each Agent, issue the following command:

  ```
  MODIFY procname,DISPLAY,MEMBERS
  ```

  Each member that has an IP address owned by this Agent should appear in the display.

- Verify that the target applications that you want to load balance to are actually available for load balancing. On the Advisor, for each load balancer connected to the Advisor, issue the following command:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=assigned_lb_index
  ```

  Check for the AVAIL flag for each member in the display. The flag is either YES, meaning the member is available for load balancing, or NO, meaning it is not available for load balancing. To be available for load balancing, all of the following must be true:

  - The Agent owning the member's IP address must be active and communicating with the Advisor.
  - The application must be active, if the member represents an application member.
  - The member must not be quiesced by the Agent operator or the load balancer. The z/OS Agent operator is able to quiesce any member that is owned by that Agent. Also, depending upon load balancer implementation, it might be possible for the load balancer administrator to quiesce individual members.

If one of the above conditions is false, correct the situation and repeat the display command until you are satisfied that all members that you intend to have available for load balancing are displayed as being available.

Steps for configuring the z/OS Load Balancing Advisor in multiple TCP/IP stack (CINET) environments

Before you begin: This section is intended to guide you through the special considerations needed if you are implementing the Advisor or Agent in a CINET environment. This section assumes you have read and followed the material in "Steps for configuring the z/OS Load Balancing Advisor" on page 1010; this section serves only as an adjunct to that section.

Prior to using this section, you should also examine your CINET environment and decide how you would like the Advisor and Agents to interact with multiple stacks in the same system. For example, if you have deployed CINET to enable multiple networks to be attached to a z/OS system while retaining some isolation, you might not want the Advisor or Agents to be able to access all TCP/IP stacks on the system. In other scenarios, you might decide that the Advisor and Agent functions should be available across multiple stacks in the same system. This section focuses on these latter scenarios, where you want to use all available stacks on a given system.

Perform the following steps to configure the z/OS Load Balancing Advisor in multiple TCP/IP stack (CINET) environments.

1. Evaluate TCP/IP workloads to be load balanced and select a load balancing solution. (optional)
2. Decide who will have authority to start the Advisor. (optional)
3. Decide who will have authority to start the Agents. (optional)
4. Authorize the Agents to use WLM services
5. Optionally configure the Advisor and Agents to automatically restart in case of application or system failure. (optional)
6. Configure and start syslogd.
7. Configure one Advisor per sysplex.
8. Configure one Agent per z/OS system in the sysplex.
9. Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on. (optional)
10. Start the TCP/IP stacks that the Advisor and the Agents will use.
11. Start the target applications that will be the targets of load balancing.
12. Customize WLM policies for the Advisor and Agents. (optional)
13. Start one Agent on each sysplex system that you want to participate in this method of workload balancing.
14. Start the one instance of the Advisor in the sysplex.
15. Configure the external load balancers.
16. Start the load balancers.
17. Verify that the Advisor system is functioning properly. (optional)

These are the same steps described in "Steps for configuring the z/OS Load Balancing Advisor" on page 1010. For special CINET considerations, see the following subsections.

Step 1: Evaluate TCP/IP workloads to be load balanced and select a load balancing solution (optional)

There are no special considerations for CINET.

Step 2: Decide who will have authority to start the Advisor (optional)

There are no special considerations for CINET.

Step 3: Decide who will have authority to start the Agents (optional)

There are no special considerations for CINET.

Step 4: Authorize the Agents to use WLM services

There are no special considerations for CINET.

Step 5: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)

If you are considering using the AUTOLOG statement to restart the Advisor in a CINET environment, and you placed the Advisor in the AUTOLOG statement list of each TCP/IP stack, each stack attempts to start the Advisor during initialization. Only the first one will succeed, as only a single instance can be active at any time within a system or within a sysplex.

Step 6: Configure and start syslogd

There are no special considerations for CINET.

Step 7: Configure one Advisor per sysplex

You must define listening sockets and ports, as described in "Define listening sockets/ports (required)" on page 1015.

Rule: If you use a unique application-instance DVIPA for the Advisor in a CINET environment, all TCP/IP stacks on that system must code the VIPARANGE statement for that DVIPA. Alternatively, and less desirably, you can establish stack affinity to one of the TCP/IP stacks that are coded with the VIPARANGE statement for that DVIPA, if you do not have VIPARANGE coded for that DVIPA on all of the TCP/IP stacks on that system. This alternative, of course, does not enable the Advisor to be moved to another TCP/IP stack in the event of failure, unless you are able to restart the Advisor with a different start procedure that can establish stack affinity to another TCP/IP stack that has the DVIPA defined in a VIPARANGE statement. For information on the use of the _BPXK_SETIBMOPT_TRANSPORT environment variable that can be used to establish stack affinity, see "Generic server versus server with affinity for a specific transport provider" on page 73.

The lb_connection_v6 statement does for IPv6 what the lb_connection_v4 statement does for IPv4. You can specify either or both of these statements. If you run the Advisor on a CINET system, be aware that the address or addresses you choose

for these statements tie the Advisor to the stack owning those addresses. Consequently, termination of that stack results in termination of the Advisor.

Step 8: Configure one Agent per z/OS system in the sysplex

If the system where the Agent is running is a CINET system, the address in the host_connection statement (and therefore, also in the Advisor's agent_id_list statement) should be a dynamic VIPA (DVIPA) to facilitate movement of the Agent to another TCP/IP stack on that system.

Rule: If you use a unique application-instance DVIPA (coded with VIPARANGE) for the Agent in a CINET environment, all TCP/IP stacks on that system must code the VIPARANGE statement for that DVIPA. When configured in this manner, if the TCP/IP stack that currently owns the DVIPA fails, the Agent remains up and automatically attempts to bind again to the DVIPA. The next available default TCP/IP stack becomes the new owner of the DVIPA. Alternatively, and less desirably (see following restriction), you can establish stack affinity to one of the TCP/IP stacks that are coded with the VIPARANGE statement for that DVIPA, if you do not have VIPARANGE coded for that DVIPA on all of the TCP/IP stacks on that system. Of course, this does not enable the Agent to be automatically moved to another TCP/IP stack in the event of failure. To recover the Agent in this type of configuration, manual intervention (or automation) is required. Because the Agent remains active if its TCP/IP stack fails, you must manually terminate the Agent. Then you must restart the Agent with a different start procedure that can establish stack affinity to another available TCP/IP stack. For information on the use of the _BPXK_SETIBMOPT_TRANSPORT environment variable that can be used to establish stack affinity, see "Generic server versus server with affinity for a specific transport provider" on page 73.

Restriction: When running in a CINET environment, establishing stack affinity between the Agent and one of the TCP/IP stacks on that system enables only resources on that TCP/IP stack to participate in workload balancing. The Agent running on that system is not aware of resources on the other TCP/IP stacks on that system. If you want to enable resources on all TCP/IP stacks in a CINET environment to participate in workload balancing, do not establish stack affinity with the Agent.

Step 9: Customize the TCP/IP profiles of the TCP/IP stacks that the Advisor and Agents will run on (optional)

The Advisor can use multiple TCP/IP stacks in a CINET environment. The addresses specified in the lb_connection_v4 and lb_connection_v6 statements can belong to different TCP/IP stacks. Moreover, because the socket that listens for Agent connections uses the IPv4 or IPv6 unspecified address, the TCP/IP stack or stacks that incoming Agent connections utilize depends upon the IP addresses specified in the Agents' agent_id_list statements. To simplify your configuration and to make Advisor outages that are the result of a TCP/IP stack failure or termination more predictable and recoverable, all incoming connections to the Advisor should use a single TCP/IP stack. Therefore, the addresses you specify in the lb_connection_v4 and lb_connection_v6 statements should belong to the same TCP/IP stack, and you should configure all load balancers and Agents to use these same IP addresses when connecting to the Advisor. The addresses you specify should be dynamic VIPAs to enable the movement of the Advisor in case of failure. This implies that these dynamic VIPAs should be defined in the TCP/IP profiles of all the stacks using a VIPARANGE statement. If the Advisor is restarted as a result of failure in a given TCP/IP stack, the dynamic VIPAs are then activated on another TCP/IP stack in that system. If you decide to use the IPv4 or

IPv6 unspecified addresses for the lb_connection_v4 and lb_connection_v6 statements, you should use the BIND parameter on the PORT reservation statement to bind these sockets to the dynamic VIPA on the one TCP/IP stack you have decided to use.

Step 10: Start the TCP/IP stacks that the Advisor and the Agents will use

In a CINET environment, certain configurations might necessitate manual intervention for recovery if the Agent's TCP/IP stack fails. For more information, see the rule in "Step 8: Configure one Agent per z/OS system in the sysplex" on page 1029.

Step 11: Start the target applications that will be the targets of load balancing

There are no special considerations for CINET.

Step 12: Customize WLM policies for the Advisor and Agents (optional)

There are no special considerations for CINET.

Step 13: Start one Agent on each sysplex system you want to participate in this method of workload balancing

There are no special considerations for CINET.

Step 14: Start the one instance of the Advisor in the sysplex

There are no special considerations for CINET.

Step 15: Configure the external load balancers

There are no special considerations for CINET.

Step 16: Start the load balancers

There are no special considerations for CINET.

Step 17: Verify that the Advisor system is functioning properly (optional)

There are no special considerations for CINET.

Operating the z/OS Load Balancing Advisor

When the Advisor and Agent are operational, you can monitor and customize the following:

- Change the logging level of the Advisor and Agents to suit your needs (optional)
- Interpret Agent and Advisor display information
- Stop distributing new workload requests (QUIESCE) to particular members or resume distribution (ENABLE) to those members (optional)

Changing the logging level of the Advisor and Agents

Optionally, you can change the logging level of the Advisor and Agents to suit your needs. The amount of information that is logged by the Advisor and Agents can be modified dynamically using the following command:

```
MODIFY procname,DEBUG,LEVEL=debug_level
```

However, modifying the logging level is not a step that should be taken lightly. The IBM default of 7 should not be changed unless instructed to do so by an IBM Service representative. For things to consider before modifying this value, see "Customizing optional statements" on page 1017.

Interpreting Agent and Advisor display information

Successfully interpreting the various flags and indicators that appear in Advisor and Agent displays can help you identify configuration problems, or might help explain why workloads are not being distributed as expected. For information on each field in the Advisor and Agent displays, refer to *z/OS Communications Server: IP System Administrator's Commands*. Some portions of the displays are described in more depth in Table 42, Table 43 on page 1033, and the subsections that follow.

Table 42. Summary of selected Advisor display output fields and flags

Flag or field name	Location	Description
AVAIL	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Indicates whether the member is available for workload balancing requests.
BASEWLM	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	WLM system weight recommendations were configured or defaulted for the members of this group, and are being used as a component of the net weight.
BASEWLM*	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Server-specific WLM recommendations were configured or defaulted for the members of this group. However, WLM system weight recommendations are actually being used as a component of the net weight.
CS WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Communications Server weight. A value calculated by the Agent that owns the member, representing the health of the application with respect to its ability to process the work that has recently been received.
LB INDEX	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=nn commands, an output field of the load balancer area	A unique identifier assigned to a load balancer for the purpose of referencing the load balancer in subsequent operator commands
LBQ	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The load balancer has quiesced the member.

Table 42. Summary of selected Advisor display output fields and flags (continued)

Flag or field name	Location	Description
NET WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The weight the Advisor passes to the load balancer, representing the desirability of the member to receive additional workload requests relative to the other members of the same group. The net weight is calculated using the WLM weight and Communications Server weight.
NOCHANGE	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=nn command output, the FLAGS field of the load balancer area	The load balancer has requested that it be sent only information about members that have changed their status or weights since the last time the load balancer received information on its members.
NODATA	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	A transient flag indicating that the responsible Agent has not had enough time to calculate a Communications Server weight.
NOTARGETAPP	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The target application of this application member is not active.
NOTARGETIP	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	This is an unusable system member.
NOTARGETSYS	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The Advisor is not aware of the system that owns the IP address of the member.
OPQ	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The z/OS operator has quiesced the member.
PUSH	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=nn command output, the FLAGS field of the load balancer area	The load balancer has requested to receive information from the Advisor on a scheduled basis, rather than having to poll the Advisor for the information.
SERVERWLM	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Server-specific WLM recommendations are being used for the members of this group as a component of the net weight.
TRUST	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=nn command output, the FLAGS field of the load balancer area	The load balancer will allow other system components besides itself to register members with the load balancer. The z/OS Load Balancing Advisor does not currently exploit this feature.

Table 42. Summary of selected Advisor display output fields and flags (continued)

Flag or field name	Location	Description
WLM WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	WorkLoad Manager weight. The value received from WLM on the system where the member resides, representing the displaceable capacity of that system relative to other systems in the sysplex (system weight), or the value received from WLM representing how well the server is performing relative to its WLM policies (server-specific weight).

Table 43. Summary of selected Agent display output flags

Flag or field name	Location	Description
ANY	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,MEMBERS command output	The target application is bound to the unspecified address, 0.0.0.0 for IPv4 or :: for IPv6.
V6	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,MEMBERS command output	The target application was bound using the IPV6_V6ONLY socket option.

MODIFY *procname*,DISPLAY,LB

This command displays all load balancers that are currently connected to the Advisor. In Figure 101, the line numbers appearing in the left margin are for reference purposes only, and are used in the description following the figure.

```
1  MODIFY LBADV,DISPLAY,LB
2  EZD1242I LOAD BALANCER SUMMARY
3  LB INDEX      : 00        UUID      : 637FFF175C
4   IPADDR..PORT : 10.42.154.105..50005
5   HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
6  LB INDEX      : 01        UUID      : 207FFF175C
7   IPADDR..PORT : 10.42.105.60..50006
8   HEALTH       : 7F        FLAGS     : PUSH TRUST
9  2 OF 2 RECORDS DISPLAYED
```

Figure 101. Sample output for the MODIFY procname,DISPLAY,LB command

LB INDEX: The LB index is generated by the Advisor to uniquely identify a load balancer, and is used in other MODIFY commands to display details of a particular load balancer. The LB index that is assigned to the next load balancer that connects can be difficult to predict at times, in case your automation attempts to predict them. Each time a load balancer connects, it is assigned a different LB index. At first, the numbers are assigned in order from 0 to 99, and then the numbers are reused. The next number assigned after 99 is based on a least-recently-used algorithm. Thus, of the load balancer connections that used the numbers 0-99, the index of the load balancer connection that ended first would be the next number assigned as the next LB index. This is done to prevent the confusion that could result if new load balancer connections obtained the lowest available index. If that were the case, it might be difficult to ascertain whether different load balancer

displays referred to the same or different load balancer connection instances. If the Advisor is brought down and back up, the indexes start from zero again. Lines 3 and 6 in Figure 101 on page 1033 show two LB indexes. For more information on the LB INDEX field, refer to *z/OS Communications Server: IP System Administrator's Commands*.

NOCHANGE, PUSH, TRUST: All of these flags are set only by the load balancer. Whether they appear or not depends upon whether the particular load balancer implementation supports them, and if they are configured in the load balancer. The z/OS administrator has no control over the settings of these flags.

The NOCHANGE flag can affect the amount of data transferred between the load balancer and the Advisor. If this flag is set, only data that has changed since the last time data was sent to the load balancer is included. Consult the load balancer documentation to determine whether this setting is supported.

The PUSH flag can have an effect on how soon the load balancer is informed of certain events. If the PUSH flag is not on, the load balancer must poll the Advisor periodically for updates. If this flag is on, certain events can be communicated to the load balancer earlier than would be possible if polling were in effect. Those events include quicker notification of a target application being taken out of service, and quicker notification of when the member's IP address has been moved to another system in the sysplex (VIPA takeover) or removed entirely. When the PUSH flag is on, the Advisor also sends the load balancer updated information about weights and status at least every update interval, if new information is available. Consult the load balancer documentation to determine whether this setting is supported.

The TRUST flag indicates that the load balancer allows other system components besides itself to register members with the load balancer. The z/OS Load Balancing Advisor does not currently exploit this feature.

Line 5 of Figure 101 on page 1033 shows all of these flags. For more information on the NOCHANGE, PUSH, and TRUST flags, refer to *z/OS Communications Server: IP System Administrator's Commands*.

MODIFY *procname*,**DISPLAY,LB,INDEX=***lbindex*

This command displays details about a particular load balancer, including its registered groups and members. In Figure 102 on page 1035, the line numbers appearing in the left margin are for reference purposes only, and are used in the description following the figure.

```
1 MODIFY LBADV,DISP,LB,INDEX=0
2 EZD1243I LOAD BALANCER DETAILS
3 LB INDEX      : 00        UUID      : 637FFF175C
4   IPADDR..PORT : 10.42.154.105..50005
5   HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
6   GROUP NAME   : SYSTEMFARM
7    GROUP FLAGS : BASEWLM
8    IPADDR..PORT: 10.42.105.154..0
9     SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : YES
10    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
11    FLAGS      :
12   IPADDR..PORT: 10.42.105.60..0
13    SYSTEM NAME: VIC007    PROTOCOL  : 000  AVAIL     : YES
14    WLM WEIGHT : 00045     CS WEIGHT : 100  NET WEIGHT: 00002
15    FLAGS      :
16   IPADDR..PORT: 10.42.105.22..0
17    SYSTEM NAME: N/A       PROTOCOL  : 000  AVAIL     : NO
18    WLM WEIGHT : 00000     CS WEIGHT : 000  NET WEIGHT: 00000
19    FLAGS      : NOTARGETSYS
20   IPADDR..PORT: 10:1::4:5..0
21    SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : NO
22    WLM WEIGHT : 00021     CS WEIGHT : 000  NET WEIGHT: 00000
23    FLAGS      : NOTARGETIP
24  GROUP NAME   : UDP_SERVER_FARM
25   GROUP FLAGS : SERVERWLM
26   IPADDR..PORT: 10.42.105.154..7777
27    SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
28    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
29    FLAGS      :
30   IPADDR..PORT: 2001:DB8::10:5:6:2..7777
31    SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
32    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
33    FLAGS      :
34   IPADDR..PORT: 10.42.105.60..7777
35    SYSTEM NAME: VIC007    PROTOCOL  : UDP  AVAIL     : YES
36    WLM WEIGHT : 00045     CS WEIGHT : 100  NET WEIGHT: 00002
37    FLAGS      :
38  GROUP NAME   : DNS_GROUP
39   GROUP FLAGS : BASEWLM*
40   IPADDR..PORT: 10.42.103.75..53
41    SYSTEM NAME: MVSW      PROTOCOL  : TCP  AVAIL     : NO
42    WLM WEIGHT : 00064     CS WEIGHT : 100  NET WEIGHT: 00000
43    FLAGS      : LBQ OPQ
44   IPADDR..PORT: 10.42.105.60..53
45    SYSTEM NAME: VIC007    PROTOCOL  : TCP  AVAIL     : NO
46    WLM WEIGHT : 00045     CS WEIGHT : 000  NET WEIGHT: 00000
47    FLAGS      : NOTARGETAPP
48   IPADDR..PORT: 10.42.105.154..53
49    SYSTEM NAME: MVS209    PROTOCOL  : TCP  AVAIL     : YES
50    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00021
51    FLAGS      : NODATA
52 10 OF 10 RECORDS DISPLAYED
```

Figure 102. Sample output for the MODIFY procname,DISPLAY,LB,INDEX=lbindex command

Group flags - BASEWLM, BASEWLM*, and SERVERWLM: The BASEWLM flag indicates that WLM system weight recommendations were configured or defaulted for this group, and are being used to calculate the net weight. The BASEWLM* flag indicates that server-specific WLM recommendations were configured for this group, but WLM system weight recommendations are being used instead. This occurs when at least one of the Agents owning members within the group does not support server-specific WLM recommendations. The SERVERWLM flag indicates that server-specific WLM recommendations are being used to calculate the net weight for each member in the group. Line 7 in Figure 102 shows the BASEWLM flag, line 39 shows the BASEWLM* flag, and line 25 shows the

SERVERWLM flag. For more information on the BASEWLM, BASEWLM*, and SERVERWLM flags, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member flags - LBQ and OPQ: The LBQ flag indicates that the load balancer has quiesced the member. For details on what this entails, see "Stopping or resuming workload distribution to particular members (QUIESCE and ENABLE)" on page 1039. Do not confuse this with the OPQ flag, which indicates that the z/OS operator has quiesced the member at the z/OS Agent. In both cases, the member is ineligible for future workloads through the external load balancer. Line 43 in Figure 102 on page 1035 shows the LBQ flag and the OPQ flag. For more information on the LBQ and OPQ flags, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member flags - NOTARGETSYS, NOTARGETIP, and NOTARGETAPP: These flags indicate that the Advisor will advise the load balancer that the member should not currently receive new workload requests because a resource is unavailable. If the IP address in the member is not present on any TCP/IP stack within the sysplex, NOTARGETSYS is displayed for that member. This flag can also appear if the Agent owning the IP address in the member has lost contact with the Advisor or has yet to be started. There might be rare instances where the load balancer might decide to go ahead and route workload requests to members that have the NOTARGETSYS flag displayed, if it has no better candidates within the group to route workload requests to. If the member represents an application member and the application is not active, NOTARGETAPP is displayed for that member. If the member is a system member and the IP address is unusable, NOTARGETIP is displayed. This includes distributed VIPAs (DVIPAs), deprecated IPv6 addresses, and unavailable IPv6 addresses. If you ever see the NOTARGETIP flag, you should update the IP address in the member at the load balancer. If application members are coded with any of these addresses, you will always see NOTARGETAPP displayed for these members. Line 19 in Figure 102 on page 1035 shows the NOTARGETSYS flag, line 23 shows the NOTARGETIP flag, and line 47 shows the NOTARGETAPP flag. For more information on the NOTARGETSYS, NOTARGETIP, and NOTARGETAPP flags, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member flag - NODATA: This flag is transient and indicates that not enough time has elapsed for the reporting Agent to calculate a Communications Server weight for the member. Therefore, only the WLM weight is used to calculate the net weight, until such time that the Agent can report a Communications Server weight. Until that time, the CS WEIGHT is displayed as 100. When a Communications Server weight has been calculated and transmitted to the Advisor, the NODATA flag is turned off. This flag appears when new members are registered by the load balancer, when the target application that the member represents first becomes active, or shortly after a target application has been moved within the sysplex. Line 51 in Figure 102 on page 1035 shows the NODATA flag. For more information on the NODATA flag, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member field - AVAIL: This field is always displayed for each registered member, and has a value of YES or NO. YES indicates that the member is available for workload request distribution. NO indicates that the member is not available for workload request distribution. For the member to be available for workload request distribution, the target application must be active (if the member represents an application member), an Agent must be active on the target system and connected to the Advisor, and the member must not be quiesced by the z/OS

operator or by the load balancer. Line 9 in Figure 102 on page 1035 shows an example of the AVAIL field set to YES, and line 17 shows an example of the AVAIL field set to NO. For more information on the AVAIL field, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member field - NET WEIGHT: This field is always present for each registered member, and is the only weight that the load balancer actually receives. It is calculated by applying the Communications Server weight as a percentage of the WLM weight, and then the weight is normalized within its group. Normalization involves reducing the weight values while largely preserving the ratios between the weights. Normalization is performed within a group only if there is more than one available member within the group. Net weights can range from 0 to 64. A higher net weight means that member is capable of receiving more work than a member within the same group that has a lower weight. There are certain situations where the net weight can be 0 when neither the WLM weight or the Communications Server weight is 0. This can happen if the member has been quiesced by the z/OS operator or the load balancer. Conversely, there is one case where the WLM weight or the Communications Server weight can be zero, but the net weight is nonzero. This can happen if the net weight of every member in the group calculates to zero, and at least one member of the group is available. In this case, the net weights of all of the available members in the group are changed to 1 to force round-robin distribution among the members in the group, rather than to stop sending new workloads to the group entirely. Line 10 in Figure 102 on page 1035 shows an example of the NET WEIGHT field set to a nonzero value, while line 18 shows an example of the field set to zero. For more information on the NET WEIGHT field, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member field - WLM WEIGHT: This field is always present for each registered member, and represents the desirability of the system owning the member, relative to the other systems in the sysplex (system weight), or a measure of how well the individual application is meeting its WLM policies (server-specific weight). Like the Communications Server weight and net weight, a higher value means it is more desirable. WLM weights can range from 0 to 64. The WLM weight is used as a key component of the net weight. Line 10 in Figure 102 on page 1035 is one of many lines that contain the WLM WEIGHT field. For more information on the WLM WEIGHT field, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member field - CS WEIGHT: This CS WEIGHT field is always present for each registered member, and represents the health of the server with respect to its ability to satisfy recent requests. As with the WLM weight and the net weight, the higher the value the better the health. The Communications Server weight can range from 0 to 100, and is used as a component of the net weight. Line 10 in Figure 102 on page 1035 is one of many lines that contains the CS WEIGHT field. For more information on the CS WEIGHT field, refer to *z/OS Communications Server: IP System Administrator's Commands*.

MODIFY *procname*,DISPLAY,MEMBERS,DETAIL

This command displays details about members that are owned by the Agent. In Figure 103 on page 1038, the line numbers appearing in the left margin are for reference purposes only, and are used in the description following the figure. For information about every field displayed by the Agent MODIFY command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

```
1  MODIFY LBAGENT,DISPLAY,MEMBER,DETAILS
2  EZD1245I MEMBER DETAILS
3  LB INDEX      : 00        UUID      : 637FFF175C
4   GROUP NAME   : SYSTEMFARM
5    GROUP FLAGS : BASEWLM
6    IPADDR..PORT: 10.42.105.154..0
7     TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : 000
8     FLAGS      :
9     JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
10    IPADDR..PORT: 10:1::4:5..0
11    TCPNAME    : TCPCS5    MATCHES   : 000  PROTOCOL  : 000
12    FLAGS      :
13    JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
14  GROUP NAME   : UDP_SERVER_FARM
15   GROUP FLAGS : SERVERWLM
16   IPADDR..PORT: 10.42.105.154..7777
17    TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : UDP
18    FLAGS      : ANY
19    JOBNAME    : TESTD1    ASID      : 0035 RESOURCE  : 000000A3
20   IPADDR..PORT: 2001:DB8::10:5:6:2..7777
21    TCPNAME    : TCPCS2    MATCHES   : 001  PROTOCOL  : UDP
22    FLAGS      : ANY V6
23    JOBNAME    : TESTD2    ASID      : 002A RESOURCE  : 00000031
24 4 OF 4 RECORDS DISPLAYED
```

Figure 103. Sample output for the MODIFY procname,DISPLAY,MEMBERS,DETAIL command

Member flag - ANY: The ANY flag means that the application represented by the port in the member is bound to INADDR_ANY or IN6ADDR_ANY. This means that in an INET (one TCP/IP stack) environment, any externally available IP address owned by that TCP/IP stack might be used to reach the target application, not just the IP address coded in this member. Therefore, there is the potential that multiple members might exist in the load balancer or other load balancers that actually represent the same application, if members were coded with the same port, protocol, and an IP address owned by the same TCP/IP stack. You need to be aware of this if you want to issue operator commands to quiesce that application. If this were the case, quiescing the application at the port level, but specifying the individual IP address of the member, might not quiesce all new workload requests to that application. Quiescing the application at the port level without specifying an IP address would be required to accomplish that task. If the application is running in a CINET (multiple TCP/IP stack) environment, any externally available IP address on the z/OS system can be used to reach the target application, unless the application has stack affinity. If the application has stack affinity, the Advisor only indicates the member is available if the IP address coded in the member belongs to the TCP/IP stack that the application has affinity to. Line 18 in Figure 103 shows an example of this flag. For more information on the ANY flag, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Member flag - V6: The V6 flag indicates that the application that the member refers to has set the IPV6_V6ONLY socket option. This socket option disallows connections to the server application using an IPv4 address as the destination IP address when the server application has bound to IN6ADDR_ANY. If a member is coded with an IPv4 address and intends to represent an application that has the IPV6_V6ONLY socket option set, the member will not be available for workload balancing and the V6 flag is displayed for this member. Conversely, for any member that represents this target application and is coded with an IPv6 address that can be used to reach the target application, the member will be available for workload balancing and the V6 flag is displayed. Line 22 in Figure 103 shows an example of this flag. For more information on the V6 flag, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Stopping or resuming workload distribution to particular members (QUIESCE and ENABLE)

At least one and possibly two methods exist to stop sending new workload requests to particular members, referred to as quiescing. Quiescing does not disrupt existing connections with the target applications, but does prevent new workload requests from being distributed to those members from load balancers. Requests sent directly to applications that do not go through the load balancer are unaffected by quiescing. Quiescing certain members can be useful for a planned outage of a particular system in the sysplex, a particular TCP/IP stack, a particular application, or a homogeneous group of applications, such as all HTTP servers. The first method of quiescing is done using the MODIFY *procname*,QUIESCE operator command available at each Agent. The second potential method is through the load balancer administrator, if the load balancer implementation supports this function. Only the MODIFY *procname*,QUIESCE command is described in detail in this section.

The MODIFY *procname*,QUIESCE command is available only on the Agents, because they own the IP addresses that belong to TCP/IP stacks on that z/OS system. Therefore, the scope of the quiesce operation cannot affect members that are not owned by the Agent that is issuing the command.

There are three major scopes or target options for the MODIFY *procname*,QUIESCE command:

- SYSTEM

  Every member owned on that z/OS system can be quiesced
- TCPNAME=*tcpname*

  Every member on one of the Agent's TCP/IP stacks can be quiesced
- PORT=*portnum*

  All members using a particular port can be quiesced

The last target option, quiescing by port, enables you to refine the quiesce to an individual member instead of quiescing all members using the port. For more information on how to specify individual members using the Agent's MODIFY *procname*,QUIESCE command, refer to *z/OS Communications Server: IP System Administrator's Commands*.

When ready for new workloads, use the MODIFY *procname*,ENABLE command to make the quiesced members available again. Like the quiesce command, this command is only an Agent command, and also has the same target options as the quiesce command.

The quiesce and enable commands are hierarchical. The system level (SYSTEM target option) is at the top of the hierarchy, the stack level (TCPNAME=*tcpname* target option) is the next highest, and the member level (PORT=*portnum* target option) is the lowest.

Rules:

- In the quiesce and enable hierarchy, a member quiesced at one level of the hierarchy cannot be enabled at a lower or higher level of the hierarchy. For example, if the MODIFY *procname*,QUIESCE,TCPNAME=*tcpname* command has been issued, a MODIFY *procname*,ENABLE,SYSTEM command or a MODIFY *procname*,ENABLE,PORT=*portnum* command will not be accepted.

- An enable command must be issued at the same level of the hierarchy as the last quiesce command that affected the member.
- When a member has been quiesced at one level of the hierarchy, it can be quiesced at a higher level of the hierarchy. This promotes the quiesce level of the member from the lower hierarchy level to the higher hierarchy level, and any history of it being quiesced at the lower level of the hierarchy is erased. When subsequently issuing an enable command in these circumstances, the enable command must be issued at the higher, promoted level of the hierarchy. For example, if member A was quiesced at the port level, a subsequent command to quiesce all members of the TCP/IP stack would be accepted. Furthermore, only a subsequent enable command at the TCP/IP stack level would be accepted to re-enable the member.

Table 44 shows which quiesce and enable commands are valid for a member after a previous quiesce command has affected that same member. A dot at the intersection of the prior command column and the current command row indicates that the current command would be valid for that member. Absence of a dot indicates that the current command would not be valid after the prior command had affected that member.

Table 44. Allowed quiesce and enable command sequences for members

		Prior command		
		QUIESCE,SYSTEM	QUIESCE,TCPNAME=	QUIESCE,PORT=
Current command	QUIESCE,SYSTEM		●	●
	QUIESCE,TCPNAME=			●
	QUIESCE,PORT=			
	ENABLE,SYSTEM	●		
	ENABLE,TCPNAME=		●	
	ENABLE,PORT=			●

Rules:

- A quiesce command is rejected if any member it applies to has already been quiesced by a command at a higher level of the hierarchy. For example, if you are running in a CINET configuration and you quiesce all members under stack A, which includes a member on stack A that used port 80, you cannot subsequently issue a quiesce command at the port level hoping to quiesce members using port 80 on stacks B and C, because the member on stack A that used that port was already quiesced at the stack level. Because the command would fail for one member, the entire command fails for all members.
- Quiesce commands at the system and stack level apply to currently registered members and also members that are registered in the future, provided that the IP addresses of those future members are owned by the Agent that issued the command. For quiesce commands issued at the stack level, the specified stack must exist at the time the command is issued.
- Quiesce commands at the port level can apply to members registered in the future, if a member currently exists at that port number. For example, if a member is registered by one load balancer at port 80 and the Agent operator quiesces all members at port 80, and then another load balancer registers that same member (same IP address, port, and protocol), the newly registered member would inherit the quiesce performed at the port level.

When a member represents a shareport group (that is, multiple server application instances sharing the same TCP port on the same TCP/IP stack), members cannot be defined to distinguish between the individual server application instances. That is, the combination of the IP address, port, and protocol represents all of the applications sharing the port. Therefore, you cannot selectively quiesce workload requests to only some of the applications sharing the port. Consequently, if you quiesce a member that represents a shareport group, all of the application instances in the group are quiesced.

If an Agent is stopped or fails and is restarted, the quiesce states of any members it might have previously owned are lost. If this occurs, reenter the appropriate quiesce commands to regain the quiesce states that existed during the previous instance of the Agent.

If a member is moved from one system in a sysplex to another using sysplex functions such as VIPA backup, the quiesce state does not move with it. For example, if quiesced member A is using IP address 1.1.1.1 on system X and system X fails, IP address 1.1.1.1 could be moved to system Y. Member A would no longer be quiesced on the new system and, assuming that the application that member A is assigned to is available and active on system Y, new workload requests would be distributed to this application.

If you plan to take a sysplex system out of service, simply stopping the Agent running on that system is not always a wise alternative to issuing a system-level quiesce command on that system's Agent. If an Agent is simply shut down, there are rare cases where a load balancer might choose to continue routing new workload requests to the applications on that system.

z/OS Load Balancing Advisor configuration example

This section includes a specific configuration example of the z/OS Load Balancing Advisor, two Load Balancing Agents, and some customization of PROFILE.TCPIP considerations.

In this example, as shown in Figure 104 on page 1042, load balancer LB1 distributes workload requests to two z/OS systems in a sysplex, SYSA and SYSB. SYSA is a CINET configuration with two TCP/IP stacks. SYSB is an INET configuration with one TCP/IP stack. The load balancer is connected to a LAN that also connects to each target TCP/IP stack in the sysplex, including the TCP/IP stack where the Advisor is running. All addresses in this example are IPv4, but the Advisor and Agents are enabled for IPv6.

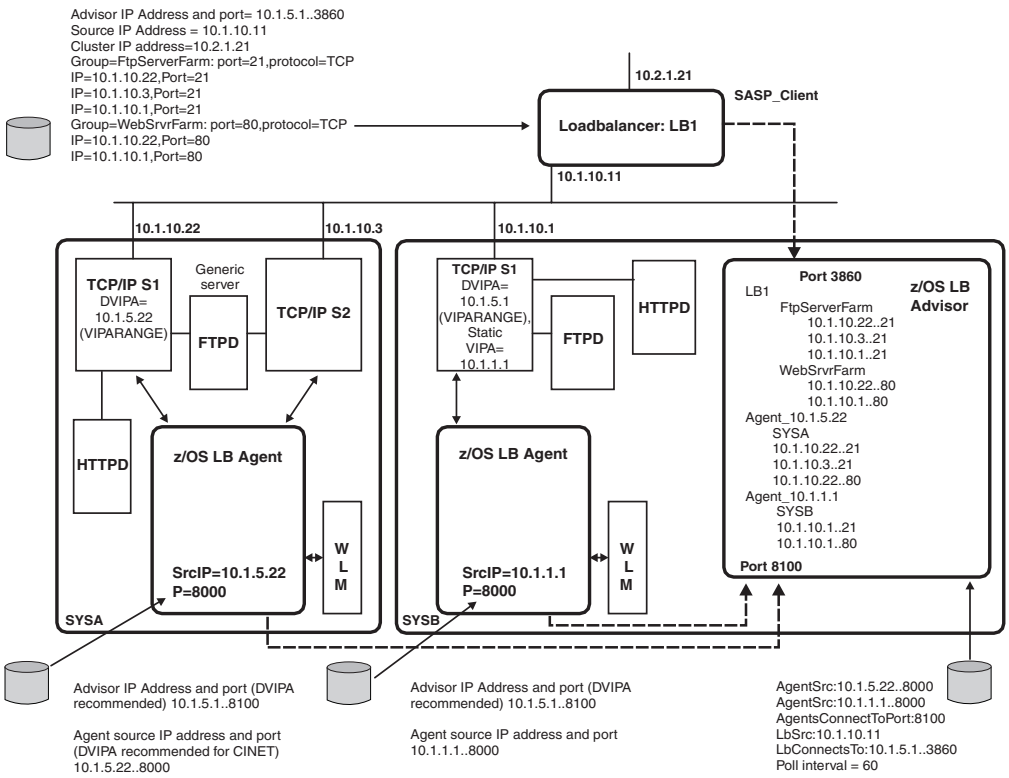


Figure 104. z/OS Load Balancing Advisor configuration example

Load balancer configuration details

The load balancer distributes two types of workload requests in this example, FTP and HTTP traffic, both of which use the TCP protocol. (Keep in mind that UDP workload requests can also be distributed if desired.) On SYSA, one FTP server is running, which is shared as a generic server between the two TCP/IP stacks. Also on SYSA, an HTTP server is bound to stack S1. On SYSB, an FTP and HTTP server are running.

The load balancer must be configured with the IP address and port of the Advisor's load balancing connection socket. In this example, the TCP/IP stack on SYSB has defined 10.1.5.1 as a dynamic VIPA. In addition, this same address and port is defined to the Advisor on the lb_connection_v4 statement in the Advisor's configuration file. The load balancer also uses one of its interfaces to communicate with the Advisor. The IP address assigned to this interface must be coded in the Advisor's lb_id_list statement. In this example, the load balancer uses the interface assigned to the address 10.1.10.11. For information on how to determine which interface the load balancer will use to communicate with the Advisor, consult the specific load balancer documentation.

The load balancer advertises its cluster IP address, 10.2.1.21, so that clients that want to connect to specific applications in the sysplex will connect to this cluster IP address as a proxy. Once the connection reaches the load balancer, to determine the actual target of the request, the load balancer consults the information that the Advisor has provided as well as possibly examining the content of the packet. The load balancer might substitute the target IP address in the packet header with the IP address of the member that is best suited to receive the new workload requests,

or simply forward the packet as is using the proper MAC address. For example, if a connection request came to the load balancer (10.2.1.21) for port 21 using protocol TCP, the load balancer would forward the packet to either 10.1.10.22, 10.1.10.3, or 10.1.10.1, depending upon which member is the better candidate at that point in time.

To distribute FTP workload requests to the sysplex, a group called FtpServerFarm is defined to the load balancer. The load balancer maps the cluster IP address, 10.2.1.21, port 21, and protocol TCP to this group. In other words, if the load balancer receives a TCP connection with destination 10.2.1.21, port 21, it consults this group to find a member to which it can forward the connection. Within this group are three members that can handle FTP connections, representing target applications within the sysplex. One target can be reached at 10.1.10.22 on port 21, the second at 10.1.10.3 on port 21, and the third at 10.1.10.1 on port 21. The target applications represented by these members do not necessarily all have to be available at all times. The load balancer avoids trying to forward connections to target applications that are not currently available. Therefore, the list of target applications represented by the members in the group should be the entire set of possible members that could handle this workload, now or in the foreseeable future.

To distribute HTTP workload requests to the sysplex, a group called WebSrvrFarm is defined to the load balancer. The load balancer maps the cluster IP address, 10.2.1.21, port 80, and protocol TCP to this group. In other words, if the load balancer receives a TCP connection with destination 10.2.1.21, port 80, it consults this group to find a member to which it can forward the connection. Within this group are two members that can handle HTTP connections, representing target applications within the sysplex. One target can be reached at 10.1.10.22 on port 80, and the second can be reached at 10.1.10.1 on port 80.

Advisor configuration details

There are two aspects of Advisor configuration, the Advisor configuration file itself, and the underlying PROFILE.TCPIP changes that go along with the remainder of the z/OS Load Balancing Advisor system configuration.

Following is an example Advisor configuration file:

```
debug_level 7                     # Error, Warning, Events-- the default

update_interval 60                # Agents update every  minute-- the default

wlm serverwlm                     # Request server-specific WLM weights

port_list
{
   21 wlm basewlm                 # Use system WLM weights for FTP
}

lb_connection_v4 10.1.5.1..3860   # DVIPA load balancer connects to

lb_id_list
{
   10.1.10.11                     # Load balancer's SASP client interface
}

agent_connection_port 8100        # Port Agents connect to

agent_id_list
```

```
{
   10.1.1.1..8000                # This system's Agent source
   10.1.5.22..8000               # SYSA's Agent source
}
```

In this example Advisor configuration file, the debug level is set to 7 in the optional debug_level statement. The value of 7 is the default value, so this statement is redundant but is shown for completeness. At the default level of 7, messages are written to the log if they are at error, warning, or event level. Messages at other debug levels, such as info or debug, are not written to the log file.

The optional update_interval statement is set to 60 seconds, which is also the default. This means that each Agent updates the Advisor with new information every minute. For some load balancer implementations, depending upon load balancer configuration, it might also determine how often the load balancer is updated with new information from the Advisor.

The wlm statement specifies the default WLM recommendation type to be used for all groups when calculating the net weights. The value of severwlm indicates that server-specific WLM recommendations will be requested of each Agent, unless overridden by the port_list statement (see next paragraph). Although server-specific WLM is the WLM recommendation type to be used for all groups except FTP in this example, there is still a possibility that WLM system weights might have to be used for some or all groups. For further details on the serverwlm value on the wlm statement, refer to *z/OS Communications Server: IP Configuration Reference*.

The port_list statement contains one port number, 21. The wlm keyword indicates that the WLM recommendation type will be overridden for all members using port 21 (FTP), to use WLM system weights rather than server-specific WLM weights when calculating the net weight. Multiple port numbers can appear in the port_list statement on separate lines, if you want to use WLM system weights for other groups of members.

The lb_connection_v4 statement includes DVIPA 10.1.5.1 and port 3860 (the default) as the address and port that load balancers use to connect to the Advisor. The load balancer specifies this address and port when defining the location of the Advisor.

The lb_id_list statement contains the address 10.1.10.11, which is the source IP address of the load balancer when the load balancer connects to the Advisor as a SASP client. If more than one load balancer is used to distribute workload requests to the sysplex, each load balancer needs to be represented in this statement list.

The agent_connection_port statement specifies that port 8100 is used to listen for connections from Agents in the sysplex. This same port appears on each Agent's advisor_id statement. This port is reserved on the TCP/IP stack that the Advisor runs on, and on any TCP/IP stacks that the Advisor could be moved to in the event of failure. Using this port, the Advisor opens a listening socket on the IPv4 or IPv6 unspecified address (0.0.0.0 or ::, respectively), depending upon the TCP/IP stack's IPv6 capability.

The agent_id_list statement contains the source IP address and port of each Agent in the sysplex. The 10.1.1.1 address and the associated port of 8000 represent the source IP address and port that the Agent on SYSB uses to communicate with the Advisor. This same address and port combination appears on the agent_connection

statement in the Agent's configuration file on SYSB. The 10.1.5.22 address and the associated port of 8000 represent the source IP address and port that the Agent on SYSA uses to communicate with the Advisor. This same address and port combination appears on the agent_connection statement in the Agent's configuration file on SYSA.

Agent configuration file on SYSB

Following is the example Agent configuration file on SYSB:

```
debug_level 7                   # Error, Warning, Events

advisor_id 10.1.5.1..8100       # DVIPA of Advisor Agent connects to

host_connection 10.1.1.1..8000  # Source address and port this Agent
                                #   uses to connect to the Advisor
```

In this example Agent configuration file for SYSB, the debug level is set to 7 in the optional debug_level statement. The debug_level statement for an Agent functions similarly to the way it functions for the Advisor.

The advisor_id statement is configured with the address 10.1.5.1 and port 8100. This tells the Agent which address and port the Advisor is using for connections from Agents. The address 10.1.5.1 is configured as a DVIPA on the Advisor's TCP/IP stack. The port of 8100 also appears on the agent_connection_port statement in the Advisor's configuration file, and is also reserved on the Advisor's TCP/IP stack.

The host_connection statement is configured with the address 10.1.1.1 and port 8000. This is the source IP address and port that this Agent uses when connecting to the Advisor. The address is defined as a static VIPA on the TCP/IP stack that the Agent will run on. This address and port must also be specified in the agent_id_list statement in the Advisor's configuration file. The port, 8000, is also reserved in PROFILE.TCPIP of stack S1 on system SYSB.

Agent configuration file on SYSA

Following is the example Agent configuration file on SYSA:

```
debug_level 7                       # Error, Warning, Events

advisor_id 10.1.5.1..8100           # DVIPA of Advisor Agent connects to

host_connection 10.1.5.22..8000     # Source DVIPA and port this Agent
                                    #   uses to connect to the Advisor
```

This Agent configuration file is for the Agent running on SYSA. The debug_level and advisor_id statements are identical to the Agent configuration file on SYSB. The host_connection statement is configured with the address 10.1.5.22 and port 8000. This is the source IP address and port that this Agent uses when connecting to the Advisor. This address and port must also be specified on the agent_id_list statement in the Advisor's configuration file. The port, 8000, is also reserved in PROFILE.TCPIP of stack S1 and S2 on system SYSA. This address is defined as a dynamic VIPA on both TCP/IP stacks on SYSA, in the event that one of the TCP/IP stacks fails.

Customization of PROFILE.TCPIP

Each TCP/IP profile in the sysplex must be updated to accommodate the z/OS Load Balancing Advisor.

The updated portion of PROFILE.TCPIP for stack S1 on system SYSB follows:

```
VIPADYNAMIC
   ;Address LB & Agents use to reach Advisor fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

DEVICE VIPA41 VIRTUAL 0    ; Static VIPA for Agent's source address
   LINK LVIPA41 VIRTUAL 0 VIPA41
   HOME 10.1.1.1 LVIPA41

PORT
   3860 TCP LBADV          ; SASP Workload Advisor (LB connections)
   8100 TCP LBADV          ; SASP Workload Advisor (Agent connections)
   8000 TCP LBAGENT        ; SASP Workload Agent (Advisor connection)
```

In this example, the address 10.1.5.1 would be within the subnetwork that has been reserved for dynamic VIPAs in the VIPARANGE statement. The load balancer and the Agents use this address to reach the Advisor. Using a dynamic VIPA (DVIPA) facilitates the movement of the Advisor to another TCP/IP stack in the event of failure. This address is defined in the lb_connection_v4 statement in the Advisor's configuration file, in the load balancer as the location of the Advisor [known generically as the SASP Global Workload Manager (GWM)], and on the advisor_id statement in each of the Agent's configuration files.

The address 10.1.1.1 is a static VIPA. The Agent on this system uses this address as its source IP address. Since SYSB is a single stack system (INET), a static VIPA is sufficient. If this were a CINET system like SYSA, a DVIPA would be best. This address appears on the agent_id_list statement in the Advisor's configuration file, as well as on the agent_connection statement in the Agent's configuration file on SYSB.

The ports used for the Advisor and Agent are reserved, as advised. Port 3860 is reserved for the Advisor and is used to communicate with load balancers. This port appears on the lb_connection_v4 statement in the Advisor's configuration file. Port 8100 is also reserved for the Advisor, and is the port that the Agents use to connect to the Advisor. This port appears on the agent_connection_port statement in the Advisor's configuration file, as well as on the advisor_id statement in each of the Agents' configuration files. Port 8000 is reserved for the Agent on this system and is used as the source port for the connection with the Advisor. This port appears on the agent_id_list statement in the Advisor's configuration file, as well as on the agent_connection statement in the Agent's configuration file on this system.

The updated portion of PROFILE.TCPIP for stack S1 on system SYSA follows:

```
VIPADYNAMIC
   ;Address Agent uses as source will fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

PORT
   3860 TCP LBADV          ; SASP Workload Advisor LB connections,
                           ;  in case Advisor is moved to this stack
   8100 TCP LBADV          ; SASP Workload Advisor Agent connections,
                           ;  in case Advisor is moved to this stack
   8000 TCP LBAGENT        ; SASP Workload Agent Advisor connection
```

In Figure 104 on page 1042, the DVIPA that the Agent uses as a source IP address on this system is shown belonging to stack S1. It could just as easily belong to stack S2, but for the purposes of this example the DVIPA belongs to stack S1.

In this updated portion of PROFILE.TCPIP, the address that the Agent uses as a source address when connecting to the Advisor, 10.1.5.22, is within the subnetwork that has been reserved for dynamic VIPAs on the VIPARANGE statement. Using a dynamic VIPA (DVIPA) facilitates the movement of the Agent to another TCP/IP stack on the same system in the event of failure. This address is defined on the host_connection statement in this Agent's configuration file, and in the agent_id_list statement in the Advisor's configuration file. The DVIPA that the Advisor would use, should the Advisor be moved to this stack, would also fall within this subnetwork.

The port that is used for the Agent is reserved, as advised. Additionally, ports that the Advisor would use if it were to be moved to this TCP/IP stack are also reserved.

The updated portion of PROFILE.TCPIP for stack S2 on system SYSA follows:

```
VIPADYNAMIC
   ;Address Agent uses as source will fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

PORT
   3860 TCP LBADV          ; SASP Workload Advisor LB connections,
                           ;  in case Advisor is moved to this stack
   8100 TCP LBADV          ; SASP Workload Advisor Agent connections,
                           ;  in case Advisor is moved to this stack
   8000 TCP LBAGENT        ; SASP Workload Agent Advisor connection
```

This updated portion of the TCP/IP profile is identical to that of stack S1 on SYSA. This TCP/IP stack is capable of supporting the Advisor and the Agent running on this z/OS system, should it be necessary to move either to this TCP/IP stack.

Example displays

The following display is from the SYSB Advisor:

```
F LBADV,DISPLAY,LB,INDEX=03
EZD1243I LOAD BALANCER DETAILS 738
LB INDEX      : 03        UUID      : 4C4231
 IPADDR..PORT : 10.1.10.11..50004
 HEALTH       : 7F        FLAGS     : PUSH
 GROUP NAME   : FTPSERVERFARM
  GROUP FLAGS :  BASEWLM
  IPADDR..PORT: 10.1.10.22..21
   SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
   FLAGS      :
  IPADDR..PORT: 10.1.10.3..21
   SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
   FLAGS      :
  IPADDR..PORT: 10.1.10.1..21
   SYSTEM NAME: SYSB      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00031     CS WEIGHT : 100  NET WEIGHT: 00001
   FLAGS      :
 GROUP NAME   : WEBSRVRFARM
  GROUP FLAGS : SERVERWLM
  IPADDR..PORT: 10.1.10.22..80
   SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
   FLAGS      :
  IPADDR..PORT: 10.1.10.1..80
```

```
   SYSTEM NAME: SYSB      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00031     CS WEIGHT : 100  NET WEIGHT: 00001
   FLAGS      :
5 OF 5 RECORDS DISPLAYED
```

The following display is from the SYSB Agent:

```
F LBAGENT,DISPLAY,MEMBERS,DETAIL
EZD1245I MEMBER DETAILS 741
LB INDEX      : 03        UUID      : 4C4231
 GROUP NAME   : FTPSERVERFARM
  IPADDR..PORT: 10.1.10.1..21
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : FTPD1     ASID      : 001D RESOURCE  : 0000001A
 GROUP NAME   : WEBSRVRFARM
  IPADDR..PORT: 10.1.10.1..80
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : HTTPD6    ASID      : 0030 RESOURCE  : 00000053
2 OF 2 RECORDS DISPLAYED
```

The following display is from the SYSA Agent:

```
F LBAGENT,DISPLAY,MEMBERS,DETAIL
EZD1245I MEMBER DETAILS 598
LB INDEX      : 03        UUID      : 4C4231
 GROUP NAME   : FTPSERVERFARM
  IPADDR..PORT: 10.1.10.22..21
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : FTPD1     ASID      : 002C RESOURCE  : 0000007D
  IPADDR..PORT: 10.1.10.3..21
   TCPNAME    : S2        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : FTPD1     ASID      : 002C RESOURCE  : 00000047
 GROUP NAME   : WEBSRVRFARM
  IPADDR..PORT: 10.1.10.22..80
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : HTTPD5    ASID      : 0033 RESOURCE  : 0000005D
3 OF 3 RECORDS DISPLAYED
```

Chapter 20. Network management

This chapter describes how to configure:

- Simple Network Management Protocol agent (osnmpd)
- z/OS UNIX snmp or osnmp command
- NetView SNMP command
- SNMP subagents
- Open Systems Adapter (OSA) support
- Trap forwarder daemon

Before you configure, read "Understanding search orders of configuration information" on page 28. It covers important information about data set naming and search sequences.

SNMP overview

SNMP is a set of protocols that describes management data and the protocols for exchanging that data between heterogeneous systems. The protocols include both the description of the management data, defined in the Management Information Base (MIB), and the operations for exchanging or changing that information. By implementing common protocols, management data can be exchanged between different platforms with relative ease.

SNMP defines an architecture that consists of:

- Network management applications
- Network management agents and subagents
- Network elements, such as hosts and gateways

The SNMP network management application can ask agents for specific information about network elements. Conversely, agents can tell the network management application when something happens to one or more network elements. The protocol used between the network management application and agents is SNMP. The transport protocol for SNMP requests is the User Datagram Protocol (UDP).

SNMP defines both the network management data and the ways in which the data is retrieved or changed by the network management application. Examples of network management data include device definitions, counts of packets received at the IP layer, TCP connection data, and so forth. The information about the network elements is stored in the Management Information Base (MIB), which is supported by the SNMP agent and its subagents. A MIB variable (or MIB object) is a specific instance of data in a collection of objects related to a common management area. The collection is called a MIB module. Each MIB object is identified by an object identifier (OID), which is a dotted-decimal value, and by a textual name. For example, the sysUpTime MIB object, which indicates how long ago the SNMP agent initialized, is defined as follows:

- Textual name

  sysUpTime
- Object identifier (OID)

  1.3.6.1.2.1.1.3

The z/OS Communications Server Network Management agent and subagents, also called SNMP agent and SNMP subagents, support many standard (RFC-based) MIB modules. The SNMP TCP/IP subagent also supports enterprise-specific MIB modules. For information on MIB modules supported by the z/OS Communications Server SNMP agent and subagents, see the SNMP agent capabilities statement, shipped as HFS file /usr/lpp/tcpip/samples/mvstcpip.caps. Additionally, enterprise-specific MIBs are documented in the /usr/lpp/tcpip/samples HFS directory. See "TCP/IP subagent" on page 1051 for more information on enterprise-specific MIB modules supported by the TCP/IP subagent. For a complete list of MIB objects supported by the SNMP agent and subagents shipped with z/OS Communications Server, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Network management application

The stations that monitor network elements run network management applications. In z/OS Communications Server, the **snmp** (or **osnmp**) command provides SNMP network management from the z/OS UNIX shell. The NetView SNMP command provides network management from the NetView command line. The z/OS UNIX **snmp** and **osnmp** commands perform exactly the same function. Both commands can be used to retrieve or change data from SNMP and monitor for asynchronous events known as notifications. An unconfirmed notification is called a trap. A confirmed notification is called an inform.

For information about the syntax and use of the snmp and NetView SNMP commands, refer to *z/OS Communications Server: IP System Administrator's Commands*.

SNMP agent

In z/OS Communications Server, the SNMP agent is a z/OS UNIX application. It supports SNMPv1, SNMPv2c, and SNMPv3. SNMPv2c offers protocol enhancements such as the GETBULK operation. SNMPv3 provides a network management framework that allows the use of user-based security in addition to, or instead of, the community-based security supported in SNMPv1 and SNMPv2c. The view-based access control model supported in SNMPv3 allows granular access control for MIB objects with either the user-based or community-based security models. SNMPv3 also enables dynamic changes to the SNMP agent configuration.

Overview of z/OS Communications Server SNMP version 3

SNMP version 3 (SNMPv3) is the standards-based solution to previous SNMP security. SNMPv3 is defined in RFCs 3411 through 3415 [see Appendix F, "Related protocol specifications (RFCs)," on page 1223]. The SNMPv3 architecture is modularized so that portions of it can be enhanced over time without requiring the entire architecture to be replaced. SNMPv3 defines a framework that consists of (among other things):

- A message processing model (SNMPv3)
- A security model (user-based security)
- An access control model (view-based access control)

The framework is structured so that multiple models can be supported concurrently and replaced over time. For example, although there is a new message format for SNMPv3, messages created with the SNMPv1 and SNMPv2c formats can still be supported. Similarly, the user-based security model can be supported concurrently with community-based security models previously used.

For information about migrating z/OS SNMP configuration files from SNMPv1 and SNMPv2c to SNMPv3, refer to Migrating z/OS SNMP to SNMPv3.

SNMP subagents

A subagent extends the set of MIB variables supported by an SNMP agent. z/OS Communications Server supports the following subagents:

- TCP/IP subagent
- OMPROUTE subagent
- TN3270 Telnet subagent
- Service Level Agreement (SLA) subagent
- Network SLAPM2 subagent

For a complete list of MIB objects, refer to *z/OS Communications Server: IP System Administrator's Commands*.

The OSA-Express Direct subagent is also shipped with the z/OS Communications Server product, but is supported by the zSeries OSA Support Group.

TCP/IP subagent

The TCP/IP subagent in z/OS Communications Server is a z/OS UNIX application that runs in its own task in the TCP/IP address space. This subagent supports many standard (RFC-based) MIB objects. In addition, it supports MIB objects in the following enterprise-specific MIB modules:

- The IBM 3172 enterprise-specific MIB.
- The IBM MVS TCP/IP enterprise-specific MIB. This MIB defines objects to extend standard MIB tables, supports retrieval and change of TCP/IP address space configuration parameters, and provides management support for the environments where Asynchronous Transfer Mode (ATM) is used.

For a more detailed list of all the data supported by the TCP/IP subagent, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Both the TCP/IP subagent and the OSA-Express Direct subagent support management data for OSA-Express features. You should use the OSA-Express Direct subagent to obtain this management data, for the following reasons:

- Because the OSA-Express Direct subagent communicates directly with the OSA-Express features, it does not require the OSA/SF and IOASNMP applications.
- The OSA-Express Direct subagent provides more OSA management data than the TCP/IP subagent.
- Because of the support provided by the OSA-Express Direct subagent, there will be no new enhancements to the OSA management data provided by the TCP/IP subagent, and support for this data will eventually be removed in a future release.

If you are using the TCP/IP subagent's OSA management data support, and decide to switch to using the OSA-Express Direct subagent instead, you no longer need to start the OSA/SF address space and the OSA IOASNMP application. For more information about this subagent, see "OSA-Express Direct subagent" on page 1053. For information on the TCP/IP subagent's dependency on OSA/SF and IOASNMP, see "Step 4: Configure the Open Systems Adapter (OSA) support" on page 1074.

The TCP/IP subagent support evolves with each release. New MIB objects might be supported, and, occasionally, old ones might be removed for functions that are no longer relevant. This is particularly true for the MIB objects in the IBM MVS TCP/IP enterprise-specific MIB. This MIB's definition is shipped as HFS file /usr/lpp/tcpip/samples/mvstcpip.mi2. For details of the changes for each release, refer to the REVISION sections of this MIB module file.

The TCP/IP subagent in z/OS Communications Server provides SET support, enabling remote configuration of some TCP/IP address space parameters. The TCP/IP subagent is configured and controlled by the SACONFIG statement in the PROFILE.TCPIP data set. Systems where SNMP support is not required can disable the subagent and save system resources.

OMPROUTE subagent

The OMPROUTE subagent implements the Open Shortest Path First (OSPF) MIB variable containing OSPF protocol and state information.

The OMPROUTE subagent supports selected MIB objects defined in RFC 1850.

For a detailed description of the OMPROUTE subagent, refer to the *z/OS Communications Server: IP Configuration Reference*.

TN3270 Telnet subagent

The SNMP TN3270 Telnet subagent provides Telnet transaction data for monitored Telnet connections using the SNMP protocol. For more information about configuring the TN3270 Telnet subagent, refer to the TNSACONFIG statement in *z/OS Communications Server: IP Configuration Reference*.

Service Level Agreement (SLA) subagent

The Service Level Agreement (SLA) subagent allows network administrators to retrieve data and determine if the current set of SLA policy definitions are performing as needed or if adjustments need to be made.

Recommendation: Support for the SLA subagent (pagtsnmp) will be dropped in a future release. Therefore, using the Network SLAPM2 subagent (nslapm2) is recommended.

Network SLAPM2 subagent

The Network SLAPM2 (nslapm2) subagent contains the information that can be used to analyze network performance for respective policy.

In the future, the old SLA subagent (pagtsnmp) will become obsolete. Use the Network SLAPM2 subagent instead of the old SLA subagent for the following reasons:

- The Network SLAPM2 subagent allows better problem determination by a network manager.
- The Network SLAPM2 subagent is more scalable and has easier indexing.
- The Network SLAPM2 subagent can support end-to-end SLA performance problem isolation.
- The Network SLAPM2 subagent contains more counts and monitored values (such as loss count, accept queue delay, connection delay).
- The Network SLAPM2 subagent is simpler in that it does not require subcomponent information, which is maintained currently on a per TCP connection basis.

- The Network SLAPM2 subagent contains 64-bit counts instead of the existing 32-bit counts to avoid wrapping.
- The Network SLAPM2 subagent reduces system overhead due to the restructure of the MIB and the interface for subagent to get performance data.
- The Network SLAPM2 subagent provides transparency in supporting either IPv4 or IPv6, since this MIB no longer keeps track of IP addresses.

OSA-Express Direct subagent

The OSA-Express Direct subagent supports management data for OSA-Express features. The OSA-Express Direct subagent is shipped with the z/OS Communications Server product, but is supported by the zSeries OSA Support Group. The MVS started procedure name of this subagent is IOBSNMP.

The TCP/IP subagent also supports management data for OSA-Express features, but you should use the OSA-Express Direct subagent to obtain this management data for the following reasons:

- Because the OSA-Express Direct subagent communicates directly with the OSA-Express features, it does not require the OSA/SF and IOASNMP applications.
- The OSA-Express Direct subagent provides more OSA management data than the TCP/IP subagent.
- Because of the support provided by the OSA-Express Direct subagent, there will be no new enhancements to the OSA management data provided by the TCP/IP subagent, and support for this data will eventually be removed in a future release.

If you are using the TCP/IP subagent's OSA management data support, and decide to switch to using the OSA-Express Direct subagent instead, you no longer need to start the OSA/SF address space and the OSA IOASNMP application. For information on the TCP/IP subagent's dependency on OSA/SF and IOASNMP, see "Step 4: Configure the Open Systems Adapter (OSA) support" on page 1074. For a complete understanding of the management data provided by the OSA-Express Direct subagent, refer to *zSeries OSA-Express Customer's Guide and Reference*.

Key generation commands

The pwtokey and pwchange commands are provided to enable generation and change of keys used for authentication and encryption with SNMPv3.

For more information about pwtokey, refer to *z/OS Communications Server: IP Configuration Reference*. For information about pwchange, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Distributed Protocol Interface

The DPI is an application interface used by the SNMP agent to communicate with subagents. With DPI, you can dynamically add, delete or replace management variables supported by the SNMP agent and its subagents. z/OS Communications Server provides DPI V2.0 for z/OS UNIX C socket users and DPI V1.1 for traditional C socket users. DPI V2.0 provides additional function, making it easier to write subagents and simplifying the task of developing and administering your application.

For more information about the DPI, refer to the *z/OS Communications Server: IP Programmer's Guide and Reference*.

Trap forwarder daemon

The trap forwarder daemon forwards traps from the SNMP agent to network management applications. It listens for traps on a port, typically 162, and forwards them to all configured managers.

For more information about the trap forwarder daemon, refer to *z/OS Communications Server: IP Configuration Reference*.

Processing an SNMP request

Figure 105 illustrates the interface between TCP/IP and the implementation of SNMP.

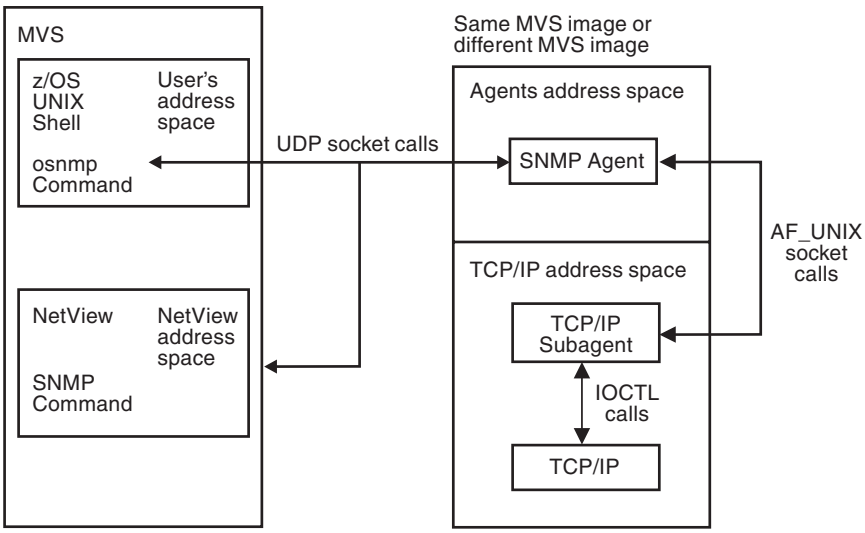


Figure 105. Overview of SNMP support

This list illustrates the sequence of events from the time you issue an SNMP command until you receive the response:

1. The user issues a NetView SNMP or z/OS UNIX snmp command.
2. The command processor validates and encodes the request in a Protocol Data Unit (PDU), and sends it to the SNMP agent.
3. The SNMP agent validates the request and, if necessary, sends it to an SNMP subagent. Requests for agent-oriented objects are handled by the agent and all others are handled by a subagent. To determine which objects are handled by the agent and which by a subagent, refer to the Management Information Base Appendix in *z/OS Communications Server: IP System Administrator's Commands*.
4. The agent sends the response to the originator of the request. The command processor displays the response.

Note: Although not shown in Figure 105, other subagents, such as the OMPROUTE subagent, the Network SLAPM2 subagent, the SLA subagent, and the TN3270 Telnet subagent shipped as part of z/OS Communications Server, also communicate with the SNMP agent using AF_UNIX socket calls or TCP socket calls from their own address spaces.

The SNMP agent and the SNMP subagents record trace information via the z/OS UNIX syslog daemon using the *daemon* facility. For detailed information regarding

syslogd and specifying the daemon facility in the /etc/syslog.conf configuration file, see "Logging of system messages" on page 56.

Deciding on SNMP security needs

The SNMP agent supports SNMPv1, SNMPv2c, and SNMPv3 security. SNMPv1 and SNMPv2c are community-based security, where a community name (or password) is passed with a request. If the community name is recognized as one that can be used by the IP address from which the request originates, the SNMP agent processes the request.

SNMPv3 provides a more powerful and flexible framework for message security and access control. Message security involves providing:

- Data integrity checking, to ensure that the data was not altered in transit
- Data origin verification, to ensure that the request or response originates from the source from which it claims to have come
- Message timeliness checking and, optionally, data confidentiality, to protect against eavesdropping

Access control is the ability to control exactly what data an individual user can read or write.

The SNMPv3 architecture introduces the User-Based Security Model (USM) for message security and the View-Based Access Control Model (VACM) for access control. The architecture supports the concurrent use of different security, access control, and message processing models. For example, community-based security can be used concurrently with USM.

USM uses the concept of a user for which security parameters (levels of security, authentication and privacy protocols, and keys) are configured at both the agent and the manager. Messages sent using USM are better protected than messages sent with community-based security, where passwords are sent in the clear and displayed in traces. With USM, messages exchanged between the manager and the agent have data integrity checking and data origin authentication. Message delays and message replays (beyond what happens normally due to a connectionless transport protocol) are protected against with the use of time indicators and request IDs. Data confidentiality, or encryption, is also available.

The use of VACM involves defining collections of data (called views), groups of users of the data, and access statements that define which views a particular group of users can use for reading, writing, or receipt in a notification.

SNMPv3 also introduces the ability to dynamically configure the SNMP agent using SNMP SET commands against the MIB objects that represent the agent's configuration. This dynamic configuration support enables addition, deletion, and modification of configuration entries either locally or remotely. Remote modification of user keys can be especially useful.

Decide on your security needs—community-based or user-based.

If you are satisfied with the security of your existing configuration, you can continue to use community-based security with no migration. If you would like to take advantage of USM or VACM, you will need to migrate your configuration. Note that USM can be used only when both the SNMP agent and the manager requesting the data support USM, as the z/OS Communications Server SNMP agent and the snmp command do. VACM can be used even for community-based

requests, but doing so requires migration of existing community name and trap destination definitions. Following is a list of the advantages and disadvantages of using each type of security.

Table 45. Security advantages and disadvantages

SNMPv1/SNMPv2c advantages	**SNMPv3 disadvantages**
Widely implemented on many platforms.	Not yet implemented on many platforms.
Easy to configure.	More robust configuration options.
SNMPv1/SNMPv2c disadvantages	**SNMPv3 advantages**
Legacy standards-based administrative model.	New standards-based administrative model.
SNMPv1 and SNMPv2c allow particular IP addresses to access all data or no data.	SNMPv3 allows a particular user to access particular data.
Not very robust (password sent in PDU).	Robust (data integrity and data origin authentication).
Any user that can read data can also change the data (for objects defined as read-write).	The ability to change data can be limited to specific users.
No data confidentiality.	Encryption available.
Configuration changes require restarting of SNMP agent.	Configuration changes for USM and VACM can be made dynamically, either locally or remotely.

For more information about security, see "Creating user keys" on page 1063.

Step 1: Configure the SNMP agent (OSNMPD)

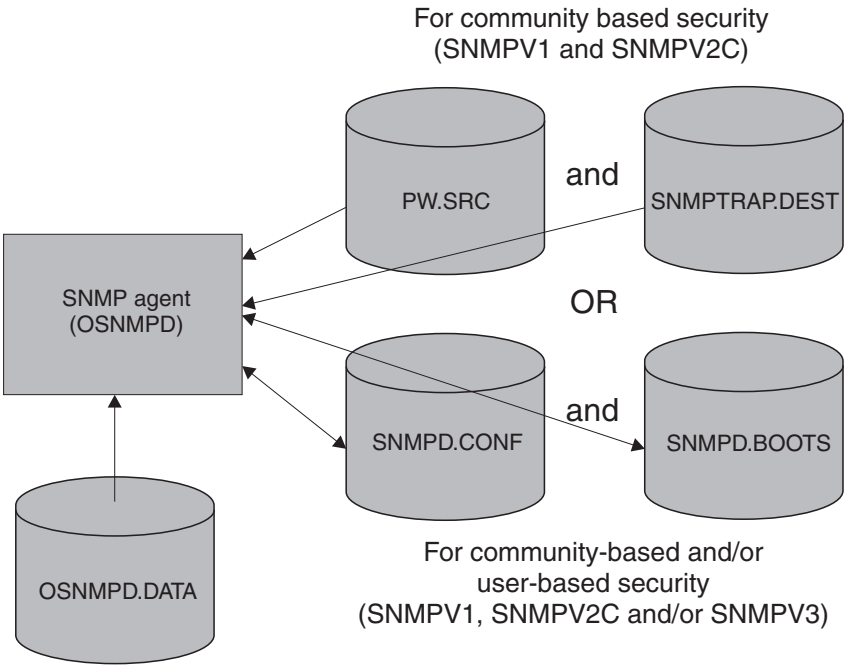


Figure 106. Configuration files for SNMP agent

Configure the SNMP agent based upon your security need. The SNMP agent accepts both SNMPv1 and SNMPv2c requests for community-based security. The SNMP agent can be configured to also use the User-based Security Model and the View-based Access Control Model. To configure the SNMP agent, perform the following tasks:

- "Provide TCP/IP profile statements"
- Depending upon whether you want to use USM and VACM, do one of the following:
  - If you are using community-based security and do not need USM or VACM, see "Provide community-based security and notification destination information" on page 1058.
  - If you want the flexibility of using USM or VACM or community-based security, see "Provide community-based and user-based security and notification destination information" on page 1061.
- "Provide MIB object configuration information" on page 1065
- Refer to *z/OS Communications Server: IP Configuration Reference* for more information about OSNMPD parameters.

Provide TCP/IP profile statements

Update the following configuration statements in *hlq*.PROFILE.TCPIP:
AUTOLOG
PORT

There are two primary TCP/IP ports used by the SNMP agent, one for receiving incoming requests and one for sending traps to managers.

The default port used by the SNMP agent to receive incoming requests is 161. If you want the agent to use port 161 for this purpose and want to insure that no other application uses this port, you must specify the following PORT statement in your profile data set:

```
PORT
  161 UDP OSNMPD              ; SNMP Agent port for SNMP requests
```

If the agent will be started from the z/OS shell, reserve the port instead for z/OS UNIX by typing *OMVS* instead of *OSNMPD*.

If you want to define a port other than 161 for SNMP requests, you must do the following:

1. Start the agent with a `-p` parameter.
2. Configure management applications to use the new port:
   - For the snmp command, make an entry in the OSNMP.CONF file with the correct port number. For details on creating this entry, see the description for *targetAgent* in the OSNMP.CONF statement in the *z/OS Communications Server: IP Configuration Reference*.
   - Where supported, configure other management applications to use the new port.
3. Configure subagents to use the new port:
   a. Specify the port number to use on the SACONFIG profile statement for the TCP/IP subagent.
   b. Specify the port number to use on the ROUTESA_CONFIG profile statement for the OMPROUTE subagent.
   c. Specify the port number to use on the -p parameter when starting the SLA subagent.

d. Specify the port number to use on the -p parameter when starting the Network SLAPM2 subagent.
e. Specify the port number to use on the TNSACONFIG profile statement for the TN3270 Telnet subagent.
f. If you are using DPI subagents other than those supplied with z/OS Communications Server, set the SNMP_PORT environment variable to enable user-written subagents to connect to the agent.

The SNMP agent uses port 162, by default, for sending traps to the managers specified in SNMPTRAP.DEST or SNMPD.CONF file. Port 162 should be reserved for the management application primarily responsible for trap processing. If your environment requires multiple management applications at the same IP address to receive traps, consider using the Trap Forwarder Daemon. See "Step 5: Configure the trap forwarder daemon" on page 1078 for more details. If the SNMP query engine is typically used for processing traps and other applications, such as snmp, that are only occasionally used, the following port reservations are recommended.

```
PORT
  162 UDP SNMPQE  ; SNMPQuery Engine
```

You must also reserve additional ports for use by the snmp command by specifying

```
nnnnn UDP OMVS
```

where *nnnnn* is a number in the range 0–65535 and *nnnnn* is used as the -p parameter value on the snmp trap command.

If you want the SNMPQE and OSNMPD address spaces to be started automatically when the TCPIP address space is started, then include SNMPQE and OSNMPD in the AUTOLOG statement:

```
AUTOLOG
  SNMPQE                    ; SNMP Query Engine
  OSNMPD                    ; SNMP Agent
ENDAUTOLOG
```

Provide community-based security and notification destination information

If you are using only community-based security without the view-based access control model, do the following to configure the security and trap destinations.

Provide community name information

SNMP agents are accessed by remote network management stations and by SNMP subagents. To allow network management stations to send inquiries to the SNMP agent, and SNMP subagents to connect to the SNMP agent, you can provide PW.SRC information that defines a list of community names and IP addresses that can use these community names. The community name operates as a password when accessing objects on, or connecting to, a destination SNMP agent. The subagents pass the community name to the agent on the connect request.

All of the z/OS Communications Server SNMP subagents connect to the agent using the IPv4 primary interface IP address of the stack with which the subagent is associated, and a community name. As long as SOURCEVIPA is not in effect on the IPCONFIG profile statement, this IP address is the source IP address that the agent uses, along with the community name, to verify the subagent's authority to connect to the SNMP agent. The IPv4 primary interface IP address is either the first IP address in the HOME list or the IP address specified on a

PRIMARYINTERFACE TCP/IP profile statement. If SOURCEVIPA is in effect, the IP address used by the agent to verify the subagent's authority is the virtual IP address associated with the IPv4 primary interface IP address. For information on determining which virtual IPv4 address is associated with a physical IPv4 address, refer to the HOME statement in *z/OS Communications Server: IP Configuration Reference*. Check the Netstat HOME/-h output to verify the IPv4 primary interface address of the stack.

The PW.SRC information is optional. If no PW.SRC information is found and no community name is specified for the -c parameter at agent invocation, then the SNMP agent will accept requests with a community name of 'public' from any IP address. If a PW.SRC file exists, but is empty, and if no community name is specified on the `-c` parameter at the agent invocation, then no requests will be accepted by the agent.

Note: Verify that there is no SNMPD.CONF file because this file can only be used with SNMPv3. If an SNMPD.CONF file is found, the PW.SRC file will not be used.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

PW.SRC example: The PW.SRC statements could be specified as follows:

```
passwd1 9.0.0.0      255.0.0.0
passwd2 129.34.81.22 255.255.255.255
IPv6passwd3 12ab::0 16
IPv6passwd4 39B3::F430:03EE 128
```

The PW.SRC statements specify community names and hosts that can use each community name. The format of a statement is:

community_name desired_network snmp_mask

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

The community name of an incoming SNMP request is compared to the known community names. If a match is found, then the IP address of the incoming request is logically ANDed with the *snmp_mask* of the PW.SRC statement. The result of the logical ANDing process is compared with the *desired_network*. If they match, the request is accepted.

In the case of a password definition to be used by an IPv6 address or range of IPv6 addresses, the *snmp_mask* can be specified as a prefix value. The prefix specifies the number of bits to be used to construct an IPv6 address mask.

In the preceding example, if a request for *community_name* `passwd1` is received from the IP address 9.34.22.122, IP address 9.34.22.122 is ANDed with 255.0.0.0. The result is 9.0.0.0, which equals the specified *desired_network* for `passwd1`, so this request is accepted. In `passwd2`, if the *community_names* match, only requests from host 129.34.81.22 are accepted. The password `IPv6passwd3` can be used by any IPv6 address that starts with 12ab.

If the *community_name* values do not match, or the IP address ANDed with the *snmp_mask* does not match, an AUTHENTICATION_FAILURE trap is sent if both of the following are true:

- A destination entry exists in SNMPTRAP.DEST.
- Authentication failure traps have been enabled. These traps are enabled by setting MIB object "snmpEnableAuthenTraps.0" to 1.

A *desired_network* and *snmp_mask* of all zeros allows anyone with the correct *community_name* to make requests. However, the passwords for IPv4 addresses and the passwords for IPv6 addresses are stored and handled separately. Defining a password for use by both IPv4 and IPv6 addresses requires two entries in PW.SRC. Likewise, defining a password to be used by all addresses (both IPv4 and IPv6) requires two entries as follows:

```
passwd5 0.0.0.0 0.0.0.0
passwd5 0::0 0
```

Note: By default, the SNMP agent and the snmp command send packets such that a VIPA address will be used as the originating address in the packet, if SOURCEVIPA is configured. This is a change introduced in V2R10; previously, the SNMP agent and the snmp command set a socket option to cause the physical interface addresses to be used as the originating addresses on packets they sent. That meant the PW.SRC file had to contain all of the possible physical interface addresses that might be used, rather than a smaller number of VIPA addresses. A customer can override this change in behavior, if desired. This can be done for the SNMP agent by invoking it with the -a option. Similarly, you can do the same for the UNIX snmp command by either invoking it from the command line with the -a option, or by coding NOSVIPA in the command's OSNMP.CONF configuration file.

Provide trap destination information

Traps are unsolicited messages that are sent by an SNMP agent to an SNMP network management station. An SNMP trap contains information about a significant network event. The management application running at the management station interprets the trap information sent by the SNMP agent.

Note: When the SNMP agent starts, it retrieves an IP address for itself. If it retrieves an IPv6 colon-hexadecimal address, when it sends traps the source IP address in each trap will be 0.0.0.0.

For a detailed description of the SNMP trap types provided by z/OS CS, refer to *z/OS Communications Server: IP System Administrator's Commands*.

The SNMP agent Distributed Protocol Interface allows subagents other than those shipped with z/OS Communications Server (which might be running on another host) to generate SNMP traps. This can allow for support of other types of traps. For more information about SNMP DPI, see the *z/OS Communications Server: IP Programmer's Guide and Reference*.

To use traps, you must provide SNMPTRAP.DEST information defining a list of managers to which traps are sent. The SNMPTRAP.DEST information is optional. If no trap destination file is found, then the SNMP agent sends traps to the IP address of the SNMP agent and issues a warning message indicating that defaults are in effect. If a trap destination file exists, but is empty, no traps are sent.

Note: Verify that there is no SNMPD.CONF file. If an SNMPD.CONF file is found, the SNMPTRAP.DEST file will not be used.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

SNMPTRAP.DEST example: The SNMPTRAP.DEST statements could be specified as follows:

```
# SNMP Trap Destination information
124.34.216.1 UDP
39B3::F430:03EE UDP
MVSSYS2      UDP
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

Provide community-based and user-based security and notification destination information

If you want to use user-based security, either concurrently with or instead of community-based security, you must configure security definitions and notification destinations. To allow SNMP subagents to connect to the SNMP agent using user-based security, you must configure community-based security definitions. The SNMP subagents pass the community name to the agent on the connect request. The community name operates as a password when the SNMP subagents connect to the SNMP agent.

All of the z/OS Communications Server SNMP subagents connect to the agent using the IPv4 primary interface IP address of the stack with which the subagent is associated, and a community name. As long as SOURCEVIPA is not in effect on the IPCONFIG profile statement, this IP address is the source IP address that the agent uses, along with the community name, to verify the subagent's authority to connect to the SNMP agent. The IPv4 primary interface IP address is either the first IP address in the HOME list or the IP address specified on a PRIMARYINTERFACE TCP/IP profile statement. If SOURCEVIPA is in effect, the IP address used by the agent to verify the subagent's authority is the virtual IP address associated with the IPv4 primary interface IP address. For information on determining which virtual IPv4 address is associated with a physical IPv4 address, refer to the HOME statement in *z/OS Communications Server: IP Configuration Reference*. Check the Netstat HOME/-h output to verify the IPv4 primary interface address of the stack.

SNMPv3 provides the ability to configure the agent dynamically, from either a local or remote host, and to make changes in the configuration while the SNMP agent is running. Doing SNMP agent configuration dynamically requires a good understanding of how the SNMP SET commands can be issued to create new rows or to change or delete existing rows, as well as familiarity with the SNMP engine configuration tables defined in RFCs 2576 and 3411 through 3415.

As an alternative to dynamically configuring the SNMP agent, z/OS Communications Server supports a configuration file to be read at agent initialization called the SNMPD.CONF file. Dynamic configuration changes made with SNMP SET commands to the SNMP agent configuration entries will be written out to the SNMPD.CONF file, so they will continue to be in effect even after the SNMP agent is restarted.

SNMPD.CONF file

The SNMPD.CONF file defines the SNMP agent security and notification destinations. If the SNMPD.CONF file exists, the agent can support SNMPv1, SNMPv2c, and SNMPv3 requests. If no SNMPD.CONF file exists, the agent will support only SNMPv1 and SNMPv2c requests.

Note: If the SNMPD.CONF file is found, the PW.SRC file and the SNMPTRAP.DEST files are not used.

SNMPD.CONF dynamic configuration: If the SNMPD.CONF information is located in an MVS data set rather than an HFS file, special considerations must be made to support dynamic configuration changes to the SNMP agent's configuration. If dynamic configuration changes are made, the file is rewritten to reflect the changes. Therefore, consider the following when allocating the SNMPD.CONF file to an MVS data set:

- The record length (LRECL) should be 512 bytes to accommodate the longest possible entry.
- The use of a member of a partitioned data set is tolerated but not recommended. Because the file might be rewritten often, frequent compression of the partitioned data set may become necessary. In addition, locking on the file is done at the data set level, not at the member level, so other members of the partitioned data set would not be usable while the SNMP agent was running (once a dynamic configuration change had been made).

SNMPD.CONF example: A sample SNMPD.CONF file is shipped as /usr/lpp/tcpip/samples/snmpd.conf.

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

The sample OSNMP.CONF file used by the snmp command contains entries that match the sample SNMPD.CONF data set. See "Configure the z/OS UNIX snmp command" on page 1067 for additional information on configuring the snmp command.

Note: By default, the SNMP agent and the snmp command send packets such that a VIPA address will be used as the originating address in the packet, if SOURCEVIPA is configured. This is a change introduced in V2R10; previously, the SNMP agent and the snmp command set a socket option to cause the physical interface addresses to be used as the originating addresses on packets they sent. That meant the SNMPD.CONF file had to contain all of the possible physical interface addresses that might be used, rather than a smaller number of VIPA addresses. A customer can override this change in behavior, if desired, by invoking the SNMP agent with the -a option or by using either the -a option or the NOSVIPA option in the snmp command's OSNMP.CONF configuration file.

SNMPD.BOOTS

The SNMP agent uses the SNMPD.BOOTS configuration file to support SNMPv3 security. This file contains agent information used to authenticate the SNMPv3 requests. The SNMPD.BOOTS keeps the agent identifier and the number of times the agent reboots. If no SNMPD.BOOTS file exists when the agent is started, the agent creates one. You may want to add comments to the beginning of this file. If a file does exist, the agent uses the values specified in the file for setting its engineID and engineBoots values. If the file exists but contains incorrect values for engineID or engineBoots, the agent issues a message and terminates.

Notes:

1. The recommended approach is to allow the SNMP agent to create the file.
2. If the SNMPD.BOOTS file is not provided, the SNMP agent creates the file. If multiple SNMPv3 agents are running on the same MVS image, use the environment variable to specify different SNMPD.BOOTS files for the different agents. For security reasons, ensure unique engineIDs are used for different SNMP agents.

Creating user keys

Authentication

Authentication is generally required for SNMPv3 requests to be processed (unless the security level requested is 'noAuth'). When authenticating a request, the SNMP agent verifies that the authentication key sent in an SNMPv3 request can be used to create a message digest that matches the message digest created from the authentication key defined for the user.

The snmp command uses the authentication key found on an entry in the OSNMP.CONF configuration file. It needs to correlate with the authentication key specified on a USM_USER entry for that user in the agent's SNMPD.CONF configuration file.

As an alternative to storing authentication keys in the client configuration file, the snmp command allows user passwords to be stored. If the snmp command is configured with a password, the code generates an authentication key (and privacy key if requested) for the user. These keys must, of course, produce the same authentication values as the keys configured for the USM_USER in the agent's SNMPD.CONF file or configured dynamically with SNMP SET commands. However, the use of passwords in the client configuration file is considered less secure than the use of keys in the configuration file.

The authentication key is generated from two pieces of information:

- The specified password.
- The identification of the SNMP agent at which the key will be used. If the agent is an IBM agent and its engineID was generated using the vendor-specific engineID formula, the agent may be identified by IP address or host name. Otherwise, the engineID must be provided as the agent identification.

A key that incorporates the identification of the agent at which it will be used is called a localized key. It can be used only at that agent. A key that does not incorporate the engineID of the agent at which it will be used is called nonlocalized.

Keys stored in the snmp command's configuration file, OSNMP.CONF, are expected to be nonlocalized keys. Keys stored in the SNMP agent's configuration file, SNMPD.CONF, can be either localized or nonlocalized, though the use of localized keys is considered more secure.

Encryption

Keys used for encryption are generated using the same algorithms as those used for authentication. However, key lengths may differ. For example, an HMAC-SHA authentication key is 20 bytes long, but a localized encryption key used with HMAC-SHA is only 16 bytes long.

z/OS Communications Server provides a facility called *pwtokey* that enables conversion of passwords into localized and nonlocalized authentication and privacy keys. The pwtokey procedure takes as input a password and an identifier of the agent and generates authentication and privacy keys. Since the procedure used by the pwtokey facility is the same algorithm used by the snmp command, the person configuring the SNMP agent can generate appropriate authentication and privacy keys to put in the SNMPD.CONF file for a user, given a particular password and the IP address at which the agent will run.

Use the pwtokey command to convert passwords into authentication and privacy keys. Refer to *z/OS Communications Server: IP System Administrator's Commands*.

Provide security product access to agent from subagents

An SNMP subagent can connect to the z/OS Communications Server SNMP agent by using the DPI API (the DPI API is documented in the *z/OS Communications Server: IP Programmer's Guide and Reference*) and specifying either a z/OS UNIX or a TCP connection. Subagents using a z/OS UNIX connection are required to have superuser authority. For subagents specifying a TCP connection, you can utilize the installation's SAF compliant security product (such as the z/OS Security Server (RACF)) to control which of the SNMP subagents are permitted to connect to the SNMP agent. One security product resource name can be created per TCP/IP stack per MVS image. The security product resource name is specified in the following format:

```
EZB.SNMPAGENT.sysname.tcpprocname
```

where *sysname* is the name of the MVS system image and *tcpprocname* is the TCP/IP started procedure name.

The profile must be created under the SERVAUTH class. After creating the profiles, use the security product to define the user IDs of those subagents which should be permitted to connect via TCP to the SNMP Agent. Authorization failures are documented by security product failure messages and SNMP agent traces.

Note: If you use this authorization function, only SNMP subagents which are associated with the same TCP/IP stack as the SNMP agent will be permitted to connect to the agent. Local SNMP subagents associated with other TCP/IP stacks, or remote SNMP subagents, will not be permitted to connect. Also, any subagents which connected to the SNMP agent before the agent security product resource name was created will not have been authorized via the security product.

You can use the control statements in the sample JCL job provided in SEZAINST(EZARACF) to define this authorization. For example, if you wanted to permit any SNMP subagents associated with a user ID of USER2 to connect to the SNMP agent you could use the following definitions:

```
RDEFINE SERVAUTH EZB.SNMPAGENT.MVSA.TCP1 UACC(NONE)
PERMIT EZB.SNMPAGENT.MVSA.TCP1  ACCESS(READ) CLASS(SERVAUTH) ID(USER2)
```

Allowing subagents with duplicate identifiers to connect

When an SNMP subagent connects to the SNMP agent, an SNMP object identifier (OID) value is used as the subagent identifier. For subagents connecting using the DPI V2.0 API, the subagent supplies its OID during the connection process. All the Communications Server subagents use the DPI V2.0 API when connecting to the agent.

For subagents connecting using the DPI V1.1 API, the agent assigns a constant OID value to the subagent. Therefore, if more than one DPI V1.1 subagent connects to the agent, all the DPI V1.1 subagents are identified with the same OID value.

The SUBAGENT-MIB, supported by the SNMP agent, defines the MIB object, saAllowDuplicateIDs, that can be used to configure whether the agent should allow subagents with duplicate OID values to connect. By default, the SNMP agent sets the value of this object to 1, which allows subagents with duplicate OID values to connect. You can configure the value of this MIB object using the OSNMPD.DATA configuration file. For more information about this file, see "Provide MIB object configuration information." If you do not want the agent to allow subagents with duplicate OID values to connect, set the value of this MIB object to 2.

Provide MIB object configuration information

An installation can set values for selected MIB objects by providing OSNMPD.DATA information. A sample of OSNMP.DATA is installed as HFS file /usr/lpp/tcpip/samples/osnmpd.data. Refer to *z/OS Communications Server: IP Configuration Reference* for syntax information. If no OSNMPD.DATA file is found, the defaults for these MIB objects are as follows:

Object **Default**

dpiPathNameForUnixStream
The default is /tmp/dpi_socket.

sysDescr If the environment variable HOSTNAME exists, its value is used. Otherwise, the default value identifies the z/OS system under which the agent is running. The maximum length of this object is 255 octets.

sysContact "SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysLocation "SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysName "SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysObjectId 1.3.6.1.4.1.2.3.13

Note: sysObjectID is defined as the vendor's authoritative identification of the network management subsystem contained in the entity. That is, it is intended to uniquely identify the SNMP agent. Changing this value is not recommended and will be disabled in a subsequent release.

sysServices A single octet that defaults to 0. See the RFC 1907 description for this object.

snmpEnableAuthenTraps
Default value is 2, which means traps are disabled.

saDefaultTimeout
5 seconds.

saMaxTimeOut
600 seconds.

saAllowDuplicateIDs

Default is 1, which means multiple instances of a subagent (that is, where the subagent identifier for all the subagents is the same) are allowed to connect to the SNMP agent. To prevent multiple instances of a subagent from connecting to the SNMP agent, set the value to 2.

Note: Because a subagent identifier cannot be specified for subagents connecting using the DPI V1.1 API, the SNMP agent assigns the same constant identifier for all DPI V1.1 subagents. Therefore, this object must be set to 1 to allow multiple DPI V1.1 subagents to run concurrently. For more information about subagent identifiers, see "Allowing subagents with duplicate identifiers to connect" on page 1064.

For information about where these MIB objects are defined, refer to the *z/OS Communications Server: IP User's Guide and Commands*.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

Start the SNMP agent (OSNMPD)

The SNMP agent (OSNMPD) runs in a separate address space that executes load module EZASNMPD. OSNMPD can be started with or without parameters. When starting OSNMPD from MVS, add the parameters to the PARMS= keyword on the EXEC statement of the OSNMPD cataloged procedure. When starting OSNMPD from z/OS UNIX, specify the desired parameters on the osnmpd command. Refer to *z/OS Communications Server: IP Configuration Reference* for the command syntax.

If the SNMP agent encounters any errors processing its configuration files, error messages are written to the syslog daemon, not to the console.

Sample JCL procedure for starting OSNMPD from MVS

Update cataloged procedure OSNMPD by copying the sample in SEZAINST(OSNMPDPR) to your system or recognized PROCLIB. Change the data set names as required to suit your local configuration. The OSNMPD address space requires access to the IBM C/370 Library during execution.

Parameters may be passed to the agent on the `PARMS=` keyword on the EXEC statement of the OSNMPD cataloged procedure. Refer to *z/OS Communications Server: IP Configuration Reference* for the command syntax and parameter information. Any agent parameters you wish to specify may be added as shown in the following example:

```
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM='POSIX(ON) ALL31(ON)/ -c abc -d 255 -p 761'
```

In this example, the agent will use port 761 to accept requests, community name 'abc' will be added to the list of community names supported by the agent, and all agent traces will be activated. For more information on tracing, see the *z/OS Communications Server: IP Diagnosis Guide*.

Starting OSNMPD from z/OS UNIX

To run the SNMP agent in background, you must add an ampersand (&) to the command and the issuer of the command must be in z/OS UNIX superuser mode. For a detailed explanation of the osnmpd parameters, refer to *z/OS Communications Server: IP Configuration Reference*.

Any agent parameters you wish to specify may be added as shown in the following example:

```
osnmpd -c abc -d 255 -p 761
```

In this example, the agent will use port 761 to accept requests, community name 'abc' will be added to the list of community names supported by the agent, and all agent traces will be activated. For more information on tracing, see *z/OS Communications Server: IP Diagnosis Guide*.

Step 2: Configure the SNMP commands

The two SNMP client applications provided with z/OS Communications Server are:

- snmp command in the z/OS shell
- SNMP command from the NetView environment

The SNMP command in the NetView environment requires the use of the NetView product. It supports SNMP version 1. The snmp command in the z/OS shell supports SNMP versions 1, 2, and 3. Depending on your requirements, you might decide to configure either or both of these clients, or to use an SNMP client on another platform.

Configure the z/OS UNIX snmp command

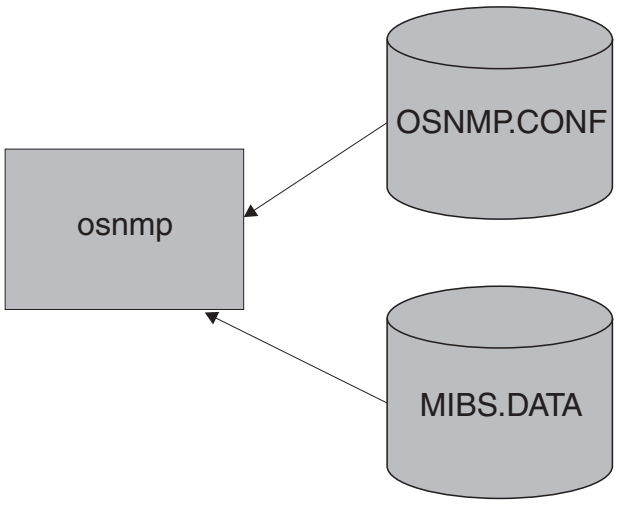


Figure 107. Configuration files for snmp

The z/OS UNIX **snmp** command is used to send SNMP requests to SNMP agents on local or remote hosts. You can also use the synonym, **osnmp**, as the name of this command. The requests can be SNMPv1, SNMPv2, or SNMPv3. For SNMPv2 and SNMPv3 requests, the OSNMP.CONF configuration file is required. The *winSNMPname* specified on an OSNMP.CONF statement can be used as the value of the -h parameter on the **snmp** command. For a detailed explanation of the

parameters you can specify on the **snmp** command, see *z/OS Communications Server: IP System Administrator's Commands*.

To configure the **snmp** command, perform the following tasks:

- The **snmp** command needs to be able to resolve the name of the host, on which the command is executing, to the host's IP address. You can provide this information either by configuring a domain name server or by configuring local host files. For information about the search order used to locate local host files, see the local host tables entry in Table 3 on page 41. As part of its support for the SNMPv3 protocol, the command also uses this IP address when creating its SNMPv3 engineID value.
- Provide snmp configuration information
- Provide user MIB object information

Provide snmp configuration information

The OSNMP.CONF file is used to define target agents and, for SNMPv3, the security parameters to be used in sending requests to them.

The contents of the file, regardless of location, are the same. Only the first file found is used. A sample of this file is installed as HFS file /usr/lpp/tcpip/samples/snmpv2.conf. This sample should be copied and modified for your installation. Refer to *z/OS Communications Server: IP Configuration Reference* for more information.

Examples:

- Example 1:

  The following entry defines an SNMPv2c node for snmp:

  ```
  mvs1     9.67.113.79 snmpv2c
  ```

  where *mvs1* is the name used with the -h parameter on the snmp command and *9.67.113.79* is the IP address of the SNMPv2c agent.

- Example 2:

  The following entry defines an SNMPv3 node:

  ```
  v3mak   127.0.0.1 snmpv3 u1 - - AuthNoPriv  HMAC-MD5 7a3e34265e0e029f27d8b4235ecfa987 - -
  ```

  where *v3mak* is the name used on the -h parameter of the snmp command. An SNMP request sent using this entry uses USM user name *u1* using HMAC-MD5 authentication but no encryption.

- Example 3:

  The following entry defines an SNMPv3 node. The needed authentication and privacy keys will be generated from the password *u6password*.

  ```
  v3sap   127.0.0.1 snmpv3 u6 u6password - AuthNoPriv   HMAC-SHA -   - -
  ```

  The USM user is *u6*. The authentication protocol is HMAC-SHA, and no encryption is used.

- Example 4:

  The following entry defines an SNMPv3 node:

  ```
  v3mpk_ipv6 ::1 snmpv3 u1 - - AuthPriv HMAC-MD5 7a3e34265e0e029f27d8b4235ecfa987 DES eac02a0d9fe90eca7911fdcaba20deae
  ```

  where *v3mpk_ipv6* is the name used on the -h parameter of the snmp command. An SNMP request sent using this entry uses USM user name *u1* using HMAC-MD5 authentication and CBC 56-bit DES encryption.

Provide MIB object information in MIBS.DATA

Like other SNMP managers, when you enter the z/OS UNIX **snmp** command you can specify the object identifier (OID) of the MIB object whose value you want to retrieve. If the MIB object is supported by one of Communication Server's SNMP functions, you can also use the MIB object textual name on the z/OS UNIX **snmp** command, instead of the object identifier. All the MIB objects supported by Communications Server functions are listed in the MIB objects appendix in *z/OS Communications Server: IP System Administrator's Commands*.

For example, to retrieve the SNMP agent sysUpTime MIB object, you can enter either of the following commands:

- `snmp -v get sysUpTime.0`
  `sysUpTime.0 = 289700`
- `snmp -v get 1.3.6.1.2.1.1.3.0`
  `sysUpTime.0 = 292700`

The 1.3.6.1.2.1.1.3 value is the OID for the sysUpTime MIB object.

If you have user-defined MIB objects or MIB objects from other products, and you want to use the textual names for these MIB objects on the z/OS UNIX **snmp** command, you can define these objects to the **snmp** command using the MIBS.DATA file. A sample of the MIBS.DATA file is installed as HFS file /usr/lpp/tcpip/samples/mibs.data. Copy this sample and modify it for your installation. For the search order used to locate the MIBS.DATA file, see "Understanding search orders of configuration information" on page 28.

If other products provide files using MIBS.DATA syntax, for the textual names of their MIB objects to be able to be used by the **snmp** command, append the statements from their files to your MIBS.DATA file so that these statements are accessible to the **snmp** command. For example, the IBM OSA-Express Direct subagent provides a file in MIBS.DATA syntax so that the textual names of its OSA-Express management data can be used with the z/OS UNIX **snmp** command. For information about how to obtain this file, refer to *zSeries OSA-Express Customer's Guide and Reference*.

MIBS.DATA statement syntax

The format of a statement in the MIBS.DATA file is:

```
character_object_name object_identifier object_type
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

Configure the NetView SNMP (SNMP) command

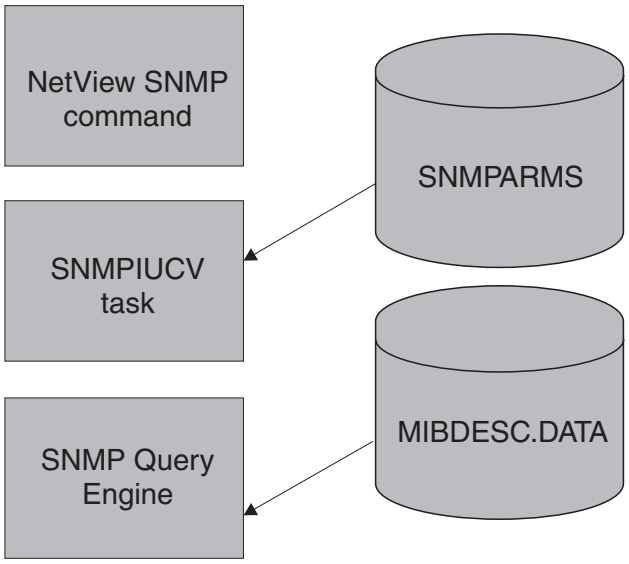


Figure 108. Configuration files for NetView SNMP

The SNMP command in the NetView environment can be used to send SNMP version 1 requests to SNMP agents on either local or remote hosts. The SNMP command requires the command processor itself, the SNMPIUCV task for inter-address space communication, and the SNMP query engine, which creates the packets sent to the SNMP agent. The NetView SNMP command and the SNMP query engine support only community-based security.

Configure the SNMP query engine

Update the SNMPQE cataloged procedure by copying the sample in SEZAINST(SNMPPROC) to your system or recognized PROCLIB. Specify SNMP parameters and change the data set names as required to suit your local configuration. The SNMPQE address space requires access to the IBM C/370 Library during execution.

The SNMP query engine (SQESERV) needs access to the *hlq*.MIBDESC.DATA data set for the MIB variable descriptions. You can find a sample of this data set in SEZAINST(MIBDESC).

MIBDESC.DATA data set: The MIBDESC.DATA data set defines the short names for MIB variables. Short names are the character representation for the ASN.1 variable names. For example, sysUpTime is the short name for 1.3.6.1.2.1.1.3.0 (the MIB variable that stores the time since the SNMP agent was last restarted). Short names are generally shown as a combination of upper and lowercase characters, though SNMP on z/OS Communications Server ignores these case distinctions. Variable names must always be in ASN.1 language when they are sent to an SNMP agent. You can always use ASN.1 language to specify the variable names in an enterprise-specific tree (assuming that the agent supports them). You can use these short names to specify the MIB variables.

When you issue an SNMP GET, GETNEXT, or SET command, and specify the variable name in ASN.1 notation, the SNMP Query Engine uses that name and sends it in the SNMP packet to the agent. When you issue an SNMP GET, GETNEXT, or SET command, and specify the short name for the variable (for

example, sysDescr), the SNMP Query Engine looks for that name in the MIBDESC.DATA data set and uses the ASN.1 name specified in the data set when it sends the SNMP packet to the agent.

The SNMPQE address space must be able to access the MIBDESC.DATA data set.

You can change the short names in the MIBDESC.DATA data set to the equivalent in your national language. You can also leave the current names and add the equivalent names in your national language. However, the SNMP MIBVNAME function returns only the first entry found in the data set that satisfies the search. In addition, all enterprise-specific variables used by hosts in your network should be added to this data set.

Entries in the data set do not need to be in a specific sequence. Each name starts on a new line. The following shows the line format.

```
short_name asn.1_name type time_to_live
```

Each variable on the line is separated by either one or more spaces or tabs. An asterisk (*) in column 1 indicates that the line is a comment line.

Following is a sample MIBDESC.DATA line with a sysDescr variable translated in Dutch and a few enterprise variables added (in this example, company ABC received 1.3.6.1.4.1.42 as the ASN.1 number for their enterprise):

```
*-------------------------------------------------------------*
* MIB Variable name | ASN.1 notation        | Type     | TTL  *
*-------------------------------------------------------------*
* Following is Dutch name for sysDescr
systeemBeschrijving  1.3.6.1.2.1.1.1.         display    900
sysDescr             1.3.6.1.2.1.1.1.         display    900
  ...
  other entries
  ...
* Following are Enterprise-Specific variables for company ABC
ABCInfoPhone         1.3.6.1.4.1.42.1.1       display    900
ABCInfoAddress       1.3.6.1.4.1.42.1.2       display    900
```

The TTL field contains the number of seconds that a variable lives in the Query Engine's internal cache. If there are multiple requests for the same variable within the TTL period, the variable value is obtained from the cache, and unnecessary network traffic is avoided.

You can define multiple short names or text names for the same variable, as shown with the Dutch translation of the sysDescr variable. In this case, the SNMP Query Engine returns the first value in the table on an SNMP MIBVNAME request. In the previous example, the SNMP Query Engine would return `systeemBeschrijving` and not `sysDescr`. The name returned is in mixed case.

When the SNMP Query Engine receives a short name or text name in a GET, GETNEXT, or SET request, it compares the name against the entries in the MIBDESC.DATA data set. This comparison is not case-sensitive. For example, a request for `SYSDESCR`, `SysDescr`, or `sysDescr` matches the `sysDescr` entry with an ASN.1 notation of `1.3.6.1.2.1.1.1.`.

When the SNMP Query Engine receives an SNMP response, it looks up the variable in the MIBDESC.DATA table Type field for information about translating the value into displayable characters. The information contained in the Type field is case-sensitive and must be specified in lowercase.

Note: If you are using SNMP to receive response or trap PDUs which contain enterprise-specific variables, the variables must be added to the MIBDESC.DATA data set.

Specifying the SNMPQE parameters: The SQESERV module can be configured to start without parameters or you can add any of the following parameters to PARMS=' in the PROC statement of the SNMPQE cataloged procedure. For example,

```
//SNMPQE   PROC MODULE=SNMPQE,PARMS='-h MVSA'
```

Refer to *z/OS Communications Server: IP Configuration Reference* for the command syntax.

Refer to *z/OS Communications Server: IP Diagnosis Guide* for more information on tracing.

Setting up authorization for SNMPQE: To create RAW sockets necessary for SNMP PING requests, the user ID associated with the SNMPQE started task must have superuser authority (z/OS UNIX UID of 0) or be permitted to BPX.SUPERUSER facility authority.

Configure NetView as an SNMP monitor

To configure the NetView interface as an SNMP monitor, perform each of the following tasks:

- Configure for SNMPIUCV
- Configure for the SNMP command processor
- Configure for the SNMP messages
- Update the SNMP initialization parameters

Configure for SNMPIUCV: SNMPIUCV is the NetView optional task that handles IUCV communication between the NetView program and the SNMP query engine. SNMPIUCV resides in the SEZADSIL data set.

Add the following TASK statement for SNMPIUCV to the DSIDMN member of the data set specified by the DSIPARM DD statement in the NetView start procedure.

```
TASK  MOD=SNMPIUCV,TSKID=SNMPIUCV,PRI=5,INIT=Y
```

This statement causes SNMPIUCV to start automatically when the NetView program is started.

If you specify INIT=N instead of INIT=Y in the TASK statement for SNMPIUCV, a NetView operator can start the SNMPIUCV task by entering the following:

```
START TASK=SNMPIUCV
```

The SNMPIUCV task tries to connect through IUCV to the SNMP query engine. If this fails, it retries the connect as specified by the SNMPQERT keyword in the SNMPARMS member of the SEZADSIP data set. The default is every 60 seconds.

Configure for the SNMP command processor: SNMP is the command processor that allows NetView operators and CLISTs to issue SNMP commands. SNMP resides in the SEZADSIL data set. This data set should be concatenated to the STEPLIB DD statement in the NetView start procedure.

Add the following statement to the DSICMD member of the data set specified by the DSIPARM DD statement in the NetView start procedure.

```
SNMP  CMDMDL  MOD=SNMP,ECHO=Y,TYPE=R,RES=Y
```

After the SNMPIUCV task is started, you can issue the SNMP command. The SNMP command passes a request to the SNMPIUCV task to forward to SNMPQE. The return code represents a request number that is associated with the request. The responses are returned asynchronously and contain this request number. The operator or CLIST must use the request number to correlate the response to the request.

Configure for the SNMP messages: The NetView SNMP messages reside in the SEZADSIM data set as DSISNM*nn*, where *nn* is the number of the member. The valid message members are DSISNM00 through DSISNM05, DSISNM10, DSISNM12, and DSISNM99. The data set containing these members should be added to the DSIMSG DD statement in the NetView start procedure.

Update the SNMP initialization parameters: SNMPIUCV reads the SNMPARMS member in the SEZADSIP data set at startup. This data set contains the initialization parameters for SNMP. The data set containing SNMPARMS should be added to the DSIPARM DD statement in the NetView startup procedure. Refer to *z/OS Communications Server: IP Configuration Reference* for detailed information for the SNMP parameter data set (SNMPARMS).

Step 3: Configure the SNMP subagents

There are several SNMP subagents shipped with z/OS Communications Server:

- The TCP/IP subagent reports information about the TCP/IP stack. For details on configuring this subagent, see "TCP/IP subagent configuration" on page 1074.
- The OMPROUTE subagent reports information specific to OSPF. The ROUTESA_CONFIG statement is used in the OMPROUTE configuration file to configure the OMPROUTE subagent. For details on ROUTESA_CONFIG, refer to *z/OS Communications Server: IP Configuration Reference*.
- The Network SLAPM2 subagent reports information about defined policies and performance statistics related to traffic using those policies. Compared to the SLA subagent, the Network SLAPM2 subagent using the NETWORK-SLAPM2-MIB has better and more relevant information on network and TCP-related performance data, reduces system overhead, and provides transparency to allow collection on both IPv4 and IPv6 addresses. For configuration information about the Network SLAPM2 subagent, see Chapter 15, "Quality of service (QoS)," on page 751.
- The SLA subagent reports information about defined service policies and performance statistics related to traffic using those policies. For configuration information about the SLA subagent, see Chapter 15, "Quality of service (QoS)," on page 751.
- The TN3270 Telnet subagent reports information about TN3270 connections and monitoring values. For information on the TNSACONFIG statement and the parameters needed to start the TN3270 Telnet subagent, refer to *z/OS Communications Server: IP Configuration Reference*.
- The OSA-Express Direct SNMP subagent is also shipped with z/OS CS but is supported by the zSeries OSA Support Group. The OSA-Express Direct SNMP subagent and the OSA MIB provided by the zSeries OSA Support Group can be used with Communications Server SNMP support to provide SNMP management data for some OSA adapters. For details regarding the OSA-Express Direct SNMP subagent and OSA MIB, refer to *zSeries OSA-Express Customer's Guide and Reference*.

These subagents use the stack's IPv4 primary interface IP address when connecting to the SNMP agent. The IPv4 primary interface IP address is either the first IPv4 IP

address specified in the HOME profile statement, or the IP address of the IPv4 interface specified on a PRIMARYINTERFACE profile statement. If the subagents cannot obtain the IPv4 primary interface IP address, they will use the IPv4 loopback IP address, 127.0.0.1, to connect to the agent. For information on using the IPv4 primary interface IP address and a community name to permit the subagents to connect to the SNMP agent, see "Provide community name information" on page 1058 or "Provide community-based and user-based security and notification destination information" on page 1061.

TCP/IP subagent configuration

There are two statements in the profile data set used to configure the TCP/IP subagent, the SACONFIG and ITRACE statements.

- SACONFIG

  Use the SACONFIG statement to configure the subagent. The SACONFIG parameters determine whether or not the subagent is automatically started at TCP/IP initialization, what port number to use to contact the agent, and other configuration values. For a detailed explanation of this statement, refer to *z/OS Communications Server: IP Configuration Reference*.

- ITRACE

  Use the ITRACE statement to determine what trace information, if any, should be recorded by the subagent. For a detailed explanation of this statement, refer to *z/OS Communications Server: IP Configuration Reference*.

Step 4: Configure the Open Systems Adapter (OSA) support

The TCP/IP subagent can retrieve SNMP management data from the Open Systems Adapter Support Facility (OSA/SF) for several OSA adapters.

The OSA product also provides an SNMP subagent, the OSA-Express Direct subagent, that supports management data for OSA-Express adapters. The OSA-Express Direct subagent can be used with Communications Server SNMP support to retrieve the management data. You should use the OSA-Express Direct subagent for OSA management data, rather than the TCP/IP subagent, because the OSA-Express Direct subagent communicates directly with the OSA-Express adapters and does not require the OSA/SF and IOASNMP applications. For more information about the management data provided by the OSA-Express Direct subagent, refer to *zSeries OSA-Express Customer's Guide and Reference*.

The TCP/IP subagent supports management data for the following OSA adapters:

- ATM management data is supported for any OSA-2 ATM or OSA-Express ATM155 adapters.
- Ethernet management data is supported for any OSA-Express Gigabit Ethernet or OSA-Express QDIO Fast Ethernet adapters.
- Tables of OSA-Express management data are supported for any OSA-Express Gigabit Ethernet, Fast Ethernet, or ATM155 adapters.

The TCP/IP subagent retrieves the data using the OSA/SF components IOAOSASF (OSA/SF application) and IOASNMP (OSA/SF socket application). For more information on these components, see "OSA/SF prerequisites" on page 1076. Specifying the SACONFIG profile statement, with the OSAENABLED and OSASF parameters, in the TCP/IP profile data set causes the TCP/IP subagent to try to connect to the OSA/SF socket application, IOASNMP, using the TCP protocol and the port number specified on the OSASF parameter. If the subagent connects successfully, the following message is issued:

```
EZZ3218I SNMP SUBAGENT: CONNECTED TO OSA/SF
```

Because of timing considerations, the TCP/IP subagent might not be able to connect to IOASNMP at initalization. If this occurs, the subagent will attempt to connect when the first request for OSA MIB data is received. Therefore, the EZZ3218I message might not always be issued during subagent initialization.

When retrieving management data from the OSA adapters, the TCP/IP subagent sends a request to IOASNMP for the data, passing the adapter's portname as an identifier. The portname is obtained from the DEVICE and LINK profile statements used to define the adapter to the TCP/IP stack. The IOASNMP socket application uses APPC to pass the request to the OSA/SF application. The OSA/SF application then retrieves the data from the adapter and returns it to IOASNMP. IOASNMP uses its TCP connection to the TCP/IP subagent to return the data to the subagent. If this configuration is active and either the IOASNMP application or the TCP/IP subagent terminates, the subagent will issue the following message:

```
EZZ3219I SNMP SUBAGENT: DISCONNECTED  FROM OSA/SF
```

To obtain the management data, the adapters must be defined to the TCP/IP stack where the subagent is active, through DEVICE and LINK statements in the TCP/IP profile.

- To retrieve ATM management data, the ATM adapter must be defined as an ATM device/link, even if it is configured for ATM LAN emulation mode and is therefore also defined to some TCP/IP instance as an LCS or MPCIPA device. For OSA-Express ATM155 adapters configured for QDIO LAN Emulation mode, you can use one of the adapter's logical port names on the PORTNAME parameter of the ATM DEVICE statement.
- To retrieve Ethernet management data, the OSA-Express adapter must be defined as an MPCIPA Ethernet link.
- To retrieve OSA-Express management data, the adapters must be defined as the following TCP/IP device/link types:
  - MPCIPA link for Gigabit Ethernet
  - MPCIPA or LCS Ethernet link for Fast Ethernet
  - ATM device/link for ATM155

If the port name is manually configured at the adapter, then management data can be retrieved from the adapter even if it is not active and not in use by any TCP/IP stack or by VTAM. If the port name is dynamically configured (e.g. MPCIPA links), then the adapter has to be active to some TCP/IP stack to retrieve the management data.

Current support consists of:

- Interface table from RFC2233 for ATM, LAN emulation links, AAL5 and ATM layer interfaces.
- ATM Channel table from the IBM MVS Enterprise-Specific MIB for OSA-2 ATM adapters.
- ATM Port and PVC tables from the IBM MVS Enterprise-Specific MIB for OSA-2 ATM and OSA-Express ATM155 adapters.
- ATM LAN Emulation tables from the IBM MVS Enterprise-Specific MIB for OSA-2 ATM and OSA-Express ATM155 adapters.
- atmInterfaceConfTable from RFC1695 for OSA-2 ATM and OSA-Express ATM155 adapters.

- IP over ATM tables from RFC2320 for OSA-2 ATM and OSA-Express ATM155 adapters.
- OSA-Express Channel and Performance tables from the IBM MVS Enterprise-Specific MIB for OSA-Express Gigabit Ethernet, Fast Ethernet, and ATM155 adapters. The performance MIB object values in the ibmMvsOsaExpChannelTable, and all of the MIB object values in the ibmMvsOsaExpPerfTable, are only available starting with zSeries processors where the adapters are, at least, at microcode level 1.31.
- OSA-Express Ethernet Port table from the IBM MVS Enterprise-Specific MIB for OSA-Express Gigabit Ethernet and Fast Ethernet adapters.
- OSA-Express Ethernet SNA table from the IBM MVS Enterprise-Specific MIB for OSA-Express Fast Ethernet adapters.
- dot3StatsTable from EtherLike-MIB in RFC2665 for OSA-Express Gigabit Ethernet and QDIO Fast Ethernet adapters.
- Asynchronous SNMP Trap generation for operational management:
  - ATM Permanent Virtual Circuit (PVC) creation - ibmMvsAtmOsasfAtmPvcCreate Trap (ATM OSA-2 only).
  - ATM Permanent Virtual Circuit (PVC) deletion - ibmMvsAtmOsasfAtmPvcDelete Trap (ATM OSA-2 only).
- Provide method for assigning an IP Address to the ATM Port.

  Use the *osnmp set* command against the ibmMvsAtmOsasfPortIpAddress MIB object to assign the IP Address.

OSA/SF prerequisites

The TCP/IP subagent provided by z/OS Communications Server will connect to OSA/SF to obtain management data. For a subagent to establish a connection to OSA/SF, two OSA/SF components must be started:

- IOAOSASF

  IOAOSASF is a sample JCL procedure that can be used to start the main OSA/SF address space. The sample has a job name of OSASF1.
- IOASNMP

  IOASNMP is a sample JCL procedure that starts the OSA/SF-provided z/OS UNIX socket application that interconnects the TCP/IP subagent with OSASF1.

These sample procedures and all entities that they call are provided with OSA/SF. For a detailed explanation of how to set up OSA/SF on your MVS system, refer to *zSeries OSA-Express Customer's Guide and Reference*. The primary purpose of OSA/SF is to manage OSA Adapters. It has been extended to support OSA management via SNMP. An instance of IOAOSASF, IOASNMP, the TCP/IP stack, the TCP/IP subagent, and the SNMP agent must be started on every MVS image where OSA management support is needed.

The recommended startup order is:

1. Start IOAOSASF and wait until it completely initializes. IOAOSASF must be started before IOASNMP.
2. Include OSNMPD (the CS SNMP agent) and IOASNMP on the AUTOLOG profile statement for your TCP/IP stack. This ensures that they will be started by autolog processing when TCP/IP is started. For additional profile statement requirements, see "Required TCP/IP profile statements" on page 1077. Start the TCP/IP stack.

On an MVS image only a single instance of either IOAOSASF or IOASNMP can (or should) be started. An attempt to start multiple copies of IOAOSASF will be rejected. Starting multiple copies of IOASNMP will yield unpredictable results.

Ensure that OSA/SF is at Version 2 Release 1 level or higher with the OSA/SF APAR OW45237 applied.

Required TCP/IP profile statements

For a detailed description of the statements mentioned here, refer to *z/OS Communications Server: IP Configuration Reference*. The following TCP/IP profile statements must be updated for OSA management support:

- SACONFIG

  On the SACONFIG statement, OSA Management support must be enabled by specifying OSAENABLED. Omission of OSAENABLED when TCP/IP is started will result in no OSA management support. The SACONFIG statement controls the operation of the subagent that runs in a TCP/IP address space as a separate task.

  The OSASF parameter specifies which port IOASNMP should use to Listen for connections from subagents to OSA/SF. For an explanation of the usage of this parameter when starting multiple TCP/IP instances, see "Multiple TCP/IP instances considerations." It is recommended that the OSASF port be reserved by also specifying it on a PORT statement.

  For example:

  ```
  SACONFIG OSAENABLED OSASF 721
  ```

- PORT

  Prereserve the port to be used in communication with OSA/SF.

  For example:

  ```
  PORT
     721  TCP IOASNMP ; OSA/SF TCP/IP Communications
  ```

  In the above example since IOASNMP runs as a z/OS UNIX application the port could have been reserved for z/OS UNIX. Review the /etc/services HFS file to insure that there are no port conflicts.

- DEVICE and LINK

  Provide ATM DEVICE and LINK statements for any OSA ATM adapter for which you want SNMP ATM management data. For example:

  ```
  DEVICE osaName ATM PORTNAME portname
  LINK linkName ATM osaName
  ```

  Provide DEVICE and LINK statements for any OSA-Express Ethernet adapter for which you want SNMP Ethernet management data. For example:

  ```
  DEVICE portname MPCIPA NONROUTER
  LINK linkName IPAQENET portname
  ```

Multiple TCP/IP instances considerations

Subagent connection to OSA/SF

When multiple z/OS Communications Server instances are active in the same MVS image, only one of the TCP/IP instances is connected to IOASNMP. In order for a TCP/IP instance to connect to IOASNMP the OSASF parameter must be specified on the SACONFIG Profile statement.

IOASNMP connects to a TCP/IP instance and acts as a server, receiving connections from those TCP/IP subagents where OSAENABLED was specified on

the SACONFIG Profile statement. The result is that all these subagents connect through the same TCP/IP to IOASNMP in order to retrieve OSA information from OSA/SF. For a depiction of this process, see Figure 109.

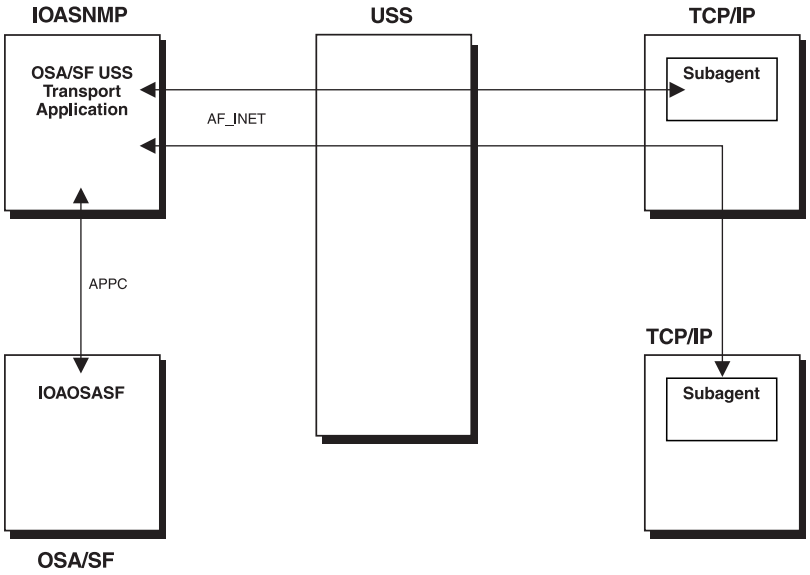


Figure 109. Subagent connection to OSA/SF

If IOASNMP loses its connection to TCP/IP it terminates and needs to be restarted.

If the currently connected TCP/IP terminates, IOASNMP will attempt to connect to another TCP/IP instance for which the OSASF parameter was specified on the SACONFIG Profile statement, using the port number specified for the OSASF parameter. The subagents will also attempt to reconnect to OSA/SF via IOASNMP using this same OSASF port number. For this reason it is recommended that the same OSASF port number be used on the SACONFIG statement of every TCP/IP instance where the OSASF parameter is specified.

Whenever a socket error occurs on the OSA/SF socket, the connected subagents will issue the following message:

```
EZZ3219I SNMP SUBAGENT: DISCONNECTED FROM OSA/SF
```

When the subagent connection is reestablished, the following message is issued:

```
EZZ3218I SNMP SUBAGENT: CONNECTED TO OSA/SF
```

Step 5: Configure the trap forwarder daemon

Most SNMPv1 agents forward traps to port 162. If a manager needs to listen for these traps, it has to bind to port 162 and listen for it. When a manager has already issued a bind it is impossible for another manager to listen for the same traps. The Trap Forwarder daemon eliminates this problem by listening for traps on port 162 and forwarding them to all the configured managers.

You can configure the Trap Forwarder daemon to receive a trap on a specified port and forward it to multiple other ports on the same host and on different hosts. This will allow multiple SNMP managers on z/OS to be able to receive all the traps sent to one port.

To configure the Trap Forwarder daemon, perform the following tasks:

1. Provide PROFILE.TCPIP statements.
2. Provide Trap Forwarder configuration.
3. Start the Trap Forwarder daemon (TRAPFWD).

Provide PROFILE.TCPIP statements

Add or update the following PORT configuration statements in *hlq*.PROFILE.TCPIP.

The default port used by the trap forwarder daemon to receive trap datagrams is UDP port 162. If you want to ensure that no other application uses this port, you must specify the following PORT statement:

```
PORT
  162 UDP TRAPFWD   ;  Trap Forwarder daemon
```

If the daemon will be started from the z/OS shell, reserve the port for z/OS UNIX by changing OMVS instead of TRAPFWD. Note that by doing so, the snmp command could make use of the port if it is started (with the trap option) before TRAPFWD is started.

Provide trap forwarder configuration information

The TRAPFWD.CONF file defines the configuration parameters for trapfwd daemon.

A sample of the TRAPFWD.CONF is shown below:

```
#
# A sample configuration file for trapfwd
#
# Syntax  : ip_address/host_name            port_number            option
#
#
9.67.113.79               1064           ADD_RECVFROM_INFO
myHost                    1066
myHost                    1067           ADD_RECVFROM_INFO
myHost                    1099
9.67.35.37                1064           ADD_RECVFROM_INFO
-                         1065
-                         1068           ADD_RECVFROM_INFO
12ab::2                   1069
```

For more information about the statement syntax, refer to *z/OS Communications Server: IP Configuration Reference*.

Starting and stopping the trap forwarder daemon

The Trap Forwarder daemon can be started from the z/OS UNIX shell or from the MVS console.

Starting the trap forwarder daemon from z/OS UNIX

This example starts the Trap Forwarder daemon on the standard port (port 162).

```
# trapfwd
```

This example starts the Trap Forwarder daemon on a particular port (port 5062).

```
# trapfwd -p 5062
```

Starting the trap forwarder daemon from an MVS console: Update cataloged procedure TRAPFWD by copying the sample in SEZAINST(EZASNTRA) to your

system. Refer to *z/OS Communications Server: IP Configuration Reference* the for more information about this cataloged procedure.

Stopping the trap forwarder daemon:

From MVS:

```
P TRAPFWD
```

If TRAPFWD was started from a cataloged procedure, procname is the member name of that procedure. If TRAPFWD was started from the z/OS shell, procname is useridX, where X is the sequence number set by the system. To determine the sequence number issue d omvs u=userid from the console. This will show the programs running under the user ID.

From UNIX:

```
kill < pid >
```

Issue the kill command to the process ID (PID) associated with TRAPFWD. To find the PID issue the *ps -ef* command from the z/OS shell.

Tracing: The modify command can be used to display the existing tracing level and also to change the tracing level.

The following example sets the trace level of the Trap Forwarder Daemon to 1.

```
MODIFY TRAPFWD,TRACE,LEVEL=1
```

The following example displays the level of tracing set for the Trap Forwarder Daemon.

```
MODIFY TRAPFWD,TRACE,QUERY
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

Dynamically refreshing configuration: The modify command can be used to dynamically refresh the configuration information. When this is done, the old configuration information is discarded completely. The configuration file is read again and the daemon is initialized.

The following example refreshes the configuration by reading the configuration file.

```
MODIFY TRAPFWD,REFRESH
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about syntax.

Chapter 21. Remote Print Server (LPD)

Read "Understanding search orders of configuration information" on page 28. It covers important information about data set naming and search sequences.

The Remote Print (LPD) server supports the Line Print Daemon and allows you to print on JES controlled printers from any host in your TCP/IP network that implements the Line Print client functions. These client functions are invoked with the LPR command. LPR is available as a TSO command, and the LPD server is implemented as a started z/OS task.

Refer to the *z/OS Communications Server: IP System Administrator's Commands* for information on starting and stopping the TCP/IP print server (LPD). When LPD is stopped by the MVS operator with the *P procname* command, LPD will terminate any TCP/IP connections currently transferring data. Before ending, LPD will finish spooling to JES any print jobs that it has received and is currently spooling. JES will handle these jobs after LPD ends.

This chapter describes how to configure the LPD server. To operate the LPD server after it is running, refer to *z/OS Communications Server: IP Configuration Reference*.

The LPD server uses the Pascal socket API, so VMCF must be started for the server to successfully initialize. If VMCF is not started, message EZY1980E will be issued and the server will terminate.

Configuring the Remote Print Server

Steps to configure the Remote Print Server:

1. Specify AUTOLOG and PORT statements in PROFILE.TCPIP.
2. Update the LPD server cataloged procedure.
3. Update the LPD server configuration data set.
4. Create a banner page (optional).

For information about operating and controlling the LPD server, refer to *z/OS Communications Server: IP Configuration Reference*.

Step 1: Configuring PROFILE.TCPIP for LPD

If you want the LPD server started automatically when the TCPIP address space is started, include the name of the member containing the LPD server cataloged procedure in the AUTOLOG list in *hlq*.PROFILE.TCPIP. For example:

```
AUTOLOG
  LPSERVE
ENDAUTOLOG
```

To ensure that port TCP 515 is reserved for the LPD server, also add the name of the member containing the LPD cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP. For example:

```
PORT
  515 TCP LPSERVE
```

See the *z/OS Communications Server: IP Configuration Reference* for more information about the AUTOLOG and PORT statements.

Step 2: Updating the LPD server cataloged procedure

Update the LPD server cataloged procedure to suit your local configuration by copying the sample to your system or recognized PROCLIB from SEZAINST(LPSPROC) and modifying the sample. Specify LPD parameters, and change the data set names as required. See "Specifying LPD server parameters" for more information on the LPD server parameters.

Specifying LPD server parameters

The system parameters required by the LPD server are passed by the PARM option on the EXEC statement of the LPD cataloged procedure. Update the following parameters as required.

LPDDATA='*data_set_name***'**
Specifies the fully qualified name of the data set containing the LPD configuration statements. This data set can be sequential or a member of a PDS.

LPDPRFX='PREFIX *your_prefix***'**
Specifies the high-level qualifier to be used for temporary data sets created by the LPD server. Include both the PREFIX keyword and your qualifier in the quoted string. The qualifier may be up to 26 characters. If it is blank, it defaults to the procedure name. The LPD task requires the authority to create and modify data sets with this prefix.

DIAG='*options***'**
Specifies any of the following diagnostic options in a quoted string of keywords separated by blanks. For example, `DIAG='VERSION TRACE'`

VERSION
Displays the version number.

TYPE Activates high-level trace facility in the LPD server. Significant events, such as the receipt of a job for printing, are recorded in the //SYSOUT DD data set specified in your LPD server cataloged procedure.

TRACE
Causes a detailed trace of activities within the LPD server to record in the //SYSOUT DD data set specified in your LPD server cataloged procedure. The detailed tracing can also be activated with the DEBUG statement in the LPD server configuration data set and with the TRACE command of the SMSG interface.

Note: The JCL PARM= statement has a limit of 100 characters.

Configuring LPDDATA

Use a DD card that requires the LPD configuration file to be in a data set.

```
//SYSLPDD  DD DISP=SHR,DSN=TCPIP.SEZAINST(LPDDATA)
```

Using this example, the DD card must be specified as SYSLPDD, but the data set name can be any valid data set name with a member specified up to 44 characters.

To use the DD card method, you must comment out or remove the LPDDATA= parameter from the PROC statement and remove the "&LPDDATA'' PARM from the LPD EXEC statement.

Note: The search order for the configuration file is:

1. LPDDATA= on the PARM= statement
2. //SYSLPDD DD statement
3. *hlq*.LPD.CONFIG

If both the LPDDATA= statement on the PARM= statement and the //SYSLPDD DD statement are specified, the data set name specified on LPDDATA= is used.

The LPD server does not limit the number of print jobs it handles per connection. This can cause a memory abend to occur if too many print jobs are sent in one connection. Certain LPR clients, such as SUN UNIX, are set up to send multiple jobs in one connection. It is recommended that the LPD start procedure be started with a region size of 6M and the LPR client send no more than 50 print jobs in one connection.

Step 3: Updating the LPD server configuration data set

The LPD configuration data set defines the local, NJE, and remote services (printers and punches) used by the LPD server. To update the LPD server configuration data set, copy the sample provided in SEZAINST(LPDDATA) and modify it to suit your installation. Refer to *z/OS Communications Server: IP Configuration Reference* for more details.

- To define a printer or punch:
  - Include a SERVICE statement with the appropriate parameters for each printer or punch your are defining.
  - Specify the type of service with the LOCAL, NJE, or REMOTE parameter in the SERVICE statement.
  - For local or NJE services, include any of the optional parameters to further define the service: EXIT, FAILEDJOB, FILTERS, LINESIZE, PAGESIZE, RACF, and SMTP.
  - For remote services, specify the destination printer and host. Any additional specifications are defined on the remote system and are not necessary in the SERVICE statement.
- To turn on LPD server tracing, include the optional DEBUG statement.
- To authorize users for the SMSG interface, include the optional OBEY statement.
- Printer names cannot contain an at sign (@).

Step 4: Creating a banner page (optional)

LPBANNER is the name of the default program that is provided in executable form in the SEZATCP data set. This program is specified on the EXIT parameter in the SERVICE statement. LPBANNER prints a separator page that contains, in large letters, a banner stating “LPD BANNER”, the user ID, the job name, and the job class. Field headings of HOST, USER, JOB, and CLASS appear in smaller letters.

The sample exit LPBANNER uses machine carriage control and is designed to be used with the SERVICE PRINTER LOCAL or SERVICE PRINTER NJE statements. To use this sample exit, both SERVICE statements require a printed LINESIZE of 78 or greater. Banner pages are usually not used with the SERVICE PUNCH or SERVICE RECFMU statements; however, if you want to have banner pages (headers) for the SERVICE PUNCH or SERVICE RECFMU devices, a user created banner exit is required.

You can either use the executable form or copy and modify the sample source provided in SEZAINST(EZAAE04S) and SEZAINST(EZAAE04T) to create a banner. If you are changing the source to create your own banner, assemble and link-edit these data sets as reentrant. You can modify and use the following JCL to do this. Changing the ENTRY and NAME to something other than LPBANNER will avoid possible maintenance problems in the future.

```
//ASMLNK JOB MSGLEVEL=(1,1),MSGCLASS=A,CLASS=A,REGION=1024K
//ASM1   EXEC PGM=ASMA90,PARM='OBJECT,XREF(FULL)'
//STEPLIB DD  DISP=SHR,DSN=HLA.OSV1R4.SASMMOD1
//*  ASSEMBLER H
//SYSLIB   DD DSN=tcpip_hlq.SEZACMAC,DISP=SHR
//         DD DSN=SYS1.MACLIB,DISP=SHR
//         DD DSN=SYS1.MODGEN,DISP=SHR
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(5,5)),DSN=&&SYSUT1
//SYSPUNCH DD DUMMY,DCB=BLKSIZE=80
//SYSPRINT DD SYSOUT=A
//SYSLIN   DD DSN=&&OBJECT(EZAAE04S),DISP=(,PASS),UNIT=SYSDA,
//            SPACE=(CYL,(5,5,1)),DCB=BLKSIZE=400
//SYSIN    DD DSN=tcpip_hlq.SEZAINST(EZAAE04S),DISP=SHR
/*
//ASM2   EXEC PGM=ASMA90,PARM='OBJECT,NODECK,XREF'
//STEPLIB  DD  DISP=SHR,DSN=HLA.OSV1R4.SASMMOD1
//*  ASSEMBLER H
//SYSLIB   DD DSN=tcpip_hlq.SEZACMAC,DISP=SHR
//         DD DSN=SYS1.MACLIB,DISP=SHR
//         DD DSN=SYS1.MODGEN,DISP=SHR
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(5,5)),DSN=&&SYSUT1
//SYSPUNCH DD DUMMY,DCB=BLKSIZE=80
//SYSPRINT DD SYSOUT=A
//SYSLIN   DD DSN=&&OBJECT(EZAAE04T),DISP=(OLD,PASS)
//SYSIN    DD DSN=tcpip_hlq.SEZAINST(EZAAE04T),DISP=SHR
/*
//LNK    EXEC PGM=IEWL,PARM='LIST,NCAL,RENT,LET'
//SYSPRINT DD SYSOUT=A
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(1,1))
//SYSLMOD  DD DSN=tcpip_hlq.SEZATCP,DISP=SHR
//AEZAMODS DD DSN=tcpip_hlq.AEZAMODS,DISP=SHR
//OBJECT   DD DSN=&&OBJECT,DISP=(OLD,DELETE)
//SYSLIN   DD *
  ORDER   EZBOECPR
  INCLUDE AEZAMODS(EZBOECPR)
  INCLUDE OBJECT(EZAAE04S)
  INCLUDE OBJECT(EZAAE04T)
  MODE    AMODE(24),RMODE(24)
  ENTRY   LPBANNER
  NAME    LPBANNER(R)
/*
```

Chapter 22. Remote procedure calls

This chapter contains information to help you configure:

- PORTMAP
- UNIX PORTMAP
- NCS
- NDB

Read "Understanding search orders of configuration information" on page 28. It covers important information about data set naming and search sequences.

Configuring the PORTMAP address space

This section describes how to configure the PORTMAP address spaces, which runs the Portmapper function.

Portmapper is the software that supplies client programs with the port numbers of server programs. Clients contact server programs by sending messages to port numbers where receiving processes receive the message. Because you make requests to the port number of a server rather than directly to a server program, client programs need a way to find the port number of the server programs they wish to call. Portmapper standardizes the way clients locate the port number of the server programs supported on a network.

Steps to configure PORTMAP:

1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the PORTMAP cataloged procedure.
3. Define the data set for well-known procedure names.

Step 1: Configuring PROFILE.TCPIP for PORTMAP

If you want the PORTMAP server to start automatically when the TCPIP address space is started, you should include PORTMAP in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  PORTMAP
ENDAUTOLOG
```

Note: If your system is using the Network File System (NFS) server, you ***must*** start the PORTMAP address space. See *z/OS Network File System Guide and Reference* for more information.

To ensure that port UDP 111 and TCP 111 are reserved for the PORTMAP server, also add the name of the member containing the PORTMAP cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP:

```
PORT
  111 UDP PORTMAP
  111 TCP PORTMAP
```

See the *z/OS Communications Server: IP Configuration Reference* for more information about the AUTOLOG and PORT statements.

Step 2: Updating the PORTMAP cataloged procedure

Update the PORTMAP cataloged procedure to suit your local conditions by copying the sample provided in SEZAINST(PORTPROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Change the data set names as required. Refer to the *z/OS Communications Server: IP Configuration Reference* for more details.

Step 3: Defining the data set for well-known procedure names

z/OS Communications Server includes a data set that contains a list of commonly used procedure names. This data set is used by the RPCINFO command to resolve remote program numbers into their equivalent program names.

To create the data set, copy SEZAINST(ETCRPC) to the default data set called *hlq*.ETC.RPC. If a user has a *user_id*.ETC.RPC data set defined, it takes precedence over the preceding data set.

Normally, you would not change this data set except to add a new application to the list. To add a new application, add a line that contains the following items:

- The *program_name* of the new application or procedure
- The *program_number* of the new application or procedure
- Any comments regarding the description of the program

The items are variable in format, each separated by a blank.

The SEZAINST(ETCRPC) data set contains the well-known procedure names. Following is the ETCRPC sample.

```
#
#   rpc   1.2   86/10/07
#
#   COPYRIGHT = NONE.
#
portmapper    100000    portmap sunrpc
rstatd        100001    rstat rup perfmeter
rusersd       100002    rusers
nfs           100003    nfsprog
ypserv        100004    ypprog
mountd        100005    mount showmount
ypbind        100007
walld         100008    rwall shutdown
yppasswdd     100009    yppasswd
etherstatd    100010    etherstat
rquotad       100011    rquotaprog quota rquota
sprayd        100012    spray
3270_mapper   100013
rje_mapper    100014
selection_svc 100015    selnsvc
database_svc  100016
rexd          100017    rex
alis          100018
sched         100019
llockmgr      100020
nlockmgr      100021
x25.inr       100022
statmon       100023
status        100024
#
# NDB Program numbers added for more useful rpcinfo output
# Values are in decimal. Corresponding hexadecimal values
# for the NDB port manager is X'20000020' and for the NDB
```

```
# servers are from X'20000021' to X'20000084'.
#
ndbportmgr    536870944
ndbserver1    536870945
ndbserver2    536870946
ndbserver3    536870947
ndbserver4    536870948
ndbserver5    536870949
ndbserver6    536870950
ndbserver7    536870951
ndbserver8    536870952
ndbserver9    536870953
ndbserver10   536870954
ndbserver11   536870955
ndbserver12   536870956
ndbserver13   536870957
ndbserver14   536870958
ndbserver15   536870959
ndbserver16   536870960
ndbserver17   536870961
ndbserver18   536870962
ndbserver19   536870963
ndbserver20   536870964
ndbserver21   536870965
ndbserver22   536870966
ndbserver23   536870967
ndbserver24   536870968
ndbserver25   536870969
ndbserver26   536870970
ndbserver27   536870971
ndbserver28   536870972
ndbserver29   536870973
ndbserver30   536870974
ndbserver31   536870975
ndbserver32   536870976
ndbserver33   536870977
ndbserver34   536870978
ndbserver35   536870979
ndbserver36   536870980
ndbserver37   536870981
ndbserver38   536870982
ndbserver39   536870983
ndbserver40   536870984
ndbserver41   536870985
ndbserver42   536870986
ndbserver43   536870987
ndbserver44   536870988
ndbserver45   536870989
ndbserver46   536870990
ndbserver47   536870991
ndbserver48   536870992
ndbserver49   536870993
ndbserver50   536870994
ndbserver51   536870995
ndbserver52   536870996
ndbserver53   536870997
ndbserver54   536870998
ndbserver55   536870999
ndbserver56   536871000
ndbserver57   536871001
ndbserver58   536871002
ndbserver59   536871003
ndbserver60   536871004
ndbserver61   536871005
```

```
ndbserver62   536871006
ndbserver63   536871007
ndbserver64   536871008
ndbserver65   536871009
ndbserver66   536871010
ndbserver67   536871011
ndbserver68   536871012
ndbserver69   536871013
ndbserver70   536871014
ndbserver71   536871015
ndbserver72   536871016
ndbserver73   536871017
ndbserver74   536871018
ndbserver75   536871019
ndbserver76   536871020
ndbserver77   536871021
ndbserver78   536871022
ndbserver79   536871023
ndbserver80   536871024
ndbserver81   536871025
ndbserver82   536871026
ndbserver83   536871027
ndbserver84   536871028
ndbserver85   536871029
ndbserver86   536871030
ndbserver87   536871031
ndbserver88   536871032
ndbserver89   536871033
ndbserver90   536871034
ndbserver91   536871035
ndbserver92   536871036
ndbserver93   536871037
ndbserver94   536871038
ndbserver95   536871039
ndbserver96   536871040
ndbserver97   536871041
ndbserver98   536871042
ndbserver99   536871043
ndbserver100  536871044
```

Starting the PORTMAP address space

If you did not include PORTMAP in the AUTOLOG statement, you can start PORTMAP with the START command. For example:

```
START PORTMAP
```

PORTMAP must be started before NDB can run.

Configuring the z/OS UNIX PORTMAP address space

This section describes how to configure the z/OS UNIX PORTMAP address space, which runs the Portmapper function.

Steps to configure z/OS UNIX PORTMAP:

1. Specify PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the PORTMAP cataloged procedure.

Step 1: Configuring PROFILE.TCPIP for UNIX PORTMAP

If you want the PORTMAP server to start automatically when the TCPIP address space is started, you should include PORTMAP in the AUTOLOG statement in the hlq.PROFILE.TCPIP data set.

```
AUTOLOG
  PORTMAP JOBNAME PORTMAP1
ENDAUTOLOG
```

To ensure that port UDP 111 and TCP 111 are reserved for the z/OS UNIX PORTMAP server, add the z/OS UNIX PORTMAP server jobname to the PORT statement in *hlq*.PROFILE.TCPIP. If you use the sample cataloged procedure, PORTMAP, to start the z/OS UNIX PORTMAP server, the jobname is PORTMAP1:

```
PORT
  111 UDP PORTMAP1     ; Portmapper Server
  111 TCP PORTMAP1     ; Portmapper Server
```

See the *z/OS Communications Server: IP Configuration Reference* for more information about the PORT statement.

Step 2: Updating the PORTMAP cataloged procedure

Update the PORTMAP cataloged procedure to suit your local conditions by copying the sample provided in SEZAINST(OPORTPRC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Change the data set names as required. Refer to the *z/OS Communications Server: IP Configuration Reference* for more details.

Starting the PORTMAP address space

There are two ways to start the portmapper as a z/OS UNIX socket application:

- From the z/OS shell
- As a started task

To start the portmapper from the z/OS shell, the user ID must be an authorized superuser. The authorized superuser ID can issue *oportmap &* to start the portmapper. For the authorization procedure, see *z/OS UNIX System Services Planning*.

You can also start PORTMAP as a started task with the START command as follows:

```
START PORTMAP
```

Note: If your system is using the Network File System (NFS) server, see *z/OS Network File System Guide and Reference* for more information.

Configuring the NCS interface

This section describes how to configure the Network Computing System (NCS).

NCS is the Remote Procedure Call (RPC) implementation of Apollo's Network Computing Architecture (NCA**).

NCS includes:

1. RPC run-time library
2. Location broker
3. Network Interface Definition Language (NIDL) compiler

The RPC run–time library and the location broker provide runtime support. Together these two elements make up the Network Computing Kernel (NCK) which includes all the software necessary to run a distributed application. The NIDL compiler is a tool for developing NCS-based applications.

In order to execute NCS applications in an MVS environment, you must start a local location broker (LLBD). One of the hosts in your TCP/IP network must also start the global location broker (GLBD). Both the LLBD and the NRGLBD maintain information about active NCS server applications.

Understanding the LLBD server

The LLBD manages the LLB database which stores information for the NCS-based servers running on this host.

Your host must run LLBD if you want to support the location broker forwarding function or allow remote access to the LLB database. The LLBD function must be started on the host before any other NCS-based servers are started and before the NRGLBD is started.

The LLB database is stored in the data set ADM@SRV.LLB@LL.DATABASE, which is not created until an entry is registered with the LLBD. This data set can be administered using the lb@admin tool.

Understanding the NRGLBD server

The NRGLBD manages the NCS global location broker (GLB) database. The GLB helps clients locate servers on a network or internet.

There are two versions of the GLB daemon: replicated GLBD and NRGLBD. The replicated GLBD is only available on Apollo, SunOS, and Ultrix systems. For other systems, such as IBM, only the NRGLBD is available. The advantage of replicated GLBD over NRGLBD is that the GLBD can be run at the same time on several network hosts, providing greater availability in the event that one of the hosts running GLBD fails or if there is a partial network failure.

You cannot use the NRGLBD on your system if:
- The replicated version of GLBD can run on some other host in your network
- Another host in your network is already running NRGLBD

The GLB database is stored in a data set ADM@SRV.LLB@LG.DATABASE. This data set is not created until an entry is registered with the NRGLBD.

The record structure for the LLBD and the NRGLBD databases is identical.

Steps to configure NCS:
1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the NRGLBD cataloged procedure in SEZAINST(NRGLBD).
3. Update the LLBD cataloged procedure in SEZAINST(LLBD).

For more information about NCS, see *Network Computing System Reference Manual*.

Step 1: Configuring PROFILE.TCPIP for NCS

If you want LLBD and NRGLBD to start automatically when the TCPIP address space is started, then you should include the names of the members containing the LLBD and NRGLBD cataloged procedures in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

The LLBD must be running before you start NRGLBD. Therefore, you must put LLBD before NRGLBD in the AUTOLOG statement.

```
AUTOLOG
  LLBD
  NRGLBD
ENDAUTOLOG
```

To ensure that port UDP 135 is reserved for the LLBD server, add the name of the member containing the LLBD cataloged procedure to the PORT statement in the *hlq*.PROFILE.TCPIP data set.

```
PORT
  135 UDP LLBD
```

Note: z/OS UNIX DCE also uses port 135 and therefore cannot be used concurrently with NCS.

See the *z/OS Communications Server: IP Configuration Reference* for more information about the AUTOLOG and PORT statements.

Step 2: Updating the NRGLBD cataloged procedure

Update the NRGLBD cataloged procedure by copying the sample provided in SEZAINST(NRGLBD) to your system or recognized PROCLIB and modifying it to suit your local conditions. Refer to the NCS chapter in *z/OS Communications Server: IP Configuration Reference* for more details.

Step 3: Updating the LLBD cataloged procedure

Update the LLBD cataloged procedure by copying the sample provided in SEZAINST(LLBD) to your system or recognized PROCLIB and modifying it to suit your local conditions. Refer to the NCS chapter in *z/OS Communications Server: IP Configuration Reference* for more information.

Configuring the Network Database (NDB) System

This section describes how to configure the Network Database (NDB) System. NDB provides access to mainframe relational database systems using TCP/IP and the remote procedure call (RPC) protocol. This allows interoperability among a variety of workstation and mainframe users with DB2 and MVS. End users in a VM, MVS, DOS, OS/2, RISC System/6000 AIX, or Sun UNIX Microsystems environment can issue SQL statements interactively or invoke NDB services from within a C application program. NDB services can then be used to pass SQL statements to DB2 and to handle responses from DB2.

The Network Database (NDB) System consists of:

- The Network Database (NDB) Port Manager
- The NDB Port Client
- 1–100 NDB Servers
- NDB Clients

Note: The NDB System requires the Portmapper to run. See "Configuring the PORTMAP address space" on page 1085 for further information.

Steps to Configure the NDB System

1. Update the NDB setup sample job (NDBSETUP).
2. Run the NDB setup sample job.
3. Update and install the DB2 sample connection exit routine (DSN3SATH), if necessary.
4. Update the PORTS cataloged procedure (PORTSPRC).
5. Update the PORTC cataloged procedure (PORTCPRC).
6. Create the NDB clients.

Note: Repeat steps 1 and 2, and step 3 if necessary, to configure each DB2 subsystem that is to be accessed by NDB servers.

Step 1: Updating the NDB setup sample job

Update the NDBSETUP sample job by copying the sample in SEZAINST(NDBSETUP) to a partitioned data set (PDS) or sequential data set and modifying it to suit your local installation. Change the DD statements as required. A copy of NDBSETUP can also be found *z/OS Communications Server: IP Configuration Reference*.

To restrict the use of NDB to selected users, modify the GRANT statement in NDBSETUP job replacing PUBLIC with a list of the specified TSO user IDs to whom you want to grant access. See the *DB2 SQL Reference* for correct statement syntax.

THE NDBSETUP job binds the DBRM DBUTIL2 into the DB2 subsystem specified. In previous releases, the DBRM was bound using the plan name DBUTIL2. For V3R2, the DBRM is bound using the plan name EZAND320. By using different plan names for each release, you can have different levels of TCP/IP simultaneously accessing the same DB2 subsystems.

Note: NDBSETUP binds the DBRM DBUTIL2 with the isolation level of cursor stability: ISOLATION(CS) on the BIND command. If a higher degree of unit of work integrity is needed than is provided by an isolation level of cursor stability, this can be changed to an isolation level of repeatable read: ISOLATION(RR) on the BIND command. Refer to the *DB2 Application Programming and SQL Guide* for information regarding isolation level settings and their effects. Check with your database administrator regarding isolation level policies at your site.

Step 2: Running the NDB setup job

Run NDBSETUP to bind the DBRM DBUTIL2 to the DB2 subsystem specified and grant the execute (run) privileges to public. Verify that it ran successfully before proceeding with this configuration.

Notes:

1. The DB2 subsystem must be Version 2 Release 3 or higher. This requirement applies to all TCP/IP servers that use DB2.
2. The DB2 subsystem must be up and running before you submit the NDBSETUP job.

3. The NDBSETUP job must be executed from a user ID with the authority to bind plans for the specified DB2 subsystem.

Step 3: Updating and installing the DB2 sample connection exit routine

Note: If you will be running only one NDB server in a single address space, you may skip this step.

Because the address space running a PORTC cataloged procedure is capable of running between 1 to 20 NDB servers, changes must be made in the DB2 sample connection exit routine, DSN3SATH ASSEMBLE. These changes tell DB2 where to look for the host userid to be used for this session (the host userid was supplied by the user at NDB client invocation).

After making the necessary changes in DSN3SATH ASSEMBLE, this routine can be assembled and link edited using the DB2 supplied job DSNTIJEX in the DSNSAMP library. See DSN3SATH ASSEMBLE modifications for NDB for the updated assembler code. For more information, see the *DB2 Administration Guide*. DSNTIJEX will place a copy of the executable in the DSNEXIT library. There is one DSNEXIT library for each DB2 subsystem and it must be updated with the recommended modifications for each DB2 subsystem that will be accessed via NDB. Run DSNTIJEX once (updating the hlq for the DSNEXIT library) for each DB2 subsystem that will be accessed via NDB.

DSN3SATH ASSEMBLE modifications for NDB

```
⋮
**********SECTION 1:  DETERMINE THE PRIMARY AUTHORIZATION ID  *********
*                                                                     *
*   IF THE INPUT AUTHID IS NULL OR BLANKS, CHANGE IT TO THE AUTHID    *
*   IN EITHER THE JCT OR THE FIELD POINTED TO BY ASCBJBNS.            *
*                                                                     *
*   THE CODE IN THIS SECTION IS AN ASSEMBLER LANGUAGE VERSION OF      *
*   THE DEFAULT IDENTIFY AUTHORIZATION EXIT.  IT IS EXECUTED ONLY     *
*   IF THE FIELD ASXBUSER IS NULL UPON RETURN FROM THE RACROUTE       *
*   SERVICE.  FOR EXAMPLE, IT DETERMINES THE PRIMARY AUTH ID FOR      *
*   ENVIRONMENTS WITH NO SECURITY SYSTEM INSTALLED AND ACTIVE.        *
*                                                                     *
***********************************************************************
         SPACE
* MOVED CHECK ON AIDLPRIM THAT WAS HERE TO PAST CHECK FOR TSO FOREGRND
         L     R7,ASCBCSCB         GET CSCB ADDRESS
         CLI   CHTRKID-CHAIN(R7),CHTSID IS IT TSO FOREGROUND ADDR SPACE
         BNE   SATH010             BRANCH IF NOT
* HERE IS NEW LOCATION OF CHECK ON AIDLPRIM
         CLI   AIDLPRIM,BLANK      IS THE INPUT PRIMARY AUTHID NULL
         BH    SATH020             SKIP IF A PRIMARY AUTH ID EXISTS
* END OF MOVED CODE - CHECK ON AIDLPRIM
         L     R7,ASCBJBNS         GET ADDRESS OF LOGON ID
         MVC   AIDLPRIM,0(R7)      MAKE IT THE PRIMARY AUTH ID
         B     SATH019             TO END OF THIS ROUTINE
SATH010  DS    0H                  NOT TSO, BUT BATCH OR STC SPACE
         L     R6,PSATOLD-PSA      CURRENT TCB ADDRESS
* START OF CODE ADDED TO SUPPORT TCP/IP NDB MULTI DB2 SUBSYSTEM FUNCT
         USING TCB,R6
         L     R7,TCBSENV          CURRENT ACEE ADDRESS
         USING ACEE,R7
         CLC   ACEEACEE,EYEACEE    IS THIS AN ACEE?
         BNE   SATH015             NO, GO USE JOB USER ID
         L     R14,TCBJSCB         ADDRESS OF JSCB
         USING IEZJSCB,R14
```

```
         CLC   JSCBPGMN,=CL8'PORTCLNT' IS IT NDB SERVER AS ?
         BNE   SATH015             NO, GO USE JOB USER ID
         MVC   AIDLPRIM,ACEEUSRI   MOVE USER ID INTO AIDL PRIMARY
         MVC   AIDLACEE,TCBSENV    MOVE ACEE POINTER INTO AIDL
         B     SATH020
         DROP  R6,R7,R14
SATH015  DS    0H
         CLI   AIDLPRIM,BLANK      IS THE INPUT PRIMARY AUTHID NULL
         BH    SATH020             SKIP IF A PRIMARY AUTH ID EXISTS
* END OF CODE ADDED TO SUPPORT TCP/IP NDB MULTI DB2 SUBSYSTEM FUNCTION
         L     R7,TCBJSCB-TCB(,R6) CURRENT JSCB ADDRESS
         L     R5,JSCBJCT-IEZJSCB(,R7) CURRENT JCT ADDRESS
         LA    R5,X'10'(,R5)       ADJUST FOR CORRECT DSECT MAPPING
         MVC   AIDLPRIM(7),JCTUSER-INJMJCT(R5)  COPY JOB USER ID
         MVI   AIDLPRIM+7,BLANK    ASSURE BLANK PADDING
SATH019  DS    0H                  END OF ROUTINE
         EJECT
⋮
```

Step 4: Updating the PORTS cataloged procedure

The PORTS procedure starts the NDB Port Manager. Update the PORTS cataloged procedure by copying the SEZAINST(PORTSPRC) sample to your system or recognized PROCLIB and modifying it to suit your local installation. Change the DD statements, as required. Refer to *z/OS Communications Server: IP Configuration Reference* for the PORTS procedure.

Note: You must start PORTS before you start PORTC.

Step 5: Updating the PORTC cataloged procedure

The PORTC procedure starts the NDB Port Client and the NDB Servers. Update the PORTC cataloged procedure by copying the sample in SEZAINST(PORTCPRC) to your system or recognized PROCLIB and modifying it to suit your local installation. Refer to the *z/OS Communications Server: IP Configuration Reference* for a copy of PORTC.

Running multiple PORTC procedures

Currently the NDB Port Manager is able to manage up to 100 NDB Servers. Each PORTC procedure is able to start up to 20 NDB servers. The number of NDB Servers that can be started in a single address space depends on a number of factors, such as how large a region size can be specified and the size of the catalog work area. These factors may limit the number of NDB Servers able to start in a single address space to less than the maximum of 20.

In order to reach the desired number of NDB Servers, you can run multiple PORTC procedures. When starting a PORTC procedure, one of the parameters specified is DB2SSID. See *z/OS Communications Server: IP Configuration Reference* for more information on the DBSSID parameter. This parameter indicates which DB2 subsystem all NDB servers running in that address space will access. Each started PORTC procedure may contain a different value for the DBSSID parameter; that is, when starting multiple PORTC procedures, each procedure may point to the same or different DB2 subsystems. To do this:

1. Copy your customized PORTC cataloged procedure to another data set or PDS member.
2. Change the name in the first statement (PROC statement). For example if your original PORTC procedure starts with //PORTC PROC, use //PORTC1 PROC and //PORTC2 PROC for the other procedures.
3. Ensure that:

- HOMEID parameter settings are the same for all PORTC procedures.
- When the NUMSRV parameter settings for all the PORTC procedures are added together, the total does not exceed 100.
- The NUMSRV value for any given PORTC does not exceed 20.

4. Start the new PORTC procedures.

Step 6: Creating the NDB clients

z/OS Communications Server provides NDB sample client code that can be compiled and run in a variety of environments. NDB clients can be created on VM, MVS, or on DOS, OS/2, AIX on RS/6000, or SUN UNIX workstations. You can find the NDB client source code in the SEZAINST data set, unless otherwise noted. They are written in the C language.

The process to create a client is similar in each case. Instructions for creating a client in each of the supported environments are specified. The MVS client code shipped with TCP/IP is ready to use without modification but you can modify it, if necessary, to suit your environment.

Note: To build an NDB client, you need access to the C runtime libraries and the TCP/IP RPC libraries. The TCP/IP products for the client platforms must be installed on those platforms before you build the NDB clients. In addition, for the DOS and OS/2 platforms, you must also install the TCP/IP Programmer's Toolkit.

Creating an NDB client in the AIX environment

1. Bring the following C source programs down to your workstation:

Program	Rename to
NDBCC	ndbc.c
NDBCLTC	ndbclt.c
NDBMAINC	ndbmain.c
NDBOUTC	ndbout.c
NDBPCLTC	ndbpclt.c
NDBXC	ndbx.c

2. Bring the following H (header) files down to your workstation:

Program	Rename to
NDBCLTH	ndbclt.h
NDBGLOBH	ndbglob.h
NDBH	ndb.h
NDBOUTH	ndbout.h
NDBPCLTH	ndbpclt.h
NDBRPCFH	ndbrpcf.h

3. Issue the command:

```
cc -o ndbclnt ndbmain.c ndbc.c ndbclt.c ndbout.c ndbpclt.c ndbx.c
```

Notes:

1. This version of the NDB sample client code was produced and tested on AIX Version 3 Release 2 for RISC System/6000.
2. The file NDBRPCFC (ndbrpcf.c) is not necessary unless the level of RPC being used does not support the RPC clnt_create function.
3. If the NDBH (ndb.h) file, an output of the RPCGEN process, is erased or corrupted and cannot be retrieved from the MVS TCP/IP library, do the following to generate a new copy before compiling the NDB sample client code.

a. Bring the following X (RPC input) file down to your workstation:

Program	Rename to
NDBXX	ndb.x

b. Issue the command:

```
rpcgen -h -o ndb.h ndb.x
```

For more information on the RPCGEN process, refer to *TCP/IP AIX Programmer's Reference* and *AIX 5L Version 5.1 Communications Programming Concepts*.

Creating an NDB client in the SUN UNIX environment

1. Bring the following C source programs down to your workstation:

Program	Rename to
NDBCC	ndbc.c
NDBCLTC	ndbclt.c
NDBMAINC	ndbmain.c
NDBOUTC	ndbout.c
NDBPCLTC	ndbpclt.c
NDBXC	ndbx.c

2. Bring the following H (header) files down to your workstation:

Program	Rename to
NDBCLTH	ndbclt.h
NDBGLOBH	ndbglob.h
NDBH	ndb.h
NDBOUTH	ndbout.h
NDBPCLTH	ndbpclt.h
NDBRPCFH	ndbrpcf.h

3. Issue the command:

```
cc -o ndbclnt -DNOPROTO ndbmain.c ndbc.c ndbclt.c ndbout.c ndbx.c
   ndbpclt.c
```

Notes:

1. This version of the NDB sample client code was produced and tested on SUN OS Version 4 Release 1 Modification 1.
2. The file NDBRPCFC (ndbrpcf.c) is not necessary unless the level of RPC being used does not support the RPC clnt_create function.
3. If the NDBH (ndb.h) file, an output of the RPCGEN process, is erased or corrupted and cannot be retrieved from the MVS TCP/IP library, do the following to generate a new copy before compiling the NDB sample client code.
   a. Bring the following X (RPC input) file down to your workstation:

Program	Rename to
NDBXX	ndb.x

   b. Issue the command:

```
rpcgen -h -o ndb.h ndb.x
```

Refer to *SUN Network Programming* for more information.

Creating an NDB client in the OS/2 environment

1. Bring the following C source programs down to your workstation:

Program	Rename to
NDBCC	ndbc.c
NDBCLTC	ndbclt.c

NDBMAINC	ndbmain.c
NDBOUTC	ndbout.c
NDBPCLTC	ndbpclt.c
NDBRPCFC	ndbrpcf.c
NDBXC	ndbx.c

2. Bring the following H (header) files down to your workstation:

Program	Rename to
NDBGLOBH	ndbglob.h
NDBH	ndb.h
NDBCLTH	ndbclt.h
NDBOUTH	ndbout.h
NDBPCLTH	ndbpclt.h
NDBRPCFH	ndbrpcf.h

3. Bring the following OS/2 files down to your workstation:

Program	Rename to
NDBCLDEF	ndbclnt.def
NDBOS2M	ndbos2.mak

   NDBOS2M has been stored as a binary file on MVS and, as such, cannot be read on the MVS host. When using FTP to move this file from MVS to OS/2, make sure the transfer type used is BINARY. The OS/2 makefile (NDBOS2M) is in the SEZAXLD2 data set.

4. Issue the command:

```
nmake -f ndbos2.mak
```

Notes:

1. This version of the NDB sample client code was produced and tested on OS/2 Warp Version 3.0, TCP/IP for OS/2 Version 3.0, including the TCP/IP for OS/2 Programmer's Toolkit, and IBM C Set ++® Version 2.0.
2. If the NDBH (ndb.h) file, an output of the RPCGEN process, is erased or corrupted and cannot be retrieved from the z/OS TCP/IP library, do the following to generate a new one before compiling the NDB sample client code.
   a. Bring the following X (RPC input) file down to your workstation:

Program	Rename to
NDBXX	ndb.x

   b. Issue the command:

```
rpcgen -h -o ndb.h ndb.x
```

See *z/OS Communications Server: IP Programmer's Guide and Reference* for more information.

Creating an NDB client in the DOS environment

1. Bring the following C source programs down to your workstation

Program	Rename to
NDBCC	ndbc.c
NDBCLTC	ndbclt.c
NDBMAINC	ndbmain.c
NDBOUTC	ndbout.c
NDBPCLTC	ndbpclt.c
NDBRPCFC	ndbrpcf.c
NDBXC	ndbx.c

   When using FTP to move these files from MVS to DOS, make sure the transfer type used is ASCII.

2. Bring the following H (header) files down to your workstation:

Program	Rename to
NDBCLTH	ndbclt.h
NDBGLOBH	ndbglob.h
NDBH	ndb.h
NDBOUTH	ndbout.h
NDBPCLTH	ndbpclt.h
NDBRPCFH	ndbrpcf.h

   When using FTP to move these files from MVS to DOS, make sure the transfer type used is ASCII.

3. Bring the following DOS makefile down to your workstation:

Program	Rename to
NDBDOSM	ndbdos.mak

   NDBDOSM has been stored as an binary file on MVS and, as such, cannot be read on the MVS host. When using FTP to move this file from MVS to DOS, make sure the transfer type used is BINARY. The DOS makefile (NDBDOSM) is in the SEZAXLD2 data set.

4. Issue the command:

   ```
   nmake -f ndbdos.mak
   ```

Notes:

1. This version of the NDB sample client code was produced and tested on IBM DOS Version 6.0, TCP/IP for DOS Version 2.1.1.4, including the TCP/IP for DOS Programmer's Toolkit [2], Microsoft Windows Version 3.1 and Microsoft C/C++ Version 7.0.

   The TCP/IP for DOS Programmer's Toolkit Version 2.1.1.4 contains a fix to the RPC library required to run the NDB DOS client.

2. TCP/IP for DOS does not support RPC server functions because it does not have a PORTMAPPER. Therefore, if the NDBH (ndb.h) file, an output of the RPCGEN process, is erased or corrupted and cannot be retrieved from the z/OS TCP/IP library, do the following to generate a new one on a workstation or mainframe that does support RPC server functions before compiling the NDB sample client code.

   a. Bring the following X (RPC input) file down to your workstation or mainframe:

Program	Rename to
NDBXX	ndb.x

   b. Issue the command:

      ```
      rpcgen -h -o ndb.h ndb.x
      ```

   c. Transfer the resulting NDB.H file to your DOS workstation

See *TCP/IP DOS Programmer's Reference*, and the programmer's reference manuals for the workstation or the mainframe being used for RPCGEN for more information.

Creating an NDB client in the VM environment

1. Bring the following C source programs to your VM user ID:

Program	Rename to

2. The TCP/IP for DOS Programmer's Toolkit Version 2.1.1.4 contains a fix to the RPC library that is needed for the NDB DOS client to run.

NDBCC	NDBC C
NDBCLTC	NDBCLT C
NDBMAINC	NDBMAIN C
NDBOUTC	NDBOUT C
NDBPCLTC	NDBPCLT C
NDBXC	NDBX C

2. Bring the following H (header) files to your VM user ID:

Program	Rename to
NDBCLTH	NDBCLT H
NDBGLOBH	NDBGLOB H
NDBH	NDB H
NDBOUTH	NDBOUT H
NDBPCLTH	NDBPCLT H
NDBRPCFH	NDBRPCF H

3. Bring the following Assembler file to your VM user ID:

Program	Rename to
NDBVMPWA	NDBVMPW ASSEMBLE

4. Bring the following VM REXX EXEC to your VM user ID:

Program	Rename to
NDBCLBD	NDBCLBD EXEC

5. Execute the EXEC by entering:

```
ndbclbd
```

Notes:

1. This version of the NDB Sample Client code was produced and tested on VM/ESA® Version 1 Release 2 using CMS Level 9, running in XA mode and in ESA mode.
2. The file NDBRPCFC (NDBRPCF C) is not necessary unless the level of RPC being used does not support the RPC clnt_create function.
3. If the NDBH (NDB H) file, an output of the RPCGEN process, should be erased or corrupted and cannot be retrieved from the MVS TCP/IP library, you will need to generate a new one before you can successfully compile the NDB sample client code. To do this:
   a. Bring the following X (RPC input) file to your VM user ID:

Program	Rename to
NDBXX	NDB X

   b. Issue the command:

```
rpcgen -h -o ndb h ndb x
```

See *TCP/IP for VM Programmer's Reference* for further information.

Creating an NDB client in the MVS environment

1. Copy the following C source programs to a PDS:

Program	Rename to
NDBCC	NDBC
NDBCLTC	NDBCLT
NDBMAINC	NDBMAIN
NDBOUTC	NDBOUT
NDBPCLTC	NDBPCLT
NDBXC	NDBX

2. Copy the following H (header) files to a PDS:

Program	Rename to
NDBCLTH	NDBCLT
NDBGLOBH	NDBGLOB
NDBH	NDB
NDBOUTH	NDBOUT
NDBPCLTH	NDBPCLT
NDBRPCFH	NDBRPCF

3. Copy the following sample JCL to a data set or PDS and customize it following the instructions found in the sample:
   NDBCLMVS
4. Execute the JCL.

Notes:

1. This version of the NDB Sample Client code was produced and tested on MVS/ESA Version 4 Release 2 Modification 2.
2. The file NDBRPCFC (NDBRPCF C) is not necessary unless the level of RPC being used does not support the RPC clnt_create function.
3. If the NDBH (NDB H) file, an output of the RPCGEN process, should be erased or corrupted and cannot be retrieved from the MVS TCP/IP library, you will need to generate a new one before you can successfully build the NDB sample client. To do this:
   a. Copy the following X (RPC input) file to a data set:

Program	Rename to
NDBXX	NDB.X

   b. Issue the command:

```
rpcgen -h -o ndb.h ndb.x
```

See the *z/OS Communications Server: IP Programmer's Guide and Reference* for further information.

Starting NDB

Follow these steps to start NDB:

1. Ensure that the Portmapper is up and running. For more information, see "Configuring the PORTMAP address space" on page 1085.
2. Run the PORTSPRC procedure to start the NDB Port Manager.
3. After PORTSPRC has started, run the PORTCPRC procedure to start the NDB Port Client and NDB Servers. Specify the start parameters as required.
4. Ensure that the required DB2 subsystem is running.
5. Invoke the NDB client.

Chapter 23. Mail servers

This chapter describes how to configure the Simple Mail Transfer Protocol (SMTP) server, z/OS UNIX sendmail, and popper.

Configuring the SMTP server

Before you configure...

Read "Understanding search orders of configuration information" on page 28. It covers important information about data set naming and search sequences.

Note: Before configuring the SMTP server, it is assumed that the necessary SYS1.PARMLIB changes have been made. Consult the Program Directory for current information about the storage estimates for this version. The Program Directory also contains information about customization of certain SYS1.PARMLIB members, which must be completed before the initial program load (IPL) for the MVS image.

The SMTP server uses the Pascal socket API, so VMCF must be started for the server to successfully initialize. If VMCF is not started, message EZY1980E will be issued and the server will terminate.

If you have specified PROFILE NOINTERCOM in your TSO user ID's profile, then there are some SMTP server messages that you will not receive.

Checklist for working within the SMTP environment

1. SMTP needs to interface with JES utilities to create, read, write and purge data from the JES spool. JES exit programs might interfere with SMTP functioning properly.
2. JES initialization parameters must be set up correctly so mail can be sent to SMTP and so that local mail can be placed on the JES spool for local users.
3. Because SMTP needs authority to create, read, write and purge data on the JES spool, any security programs such as RACF, protecting JES spool access, must have the SMTP started task name in their definitions of authorized users. If the security product being used is RACF, more information regarding RACF controls for JES spool can be found in *z/OS Security Server RACF Security Administrator's Guide*.
4. DASD management is important to have SMTP run properly, it is recommended that SMTP have its own dedicated volumes. The MAILFILEVOLUME statement may be used to specify a particular volume where newly allocated SMTP data sets reside. Refer to the *z/OS Communications Server: IP Configuration Reference* for more information on this parameter.
5. SMTP is a heavy user of data set I/O functions. It creates data sets for every piece of mail it processes. There are two data sets associated with each piece of mail:
   a. SMTPhlq.*.ADDRBLOK (control file)
   b. SMTPhlq.*.NOTE (message content)

The SMTP high level qualifier (SMTPhlq) is configured in the SMTP configuration data set using the MAILFILEDSPREFIX statement. When SMTP is executing, SMTP must have exclusive access to the data sets it has created to work properly.

To avoid contention, applications that manage DASD should only be run when SMTP is not active, or exclude the SMTP data sets or the volumes on which they reside from their processing. If contention occurs, EZA5335E will be displayed before the SMTP server terminates. Also, the SMTP high level qualifier can be used to exclude the SMTP data sets if necessary.

6. SMTP requires that the ASCII LineFeed character (x'0A') be translated to an EBCDIC LineFeed (x'25'). The translation table used by SMTP must have this translation. For information regarding customizing translation tables and for the search order used by SMTP to locate the translation table, refer to *z/OS Communications Server: IP Configuration Reference*.

Configuration process

Steps to configure SMTP:

1. Specify AUTOLOG and PORT statements in the *hlq*.PROFILE.TCPIP data set.
2. Update the SMTP cataloged procedure SEZAINST(SMTPPROC).
3. Customize the SMTPNOTE CLIST and modify parmlib members.
4. Customize the SMTP mail headers (optional).
5. Set up a TCP-to-NJE mail gateway (optional).
6. Specify configuration statements in the SMTP configuration data set.
7. Create an SMTP security table (optional).
8. Enable SMTP domain name resolution.
9. Enable sending messages to SMTP users and users on an IP Network.
10. Optionally, design SMTP exit to inspect and filter unwanted mail (*spam*)

Step 1: Specify AUTOLOG and PORT statements in hlq.PROFILE.TCPIP

If you want the SMTP server to start automatically when the TCPIP address space is started, include the name of the member containing the SMTP cataloged procedure in the AUTOLOG statement of the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  SMTP
ENDAUTOLOG
```

To ensure that port TCP 25 is reserved for SMTP, verify that the name of the member containing the SMTP cataloged procedure has been added to the PORT statement in *hlq*.PROFILE.TCPIP.

```
PORT
  25 TCP SMTP
```

For more information on the AUTOLOG and PORT statements, see *z/OS Communications Server: IP Configuration Reference*.

HOME statement in the *hlq*.PROFILE.TCPIP data set consideration: SMTP allows a maximum of 255 entries on the HOME statement.

Step 2: Update the SMTP cataloged procedure

Update the SMTP cataloged procedure by copying the sample in SEZAINST(SMTPPROC) to your system or recognized PROCLIB. Specify SMTP

parameters, and change the data set as required to suit your local configuration. Refer to *z/OS Communications Server: IP Configuration Reference* for more detailed information about the procedure.

Note: SMTP does not support HFS files.

Step 3: Customize the system CLIST and modify parmlib data sets

You must copy and customize the SMTPNOTE CLIST on every system where users will be able to send mail with the SMTPNOTE command. This includes TCP/IP nodes and each NJE node that sends mail through SMTP on a remote gateway node. SMTPNOTE uses the TSO transmit (XMIT) command to interface with SMTP.

Copy SEZAINST(SMTPNOTE) into the system CLIST data set. Since the SEZAINST data set is in a fixed format, the SMTPNOTE member may be truncated if your system CLIST library is not in a fixed format.

You should customize the following variables in the SMTPNOTE CLIST:

DDNAME
The DDNAME that SMTPNOTE will use to allocate the input data set. The allocation is done to allow shared access to the data set. The default value is set to EZBSMTPN and should only be changed if this value will cause a conflict on your system.

HOSTNAME
The name of the system on which this CLIST is installed. Typically, the name is the NJE node name of this system. The NJE node name of the system is the value of the NAME parameter on the NODE(nn) statement in the JES2PARM member of parmlib.

SMTPNODE
The NJE node on which the SMTP server runs. Typically, HOSTNAME and SMTPNODE have the same value. When SMTPNODE is used on an NJE network in conjunction with a TCP-to-NJE gateway, the value of this parameter is the NJE node name of that gateway.

SMTPJOB
The name of the address space in which SMTP runs at SMTPNODE. Usually this is SMTP. The SMTPJOB name must not be defined as a node name to JES and cannot begin with the characters R, RM, or RMT, since SMTPNOTE uses TSO XMIT to transmit the note to the SMTP address space.

TEMPDSN
The name of the temporary data set used to store the contents of the note being created. This can be any arbitrary data set name that ends with the low-level qualifier, TEXT. Do not use a fully qualified name. If you do not fully qualify the name (no quotes), the data set name will be prefixed by the *userid*. If you enclose the name in single quotes, several users can use this temporary data set.

TIMEZONE
The time zone for your system. This will appear in the ″Date:″ stamp of the RFC 822 header. See RFC 822 for valid time zone formats.

ATSIGN
Some foreign languages need to use a different character to represent the @ symbol. This input symbol is a single-byte representation of the @ symbol in their national language code page.

You should also modify the following parmlib members:

IEFSSNxx
When the SMTP server is running in the same host (or system) as the SMTPNOTE CLIST, the IEFSSNxx member can be modified in one of the following two ways:

- The following lines may be included:

```
TNF,MVPTSSI
VMCF,MVPXSSI, nodename
```

  where *nodename* is the NJE node name. The NJE node name, *nodename*, must be the same as the *hostname* and the *smtpnode* that are defined in the SMTPNOTE CLIST.
- If you are using restartable VMCF, you must make changes to IEFSSxx members in the SYS1.PARMLIB data set.

  For introductory information on restartable VMCF, refer to "Step 3: Configure VMCF and TNF" on page 108. For the MVS system changes required for restartable VMCF, refer to the TCP/IP for MVS Program Directory. For information on VMCF commands, refer to *z/OS Communications Server: IP Diagnosis Guide*.

  Note: You should define the SystemName in the IEFSSNxx parmlib member to be the same as your JES2 or JES3 (NJE) nodename. This is required for correct delivery of SMTP mail. For example, if the following line is coded in your SMTPNOTE CLIST:

```
SMTPNODE P390
```

  you need to code NAME=P390 in your IEFSSNxx parmlib member. As an alternative, instead of using the IEFSSNxx parmlib member to specify the JES node, you can use the keyword NJENODENAME within your SMTP configuration to a valid NJE node. For more information, see NJENODENAME.

IKJTSOxx
The TRANSREC statement must contain the correct nodename, or the NODESMF parameter can be coded as NODESMF((*,*)). For more information on the TRANSREC statement, refer to *z/OS MVS Initialization and Tuning Reference*.

Step 4: Customize the SMTP mail headers (Optional)

Electronic mail has a standardized syntax for text messages that are sent across networks. The standard syntax is described in RFC 822. Messages have an envelope and contents. Envelopes contain all necessary information to accomplish transmission and delivery of the message content. The fields within the envelope are in a standard format.

In most cases, the IBM-supplied mail header defaults are adequate. To understand the IBM-supplied mail header defaults, see "Default SMTP rules" on page 1111. If it is necessary for you to change them, you can use the REWRITE822HEADER statement in the SMTP configuration data set to control the way SMTP performs a rewrite of the RFC 822 mail headers. Mail headers are passed from the local

system, or NJE network, to the TCP network. Mail headers passing from the TCP network to the local system or NJE network are not affected. Only the addresses under certain RFC 822 header fields can be subject to the header rewriting rules.

The header fields affected by the REWRITE822HEADER statement are:

Field Description

From The identity of the person sending the message.

Resent-From
Indicates the person that forwarded the message.

Reply-To
Provides a mechanism for indicating any mailboxes to which responses are to be sent.

Resent-Reply-To
Indicates the person to whom you should forward the reply.

Return-Path
This field is added by the mail transport service at the time of final delivery. It contains definitive information about the address and route back to the originator of the message.

Sender
The authenticated identity of the agent that sent the message. An agent can be a person, system, or process.

Resent-Sender
The authenticated identity of the agent that has resent the message.

To Contains the identity of the primary recipient of the message.

Cc Contains the identity of the secondary (informational) recipients of the message.

Bcc Contains the identity of additional recipients of the message. The contents of this field are not included in copies sent to the primary and secondary recipients of the message but are included in the author's copy.

The SMTP rules data set: You can override the default rules for header addresses by creating an SMTP rules data set. This allows you to customize the address transformations to the needs of a particular site. If you are customizing SMTP mail headers, this task is required.

The SMTP rules data set is pointed to by the //SMTPRULE DD statement in the SMTP cataloged procedure. The SMTP rules data set consists of:

Field definition
Contains the names of all header fields whose addresses are to be rewritten.

Rules definition
Contains the rewrite rules for the header fields.

Statement syntax: In creating the SMTP rules data set you must use the following syntax conventions:

- The data set statements are free-format. Tokens can be separated by an arbitrary number of spaces, and statements can span an arbitrary number of lines. However, you must end every statement with a semicolon (;).

- A character string appearing within single quotation marks ('...') is not case-sensitive. For example, 'abc' represents 'abc', 'Abc', 'ABC', and so forth. The use of character strings is illustrated in the following sections.
- A character string appearing within double quotation marks ("...") is case-sensitive. For example, "abc" only represents "abc". It does not represent "Abc", "ABC", and so forth.

  Special characters, such as @ and % are treated the same whether enclosed by single quotation marks or double quotation marks.
- Double-hyphens ("--") are used to begin a comment. The comment extends to the end of the line.

The following sections describe the components of the SMTP rules data set.
- Format of the field definition section
- Format of the rules definition section

Format of the field definition section: The field definition section is the first section in any SMTP rules data set. It defines any applicable alias fields, and it is introduced by the following heading:

```
Field Definition Section
```

This section allows similar fields to be grouped under an alias or common name. This name, or alias, is used to represent the field list. You can define an arbitrary number of aliases representing a set of field lists.

An alias name can be any alphanumeric sequence of characters that is not a predefined keyword within the SMTP rules (see the following). However, the alias name `DefaultFields` is treated specially by the SMTP configuration interpreter. If `DefaultFields` is defined, and if a rule is written that does not specify an associated field alias, the rules interpreter assumes that `DefaultFields` is the associated field alias.

The alias definition within this section is of the following form:

```
alias_name = alias_definition; optional comment
```

where *alias_name* is the name of the alias and *alias_definition* is an expression describing which fields are to be grouped under this alias. This expression can be as simple as a single field name. For example:

```
MyAlias = 'To';
```

The aliases can be a list or set of field names. The field names ***To, From, Cc***, and ***Bcc***, in the following example are part of a set of field names referenced by the alias `MyAlias`.

```
MyAlias = 'To' 'From' 'Cc' 'Bcc' ; -- first list of fields
```

You can combine field names and previously defined aliases to create a new alias. In the following example, the set of field names defined as `MyAlias` and the field names in the new alias `YourAlias` are combined to form a third set. The new alias `TheirAlias` is the union of both aliases and contains the fields of `MyAlias` and `YourAlias`.

```
MyAlias    = 'To' 'From' 'Cc' 'Bcc';
YourAlias  = 'Errors-To' 'Warnings-To';
TheirAlias =  MyAlias YourAlias;
```

In the previous example, `TheirAlias` is an alias that represents the following fields:

```
TheirAlias:  'To' 'From' 'Cc' 'Bcc' 'Errors-To' 'Warnings-To'
```

You can perform the following operations on set members of the alias to create a subset of the initial alias:
- Union operations
- Difference operations
- Intersection operations

Union and difference operations: Certain field names can be added to or omitted from a new alias of field names by using a minus sign to omit set members and an optional plus sign to include another field name. In the mathematics of sets, when you add together 2 or more sets, they form a union. When set members are omitted, the remaining set is created by the difference operation. In the following example `HerAlias` and `HisAlias` are defined. The alias `HisAlias` is created from the union of `TheirAlias`, `HerAlias`, and the omission of `Warning-To` and `Bcc` from the sets:

```
HerAlias   = 'Reply-To' 'Sender';
HisAlias   = TheirAlias - 'Warnings-To' - 'Bcc' + HerAlias;
```

In the previous example, `HisAlias` is an alias that represents the following fields:

```
HisAlias:  'To' 'From' 'Cc' 'Errors-To' 'Reply-To' 'Sender'
```

Intersection operations: A field definition can include an intersection operation. When the intersection operation is applied to two field expressions, the resulting set contains the fields common to both. In the following example, `MyAlias` and `YourAlias` are defined. The alias `OurAlias` is created from the intersection of `MyAlias` and `YourAlias`. The asterisk (*) is the intersection operator.

```
MyAlias    = 'Bcc' 'Cc' 'From' 'Reply-To';
YourAlias  = 'Resent-From' 'Cc' 'Sender' 'To' 'Bcc';
OurAlias   = MyAlias * YourAlias; -- the intersection
```

In the previous example, `OurAlias` represents the following fields:

```
OurAlias:  'Bcc' 'Cc'
```

In the following complex example `TheirAlias` is created from the intersection of `YourAlias` with the sum of `MyAlias` plus `Resent-From`:

```
TheirAlias = (MyAlias + 'Resent-From') * YourAlias;
```

In the previous example, `TheirAlias` represents the following fields:

```
TheirAlias: 'Bcc' 'Cc' 'Resent-From'
```

The parentheses within the definition of `TheirAlias` perform the same functions as in algebra. Field expressions are evaluated from left to right, but the intersection operation has greater priority than union and difference operations. If parentheses were not used in the definition of `TheirAlias`, the result would be:

```
TheirAlias: 'Bcc' 'Cc' 'From' 'Reply-To' 'Resent-From'
```

Format of the rule definition section: The rule definition section is the next section in any SMTP RULES data set. It contains the header rewriting rules that define the intended address transformations, and it is introduced by the following heading:

```
Rule Definition Section
```

The basic form of a rewrite rule is:

```
alias :before-address-pattern  => after-address-pattern;
```

where the alias name *alias* is an optional name representing the fields for which the rule is applicable. If the alias name *alias* : is omitted from this part of the rules, then `DefaultFields` is assumed.

The sequence of tokens that define how a particular type of address is to be recognized is the *before-address-pattern* portion of the rules definition. The sequence of tokens that define how the address is to appear after the address has been rewritten is the *after-address-pattern* portion of the rules definition. The following example is the rule for converting host names:

```
A '@' NJEHostName => A '@' TCPHostName; -- convert host names
```

In the previous example, `A '@' NJEHostName` is the *before-address-pattern* portion of this rule, and `A '@' TCPHostName` is the *after-address-pattern* portion. This rule specifies that the address to be rewritten has an arbitrary local name (`A`), an at sign (@), and the NJE host name (`NJEHostName`) of the current site. The rule also specifies that the rewritten address must contain the same arbitrary local name (`A`), an at sign, and the current site's TCP host name `TCPHostName`.

SMTP rules syntax conventions: Use the following syntax convention when writing SMTP rules:

- Some keywords have special meaning to the rules interpreter. For example, `NJEHostName` keyword means the NJE host name of the present system, and `TCPHostName` keyword means the TCP host name of the present system. For more information about valid keywords see "Predefined keywords within the SMTP rules" on page 1110. Some keywords, such as `TCPHostName`, have single values. Other keywords, such as `AltTCPHostName` and `AnyDomainName`, can have many possible values. To avoid ambiguity, any keyword that can have multiple values, and is used in the *after-address-pattern* of a given rule, must appear exactly once within the *before-address-pattern* of that rule. The following rule example shows a valid syntax:

  ```
  A '@' AltTCPHostName '.' AltTCPHostName =>
     A '%' TCPHostName '@' TCPHostName;
  ```

  The following two rules have incorrect syntax because the first keyword `AltTCPHostName` must be rewritten to a keyword with specific values. The `AltTCPHostName` is attempting to be rewritten to the same `AltTCPHostName` but with unknown values, which is not valid.

  ```
  A '@' AltTCPHostName '.' AltTCPHostName =>
     A '%' AltTCPHostName '@' TCPHostName;

  A '@' TCPHostName => A '@' AltTCPHostName;
  ```

  Any rule whose *before-address-pattern* includes a keyword that has a null value is ignored during the header rewriting. Thus, if there is no `AltNJEDomain` defined in the system configuration data set, no rule that includes `AltNJEDomain` in the *before-address-pattern* is considered during the header rewriting.
- Alphanumeric identifiers that are not within single or double quotation marks, and that are not predefined keywords, are considered wildcards in the rule statement. Wildcards represent an arbitrary (non-null) sequence of characters. The identifier `A`, in the previous rule example, is a wildcard. Thus, if `host` were the NJE host name for the current site, and if `tcphost` were the TCP host name for the current site, the previous rule example recognizes `abc@host` and `d@host` as candidates for address rewriting, and rewrites them as `abc@tcphost` and `d@tcphost` respectively. To avoid ambiguity, within the *before-address-pattern* of a given rule, no two wildcards are allowed in a row, and the same wildcard cannot be used more than once. The following rules have valid syntax:

```
A '@' B TCPHostName      => A '%' B '@' TCPHostName;
A '%' B '@' NJEHostName => A B '@' TCPHostName;
```

The following rules have incorrect syntax because the first rule has 2 wildcards in a row A and B. The second rule has the same wildcard A repeated:

```
A B '@' TCPHostName      => A A '%' B '@' TCPHostName;
A '%' A '@' NJEHostName => A '@' TCPHostName;
```

- A character string appearing within single or double quotation marks tells the rules interpreter where a particular string is to appear within a header address. In the previous rule example, the '@' string in the *before-address-pattern* tells the rules interpreter that an at-sign (@) must appear between the arbitrary character string and the NJE host name. The '@' string in the *after-address-pattern* tells the rules interpreter that the address must be rewritten so an at-sign appears between the arbitrary string and the TCP host name. As previously mentioned, single quotation marks denote strings that are not case-sensitive, and double quotation marks denote case-sensitive strings.
- The character sequence "=>", with no spaces between the characters, separates the *before-address-pattern* from the *after-address-pattern*.
- The order in which the rules are specified is important; the first rule encountered whose *before-address-pattern* matches the current address is the rule to dictate the address transformation. Once a matching rule has been found for an address, no other rule is considered.

In addition to the rules themselves, there is the capability for some simple logic to decide at system configuration time which rules within the data set should become active. These conditions are specified in the form of an IF-THEN-ELSE statement, as shown in the following example:

```
IF cond THEN
   statement list
ELSE
   statement list
ENDIF
```

A statement list can consist of any number of rules or nested IF statements, or both. Each IF statement, regardless of whether it is nested, must be ended by an ENDIF keyword. As with IF statements in other programming languages, the ELSE clause is optional.

There are only two conditions recognized by an IF statement:
1. IF *predefined keyword* = '*character string*' THEN ... ENDIF
2. IF *predefined keyword* CONTAINS '*character string*' THEN ... ENDIF

The conditional operators = and CONTAINS can be prefixed by the word NOT to invert the conditions.

The *predefined keyword* must be a keyword that resolves to a single value at system configuration time. The character string in the first condition can be null. A character string cannot span more than one line.

The following example shows the use of IF statements.

```
IF NJEDomain = '' THEN
   A '@' AnyNJEHostName => A '%' AnyNJEHostName '@' TCPHostName;
ELSE
   A '@' NJEHostName '.' NJEDomain    => A '@' TCPHostName;
   A '@' NJEHostName '.' AltNJEDomain => A '@' TCPHostName;
   IF NJEDomain CONTAINS '.' THEN
      A '@' AnyNJEHostName                   =>
```

```
         A '@' AnyNJEHostName '.' NJEDomain;
      A '@' AnyNJEHostName '.' NJEDomain    =>
         A '@' AnyNJEHostName '.' NJEDomain;
      A '@' AnyNJEHostName '.' AltNJEDomain =>
         A '@' AnyNJEHostName '.' NJEDomain;
   ELSE
      A '@' AnyNJEHostName                   =>
         A '%' AnyNJEHostName '.' NJEDomain '@' TCPHostName;
      A '@' AnyNJEHostName '.' NJEDomain    =>
         A '%' AnyNJEHostName '.' NJEDomain '@' TCPHostName;
      A '@' AnyNJEHostName '.' AltNJEDomain =>
         A '%' AnyNJEHostName '.' NJEDomain '@' TCPHostName;
   ENDIF
ENDIF
```

Predefined keywords within the SMTP rules: The following predefined keywords can be used to define the header rewriting rules:

AltNJEDomain
Matches the alternative domain name of the NJE network as defined by the ALTNJEDOMAIN statement in the SMTP configuration data set.

AltTCPHostName
Matches any alternative TCP host name of the system, as defined by ALTTCPHOSTNAME statements in the SMTP configuration data set.

AnyDomainName
Matches any fully qualified domain name. Any host name with a period (.) is considered to be a fully qualified domain name.

AnyNJEHostName
Matches any (unqualified) NJE host name defined in the SMTPNJE.HOSTINFO data set.

NJEDomain
Matches the domain name of the NJE network as defined by the NJEDOMAIN statement in the SMTP configuration data set.

NJEHostName
Matches the NJE host name of the system.

SecureNickAddr
Matches an address of the form *NJE_user_id@NJE_node_id*, where *NJE_user_id*, and *NJE_node_id* are defined with a nickname in the SMTP security data set.

Note: This only matches user and node IDs that are defined with nicknames.

When SecureNickAddr is specified in the *before-address-pattern* of a rule, SMTP automatically associates the keyword SecureNickName with the corresponding nickname. This allows SecureNickName to be specified in the *after-address-pattern*.

SecureNickName
Matches a nickname defined in the SMTP security data set. When SecureNickName is specified in the *before-address-pattern* of a rule, SMTP automatically associates the keyword SecureNickAddr with the corresponding *NJE_user_id@NJE_node_id*. This allows SecureNickAddr to be specified in the *after-address-pattern*.

ShortTCPHostName
Matches the first portion of the TCP host name of the system, as defined

by the HOSTNAME statement in the TCPIP.DATA data set. For example, if the TCP host name was `mvs1.acme.com`, the value of ShortTCPHostName is `mvs1`.

TCPHostName
Matches the TCP host name of the system as defined by the concatenation of the HOSTNAME and DOMAINORIGIN statements in the TCPIP.DATA data set.

TCPHostNameDomain
Matches the domain portion of the TCP host name of the system as defined by the DOMAINORIGIN statement in the TCPIP.DATA data set. For example, if the TCP host name was `mvs1.acme.com`, the value of TCPHostNameDomain is `acme.com`.

The predefined keywords can consist of any combination of uppercase and lowercase characters; the rules interpreter does not distinguish between them.

The secure keywords are only valid when SMTP is configured to be a secure gateway.

Default SMTP rules: If the //SMTPRULE DD statement is not found, SMTP uses a default set of rules. The default set used depends on whether SMTP is configured as a secure gateway.

SMTP nonsecure gateway configuration defaults: If SMTP is not configured as a secure gateway, SMTP uses the following defaults:

```
FIELD DEFINITION SECTION
DEFAULTFIELDS = 'BCC' 'CC' 'FROM' 'REPLY-TO' 'RESENT-FROM'
                'RESENT-REPLY-TO' 'RESENT-SENDER' 'RETURN-PATH'
                'SENDER' 'TO';

RULE DEFINITION SECTION

A '@' NJEHOSTNAME => A '@' TCPHOSTNAME;

IF NJEDOMAIN = '' THEN
   A '@' ANYNJEHOSTNAME => A '%' ANYNJEHOSTNAME '@' TCPHOSTNAME;
ELSE
   A '@' NJEHOSTNAME '.' NJEDOMAIN    => A '@' TCPHOSTNAME;
   A '@' NJEHOSTNAME '.' ALTNJEDOMAIN => A '@' TCPHOSTNAME;
   IF NJEDOMAIN CONTAINS '.' THEN
      A '@' ANYNJEHOSTNAME                 =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
      A '@' ANYNJEHOSTNAME '.' NJEDOMAIN    =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
      A '@' ANYNJEHOSTNAME '.' ALTNJEDOMAIN =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
    ELSE
       A '@' ANYNJEHOSTNAME                  =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
       A '@' ANYNJEHOSTNAME '.' NJEDOMAIN    =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
       A '@' ANYNJEHOSTNAME '.' ALTNJEDOMAIN =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
    ENDIF
 ENDIF

A '@' TCPHOSTNAME       => A '@' TCPHOSTNAME;
A '@' SHORTTCPHOSTNAME  => A '@' TCPHOSTNAME;
A '@' ALTTCPHOSTNAME    => A '@' TCPHOSTNAME;
A '@' ANYDOMAINNAME     => A '@' ANYDOMAINNAME;
A '@' B                 => A '@' B '.' TCPHOSTNAMEDOMAIN;
```

SMTP secure gateway configuration defaults: If SMTP is configured as a secure gateway, SMTP uses the following defaults:

```
FIELD DEFINITION SECTION
DEFAULTFIELDS = 'BCC' 'CC' 'FROM' 'REPLY-TO' 'RESENT-FROM'
                'RESENT-REPLY-TO' 'RESENT-SENDER' 'RETURN-PATH'
                'SENDER' 'TO';

RULE DEFINITION SECTION

SECURENICKADDR    => SECURENICKNAME '@' TCPHOSTNAME;
A '@' NJEHOSTNAME => A '@' TCPHOSTNAME;

IF NJEDOMAIN NOT = '' THEN
   SECURENICKADDR '.' NJEDOMAIN    => SECURENICKNAME '@' TCPHOSTNAME;
   SECURENICKADDR '.' ALTNJEDOMAIN => SECURENICKNAME '@' TCPHOSTNAME;
   A '@' NJEHOSTNAME '.' NJEDOMAIN    => A '@' TCPHOSTNAME;
   A '@' NJEHOSTNAME '.' ALTNJEDOMAIN => A '@' TCPHOSTNAME;
   IF NJEDOMAIN CONTAINS '.' THEN
      A '@' ANYNJEHOSTNAME                  =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
      A '@' ANYNJEHOSTNAME '.' NJEDOMAIN    =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
      A '@' ANYNJEHOSTNAME '.' ALTNJEDOMAIN =>
         A '@' ANYNJEHOSTNAME '.' NJEDOMAIN;
    ELSE
       A '@' ANYNJEHOSTNAME                  =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
       A '@' ANYNJEHOSTNAME '.' NJEDOMAIN    =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
       A '@' ANYNJEHOSTNAME '.' ALTNJEDOMAIN =>
          A '%' ANYNJEHOSTNAME '.' NJEDOMAIN '@' TCPHOSTNAME;
    ENDIF
 ENDIF
 A '@' TCPHOSTNAME       => A '@' TCPHOSTNAME;
 A '@' SHORTTCPHOSTTNAME => A '@' TCPHOSTNAME;
 A '@' ALTTCPHOSTNAME    => A '@' TCPHOSTNAME;
 A '@' ANYDOMAINNAME     => A '@' ANYDOMAINNAME;
 A '@' B                 => A '@' B '.' TCPHOSTNAMEDOMAIN;
```

Examples of header rewrite rules: The following examples show how the header rewriting rules affect an SMTP mail header. The example site is not a secure gateway and is configured as follows:

```
TCPHostName      = mvs1.acme.com
ShortTCPHostName = mvs1
AltTCPHostName   = seeds.acme.com
NJEHostName     = mvs1
NJEDomain       = acmenet
AltNJEDomain    = centralnet
```

Note that the above keywords are configured according to the definitions found in "Predefined keywords within the SMTP rules" on page 1110 (for example, from TCPIP.DATA). In addition, assume that the following are known to be other NJE hosts:

```
bird
iron
```

Then the following header:

```
From: abc@mvs1 (Brendan Beeper)
To: Jenny Bird <def@bird>
Cc: ghi@iron.acmenet, j@mvs1,
```

```
 k@seeds.acme.com,
 Mailing List <owner@acmenet>,
 lmno@iron.centralnet
Subject: New Ore
```

is rewritten by the default header rewriting rules as:

```
From: abc@mvs1.acme.com (Brendan Beeper)
To: Jenny Bird <def%bird.acmenet@mvs1.acme.com>
Cc: ghi%iron.acmenet@mvs1.acme.com, j@mvs1.acme.com,
 k@mvs1.acme.com,
 Mailing List <owner%acmenet@mvs.acme.com>,
 lmno%iron.acmenet@mvs1.acme.com
Subject: New Ore
```

The next example deviates from the defaults listed in "Default SMTP rules" on page 1111. On the configuration for nonsecure gateways, if you change the rule before the 2 ENDIFs to:

```
A '@' AnyNJEHostName '.' AltNJEDomain =>
   '<@' TCPHostName ':' A '@' AnyNJEHostName '.' NJEDomain '>';
```

then the last address in the Cc: field within our header is rewritten as:

```
Cc: <@mvs1.acme.com:lmno@iron.acmenet>
```

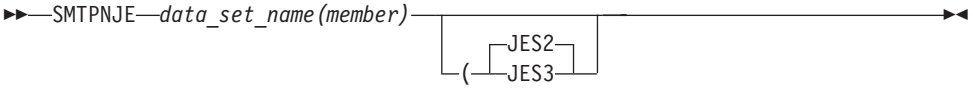
Note: Do not make the change shown in the previous example; it is intended only as a demonstration of the capabilities of the pattern-matching language.

Step 5: Set up a TCP-to-NJE mail gateway (Optional)

You can configure the SMTP server to run as a mail gateway between TCP network users and users located on a NJE network attached to the local host. This way NJE users can send mail or data sets to users on TCP hosts using SMTPNOTE. See *z/OS Communications Server: IP User's Guide and Commands* for more information about SMTPNOTE. For JES2, refer to *z/OS JES2 Initialization and Tuning Guide* and *z/OS JES2 Initialization and Tuning Reference*. For JES3, refer to *z/OS JES3 Initialization and Tuning Guide* and *z/OS JES3 Initialization and Tuning Reference*.

Follow these steps to set up your TCP-to-NJE mail gateway:

1. Add the GATEWAY statement to the SMTP configuration data set. Add other related statements, such as ALTDOMAIN, NJECLASS, NJEDOMAIN, and NJEFORMAT, as required by your configuration.
2. Issue the SMTPNJE command.



data_set_name(member)
The name of the input data set for SMTPNJE. It specifies the initialization data set of the JES2 or JES3 subsystem that is scanned for NJE nodes by SMTPNJE. The data set name is the same name as defined on ddname HASPPARM in your JES2 procedure or in the JES3IN ddname in your JES3 procedure.

member is the JES2PARM member that contains the NODE and DESTID entries for your installation.

(Required delimiter.

JES2 or JES3
Denotes whether the initialization data set being pointed to is for JES2 or JES3. If omitted, the default is JES2. For JES2, the SMTPNJE program scans for the keywords NODE and DESTID from which it extracts the information. For JES3, the keyword scanned for is NJERMT.

The SMTPNJE program creates the NJE host table data set called *user_id*.SMTPNJE.HOSTINFO. You can rename this data set and include the name of the data set on the SMTPNJE DD statement in the SMTP cataloged procedure. The //SMTPNJE DD statement is required.

3. Install the SMTP server (along with the TCPIP address space) on the gateway node. Use the GATEWAY, NJEDOMAIN, and NJEFORMAT statements in the configuration data set. Optionally, you can use either the RESTRICT or the SECURE statements to limit which users can use the gateway.

Step 6: Specify configuration statements in SMTP configuration data set

Copy the member SEZAINST(SMTPCONF) to your own SMTP configuration data set and modify it for your site using the SMTP configuration statements.

Note: If the SMTP configuration data set is a sequential data set, you cannot edit the data set while SMTP is running. If the data set is a PDS member, it can be edited while SMTP is running.

Summary of SMTP configuration statements: The SMTP configuration statements are summarized in Table 46.

Note: Refer to *z/OS Communications Server: IP Configuration Reference* for more information about these statements.

Table 46. Summary of SMTP configuration statements

Statement	*Description*
ALTNJEDOMAIN	Specifies an alternative domain name of the NJE network, if SMTP is running as a mail gateway.
ALTTCPHOSTNAME	Specifies an additional host name for the local host. Mail received for this host name is accepted and delivered locally.
ATSIGN	Enables SMTP to replace the @ symbol used in addressing strings.
BADSPOOLFILEID	Specifies the user ID on the local system where SMTP transfers unreadable spool files and looping mail.
CHECKSPOOLSIZE	Enables SMTP to check the size of the JES spool file prior to writing the data to the hlq.TEMP.NOTE file.
DBCS	Specifies that DBCS code conversion be performed on the mail.
DEBUG	Records all SMTP commands and replies.
DELETEBADSPOOLFILE	Permits SMTP to delete the spool file from the JES spool that would cause an ABENDS001 when accessed by SMTP.
DISALLOWCMD	Enables the SMTP server to discontinue support for certain specified SMTP commands.
EXITDIRECTION	Enables SMTP to call the SMTP exit provided by the customer for data coming from the JES spool.
FINISHOPEN	Specifies the SMTP wait time for connection.
GATEWAY	Specifies operation of SMTP as a gateway.
INACTIVE	Specifies the SMTP wait time before closing an inactive connection.

Table 46. Summary of SMTP configuration statements (continued)

Statement	Description
INBOUNDOPENLIMIT	Specifies a limit on the maximum number of simultaneous TCP connections over which the SMTP server will receive mail.
IPMAILERADDRESS	Specifies the IP address of an SMTP server that can resolve network addresses of unknown hosts.
IPMAILERNAME	Enables SMTP to forward non-local mail to the specified IP mailer name.
LISTENONADDRESS	Allows you to restrict which IP address is used to receive mail on a multihomed system.
LOCALCLASS	Specifies the spool data set class for local mail delivery.
LOCALFORMAT	Specifies the spool data set format for local host mail delivery.
LOG	Directs SMTP to log all SMTP traffic.
MAILER	Specifies the address of the batch SMTP server that receives mail.
MAILFILEDSPREFIX	Specifies the prefix to add to mail data sets.
MAILFILESUNIT	Specifies the unit where SMTP mail data sets reside.
MAILFILEVOLUME	Specifies the volume where newly allocated SMTP data sets reside.
MAXMAILBYTES	Specifies the maximum size of mail that is accepted over a TCP connection.
NJECLASS	Specifies the spool data set class for mail delivered on an NJE network.
NJEDOMAIN	Specifies the domain name of the NJE network if SMTP functions as a gateway.
NJEFORMAT	Specifies the spool data set format for mail delivered on the NJE network.
NJENODENAME	Specifies the node name of the local JES2 or JES3 node for mail delivered on the NJE network.
NOLOG	Turns off the logging of mail transactions.
NOSOURCEROUTE	Enables SMTP to *not* generate source routing addressing strings on certain RFC 821 SMTP commands.
OUTBOUNDOPENLIMIT	Specifies a limit on the maximum number of simultaneous TCP connections over which SMTP actively delivers mail.
PORT	Specifies an alternative port number for the SMTP server during testing.
POSTMASTER	Specifies the address (or addresses) for mail addressed to the postmaster at the local host.
RCPTREPLY452	Enables SMTP to handle reply code 452 differently for the RCPT command.
RCPTRESPONSEDELAY	Specifies how long the SMTP server delays responding to the RCPT commands.
RESOLVERRETRYINT	Specifies the number of minutes SMTP waits between attempts to resolve domain names.
RESOLVERUSAGE	Specifies whether SMTP will send queries to the domain name servers if they are configured in the TCPIP.DATA file.
RESTRICT	Specifies addresses of users who are not allowed to use SMTP mail services.
RETRYAGE	Specifies the number of days after which mail is returned as undeliverable.
RETRYINT	Specifies the number of minutes between attempts to send mail to an inactive TCP host.
REWRITE822HEADER	Prevents SMTP from rewriting RFC 822 headers with source routing.
SECURE	Specifies that SMTP operates as a secure mail gateway between TCP network sites and NJE network sites.

Table 46. Summary of SMTP configuration statements (continued)

Statement	*Description*
SMSGAUTHLIST	Specifies the addresses of users authorized to issue privileged SMTP SMSG commands.
SPOOLPOLLINTERVAL	Specifies the interval for SMTP to check the spool for incoming batch data sets.
STOPONRENF	Enables control of the SMTP server such that if a RENAME failure occurs on a data set associated with the batch connection (257), the SMTP server terminates normally. **Tip:** This statement is associated with APAR PK05655. If the associated PTF is not installed on your system, you need to apply the PTF to use this statement.
TEMPERRORRETRIES	Specifies the number of times SMTP tries to redeliver mail to a host with a temporary problem.
TIMEZONE	Sets the printable name of the local time zone.
WARNINGAGE	Specifies the number of days after which a copy of the mail is returned to the sender, indicating that the mail has so far been undeliverable and that SMTP will continue to retry delivery for RETRYAGE days.

Sample SMTP configuration data set (SMTPCONF): The sample SMTP Configuration data set can be found in SEZAINST(SMTPCONF). Refer to the *z/OS Communications Server: IP Configuration Reference* for more information on configuration data set parameters.

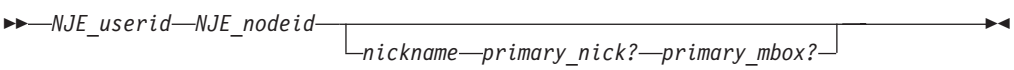
Step 7: Create an SMTP security table (Optional)

If you want to set up a secure TCP-to-NJE gateway, you need to:

- Include the SECURE statement in the SMTP configuration data set.
- Create a security data set that contains a list of NJE users who are authorized to use the gateway.
- Create a mailfiledsprefix.SECURITY.MEMO data set. The contents data set are sent to unauthorized NJE users whose mail is rejected. In this section, see "Rejected mail examples" on page 1117 for sample contents of this data set. This data set must be defined as LRECL=255 and RECFM=VB. It will be dynamically allocated by SMTP when needed.

The SMTP security data set is pointed to by //SECTABLE DD statement. The security table data set must be allocated with LRECL=255 and RECFM=VB. Records whose first nonblank character is an asterisk (*) are treated as comments and are ignored.

Use the following format when creating the list of NJE users:



NJE_userid
The NJE user ID of the authorized user.

NJE_nodeid
The NJE node ID of the authorized user.

nickname
The name by which this user is known on the TCP side of the gateway. This name must not contain any special characters, such as < > ( ) [ ] \ . , ; : @ and ".

primary_nick?
Either Y or N. If Y is specified, then mail addressed to *nickname*`@smtp-gateway` is automatically forwarded to *NJE_userid* `at` *NJE_nodeid*. Each nickname can have only one *primary_nick?* record set to Y.

primary_mbox?
Either Y or N. If Y is specified, then mail from *NJE_userid* `at` *NJE_nodeid* is converted to *nickname*`@smtp-gateway` before it is sent to the TCP recipient. Each *NJE_userid*, *NJE_nodeid* pair can only have one *primary_mbox?* record.

SMTP security data set examples: The following example shows an SMTP security data set:

```
*  Records for Jane Doe, within IBM
JDOE   ALMADEN
JDOE   AUSTIN
*  Records for John Smith, within IBM
SMITH  RALEIGH   JOHNNY  Y  N
SMITH  YORKTOWN  JOHNNY  N  Y
SMITH  DALLAS    JOHNNY  N  N
SMITH  RALEIGH   JSMITH  Y  Y
```

For example, mail sent from the following NJE network addresses through the SMTP gateway is rewritten to the following TCP network addresses. Assume the host name of the gateway is `SMTP-GATEWAY.IBM.COM`.

```
NJE Address              TCP Address
-------------------------------------------------------------------------

JDOE  at ALMADEN         JDOE%ALMADEN@SMTP-GATEWAY.IBM.COM
JDOE  at AUSTIN          JDOE%AUSTIN@SMTP-GATEWAY.IBM.COM
SMITH at RALEIGH         JSMITH@SMTP-GATEWAY.IBM.COM
SMITH at YORKTOWN        JOHNNY@SMTP-GATEWAY.IBM.COM
SMITH at DALLAS          JOHNNY%DALLAS@SMTP-GATEWAY.IBM.COM
```

Mail sent from the TCP network to the following TCP network addresses is forwarded to the following NJE network addresses. Assume the host name of the gateway is `SMTP-GATEWAY.IBM.COM`.

```
TCP Address                              NJE Address
-------------------------------------------------------------------------

JDOE%ALMADEN@SMTP-GATEWAY.IBM.COM        JDOE  at ALMADEN
JDOE%AUSTIN@SMTP-GATEWAY.IBM.COM         JDOE  at AUSTIN
JSMITH@SMTP-GATEWAY.IBM.COM              SMITH at RALEIGH
JOHNNY@SMTP-GATEWAY.IBM.COM              SMITH at RALEIGH
SMITH%DALLAS@SMTP-GATEWAY.IBM.COM        SMITH at DALLAS
```

Rejected mail examples: SMTP rejects mail to or from an unauthorized NJE user. If the mail is from the TCP network, SMTP rejects the RCPT TO command with the error:

```
550 User is not a registered gateway user
```

If the mail is from the NJE network, SMTP rejects the MAIL FROM command with the error:

```
550 User is not a registered gateway user
```

and includes the *mailfiledsprefix*.SECURITY.MEMO data set as an explanation.

The following example shows a sample *mailfiledsprefix*.SECURITY.MEMO data set:

```
The mail you sent to this SMTP gateway cannot be delivered because
you are not a registered user of this gateway.  Contact your local
administrator for instructions on how to be authorized to use this
SMTP gateway.
```

The following is an example of rejected mail that was sent to an unregistered NJE user:

```
Date: Fri, 5 Jul 91 10:55:59 EST
From: SMTP@MVS1.ACME.COM
To: DANIEL@MVS1
Subject: Undeliverable Mail

MVS1.ACME.COM unable to deliver following mail to recipient(s):
    <MATT@SMTP-GATEWAY.IBM.COM>
MVS1.ACME.COM received negative reply from host:
    SMTP-GATEWAY
550 User 'MATT@SMTP-GATEWAY' is not a registered gateway user

           ** Text of Mail follows **
Date: Fri, 5 Jul 91 10:55:56 EDT
From: <DANIEL@MVS1.ACME.COM>
To:   <MATT@SMTP-GATEWAY.IBM.COM>
Subject:  Lunch
Matt,

Do you have time to meet for lunch next week?  I want to discuss the
shipment of ACME iron birdseed.
Daniel
```

The following is an example of rejected mail that was sent from an unregistered NJE user:

```
Date: Fri, 5 Jul 91 11:35:18 EST
From: <SMTP@SMTP-GATEWAY.IBM.COM>
To: <MATT@SMTP-GATEWAY.IBM.COM>
Subject: Undeliverable Mail
Unable to deliver mail to some/all recipients.
050 MAIL FROM:<MATT@SMTP-GATEWAY.IBM.COM>
550-User 'MATT@SMTP-GATEWAY' is not a registered gateway user.
550-
550-The mail you sent to this SMTP gateway cannot be delivered because
550-you are not a registered user of this gateway.  Contact your local
550-administrator for instructions on how to be authorized to use this
550 SMTP gateway.
           ** Text of Mail follows **
HELO SMTP-GATEWAY.IBM.COM
MAIL FROM:<MATT@SMTP-GATEWAY.IBM.COM>
RCPT TO:<DANIEL@MVS1.ACME.COM>
DATA
Date: Fri, 5 Jul 91 11:34:17 EST
From: <MATT@SMTP-GATEWAY.IBM.COM>
To:   <DANIEL@MVS1.ACME.COM>
Subject:  Awaiting your message
Daniel,
When are you going to contact me about the iron birdseed and giant
electromagnet that I ordered?  I would like to meet with you soon.
Matt

 .
QUIT
```

Step 8: Enable SMTP domain name resolution

The SMTP server's RESOLVERUSAGE statement indicates if domain name resolution is to be used or not. If name resolution is not desired,

RESOLVERUSAGE NO should be specified. See "Step 9: Enable sending of non-local messages to other mail servers" on page 1120.

If the RESOLVERUSAGE statement is not specified or is specified as RESOLVERUSAGE YES, the SMTP server will resolve domain names. Resolver TCPIP.DATA statements must be configured before you can use domain name resolution for SMTP. For a description of how TCPIP.DATA statements can be specified, see "Understanding resolvers" on page 20.

For more information on the SMTP RESOLVERUSAGE statement and the TCPIP.DATA resolver statements, refer to *z/OS Communications Server: IP Configuration Reference*.

To use a domain name server, configure the TCPIP.DATA data set with the IP address of one or more name servers. If the TCPIP.DATA data set does not point to any name servers, the local site tables are used by SMTP. However, if the SMTP server is configured to use name servers, SMTP does not use the site tables.

To determine which DSN the SMTP server is using, look for message number EZA5231I in the output data set specified by the OUTPUT statement in SEZAINST(SMTPPROC).

When SMTP uses a domain name server, it asks the domain name server for the MX records for the host to which it is trying to connect. If SMTP does not find MX records for a host, it delivers mail only to the primary host listed in the A records. The MX and A records are coded in the domain name server database.

The basic idea behind MX records is to send the mail as close as possible to the final destination. The destination host may currently be inactive, for example, because it is in another time zone. SMTP needs a synchronous connection to deliver the mail, but due to the different time zones, two systems might never be active at the same time and would never be able to exchange mail.

Using MX records would allow the SMTP server to deliver the mail to an alternate host if the first one is unavailable. SMTP tries to deliver mail to the host with the lowest MX record count. If the host is not currently available, it tries the host with the next lowest count.

For example, if SMTP wants to send mail to USER@BASKET, it checks the name server for MX records and finds the following:

```
MVS20  BASKET  A
       BASKET  MX  0    MVS20
       BASKET  MX  5    MVS18
       BASKET  MX  10   VMQ
```

SMTP delivers the mail to the BASKET with the lowest count on its MX record. If `MVS20` is unable to receive the mail, SMTP then tries to deliver it to `MVS18`. If `MVS18` cannot receive the mail, it tries `VMQ`. If none of the hosts can receive the mail, SMTP stores the mail and queues it for later delivery, at which time the process repeats.

For more information about how to add MX records to your name server, consult RFC 974, "Mail Routing and the Domain System."

To receive a detailed trace on how SMTP is resolving a particular host name, you can issue the SMSG SMTP TRACE command at the console or use a SYSTCPT DD statement in the SMTP cataloged procedure. You can also add the TRACE

RESOLVER statement when configuring the TCPIP.DATA data set, but this will also trace the name resolution for all the other applications using the name server. To prevent the console log from becoming too large, only use the TRACE RESOLVER statement for debugging.

If changes to the domain name server requires you to resolve already queued mail again, use the SMSG SMTP EXPIRE command as described in the *z/OS Communications Server: IP User's Guide and Commands*. You can also query operating statistics, such as mail delivery queues of the SMTP server, by using the SMSG SMTP command. This and other administrative tasks are discussed in more detail in the *z/OS Communications Server: IP User's Guide and Commands*.

Step 9: Enable sending of non-local messages to other mail servers

Non-local mail is mail that must go through a Mail Transfer Agent (MTA) to get to another host. SMTP supports the following configuration statements to assist in forwarding non-local mail:

- IPMAILERNAME, for non-local mail destined for SMTP servers in the IP network using a hostname
- IPMAILERADDRESS, for non-local mail destined for SMTP servers in the IP network using a static IP address
- MAILER, for non-local mail destined for SMTP servers in the NJE network using the JES spool

Restriction: You cannot use the IPMAILERNAME statement, the IPMAILERADDRESS statement, or the MAILER statement with the UNKNOWN option simultaneously.

For more information regarding these configuration statements, refer to *z/OS Communications Server: IP Configuration Reference*.

The SMTP server can be configured to send all your non-local TCP/IP SMTP mail to a specified mail server, or mail relay. You might need to do this if you have installed a FIREWALL. One way to accomplish this is by using IPMAILERADDRESS, making both of the following changes to your *hlq*.SMTP.CONFIG data set:

1. Inhibit SMTP from attempting to resolve non-local hostnames by specifying the following statement in your SMTP.CONFIG data set:

   ```
   RESOLVERUSAGE NO
   ```

2. Update the SMTP.CONFIG file to redirect mail to a specific server using the IPMAILERADDRESS statement:

   ```
   IPMAILERADDRESS ip_address
   ```

   where ip_address is address of the mail server that can perform the hostname resolution.

Step 10: Design SMTP exit to inspect and filter unwanted mail (optional)

The SMTP exit facility allows an installation to better control the volume of unwanted mail (spam) that is entering the installation. SMTP makes use of the Dynamic Exit Facility (CSVDYNEX macro) provided by MVS. Refer to *z/OS MVS Programming: Authorized Assembler Services Guide*, for more information. The exit is provided by the customer to implement policies that they deem workable. Based on user-defined (and implemented) criteria, individual mail items may be rejected before they consume other resources. SMTPEXIT is provided as a programming

guide to aid in the implementation of the local policies. It can be found in SEZAINST. This exit must be REENTRANT and AMODE 31, in an authorized library. In using the SMTP exit a name token (EZBTCPIPSMTPEXIT) needs to be established in SYS1.PARMLIB(PROGxx).

If a user program is enabled, message EZA5549I is generated in the SMTP output data set when the SMTPPROC program is started. This message indicates a user exit is active.

This exit can be replaced dynamically without stopping the SMTPPROC program. The procedure for doing this follows:

1. Issue a "SMSG *smtpprocname* STOPEXIT" TSO command. The TSO user ID must be in the authorized list for SMTPPROC to issue this command. This will cause SMTP to issue the termination call to the exit and then set a flag so that the exit will not be called anymore. Processing of mail will continue as if there is no exit.
2. Remove the exit via the SETPROG EXIT operator command or by updating SYS1.PARMLIB(PROGxx) and issuing the refresh console command. Example of updating SYS1.PARMLIB follows:
   a. Include the following in SYS1.PARMLIB(PROGxx):
      ```
      EXIT DELETE EXITNAME(EZBTCPIPSMTPEXIT) MODNAME(MYEXIT) FORCE(YES)
      ```
   b. At the MVS console issue SET PROG=xx.
3. Replace with the desired new exit by adding the exit via the SETPROG EXIT operator command or by updating SYS1.PARMLIB(PROGxx). Example of updating SYS1.PARMLIB follows:
   a. In SYS1.PARMLIB(PROGxx) have this line:
      ```
      EXIT ADD EXITNAME(EZBTCPIPSMTPEXIT) MODNAME(NEWEXIT)
      ```
   b. At the MVS console issue SET PROG=xx.
4. Issue a "SMSG *smtpprocname* STARTEXIT" TSO command. This will cause SMTP to issue the initialization call to the exit. A flag is then set so the exit will be called from then on for new mail connections. Processing of new mail will continue with the exit being called. The first smtp command to be seen by a reinstated exit will be HELO. The exit will not be called in the middle of a currently processing exchange.

In designing the SMTP exit some of the following design points need to be considered. It should be noted that a remote SMTP application will be connected to the local SMTP while this exit is running. If too much time is spent in the exit, timeout situations may occur and the remote SMTP application may terminate the connection and then go into retry logic. This will seriously affect the performance of the mail system. The exit must be coded as efficiently as possible and all efforts should be taken to avoid excessive processing or waiting, e.g. I/O operations and DNS resolver calls, while within the exit. Efforts to reject mail may be more efficient if extensive scanning of the data portion of the message can be avoided. The exit may allow processing to continue or reject the entire message and does not have the ability to reject individual segments of a message. The message contents cannot be changed in any way by the exit. The exit may accept a message at any point and disable further exit calls for that message. Only commands that are currently implemented by the SMTP program will be passed to the exit program. RFC 2505 and RFC 2635 should be read and understood before undertaking such a coding effort. Multiple connections can occur simultaneously and the exit must take precautions to keep any desired state information on a connection basis. More information on SMTP commands and standards are documented in RFCs 821 and 822.

The SMTP server can be allowed to call the SMTP exit program to interrogate data coming from the JES spool as well as the inbound TCP/IP connections.

Refer to the *z/OS Communications Server: IP Configuration Reference* for more detailed information.

Configuring z/OS UNIX sendmail and popper

The following is intended to provide the administrator with specific information on how to configure sendmail on the z/OS platform. Before using this chapter, become familiar with the industry-accepted publication for sendmail, *sendmail* by O'Reilly & Associates, Inc. (ISBN 1-56592-839-3). That publication is known throughout the industry as simply the *bat book*, and this chapter consistently refers to the *bat book* for further information.

The *bat book* supports sendmail version 8.12. For new features supported in sendmail 8.12.1, this chapter refers to documents that come from sendmail resources. The most noteworthy document is *Sendmail Installation and Operation Guide*. You can find this document on the web, http://www.sendmail.org/~ca/email/doc8.12/op.html, and it is also shipped in /usr/lpp/tcpip/samples/sendmail/sendmail.ps.

Sendmail 8.12.1 has been developed with two main topics in mind, enhanced security mail filters and IPv6 support. Associated topics like Transport Layer Security (TLS), mail filters (Milter) and IPv6 are explained in the following sections.

If sendmail is to be used in a multilevel secure environment, for more details concerning sendmail configuration and setup, see "z/OS UNIX sendmail" on page 164.

Additional information about sendmail can also be found on the z/OS UNIX application Web site, http://www.s390.ibm.com/unix, as well as in documents from the sample directory that were received during the port of sendmail 8.12.1 from the http://www.sendmail.org Web site. The *Sendmail Installation and Operation Guide* document (/usr/lpp/tcpip/samples/sendmail/sendmail.ps), for instance, is the generic guide from http://www.sendmail.org, which might be helpful as a more thorough guide in a slightly different format. The README.m4 document gives more details for building a configuration file using the m4 preprocessor.

This chapter also provides information on how to configure popper on the z/OS platform. The popper function requires very little configuration. For more information on the protocol used by this UNIX application, see RFC 1939.

Overview

The simple mail architecture in which sendmail and popper fit includes a mail user agent (MUA), a mail transfer agent (MTA), and a mail delivery agent (MDA). An MUA is client software that a user invokes directly to send and receive e-mail. Examples of MUAs include Eudora, Netscape Navigator, pine and elm. An MTA is software that actually routes messages from a sender's system to the receiver's system. sendmail is an MTA. It is worth noting, however, that sendmail relies on other programs to implement non-SMTP based transport (for example, UUCP-based transport as well as local delivery to a user's mail spool file). An MDA is server software that delivers received mail to a user's MUA. Popper is an example of an MDA using the POP3 protocol.

At the sender's end of the mail delivery process, the sender's MUA transmits the message to be delivered to sendmail, as shown in Figure 110.

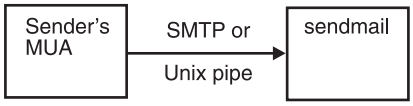


Figure 110. Sender MUA transmits the message to sendmail

This can occur in one of two ways. If the MUA is running on the local host, the message can be transmitted by executing a copy of sendmail and transmitting the message to the standard input of that process via a UNIX pipe.

Alternatively (and more commonly), a copy of sendmail will be running as a daemon, and the MUA (running on either the local host, or on a remote host) will open an SMTP connection to the sendmail daemon, transmitting the message to be delivered via that SMTP connection. In this case, sendmail is acting as an SMTP server, while the MUA is acting as an SMTP client.

In the next step, for each recipient address, sendmail transmits the message to some other SMTP server, to route the message to its final destination at the recipient's site. This is shown in Figure 111.

Figure 111. sendmail transmits the message to an intermediate SMTP server

The receiving SMTP server, in this case, might be a local hub that handles all mail at the sender's site, a remote hub handling all mail at the recipient's site, or an SMTP server at the recipient's host system.

In the next step, sendmail acts as an SMTP client, initiating an SMTP connection with some SMTP server, and then transmitting the message to be delivered to that server, via the SMTP connection.

At the receiver's end of the mail delivery process, a sendmail daemon receives the message from some SMTP client, as shown in Figure 112.

Figure 112. A sendmail daemon receives the message from an SMTP client

The sendmail daemon, acting as an SMTP server, accepts an incoming SMTP connection, and receives a message to be delivered over that SMTP connection. (This is identical to receipt of a message from an MUA, over an SMTP connection.)

Upon receiving the message, sendmail delivers it to the local recipient by appending the message to the recipient's mail spool file. To do this, sendmail requires a local mailer program, as depicted in Figure 113 on page 1124.

Figure 113. sendmail delivers the message to the local recipient

In this step, sendmail executes a specified local mailer program, such as /usr/lib/tsmail, and transmits the message to be delivered to that mailer through a UNIX pipe. The mailer program appends the message to the recipient's mail spool file. With this sendmail's role in delivery of mail is completed.

For the recipient to now read the received message, an MUA must be used. As mentioned in the previous section, depending upon the MUA, this may or may not require an additional MDA, such as popper. If the receiver's MUA has direct access to the mail spool file, the MUA may retrieve the mail directly from the spool file, as depicted in Figure 114.

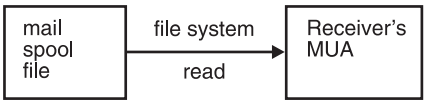


Figure 114. Receiver's MUA has direct access to the mail spool file

Alternatively (and more commonly), the MUA will establish a POP3 connection with a popper daemon, and retrieve the message over that connection. This is shown in Figure 115.

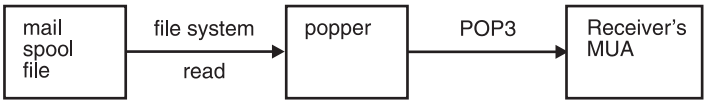


Figure 115. Receiver's MUA retrieves the message over a POP3 connection with a popper daemon

The popper daemon will also allow the receiver's MUA to manage the mail spool file, by allowing it to specify whether and which message should be deleted.

For security issues, sendmail has supported RFC 2487 (*SMTP Service Extension for Secure SMTP over TLS*) to provide private, authenticated communication over the Internet, as shown in Figure 116 on page 1125. z/OS UNIX sendmail uses System SSL instead of Open SSL.

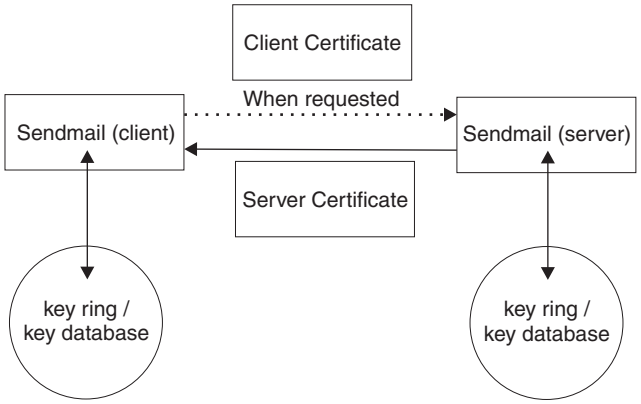


Figure 116. Using a certificate to establish a secure connection

Also, when sendmail receives mail, a mail filter provides an interface for third-party software to validate and modify messages as they pass through the mail transport system. Filters can process messages' connection information, envelope protocol elements, headers, and body contents, and modify a message's recipients, headers, and body. The MTA configuration file specifies which filters are to be applied, and in what order, allowing you to combine multiple independently-developed filters. Mail filters are separate daemons that can be run remotely or locally. How a mail filter works is shown in Figure 117.

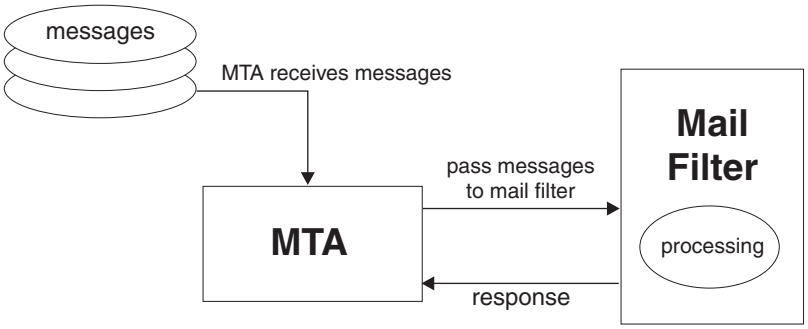


Figure 117. Mail filter processing

The sendmail samples directory

Much of the sendmail samples directory is dedicated to the automated creation of the configuration file. The /usr/lpp/tcpip/samples/sendmail/cf directory contains a sample.mc file and the subsequent sample.cf configuration file that was created by running the m4 macro preprocessor on the sample.mc file. If the /usr/lpp/tcpip/samples/sendmail directory is examined, the following directory structure can be found:

```
cd /usr/lpp/tcpip/samples/sendmail

README.m4       TRACEFLAGS      feature         mailer          sh
README.milter   TUNING          hack            milter          siteconfig
RELEASE_NOTES   cf              inetd.conf.pop  ostype
SECURITY        domain          m4              sendmail.ps
```

README.m4

Contains all the latest information regarding this latest version of sendmail from the *www.sendmail.org* site.

README.milter
This README file describes the steps needed to compile and run a filter, through reference to a sample filter that is attached at the end of this file.

RELEASE_NOTES
The sendmail release notes.

SECURITY
Gives some security-related hints on how to configure and run sendmail.

TRACEFLAGS
Describes the different trace flags used with the -d option.

TUNING
If the default configuration of sendmail does not achieve the required performance, several configuration options can be changed to increase performance. However, before those options are changed, it is necessary to understand why performance is not as good as desired. This file describes the performance implications of sendmail.

cf
Both site-dependent and site-independent descriptions of hosts. Files ending in **.mc** (Master Configuration) are the input descriptions. The output is in the corresponding **.cf** file. The general structure of these files is described below.

domain
Site-dependent subdomain descriptions. These are tied to the way your organization wants to do addressing. These descriptions are referenced using the DOMAIN *m4* macro in the **.mc** file.

feature
Definitions of specific features that some particular host in your site might want. These are referenced using the FEATURE *m4* macro. An example feature is `use_cw_file`, which tells z/OS UNIX sendmail to read an `/etc/mail/local-host-names` file on startup to find the set of local names.

hack
Local hacks, referenced using the HACK *m4* macro. This code should be used only during the transition from Berkeley.EDU names to .CS.Berkeley.EDU names.

inetd.conf.pop
The inetd changes needed to run popper.

m4
Site-independent *m4(1)* include files that have information common to all configuration files. Think of this as a "#include directory.

mailer
Definitions of mailers, referenced using the MAILER *m4* macro. The mailer types that are known in this distribution are fax, local, smtp, uucp, and usenet. For example, to include support for the UUCP-based mailers, use MAILER(uucp).

milter A directory containing a sample mail filter.

ostype Definitions describing various operating system environments (such as the location of support files). These are referenced using the OSTYPE *m4* macro.

sendmail.ps
This is a postscript file of the *Sendmail Installation and Operation Guide* provided by *www.sendmail.org* in this version of sendmail.

sh
Shell files used by the *m4* build process.

siteconfig
Local UUCP connectivity information. These normally contain lists of site information, for example:

- SITE(contessa)
- SITE(hoptoad)
- SITE(nkainc)
- SITE(well)

These are referenced using the SITECONFIG macro:

```
SITECONFIG(site.config.file, name_of_site,X)
```

where *X* is the macro or class name to use. It can be U (indicating locally connected hosts) or one of W, X, or Y for up to three remote UUCP hubs. This directory has been supplanted by the mailer table feature. Any new configurations should use that feature to do UUCP (and other) routing.

Steps for configuring z/OS UNIX sendmail

Before you begin: You need to know how to configure standard UNIX sendmail items, such as mailing lists and ~/.forward files. Only z/OS UNIX specific items are covered in this section. The *bat book* contains information on all the configurable options that can be included in the master configuration (.mc) file.

Perform the following steps to configure sendmail for z/OS UNIX:

1. Creating the configuration file

   _____

2. Creating the z/OS-specific file

   _____

3. Using sendmail databases

   _____

4. Configuring an IPv6 daemon and relay client (optional)

   _____

5. Configuring TLS support (optional)

   _____

6. Configuring Security Server (RACF or equivalent) items

   _____

7. Setting up a Milter (optional)

   _____

8. Creating the Message Submission Program (MSP) file submit.cf

   _____

9. Running sendmail as a daemon

   _____

You know you are done when you can send and, optionally receive mail. The easiest way to send mail is using the following UNIX command:

```
date | sendmail -v <user@host>
```

The -v option displays verbose error messages if any errors occur. After sending mail from another host, use the mailx UNIX command to receive mail.

Creating the configuration file

To make it easier to test sendmail, you can simply copy the /usr/lpp/tcpip/samples/sendmail/cf/sample.cf file as /etc/mail/sendmail.cf and use the copy. Later, when you are more familiar with creating your own configuration file, creating your own sendmail.cf file is recommended. However, this sample will suffice for a simple client configuration.

The basic steps to create the configuration file are:

1. Retrieving the m4 preprocessor.
2. Creating the *.mc* file.
3. Building the configuration file.

Retrieving the m4 preprocessor: Retrieve the m4 macro preprocessor from the z/OS Toys and Tools Web page at http://www.s390.ibm.com/unix/bpxa1toy.html.

The m4 macro preprocessor can be given input that will generate a z/OS UNIX sendmail configuration file. It takes as input a user-defined master configuration source file (.mc file) that can define mail delivery mechanisms using files provided in the sample directory. For more information on the .mc file, see "Creating the .mc file."

The m4 preprocessor is downloaded as m4.bin.pax.Z. To *unpax* the file, issue the following command:

```
pax -rzf m4.bin.pax.Z
```

The m4 preprocessor is created in ./bin/m4.

Creating the .mc file: The process of building a z/OS UNIX sendmail configuration file begins by creating a file of m4 statements. The suffix for this file is *.mc*.

The minimal mc file: Every **.mc** file must contain minimal information. This file defines the mail delivery mechanisms understood at this site, how to access them, how to forward e-mail to remote mail systems, and a number of tuning parameters. The following table shows which items are required and also which items are recommended. It is recommended that the starting point for these items be as shown in the sample.mc file, and an investigation of all the m4 techniques that are available to customize the .mc file for your mail server is encouraged (refer to the *bat book*).

Table 47. Required and recommended m4 items

Item	Bat book reference	Required or recommended	Description
OSTYPE()	4.2.2.1	Required	Support for your operating system
MAILER()	4.2.2.2	Required	Necessary delivery agent
DOMAIN()	4.2.2.3	Recommended	Common domain wide information
FEATURE()	4.2.2.4	Recommended	Solutions to special needs

Table 47. Required and recommended m4 items (continued)

Item	Bat book reference	Required or recommended	Description
confBIND_OPTS	N/A	Recommended	Add for IPv6 errors on name servers

Example files can be found in the `/usr/lpp/tcpip/samples/sendmail` directory. The *cf* directory contains an example of an **.mc** file. Of special interest are the files that begin with *generic*. These can serve as template statements in developing customized **.mc** files. The following is an example of a simple **.mc** file.

```
divert (-1)
divert(0) dnl
VERSIONID(`z/OS sample configuration 2002/09/20')
OSTYPE(zOS)dnl
DOMAIN(generic)dnl
MAILER(local)dnl
MAILER(smtp)dnl
```

Following is a description of these common *m4* items. For more information on these items, refer to the *bat book.*

divert

- (-1) Ignore the lines following.
- (0) Stop diverting and output immediately.

VERSIONID
Used to insert an identifier into each **.mc** and **.m4** file that will become your header.

OSTYPE()

- Support for operating system (the only ostype provided in the `/usr/lpp/tcpip/samples/sendmail/ostype` directory is zOS.m4).
- Required.

MAILER()

- Necessary delivery agent.
- Required.
- Known values include:
  - fax
  - local
  - smtp
  - uucp
  - usenet

DOMAIN()
Common domain wide information.

FEATURE()
Solution to special needs.

RELAY_DOMAIN()
Defines hosts and domain names for which the sendmail server should allow mail to be relayed. By default, the sendmail server does not allow relayed mail. If the sendmail client uses the submit.cf file and uses FEATURE(`msp'), the sendmail server should be configured to allow this relayed mail.

Building the configuration file: To build the configuration file, go to the directory containing m4/cf.m4 and issue the following command:

```
m4  ../m4/cf.m4 yourmcfile.mc > yourcffile.cf
```

where *yourmcfile* is the name of your .mc file and *yourcffile* is the name you want to give your .cf file.

The `../m4/cf.m4` specifies the master prototype configuration file `cf.m4` in the `m4` directory of the samples/sendmail directory. This is the path to the samples/sendmail directory structure from the location of your m4 executable. This can also be specified in your .mc file using *include* as follows:

```
include(`/usr/lpp/tcpip/samples/sendmail/m4/cf.m4')
```

Assuming that the m4 preprocessor is downloaded to /tmp/bin/m4 and that the master configuration (.mc) files are in /tmp/sendmail.mc and /tmp/submit.mc, the following commands will produce the configuration files:

```
$ cd /usr/lpp/tcpip/samples/sendmail/cf
$ /tmp/bin/m4 ../m4/cf.m4 /tmp/sendmail.mc > /etc/mail/sendmail.cf
$ /tmp/bin/m4 ../m4/cf.m4 /tmp/submit.mc > /etc/mail/submit.cf
```

Creating the z/OS-specific file

The purpose of the z/OS-specific file is to specify z/OS unique options. A sample file is in /usr/lpp/tcpip/samples/sendmail/cf/zOS.cf and can be copied to /etc/mail/zOS.cf with the installation information. The actual location of the file can be set by the confZOS_FILE m4 parameter. It is assumed that the administrator received the following information from the security administrator.

KeyfilePath
Directory path for the key ring files and password stash files.

ServerKeyFile
Name of the key database file or RACF key ring, used when sendmail acts as the server. If a key database is specified, it must be an existing z/OS HFS file. If a RACF key ring is specified, it must be an existing key ring and the current user ID must have READ access to the IRR.DIGTCERT.LISTRING and IRR.DIGTCERT.LIST resources in the FACILITY class.

ClientKeyFile
Name of the key database file or RACF key ring, used when sendmail acts as the client. If a key database is specified, it must be an existing z/OS HFS file. If a RACF key ring is specified, it must be an existing key ring and the current user ID must have READ access to the IRR.DIGTCERT.LISTRING and IRR.DIGTCERT.LIST resources in the FACILITY class.

ServerPWFile
Name of the file that contains the password for the key database file, used when sendmail acts as the server. It must not be given a value when a RACF key ring is specified in ServerKeyfile.

ClientPWFile
Name of the file that contains the password for the key database file, used when sendmail acts as the client. It must not be given a value when a RACF key ring is specified in ClientKeyfile.

CipherLevel
Specifies the list of SSLV3/TLSV1.0 ciphers in the order of usage preference. If it is not set, it takes on the default SSLV3 cipher specs. The

default cipher spec list when Security Level 3 FMID JCPT321 is installed is "05040A0306090201". If Security Level 3 FMID JCPT321 is not installed, the default cipher spec list is "0306090201".

GskTraceFile
Specifies the file to receive SSL Trace information, used to debug problems using the sendmail TLS interface. The GSK_TRACE_FILE environmental variable is set to the value specified. For a discussion of concerns when obtaining a System SSL trace, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*. Ensure that the file is writable by the UID that sendmail will execute under. Be aware that sensitive information might be written to this file, and use a percent sign (%) to substitute the PID into the file name and avoid multiple tasks writing to (and over) the same file. To create a readable copy of the trace information, use the System SSL gsktrace command, which takes the trace file name as input and writes readable trace output to standard output.

z/OS sendmail also supports querying for certificate revocation lists (CRLs) if an LDAP server is specified.

LdapServer
Support LDAP for X.500 certificate verification.

LdapUser
LDAP user ID to support X.500 certificate verification.

LdapPw
LDAP password to support X.500 certificate verification.

LdapPort
Port number to be used to connect to the LDAP server.

Using sendmail databases

The aliases database is the only database included in every configuration. If you want to use other databases, you need to add the database to sendmail configuration. For more information on databases, refer to /usr/lpp/tcpip/samples/sendmail/README.m4.

Configuration option: During the process of building the configuration file, the following defines and features are used to configure support for the optional databases, as well as some import files:

RELAY_DOMAIN_FILE(`path')
This option specifies the file that contains the list of hosts, domains, and addresses that mail can be relayed to.

FEATURE(`use_ct_file'[, `path'])
This feature adds trusted usernames from the files to the class variable t. If the path is not provided, sendmail 8.12 defaults to /etc/mail/trusted-users.

FEATURE(`use_cw_file'[, `path'])
This feature adds hostname aliases from the file to the class variable w. If the path is not provided, sendmail 8.12 uses /etc/mail/local-host-names as the default.

Three basic files: The following three databases are disk files used to load sendmail.cf class variables:

trusted-users file
The default path of this file is /etc/mail/trusted-users. It adds usernames,

to the list of users that are trusted to send mail under another user's name, to the class variable t in the configuration file. By default, the class variable t contains the names daemon, root, and uucp.

relay-domains file
Sendmail copies data written in this file to the class variable R. This enables relaying in the domains of the /etc/mail/relay-domains file.

local-host-names file
This file is used when systems use the sendmail server as a mailbox server to hold their mail. By using class variable w in the configuration file, the mail to the system listed in /etc/mail/local-host-names will be accepted as a local delivery.

Aliases database: Aliasing is the process of converting one recipient name into another; a generic name (such as root) into a real user name; or one name into a list of names (that is, a mailing list). Define the location of your aliases file using define(`ALIAS_FILE',`path') in your sendmail.mc file. For example:

```
define(`ALIAS_FILE',`/etc/mail/aliases')
```

For sendmail to work, aliases are required for MAILER-DAEMON and postmaster. Every aliases file must include these required aliases.

The alias for postmaster must expand to the name of a real user, based on the requirement that every site has to be able to accept mail addressed to a user named postmaster. Unless a site has real user account named postmaster, an alias is required in the aliases file. The postmaster receives mail about mail problems sent by mail-related programs and by users that are having trouble sending mail.

When mail is bounced (returned because it could not be delivered), it is sent from MAILER-DAEMON but it is shown as being the original sender who sent the mail. This alias is defined because users often inadvertently reply to the bounced mail.

Following is an example of an aliases file. Lines that begin with # are comments. Empty lines are ignored. For more information on the different forms of aliases, refer to the *bat book.*

```
# Alias for mailer daemon
MAILER-DAEMON:IBMUSER

# Following alias is required by the new mail protocol, RFC 822
postmaster:IBMUSER

# Alias to handle mail to msgs and news
nobody: /dev/null
```

Note: After the aliases file is created and before the sendmail daemon is brought up for the first time, the aliases file must be loaded by running sendmail using the *newaliases* command or with the *-bi* command-line switch.

Restriction: The section "USING LDAP FOR ALIASES, MAPS, AND CLASSES", as documented in /usr/lpp/tcpip/samples/sendmail/README.m4, is not supported on z/OS.

Configuring an IPv6 daemon and relay client (optional)

To enable IPv6 support, DaemonPortOptions in the configuration file must be set to accept IPv6 connections. For example, to listen on IPv6 interfaces, use the following m4 configuration file option to build the configuration file:

```
DAEMON_OPTIONS(`Name=MTA-v6, Family=inet6')
```

To listen on both IPv4 and IPv6 interfaces, use the following:

```
DAEMON_OPTIONS(`Name=MTA-v4, Family=inet')
DAEMON_OPTIONS(`Name=MTA-v6, Family=inet6')
```

To set a restriction for outgoing connections on a particular family, use ClientPortOptions in the configuration file. For example, the following indicates that sendmail will be a relay client for the IPv6 family.

```
CLIENT_OPTIONS(Family=inet6);
```

Note: If the following message is shown in the log file when invoking the sendmail daemon, check that your system supports IPv6. If your system has no IPv6 capability, sendmail will fail to start the daemon.

```
opendaemonsocket: daemon MTA-v6: can't create server SMTP socket"
opendaemonsocket: daemon MTA-v6: problem creating SMTP socket"
```

Configuring TLS support (optional)

Refer to *Sendmail Installation and Operation Guide* (/usr/lpp/tcpip/samples/sendmail/sendmail.ps). Note that z/OS sendmail does not use OpenSSL, but instead uses the Security Server SSL interface. Therefore, a flag confZOS_FILE is defined to indicate where this information is set. The default location is /etc/mail/zOS.cf. A sample zOS.cf file is shipped in /usr/lpp/tcpip/samples/sendmail/cf/zOS.cf. For more information on the SSL fields used in the zOS.cf file, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*. For an explanation of the fields within the zOS.cf file, see "Creating the z/OS-specific file" on page 1130.

Configuring Security Server (RACF or equivalent) items

Sendmail assumes user IDs are used when running, and these user IDs must be defined to execute sendmail correctly. The commands to define the sendmail user IDs are defined in SEZAINST(EZARACF). The commands are:

```
ADDGROUP SMMSPGRP OMVS(GID(25))
ADDGROUP SNDMGRP OMVS(GID(26))
ADDUSER MAILNULL DFLTGRP(SNDMGRP)  NOPASSWORD OMVS(UID(26) HOME('/'))
ADDUSER SENDMAIL DFLTGRP(SNDMGRP)  NOPASSWORD OMVS(UID(0) HOME('/'))
ADDUSER SMMSP    DFLTGRP(SMMSPGRP) NOPASSWORD OMVS(UID(25) HOME('/'))
RDEFINE STARTED  SENDMAIL.* STDATA(USER(SENDMAIL))
SETROPTS RACLIST(STARTED) REFRESH
```

The queue directories must have the proper read and write permission for UID 25 and 26 respectively.

In addition, there are security concerns for programs that change user ID without prompting for a password. Program control is the Security Server facility used to manage programs that change user IDs without prompting for a password. By having an installation use program control, applications not permitted to the facility are not allowed to change user IDs without prompting for a password. The commands are:

```
PERMIT BPX.DAEMON CLASS(FACILITY) ID(SENDMAIL) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

For more information on Security Server commands used to allow sendmail access to the program control facility, refer to SEZAINST(EZARACF). For complete information on the program control facility, refer to *z/OS Security Server RACF Security Administrator's Guide*.

When /usr/sbin/sendmail begins execution as a started task, or as a Mail Transmission Agent (MTA) daemon from the UNIX shell, it starts with the UID

defined for the started task or the shell. It immediately does a setuid() to the confRUN_AS_USER (sendmail uid 0) to listen on port 25 and do other setup tasks. Then, when it begins processing mail, it does a setuid() to the confDEF_USER_ID (mailnull uid 26).

In addition, when using /bin/sendmail to create mail as a Mail User Agent (MUA), the Mail Submission Agent (MSA) configuration file /etc/mail/submit.cf is used. This file must exist in a program control environment. If it does not exist, EZZ9895I is issued when sendmail does a setuid() to the confRUN_AS_USER (smmsp uid 25) to do all the mail processing. With program control, an installation must have a /etc/mail/submit.cf, which can be a copy of /usr/lpp/tcpip/samples/sendmail/cf/submit.cf. /bin/sendmail must be owned by the UID confRUN_AS_USER (smmsp uid 25), and must have the Set UID and Set GID bits set. Assuming that the default UID of 25 is used, the following commands can be used to run /bin/sendmail in a program control environment:

```
chown 25:25 /bin/sendmail
chmod 6755  /bin/sendmail
```

Rule: The chown command must be issued before the chmod command, since chown turns off the Set UID and Set GID bits. To verify this has been set, issue the following command:

```
ls -l /usr/lpp/tcpip/bin/sendmail
```

The output will be as follows:

```
-rwsr-sr-x   1 SMMSP    SMMSPGRP ...  /usr/lpp/tcpip/bin/sendmail
```

It is important to have sendmail.cf and submit.cf files to isolate tasks done by the sendmail daemon MTA and the sendmail user MUA. Also, the queue directories must have the appropriate permission bits. The default directories used by the MTA and the MSA are /var/spool/mqueue and /var/spool/clientmqueue. If these directories are not correct for your system, use the QUEUE_DIR and MSP_QUEUE_DIR flags. Ensure that the permissions for the queue directories are set up for the user IDs defined.

Setting up a Milter (optional)

To describe where filters are located, you need to add reference to the filters in the sendmail configuration file (sendmail.cf). Filters declaration in the sendmail configuration file are made in the following form:

```
X name field =value |*
```

Filters are specified with a key letter (X for external), and *name* is the name of the filter (used internally only). The *field=value* pairs, or *equates*, define attributes of the filter.

The fields and their values are:

Socket

The socket specification is in one of the following forms:

```
S = inet: port @ host | [IP address]
S = inet6: port @ host | [IP address]
S = local: path
```

The first two describe an IPv4 (inet) and IPv6 (inet6) socket listening on a certain port at a given host or IP address. The final form describes a named socket on the file system at the given path.

Flags Special flags for a filter are:

```
F=R           Reject connection if filter unavailable
F=T             Temporary fail connection if filter unavailable
```

Timeouts

To override the default timeouts used by sendmail when talking to the filters, use T = *timeout*, where *timeout* includes four fields as follows:

```
C       Timeout for connecting to a filter (if 0, use system timeout)
S       Timeout for sending information from the MTA to a filter
R       Timeout for reading reply from the filter
E       Overall timeout between sending end-of-message to filter and
        waiting for the final acknowledgment
```

The separator between each timeout field is a semicolon (;). The default timeout values, if not set in the configuration file, are as follows, where *s* is seconds and *m* is minutes:

```
T=C:5m;S:10s;R:10s;E:5m
```

A comma (,) separates equates. Following are some example filters:

```
Xfilter1, S=local:/var/run/f1.sock, F=R
Xfilter2, S=inet6:999@localhost, F=T, T=C:10m;S:1s;R:1s;E:5m
Xfilter3, S=inet:3333@localhost
```

Assuming the filters are stored in /var/spool/milter, the following commands would start the above three mail filters. Note that S= is replaced by -p, and the flags are dropped.

```
/var/spool/milter/filter1 -p local:/var/run/f1.sock &
/var/spool/milter/filter2 -p inet6:999@localhost &
/var/spool/milter/filter3 -p inet:3333@localhost &
```

Which filters are invoked, and their sequencing, is handled by the InputMailFilters option as follows, where *fnamex* are the names of the filters:

```
O InputMailFilters=fname1, fname2, fname3
```

If InputMailFilters is not defined, no filters are used. However, there is a special case. Since sendmail can run multiple daemons, you can assign different filters to different daemons. For example:

```
O DaemonPortOptions=Port=6666,Name=mmta,I=filter
.
.
.
Xfilter, S=inet:3333@localhost
```

DaemonPortOptions is used to define multiple clients and daemons, and the I field is used to assign a specific filter to the daemon. If a daemon is assigned a specific filter, it connects to that filter only. In the previous example, the daemon using port 6666 connects to the filter defined by [Xfilter, S=inet:3333@localhost]. If a daemon is not assigned a specific filter, it connects to filters defined by [InputMailFilters].

There are two suboptions for Milter:

LogLevel

Log level for input mail filter actions, defaults to LogLevel

Macros

Specifies list of macros to transmit to filters

The macros option has the following suboptions, which specify the list of macros to transmit to milters after a certain event has occurred. By default, the lists of macros are empty.

Option After this event

Connect
Session connection start

Helo HELO command

Envfrom
MAIL FROM command

Envrcpt
RCPT TO command

Following are some examples:

```
O Milter.LogLevel=12
O Milter.macros.connect=j, _, {daemon_name}
```

These filters can easily be configured in your mc file using the following commands:

```
MAIL_FILTER(`name', `equates')
INPUT_MAIL_FILTER(`name', `equates')
```

The first command, MAIL_FILTER(), simply defines a filter with the given name and equates. For example:

```
MAIL_FILTER(`archive', `S=local:/var/run/archivesock, F=R')
```

This creates the following equivalent sendmail.cf entry:

```
Xarchive, S=local:/var/run/archivesock, F=R
```

The INPUT_MAIL_FILTER() command performs the same actions as MAIL_FILTER, but also populates the m4 variable `confINPUT_MAIL_FILTERS' with the name of the filter such that the filter will actually be called by sendmail. For example:

```
INPUT_MAIL_FILTER(`archive', `S=local:/var/run/archivesock, F=R')
INPUT_MAIL_FILTER(`spamcheck', `S=inet:2525@localhost, F=T')
```

The two commands above are equivalent to the following three commands:

```
MAIL_FILTER(`archive', `S=local:/var/run/archivesock, F=R')
MAIL_FILTER(`spamcheck', `S=inet:2525@localhost, F=T')
define(`confINPUT_MAIL_FILTERS', `archive, spamcheck')
```

In general, INPUT_MAIL_FILTER() should be used, unless you need to define more filters than you want to use for `confINPUT_MAIL_FILTERS'. Note that setting `confINPUT_MAIL_FILTERS' after any INPUT_MAIL_FILTER() commands clears the list created by the prior INPUT_MAIL_FILTER() commands.

The following table shows M4 variables:

Table 48. M4 variables

M4 variable	Configuration variable	Description
confINPUT_MAIL_FILTERS	InputMailFilters	A list of filters, separated by commas, that determines which filters are contacted for incoming SMTP messages, as well as the invocation sequence. If none are set, no filters are contacted.

Table 48. M4 variables (continued)

M4 variable	Configuration variable	Description
confMILTER_LOG_LEVEL	Milter.LogLevel	Log level for input mail filter actions, defaults to LogLevel.
confMILTER_MACROS_CONNECT	Milter.macros.connect	Macros to transmit to milters when a session connection starts. Defaults to [j, _, {daemon_name}, {if_name}, {if_addr}].
confMILTER_MACROS_HELO	Milter.macros.helo	Macros to transmit to milters after HELO command. Defaults to [{tls_version}, {cipher}, {cipher_bits}, {cert_subject}, {cert_issuer}].
confMILTER_MACROS_ENVFROM	Milter.macros.envfrom	Macros to transmit to milters MAIL FROM command. Defaults to [i, {auth_type}, {auth_authen}, {auth_ssf}, {auth_author}, {mail_mailer}, {mail_host}, {mail_addr}].
confMILTER_MACROS_ENVRCPT	Milter.macros.envrcpt	Macros to transmit to milters after RCPT TO command. Defaults to [{rcpt_mailer}, {rcpt_host}, {rcpt_addr}].

For instructions on how to compile and link a mail filter program, refer to *z/OS Communications Server: IP Programmer's Guide and Reference*.

Creating the Message Submission Program (MSP) file submit.cf

Sendmail needs to run as root for several reasons. The MSP configuration file submit.cf eliminates sendmail running as root to write e-mail submitted from the command line to the queue directory.

MSP requires a set-user-ID/set-group-ID program to avoid problems with a world-writable directory. It is, however, sufficient to have a set-group-ID program and a group-writable queue directory. This can be fulfilled by a sendmail daemon that is started by root. This section explains how to use two sendmail configurations to accomplish the goal of having a sendmail binary that is not set-user-ID root, and thus is less problematic in the presence of system configuration and OS problems.

The default configuration, starting with sendmail 8.12, uses one sendmail binary that acts differently based on operation mode and supplied options. When running in a program control environment, two binaries are used, /usr/sbin/sendmail and /bin/sendmail. For information on program control, see "Configuring Security Server (RACF or equivalent) items" on page 1133.

Sendmail must be a set-group-ID (default group: smmspgrp, recommended gid: 25) program to allow for queueing mail in a group-writable directory. Two .cf files are required, sendmail.cf for the daemon and submit.cf for the submission program. For the permissions that should be used, see Table 49 on page 1140.

The SEZAINST(EZARACF) file shows sample commands to add the smmsp user and group, as follows:

```
ADDGROUP SMMSPGRP OMVS(GID(25))
ADDUSER  SENDMAIL DFLTGRP(SMMSPGRP) OMVS(UID(0) HOME('/'))
ADDUSER  SMMSP    DFLTGRP(SMMSPGRP) NOPASSWORD OMVS(UID(8000) PROGRAM('/bin/sh')
```

That is, the owner of sendmail is root, the group is smmspgrp, and the binary is set-group-ID. The client mail queue is owned by smmsp with group smmspgrp and is group writable. The client mail queue directory must be writable by smmspgrp, but it must not be accessible for others. That is, do not use world read or execute permissions. In submit.cf, the option UseMSP must be set, and QueueFileMode must be set to 0660. submit.cf is available in /usr/lpp/tcpip/samples/sendmail/cf, which has been built from /usr/lpp/tcpip/samples/sendmail/cf/submit.mc. The file can be used as is, or if you want to add more options, use /usr/lpp/tcpip/samples/sendmail/cf/submit.mc as a starting point.

Recommendation: Do not add options to submit.mc unless you are absolutely sure you need them. Options you might want to change include:

- confTIME_ZONE
- confDELIVERY_MODE is set to interactive in msp.m4 instead of the default background mode

Some features are not intended to work with the MSP. These include features that influence the delivery process (for example, mailertable, aliases), or those that are only important for an SMTP server (for example, virtusertable, DaemonPortOptions, multiple queues). Moreover, relaxing certain restrictions (RestrictQueueRun, permissions on queue directory) or adding features (for example, enabling prog/file mailer) can cause security problems.

Other things do not work well with the MSP and require tweaking or workarounds. For example, to allow for client authentication, it is not sufficient to just provide a client certificate and the corresponding key, but it is also necessary to make the key group (smmsp) readable and tell sendmail not to complain about it as follows:

```
define(`confDONT_BLAME_SENDMAIL', `GroupReadableKeyFile')
```

When FEATURE(`msp') is coded, the sendmail client will send all mail to the local mail server. If using the sendmail server as the local mail server, review the RELAY_DOMAIN() for the sendmail server. If needed, the sendmail client can be configured to send mail to a different server with this feature.

/usr/lpp/tcpip/samples/sendmail/feature/msp.m4 defines almost all settings for the MSP. Most of these should not be changed at all. Some of the features and options can be overridden if really necessary. It is a bit tricky to do this, because it depends on the actual way the option is defined in feature/msp.m4. If it is directly defined [that is, with define()], the modified value must be defined after the following line:

```
FEATURE(`msp')
```

If it is conditionally defined [that is, with ifdef()], the desired value must be defined before the FEATURE line in the .mc file. To see how the options are defined, read feature/msp.m4.

The .cf file (sendmail.cf or submit.cf) is chosen based on the operation mode. For -bm (default), -bs, and -t, it is submit.cf, if it exists. For all others, it is sendmail.cf. This selection can be changed by -Ac (to use submit.cf) or -Am (to use sendmail.cf).

The daemon must be started by root as usual, for example:

```
/usr/sbin/sendmail -L sm-mta -bd -q1h
```

Note: If you run sendmail from inetd (which, in general, is not recommended), you must specify -Am in addition to -bs.

Mail ends up in the client queue if the daemon does not accept connections or if an address is temporarily not resolvable. The latter problem can be minimized by using the following:

```
FEATURE(`nocanonify', `canonify_hosts')
define(`confDIRECT_SUBMISSION_MODIFIERS', `C')
```

However, the above might have undesired side effects, as discussed in /usr/lpp/tcpip/samples/sendmail/README.m4. In general, it is necessary to clean the queue either with a cronjob or by running a daemon as follows:

```
Cronjob: /usr/sbin/sendmail -L sm-msp-queue -Ac -q
Daemon:  /usr/sbin/sendmail -L sm-msp-queue -Ac -q30m
```

Requirement: If z/OS Security Server program control is used, the cronjob submission must be run from UID 0 and have READ access to the BPX.DAEMON facility class.

If the option UseMSP is not set, sendmail will complain during queue runs about bogus file permission. If you want a queue runner for the client queue, you probably have to change OS-specific scripts to accomplish this (check the man pages of your OS for more information). You can start this program as root, and it will change its user ID to RunAsUser (smmsp by default, recommended uid: 25). This way, smmsp does not need a valid shell.

Following is a brief summary of how the two configuration files are used:

sendmail.cf

For the MTA (mail transmission agent). The MTA is started by root as daemon as follows:

```
/usr/sbin/sendmail -L sm-mta -bd -q1h
```

SMTP connections are accepted, on ports 25 and 587 by default. It runs the main queue, /usr/spool/mqueue by default if using /usr/lpp/tcpip/samples/sendmail/cf/sample.cf.

submit.cf

For the MSP (mail submission program). The MSP is used to submit e-mails. Thus, it is invoked by programs, and maybe users. It does not run as SMTP. It uses /usr/spool/clientmqueue by default if using /usr/lpp/tcpip/samples/sendmail/cf/sample.cf, and can be started to run that queue periodically as follows:

```
/usr/sbin/sendmail -L sm-msp-queue -Ac -q30m
```

Running sendmail as a daemon

Just as sendmail can transport a mail message over a TCP/IP-based network, it can also receive mail that is sent to it over the network. To do this, it must be run in daemon mode. A daemon is a program that runs in the background independent of terminal control.

As a daemon, sendmail is run once, usually when your machine is booted. Whenever an e–mail message is sent to your machine, the sending machine talks to the sendmail daemon that is listening on your machine.

The -bd command-line switch tells sendmail to run in daemon mode. The -q1h command-line switch tells sendmail to wake up once per hour and process the queue. Command-line switches are described in *z/OS Communications Server: IP User's Guide and Commands*.

Configuration hints and tips

This section contains other required or useful information for configuring sendmail. For further information on these topics, refer to the *bat book.*

- SuperUser status is needed to start the sendmail daemon.
- The QueueDirectory option defined in the config file tells sendmail where to queue messages that are temporarily undeliverable. This directory must exist before sendmail is started.
- Sendmail is highly dependent on the Domain Name Server (DNS); it is important that the resolver be set up correctly to avoid unnecessary searching for a user. For more information on DNS, see Chapter 13, "Domain Name System (DNS)," on page 595.
- A program controlled environment is necessary for sendmail to run in daemon mode when BPX.DAEMON is enabled, since many functions of sendmail (especially daemon functions) require it to change the user ID (UID) without prompting for a password. For more information regarding security and sendmail, see *z/OS UNIX System Services Planning* as well as the *bat book.*
- The daemon must be started by root, as usual. Table 49 shows the recommended security file permissions of files that sendmail might use.

Table 49. Sendmail permission table

Path	Type	Owner	Mode	Required or configurable
/	Directory	*root*	755 dr-xr-xr-x	Required
/usr	Directory	*root*	755 dr-xr-xr-x	Required
/usr/sbin	Directory	*root*	755 dr-xr-xr-x	Required
/usr/sbin/sendmail	File	*root*	755 -r-xr-xr-x	Required
/bin/sendmail	File	*smmsp*	755 -rwsr-sr-x	Configurable[1]
/etc/mail	Directory	*root*	755 dr-xr-xr-x	Configurable
/etc/mail/sendmail.cf	File	*root*	444 -r--r--r--	Configurable
/etc/mail/submit.cf	File	*root*	444 -r--r--r--	Configurable
/var/spool/mqueue	File	*sendmail*	700 drwx------	Configurable
/var/spool/clientmqueue	File	*smmsp*	770 drwxrwx---	Configurable

1. Used only with RACF program control systems.

Rule: When sendmail is attempting to canonify a hostname, some broken nameservers will return SERVFAIL (a temporary failure) on T_AAAA (IPv6) lookups. To allow sendmail to accept this behavior, ResolverOptions in the configuration file is set to WorkAroundBrokenAAAA by default.

If a system has thousands of users defined in the Users list, the administrator might consider enabling the UNIXMAP class. This increases the speed of the security checks performed by sendmail. APAR OW30858 provides details about what is needed to enable the UNIXMAP class. For additional information on enabling the UNIXMAP class, refer to *z/OS Security Server RACF Security Administrator's Guide*.

Environment variables

Table 50 provides a list of environment variables that can be explicitly set by sendmail.

Table 50. Environment variables for sendmail

Environment variable	Description
GSK_TRACE	Specifies a bit mask enabling system SSL trace options.
GSK_TRACE_FILE	When set to the name of a file in a directory, enables the system SSL trace.
HOME	The system initializes this variable at login time to a path name of the user's home directory.
HOSTALIASES	The host aliases data set or file.
MIB_DESC	Specifies the location of the MIBDESC.DATA configuration file.

Configuring popper

POP3 resides on port 110. You can define additional ports if there is a need for additional command-line options for popper. For information on the options that might be suitable for your site, see the *z/OS Communications Server: IP User's Guide and Commands*.

z/OS UNIX popper will most likely be used by those whose local mailer requires a POP3 server. Typically their administrator will provide them with the address or name of the z/OS running the POP3 server, with instructions on where this information should be used.

The authentication method used in this implementation of popper is user ID and password.

The popper implementation can be divided into four steps:

- Update the /etc/services file.
- Update the /etc/inetd.conf file.
- Create the directory for the temporary maildrop file.
- Start inetd.

Update the /etc/services file

You need a port for the POP3 server defined in /etc/services. Add the following line to your /etc/services file. Note that the well-known port 110 is being used in this example:

```
pop3                110/tcp popper
```

Update the /etc/inetd.conf file

Since popper is invoked by INETD, add the following information to your /etc/inetd.conf file, where -d indicates to run popper in debugging mode:

```
pop3       stream tcp nowait bpxroot     /usr/sbin/popper popper -d
```

The debugging option is used to confirm proper installation, as shown in "Correct connection."

Create the directory for the temporary maildrop file

When popper is invoked through a Mail User Agent (MUA) client request, such as GET NEW MESSAGES, popper starts to read the user's mail file system (/usr/mail/username) where sendmail has stored the data, and puts this data (if there are messages) in a temporary maildrop file /usr/mail/popper/.username.pop. The contents of this file are transmitted to the remote client using the existing POP3 TCP connection.

Since the directory does not exist, create it as follows:

```
/usr/mail/popper/
chmod 777 /usr/mail/popper
```

The popper uses this directory to create and fill the maildrop file. If this directory is not specified or if permissions are incorrect, you will get an error message similar to the following:

```
EZZ7605I:Unable to open temporary maildrop '/usr/mail/popper/.username.pop'
```

Also, inside the POP3 session, you will get an error message similar to the following:

```
-ERR System error, can't open temporary file, do you own it?
```

Start inetd

After updating the inetd.conf file, start INETD. For more information on inetd, see Appendix A, "Setting up the inetd configuration file," on page 1165. Also refer to *z/OS Communications Server: IP Configuration Reference*.

Correct connection

Following is an example of a working connection to the popper server:

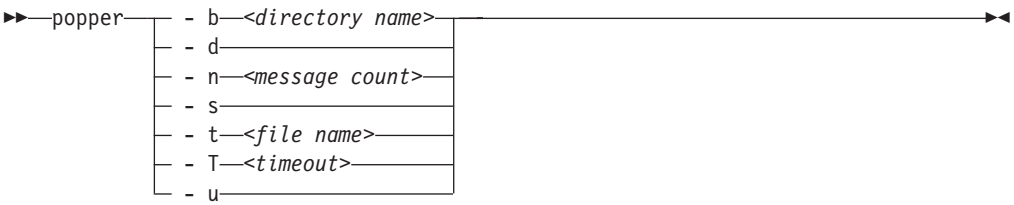
```
Debugging turned on
(v2.3)Servicing request from "hostname.domainname"at xxx.xxx.xxx.xxx
1
+OK QPOP (version 2.3)at hostname.domainname starting.
Sending line "+OK QPOP (version 2.3)at hostname.domainname starting."
Received:"USER username"
+OK Password required for username.
Sending line "+OK Password required for username."
Received:"pass xxxxxxxxx"
Creating temporary maildrop '/usr/mail/popper/.username.pop'
uid =4029,gid =1
Checking for old .username.pop file
Old .username.pop file not found,errno (0)
Msg 1 being added to list
Msg 1 uidl c313e341d7a5167b833153eb4bbfea25 at offset 0 is 700 octets long and h
Msg 2 being added to list
Msg 2 uidl c50b65c95f934fb6c22dc23573be88a1 at offset 748 is 2072 octets long an
Msg 3 being added to list
Msg 3 uidl 91c0636d1513127489f49995e6d8f1e5 at offset 2842 is 3998 octets long a
Msg 4 uidl 61021c4ea6471f83f518d8c64d8c1740 at offset 6772 is 453814 octets long
Msg 5 being added to list
```

```
Msg 5 uidl da701c81e2b2df3a60121a5ca1cdd76b at offset 454502 is 928 octets long
Msg 6 being added to list
Msg 1 uidl c313e341d7a5167b833153eb4bbfea25 at offset 0 is 700 octets long and h
Msg 2 uidl c50b65c95f934fb6c22dc23573be88a1 at offset 748 is 2072 octets long an
Msg 3 uidl 91c0636d1513127489f49995e6d8f1e5 at offset 2842 is 3998 octets long a
Msg 4 uidl 61021c4ea6471f83f518d8c64d8c1740 at offset 6772 is 453814 octets long
Msg 5 uidl da701c81e2b2df3a60121a5ca1cdd76b at offset 454502 is 928 octets long
Msg 6 uidl 0a48082f727723c8f47b306e26b49652 at offset 455469 is 691 octets long
+OK username has 6 messages (462203 octets).
Sending line "+OK username has 6 messages (462203 octets)."
Received:"STAT"
6 message(s)(462203 octets).
+OK 6 462203
Sending line "+OK 6 462203"
Received:"LIST"
+OK 6 messages (462203 octets)
Received:"UIDL"
+OK uidl command accepted.
Sending line "+OK uidl command accepted."
Sending line "1 c313e341d7a5167b833153eb4bbfea25"
Sending line "2 c50b65c95f934fb6c22dc23573be88a1"
Sending line "3 91c0636d1513127489f49995e6d8f1e5"
Sending line "4 61021c4ea6471f83f518d8c64d8c1740"
Sending line "5 da701c81e2b2df3a60121a5ca1cdd76b"
Sending line "6 0a48082f727723c8f47b306e26b49652"
Received:"QUIT"
Performing maildrop update...
Checking to see if all messages were deleted
Opening mail drop "/usr/mail/username"
Creating new maildrop "/usr/mail/username"from "/usr/mail/popper/.username.pop"
Copying message 1.
Copying message 2.
Copying message 3.
Copying message 4.
Copying message 5.
Copying message 6.
+OK Pop server at hostname.domainname signing off.
Sending line "+OK Pop server at hostname.domainname signing off."
(v2.3)Ending request from "username"at (hostname.domainname)xxx.xxx.xxx.xxx
```

Popper command—administering received mail

If the receiver's MUA does not have direct access to the mail spool file, use Popper to access the mail spool on the local host. z/OS Popper will be used when a POP3 server is needed.

Syntax:



Parameters: The following command line options can be used when invoking Popper.

-b *<directory name>*
Specifies the name of the directory in which bulletins are found. If not specified, /usr/mail/bulletins is used as the default.

-d Requests additional debugging messages be turned on.

-n *<message count>*
Specifies the number of old bulletins to be delivered to new users. If not specified, no bulletins are delivered.

-s Requests statistics logging be turned on.

-t *<file name>*
Specifies a trace file for all message logging. If not specified, messages are logged via the syslog facility.

-T *<timeout>*
Specifies the time, in seconds, before an idle POP3 connection is terminated. RFC 1939 specifies a minimum timeout of 600 seconds, but in practice such a long timeout does not work well. (When a connection gets aborted, the user is locked out of his mailbox for the timeout period.) If not specified, 120 seconds is used as the default timeout period.

-u Requests the user's mailbox be updated on abort. RFC 1939 specifies that mailboxes should not be updated (that is, no messages should be deleted) if a connection is aborted abnormally. This option forces an update to occur despite the aborted connection. If not specified, no update will occur on aborted connections.

Chapter 24. TIMED daemon

TIMED is a daemon that is used to provide the time in response to UDP requests. TIMED gives the time in seconds since midnight January 1, 1900. You can start TIMED from the z/OS shell or as a started procedure. TCP/IP must be started prior to starting TIMED.

Note: TIMED is different from the TIME daemon available as an internal daemon of INETD. INETD cannot be used to start and perform as a listener for TIMED.

Starting TIMED from z/OS shell

TIMED is installed in the /usr/lpp/tcpip/sbin/ directory.

To start the TIMED server from the command line, type the timed command.

```
timed [-l] [-p port]
```

Following are the parameters used for the timed command:

-l Logs all the incoming requests and responses to the system log. Logged information includes the IP address of the requestor.

-p port
The TIMED server usually receives requests on well-known port 37. TIMED uses UDP only. You can specify the port in which requests are to be received.

Starting TIMED as a procedure

The following sample shows how to start TIMED as a procedure.

```
//TIMED    PROC
//*
//*   5694-A01 (C) Copyright IBM Corp. 1997, 2002
//*   Licensed Materials - Property of IBM
//*   This product contains "Restricted Materials of IBM"
//*   All rights reserved.
//*   US Government Users Restricted Rights -
//*   Use, duplication or disclosure restricted by
//*   GSA ADP Schedule Contract with IBM Corp.
//*   See IBM Copyright Instructions.
//*
//* Function: Time server start procedure
//* SMP/E distribution name: EZATTMDP
//*
//TIMED    EXEC PGM=TIMED,REGION=0K,TIME=NOLIMIT,
//         PARM='POSIX(ON),ALL31(ON),TRAP(OFF)/'
//*STEPLIB DD DISP=SHR,DSN=TCP.SEZALOAD,
//*           VOL=SER=,UNIT=
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
```

```
//SYSOUT   DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
//SYSABEND DD SYSOUT=*
//         PEND
```

Chapter 25. SNTPD daemon

SNTPD is a TCP/IP daemon that is used to synchronize time between a client and a server. Simple Network Time Protocol (SNTP) is a protocol for synchronizing clocks across a WAN or LAN through a specific formatted message. An External Time Reference (ETR), named stratum 0, is chosen as the highest timer reference used for synchronization. A stratum 1 server is attached to and receives the time from the stratum 0 timer. For example, the z/OS sysplex timer could be a stratum 0 timer, and z/OS Communications Server would be a stratum 1 server. A client attached to the stratum 1 server can also be a stratum 2 server, receiving the time from the stratum 1 server, and so on. SNTP uses UDP packets for data transfer with the well-known port number 123. RFC 2030 (Mills 1996) describes SNTP. You can start SNTPD from the z/OS shell or as a started procedure. Each of these methods is described below. TCP/IP must be started prior to starting SNTPD.

Steps for starting SNTPD from the z/OS shell

Before you begin: Ensure the existence of the following files. The HFS files used by z/OS UNIX SNTPD and their locations in the HFS are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/usr/sbin/sntpd
This is a symbolic link to /usr/lpp/tcpip/sbin/sntpd, which is a sticky-bit file. The SNTPD member of SEZALOAD contains the executable code for the SNTP server.

/usr/lib/nls/msg/C/sntpdmsg.cat
The message catalog used by the z/OS UNIX SNTPD server.

When restricting low port usage, the port used by SNTPD (default value of 123) should either:

- Be reserved for the name of the SNTPD start procedure
- Use the SAF parameter on the PORT statement to restrict access to the SNTPD port

Note: There is no configuration file specifically for SNTPD.

Perform the following step to start SNTPD from the z/OS shell:

1. Type sntpd & on the command line. This will start SNTPD and send it to the background.

   Following are the optional parameters used for the sntpd command:

   -d Enables debugging. Debug messages go to syslog daemon.

   -df *pathname*
   Enable debugging. Debug messages go to the specified file location.

   -pf *HFS-pathname*
   HFS path for pid file. For example:

   ```
   -pf /var
   ```

-m *nnnnn*
Acts in multicast mode. Sends multicast updates (TTL=1) on all interfaces every *nnnnn* seconds. Listens to multicast requests and responds with unicast replies. The valid range for -m is 1 to 16284.

-b *nnnnn*
Acts in broadcast mode. This parameter sends local broadcasts on all interfaces every *nnnnn* seconds. It also specifies to listen to broadcast requests and respond with unicast replies. The valid range for -b is 1 to 16284.

-s *nn* Use *nn* as the stratum level in all replies sent by the server. The valid range for *nn* is 1 to 15. If -s is not specified or a value is specified that is not valid, the default stratum level of 1 is used. The stratum level indicates the relative accuracy of the local clock compared to the clocks of other SNTP servers in the network. The value 1 is most accurate and 15 is least accurate.

-? Display the syntax of the command usage and options.

You know that SNTPD has started when the following message appears on the MVS console:

```
EZZ9600I SNTP SERVER READY.
```

Steps for starting SNTPD as a procedure

Before you begin: Obtain a copy of this sample procedure from SEZAINST and store it in one of your PROCLIB concatenation data sets.

Perform the following step to start SNTPD as a procedure:

1. Invoke the procedure using the system operator start command. The following sample, SEZAINST(SNTPD), shows how to start SNTPD as a procedure:

```
//SNTPD    PROC
//*
//* TCP/IP FOR MVS SIMPLE NETWORK TIME PROTOCOL
//* SMP/E DISTRIBUTION NAME: EZASNPRO
//*
//*   5694-A01 (C) COPYRIGHT IBM CORP. 2002.
//*   LICENSED MATERIALS - PROPERTY OF IBM
//*   THIS PRODUCT CONTAINS "RESTRICTED MATERIALS OF IBM"
//*   ALL RIGHTS RESERVED.
//*   US GOVERNMENT USERS RESTRICTED RIGHTS -
//*   USE, DUPLICATION OR DISCLOSURE RESTRICTED BY
//*   GSA ADP SCHEDULE CONTRACT WITH IBM CORP.
//*   SEE IBM COPYRIGHT INSTRUCTIONS.
//*
//* FUNCTION: SNTP DAEMON START PROCEDURE
//*
//SNTPD    EXEC PGM=SNTPD,REGION=4096K,TIME=NOLIMIT,
//         PARM='POSIX(ON),ALL31(ON),TRAP(OFF)/ -d'
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT   DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
//SYSABEND DD SYSOUT=*
//*
```

You know that SNTPD has started when the following message appears on the console:

```
EZZ9600I SNTP SERVER READY.
```

Stack affinity

If you are running in a multiple stack environment and want SNTPD to use only a single stack, the environment variable _BPXK_SETIBMOPT_TRANSPORT can be used.

Chapter 26. Remote Execution

This chapter describes how to configure and operate both the Remote Execution server and the UNIX Remote Execution server. z/OS Communications Server supports remote execution daemons in both the UNIX and TSO environments.

To execute commands under the UNIX shell, use the RSH command. To execute commands under TSO, use the REXEC command. These requests are serviced by the UNIX daemons, orshd and orexecd, and the TSO RXSERVE daemon.

The differences between the UNIX daemons and the TSO RXSERVE daemon are as follows:

- The UNIX daemons are initiated through the INETD server and can be configured to support a port other than their well-known port.
- The TSO daemon must be active and will only service REXEC and RSH requests on their well-known ports.

Only the UNIX daemons or the TSO daemon can be active at any one time.

UNIX REXEC

The UNIX Remote Execution Protocol Daemon (REXECD) is the server for the REXEC routine. REXECD allows execution of z/OS UNIX commands with authentication based on user names and passwords.

The Remote Shell Server (RSHD) is the server for the remote shell (RSH) client. The server provides remote execution facilities with authentication based on privileged port numbers, user IDs, and passwords.

See "Configuring the z/OS UNIX Remote Execution servers" on page 1155 for more information about configuring this server.

TSO REXEC

The TSO Remote Execution server allows execution of a TSO command that has been received at a remote host. This server runs the Remote EXEcution Command Daemon (REXECD) which supports both the Remote Execution (REXEC) and Remote Shell (RSH) protocols.

The TSO Remote Execution server has affinity for a specific transport in a CINET environment. Configure and execute a unique instance of the server for each TCP/IP stack requiring TSO remote execution services.

This chapter describes how to configure and operate the Remote Execution server.

Configuring the TSO Remote Execution server

Steps to configure the TSO Remote Execution server:

1. Update AUTOLOG and PORT statements in the PROFILE.TCPIP data set.
2. Determine whether the Remote Execution client will send a Remote Execution (REXEC) command or Remote Shell (RSH) command.

3. Permit remote users to access MVS resources. (Required only if the client is not sending a password.)
4. Update the Remote Execution cataloged procedure.
5. Create a user exit routine (optional).
6. Permit access to JESSPOOL files.

Step 1: Configuring PROFILE.TCPIP for TSO Remote Execution server

If you want the Remote Execution server to start automatically when the TCPIP address space is started, include the name of the member containing the RXSERVE cataloged procedure in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  RXSERVE
ENDAUTOLOG
```

To ensure that port 512 is reserved for the Remote Execution protocol and port 514 for the Remote Shell protocol, add the name of the member containing the Remote Execution cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP:

```
PORT
  512 TCP RXSERVE
  514 TCP RXSERVE
```

Refer to *z/OS Communications Server: IP Configuration Reference* for more information about the AUTOLOG and PORT statements.

Step 2: Determine whether Remote Execution client will send REXEC or RSH commands

The Remote Execution client can send commands to the TSO Remote Execution server by the following methods:

1. Sending the Remote Execution (REXEC) command
2. Sending the Remote Shell (RSH) command with a user ID and password separated by a slash (/) character with the -l option on the RSH command
3. Sending the Remote Shell (RSH) command without a password

With methods 1 and 2, the TSO Remote Execution server executes the request and passes the password to MVS for verification. (REXEC commands require a password.) When these methods are used, skip Step 3.

With method 3, to enable an RSH client to send RSH commands to the TSO Remote Execution server without specifying a password, Step 3 is required.

Step 3: Permit remote users to access MVS resources (optional)

This step is necessary only if your installation allows users to issue remote execution commands without the requirement of specifying a password on the remote execution client.

Use the following steps to ensure that the server can correctly access necessary MVS resources. You can use z/OS Security Server (RACF) or an equivalent security program.

1. Verify that your system has been configured for allowing surrogate job submission as described in *z/OS Security Server RACF Security Administrator's Guide* (SC28-1915) or by using an equivalent security program.
2. Authorize the TSO Remote Execution server to submit jobs for the MVS user ID specified with the -l option of the RSH command. This can be done with the RACF facility as described in *z/OS Security Server RACF General User's Guide* (SC28-1917), or by using an equivalent security program.
3. Define an *mvs_userid*.RHOSTS.DATA data set and authorize the TSO Remote Execution server userid permission to read this data set. This can be done with the RACF facility as described in *z/OS Security Server RACF General User's Guide* (SC28-1917), or by using an equivalent security program.

   Note: This is the userid used to start the RXSERVE address space.

   This data set identifies the Remote Execution clients that can execute MVS commands remotely by sending an RSH command.

   When a Remote Execution client sends an RSH request to the TSO Remote Execution server, the request includes the local user ID of the client user (*local_userID*) and, if the client user specified the -l option of the RSH command, the request also contains the user ID to use on the remote host (*mvs_userid*). If the client does not specify the -l option, the user ID to be used on the remote host is assumed to be the same as the *local_userID*.

   When the TSO Remote Execution server receives an RSH command without a password, the server looks for a data set called *mvs_userid*.RHOSTS.DATA. The *mvs_userid*.RHOST.DATA data set contains one or more entries. Each entry consists of two parts, a fully qualified name of the client user's host and a *local_userID* associated with that host. The *local_userID* is case sensitive. If the data set exists, the server reads it and looks for an entry with a host name that matches the client user's host. If the user ID specified on this entry in the RHOSTS.DATA data set matches the *local_userID* passed on the RSH command, the RSH command continues processing. If the entry does not exist, the server responds to the client with message `EZA4386E Permission denied`.

   In the following example of an RHOSTS.DATA data set, the MVS client user `mvsuser` is allowed to issue the RSH command without a password from host `rs60007` with a local AIX user ID of `mvsuser`.

   Example of mvsuser.RHOSTS.DATA data set:

   ```
   rs60007.itso.ral.ibm.com mvsuser
   ```

4. Users may be authenticated using Kerberos or GSS. If the username in the Kerberos or GSS credentials matches the local user ID (*local_userID*) of the client supplied by the RSH client, then no password is required.

Step 4: Update the TSO Remote Execution cataloged procedure

Update the TSO Remote Execution cataloged procedure by copying the sample provided in SEZAINST(RXPROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Specify the TSO Remote Execution server parameters and modify the JCL as required for your installation.

Tip: You can update the TSO Remote Execution server operating parameters during execution with the MODIFY command. All but MAXCONN, IPV6, and SECLABEL can be changed.

Step 5: Create a user exit routine (optional)

Optionally, you can provide a user exit routine. This routine can be used to alter the JOB and EXEC statement parameters to meet installation-specific requirements such as which system should process the job and/or environment unique accounting information prior to submission of the TSO batch job.

The user exit should have the AMODE(31) and RMODE(24) attributes to provide addressability to the input parameters.

On entry to the user exit, register 1 points to the following parameter list:

Offset	Description
0	A pointer to a mixed AF_INET or AF_INET6 address structure **Rule:** The address family must be examined to determine which type of address structure is being used.
4	A pointer to JOB statement parameters
8	A pointer to EXEC statement parameters
12	A pointer to an optional JES control statement

The INET address consists of the following fields:

Offset	Description
0	2 bytes (AF_INET or AF_INET6)
2	2 bytes (server port)
4	4 or 16 bytes (client AF_INET or AF_INET6 address) If the server is IPv6 enabled and an IPv4 client connects, the IP address is the 4-byte IPv4 address and not the IPv4-mapped IPv6 address.

The JOB statement parameters can be up to 1024 characters in length and are ended by X'00'. You can modify the parameters with the exit routine. Upon entry, the parameters are set to:

- *user_ID*
- USER=*user_id*
- PASSWORD=*password*
- MSGCLASS=*msgclass*
- SECLABEL=*seclabel*

msgclass is as specified in the Remote Execution cataloged procedure. *user_id* and *password* are as received from the requesting client. *seclabel* is the security label used on the job submission.

For RSH commands without passwords, note that the PASSWORD= parameter is not present. The *userid* in the first positional parameter can be processed by an installation-written JES exit.

The EXEC statement parameters can be up to 256 bytes in length and are ended by X'00'. These parameters can be modified by the exit routine. On entry, it contains the EXEC statement for the procedure specified in the TSOPROC parameter of the Remote Execution server or the default IKJACCNT procedure if TSOPROC is not specified.

The JES control statement parameter can be up to 256 bytes in length and is ended by X'00'. Upon entry, the parameter field is set to X'00'. Any JES control statement added by the user exit will be put between the JOB and the EXEC statement.

The modified JOB and EXEC statements are submitted as a TSO batch job.

The user exit is shipped as a sample in the RXUEXIT member of the SEZAINST data set. Refer to the REXEC chapter in *z/OS Communications Server: IP Configuration Reference* for more information about this sample.

Step 6: Permit access to JESSPOOL files

If the SAF resource class JESSPOOL is defined, ALTER access is required to access output files. Alternatively, use a jobname of *prefix**, where *prefix* is defined in RXSERVE.

Configuring the z/OS UNIX Remote Execution servers

Installation information

This section describes the HFS files used by z/OS UNIX REXECD and RSHD.

HFS files for z/OS UNIX REXECD

Note: The userid associated with the daemon in /etc/inetd.conf requires superuser authority. Refer to *z/OS UNIX System Services Planning* for a description of the kinds of authority defined for daemons.

The HFS files used by z/OS UNIX REXECD and their locations in the HFS are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/orexecd
The server.

If BPX.DAEMON is specified, then the sticky bit must be set on, and /usr/sbin/orexecd, and orexecd can reside in an authorized MVS data set.

/usr/lib/nls/msg/C/rexdmsg.cat
The message catalog used by the z/OS UNIX REXECD server.

Note: This is not an actual member at this location, but it is a symbolic link to the part in /usr/lpp/tcpip/nls/msg/C/*.

Where the server looks for the message catalog (rexdmsg.cat) depends on the value of NLSPATH and LANG environment variables. If you want to store the msg.cats elsewhere, you need to change the NLSPATH or the LANG environment variables. If rexdmsg.cat does not exist, the software will default to the messages hard-coded within the software. These messages duplicate the English message catalog that is shipped with the product.

HFS files for z/OS UNIX RSHD

The HFS files used by z/OS UNIX RSHD and their locations in the HFS are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/orshd
The server.

If BPX.DAEMON is specified, the sticky bit must be set on, and /usr/sbin/orshd, and orshd can reside in an authorized MVS data set.

/usr/sbin/ruserok
An optional user exit that will authenticate users logging into the z/OS UNIX RSHD server with a null password. See "Setting up the z/OS UNIX RSHD installation exit" below for more information.

Note: This exit is required to allow support for null passwords with RSH.

/usr/lib/nls/msg/C/rshdmsg.cat
The message catalog associated with the z/OS UNIX RSHD client is stored here. If this file does not exist, the software will default to the messages hard-coded within the software. These messages duplicate the English message catalog that is shipped with the product.

Note: The message catalog is not actually stored here. This is a symbolic link, and the actual member is in /usr/lpp/tcpip/nls/msg/C/*.

Setting up the z/OS UNIX RSHD installation exit

When the -r option is enabled, if there is no password specified on the RSH command from the client, z/OS UNIX RSHD will drive the installation exit. When the installation exit is driven, RSHD looks for a program in /usr/sbin named ruserok. This is the only name that it will look for. If /usr/sbin/ruserok is not found, the request will fail.

When the z/OS UNIX RSHD server invokes /user/sbin/ruserok, it will pass parameters in the following order:

1. Host name or the host IP address
2. Local user's UID
3. Remote userid
4. Local userid

If z/OS UNIX RSHD receives a return code of zero from the installation exit, z/OS UNIX RSHD continues. Any nonzero return code from the installation exit will cause RSHD to issue message EZYRS25E to the client and terminate all connections. The following code fragment can be used as an example to begin building a working ruserok installation exit:

```
int main(argc, argv)
     int argc;
     char *argv[];
  char *rhost1;    /* "hostname" or "hostname.domain" of client
                      obtained by caller:
                      gethostbyaddr(getpeername()) or the host
                      ip address used by the gethostbyaddr if
                      it failed to return a "hostname" */
```

```
int locuid;      /* uid of the user name on local system */
char *cliuname;  /* user name on client's system */
char *servuname; /* user name on this (server's) system */
int rc = 4;

rhost1 = argv[1];
locuid = atoi(argv[2]);
cliuname = argv[3];
servuname = argv[4];
.
<authenticate user and set rc=0 if valid>
.
return(rc);
```

Configuring TSO and z/OS UNIX Remote Execution servers to use the same port

Since the remote execution servers are generic servers, they attempt to bind to INADDR_ANY when they are started. This allows them to listen on all defined IP addresses. However, this prevents both the TSO and z/OS UNIX Remote Execution servers from listening on the same port, and one of the servers would have to use a nonstandard port. Using the BIND parameter on the PORT reservation statement in the TCPIP profile data set allows both the TSO and z/OS UNIX Remote Execution servers to bind to the same ports using different IP addresses. The following steps illustrate how this can be done. For more information on the PORT reservation statement, see *z/OS Communications Server: IP Configuration Reference*.

1. Define a VIPA address to the TCPIP profile data set. This example shows a static VIPA address, but either a static or dynamic VIPA can be used.

   ```
   DEVICE VIPAD1 VIRTUAL 0
   LINK   VIPA1  VIRTUAL 0 VIPAD1

   HOME
   134.134.134.36       VIPA1
   ```

2. Add PORT statements to the TCPIP profile for both the TSO and z/OS UNIX Remote Execution servers. One of the servers will bind to the VIPA address. The other can bind to INADDR_ANY by not specifying the BIND parameter. In this example, the z/OS UNIX Remote Execution servers will bind to the VIPA address. Also update the /etc/services file so that exec uses 512 and shell uses 514.

   ```
   512 TCP OMVS BIND 134.134.134.36 ; z/OS Unix REXECD
   514 TCP OMVS BIND 134.134.134.36 ; z/OS Unix RSHD
   512 TCP RXSERVE                  ; TSO REXECD
   514 TCP RXSERVE                  ; TSO RSHD
   ```

   It is important that the server with the BIND parameter is listed before the one without the BIND parameter. This setup directs all requests to ports 512 or 514 with a destination IP address of 134.134.134.36 to the z/OS UNIX Remote Execution servers. Requests to ports 512 or 514 with a local destination IP address that is not 134.134.134.36 are directed to the TSO Remote Execution server.

To verify this setup:

1. Start the stack with the TCPIP profile changes described above and start RXSERVE and INETD.

   Note: INETD listens for the REXEC and RSH servers under z/OS UNIX.

2. Issue NETSTAT and it should show that both the REXEC servers are listening on port 512 and both RSH servers are listening on port 514. INETD, which listens for the z/OS UNIX Remote Execution servers, only listens on the VIPA address.

```
MVS TCP/IP NETSTAT CS V1R2       TCPIP NAME: TCPCS           21:34:41
User Id  Conn     Local Socket           Foreign Socket         State
-------  ----     ------------           --------------         -----
INETDCS1 0000000D 134.134.134.36..514    0.0.0.0..0             Listen
INETDCS1 0000000E 134.134.134.36..512    0.0.0.0..0             Listen
RXSERVE  00000019 0.0.0.0..514           0.0.0.0..0             Listen
RXSERVE  00000018 0.0.0.0..512           0.0.0.0..0             Listen
```

Chapter 27. Miscellaneous (MISC) server

The Miscellaneous (MISC) server is a server that can be used to test and debug applications.

The MISC server supports the 3 protocols described in RFCs 862, 863, and 864:
- Discard
- Echo
- Character Generator

Discard protocol

The MISC server simply throws away any data it receives. A TCP-based server listens for TCP connections on TCP port 9. If a connection is established, the data is discarded and no response is sent. A UDP-based server listens for UDP datagrams on UDP port 9. When a datagram is received, it is discarded and no response is sent.

Echo protocol

The MISC server returns to the originating application any data that it receives. A TCP-based server listens for TCP connections on TCP port 7. Once a connection is established, any data that is received is sent back to the originating application. A UDP-based server listens for UDP datagrams on UDP port 7. When a datagram is received, the data it contained is sent back as an answering datagram.

Character generator protocol

The MISC server sends a repetitive stream of character data without regard to its content. A TCP-based server listens for TCP connections on TCP port 19. When a connection is established, a stream of data is sent to the connecting application. Any data that is received is thrown away. A UDP-based server listens for UDP datagrams on port 19. When a datagram is received, an answering datagram is sent that contains a random number (between 0 and 512) of characters. The data in the received datagram is ignored.

The data that is generated follows an ordered sequence. It repeats a pattern of 94 printable ASCII characters in a ring, so that character number 0 follows character number 94.

Following is an example of the repeated pattern.

```
!"#$%&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefgh
"#$%&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghi
#$%&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghij
$%&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijk
%&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijkl
&'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklm
'()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmn
()*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmno
)*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnop
*└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopq
└,-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqr,-
./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrs
-./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrst
./0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstu
/0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuv
0123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvw
123456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwx
23456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxy
3456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz
456789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{
56789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|
6789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}
789:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~
89:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~
9:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~ !
:;◄=►?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~ !"
```

Configuring the MISC server

1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the MISC server cataloged procedure (MISCSERV).

Step 1: Configuring PROFILE.TCPIP for the MISC server

To allow the MISC server to start automatically when TCPIP is initialized, include the member name of the MISC server cataloged procedure in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP.

```
AUTOLOG
  MISCSERV
ENDAUTOLOG
```

The AUTOLOG entry in *hlq*.PROFILE.TCPIP is optional. You can choose to start the MISC server manually, when it is needed, using the START command:

```
START MISCSERV
```

The MISC server requires ports 7, 9, and 19 for both TCP and UDP. To ensure that these ports are reserved for the MISC server, verify that they are assigned to the member containing the MISC server cataloged procedure in the PORT statement in PROFILE.TCPIP.

```
PORT
    7 UDP MISCSERV
    7 TCP MISCSERV
    9 UDP MISCSERV
    9 TCP MISCSERV
   19 UDP MISCSERV
   19 TCP MISCSERV
```

For more information on these statements, see the *z/OS Communications Server: IP Configuration Reference*.

Step 2: Updating the MISC server cataloged procedure (MISCSERV)

Update the MISC server cataloged procedure by copying the sample in SEZAINST(MISCSERV) to your system or recognized PROCLIB and modifying the parameters and data set names to suit your local conditions.

MISC server cataloged procedure (MISCSERV)

```
//MISCSERV PROC MODULE=MISCSRV,PARMS=''
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: SEZAINST(MISCSERV)
//*
//*        Licensed Materials - Program Property of IBM.
//*        "Restricted Materials of IBM"
//*        5694-A01 (C) COPYRIGHT IBM CORP. 1994, 2003
//*        Status = CSV1R5
//*        Distribution library SEZAINST(MISCSERV)
//*
//MISCSERV EXEC PGM=&MODULE,
//           REGION=4096K,TIME=1440,
//           PARM='&PARMS'
//*
//*        The C runtime libraries should be in the system's link list
//*        or add them to the STEPLIB definition here. If you add
//*        them to STEPLIB, they must be APF authorized.  Change
//*        the name as appropriate for your installation.
//*
//STEPLIB  DD DISP=SHR,
//         DSN=TCPIP.SEZATCP
//SYSPRINT DD SYSOUT=*
//SYSIN    DD DUMMY
//SYSMDUMP DD SYSOUT=*
//*
//*        MSMISCSR identifies an optional data set for NLS support.
//*        It specifies the MISC server message repository.
//*
//*MSMISCSR DD DISP=SHR,
//*         DSN=TCPIP.SEZAINST(MSMISCSR)
//*
//*       SYSTCPD explicitly identifies which data set is to be
//*       used to obtain the parameters defined by TCPIP.DATA
//*       when no GLOBALTCPIPDATA statement is configured.
//*       See the IP Configuration Guide for information on
//*       the TCPIP.DATA search order.
//*       The data set can be any sequential data set or a member of
//*       a partitioned data set (PDS).
//*
//SYSTCPD  DD DISP=SHR,
//         DSN=TCPIP.SEZAINST(TCPDATA)
```

Figure 118. MISC server cataloged procedure (MISCSERV)

Specifying the MISC server parameters

The MISC server generates periodic messages whenever a client sends data to ports 7, 9, or 19. If this server runs continually for a long period of time, considerable amounts of spool space can be consumed. Therefore, the MISC server has all tracing turned off by default.

You can enable the trace options for any of the three MISC server protocols using the PARMS= parameter on the PROC statement of the cataloged procedure. These options will be in effect when the server starts.

TRACE
Turns on tracing for any of the specified protocols and must be followed by one or more of these three keywords:

ECho Specifies tracing for the echo protocol on port 7.

DIscard
Specifies tracing for the discard protocol on port 9.

CHargen
Specifies tracing for the character generator protocol on port 19.

DEbug
Specifies tracing for problem determination .

For example, the following statement turns tracing on for the echo and discard protocols.

```
//MISCSERV PROC MODULE=MISCSRV,PARMS='TRACE ECHO DISCARD'
```

Part 3. Appendixes

Appendix A. Setting up the inetd configuration file

inetd is a generic listener program used by such servers as z/OS UNIX telnet server and z/OS UNIX rexec server. Other servers such as z/OS UNIX ftp server have their own listener program and do not use inetd.

inetd.conf is an example of the user's configuration file. It is stored in the /etc directory. Ensure that the inetd services required on your system are enabled using configuration statements like those in the following example:

```
#=====================================================================
# service | socket | protocol | wait/ | user | server  | server program
# name    | type   |          | nowait|      | program |   arguments
#=====================================================================
#
shell    stream    tcp        nowait  OMVSKERN /usr/sbin/orshd orshd -l
exec     stream    tcp        nowait  OMVSKERN /usr/sbin/orexecd orexecd -LV
otelnet  stream    tcp        nowait  OMVSKERN /usr/sbin/otelnetd otelnetd  -LV
| # Add the following line to enable Kerberos for orshd
| kshell   stream    tcp        nowait  OMVSKERN /usr/sbin/orshd orshd -l -k KRB5
```

Figure 119. Adding applications to /etc/inetd.conf

If the rshd, rexecd, or otelnetd service is to support IPv6 clients, then *tcp6* should be specified instead of *tcp*. Kerberos is not supported for IPv6-enabled services, such as z/OS UNIX Telnet, z/OS UNIX rsh, and z/OS UNIX rexec.

For IPv4 connection partners, the terminal ID passed from INETD to RACF (or an equivalent security program) is an 8-byte hexadecimal character string containing an IPv4 address. For example, the IP address 163.97.227.17 is translated to X'A361E311'. RACF interprets this as a terminal logon address and rejects it if it is not previously defined.

For IPv6 connection partners, only IPv4-mapped IPv6 addresses are handled in this way. The IPv4 address portion of the IPv6 address is placed in the terminal ID for RACF validation. No other IPv6 address format is supported through terminal ID RACF validation.

To establish a relationship between the servers defined in the /etc/inetd.conf file and specific port numbers in the z/OS UNIX environment, insure that statements have been added to ETC.SERVICES for each of these servers. See the sample ETC.SERVICES installed in the /usr/lpp/tcpip/samples/services directory for how to specify ETC.SERVICES statements for these servers.

The traces for both the z/OS UNIX rexec and rsh servers are enabled through options in the inetd configuration file (/etc/inetd.conf):

```
#====================================================================
# service | socket | protocol | wait/ | user | server  | server program
# name    | type   |          | nowait|      | program |   arguments
#====================================================================
#
shell      stream    tcp       nowait OMVSKERN /usr/sbin/orshd orshd -d   1
exec       stream    tcp       nowait OMVSKERN /usr/sbin/orexecd orexecd -d   2
```

Figure 120. Setting traces in /etc/inetd.conf

The traces are turned on for both servers by passing a -d argument to the server programs. **1** is the RSHD server and **2** is the REXECD server. All commands executed after the debug flags have been turned on in the inetd configuration file and the inetd server has reread the file will produce trace output.

The trace is written in formatted form to the syslogd facility name daemon with a priority of debug. The trace data can be routed to a file in your Hierarchical File System by specifying the following definition in your syslogd configuration file (/etc/syslogd.conf):

```
  #
  # All ftp, rexecd, rshd
  # debug messages (and above
  # priority messages) go
# to server.debug.a
#
daemon.debug              /tmp/syslogd/server.debug.a
```

In this example, the trace data is written to /tmp/syslogd/server.debug.a in your Hierarchical File System. For more information on syslogd, refer to "Logging of system messages" on page 56.

For more information about inetd, refer to *z/OS UNIX System Services Planning* or *z/OS UNIX System Services Command Reference*.

Appendix B. TLS/SSL security

This appendix is a TLS/SSL reference for the z/OS TN3270 Telnet server, the FTP server and the Digital Certificate Access Server (DCAS). The gskkyman utility and RACF are used as examples for certificate and key ring creation and management. References to RACF apply to any other SAF-compliant security products which contain the required support. Host On Demand V4 running on NT is used as a sample Telnet client.

An overview of Secure Socket Layer (SSL) is given first, followed by the detailed steps needed to perform authentication and encryption at the following levels:

- Using gskkyman
  - Server Authentication only
  - Client Authentication Level 1
- Using RACF
  - Server Authentication only
  - Client Authentication Level 1
  - Client Authentication Level 2
  - Client Authentication Level 3

Additional information about the concepts of cryptography and SSL can be found at the following Web sites:

- http://home.netscape.com/eng/ssl3/
- http://www.verisign.com/repository/crptintr.html

Secure Socket Layer overview

SSL provides data privacy and integrity as well as server and client authentication based upon a Public Key Infrastructure (PKI) method. PKI requires that the server organization generate a public key/private key pair that can be used during negotiations. PKI requires that data encrypted with the public key be decrypted by only the private key and that data encrypted with the private key be decrypted by only the public key. This is considered an asymmetric encryption method because different keys are used at each end of the secure connection. The Server sends its public key to the client when the client requests a connection.

The client and server encrypt SSL parameter negotiations using the PKI method of encryption. One of the most important items negotiated is the encryption algorithm to be used during data transmission. The algorithm chosen will be one that uses the same key at each end of the secure connection. This is known as a symmetric encryption method and is about 1000 times faster than the asymmetric PKI method used during SSL parameter negotiation. The encryption key used by the symmetric encryption method is created and exchanged during SSL negotiation protected by the PKI encryption method.

Some client-server connections support negotiations to determine if the client wants or supports SSL prior to beginning the SSL handshake. Most servers and clients can be configured to immediately start the SSL handshake process or to negotiate whether or not to perform the SSL handshake. Refer to the security

information in the appropriate server or client chapters for information on whether negotiated TLS/SSL is supported and how it is implemented.

The SSL protocol begins with the handshake. During the handshake:

- Server authentication is done by the client.
- Optional client authentication is done by the server.
- An encryption algorithm and single encryption key are chosen to encrypt and decrypt session data between the client and server.

Server authentication

When using SSL to secure communications, the SSL authentication mechanism known as Server Authentication is used. This is the minimum amount of security provided by SSL and allows the client to validate that the Server is what it says it is.

To ensure that someone has not stolen the server's private and public keys and is pretending to be the server, the server sends additional information with the public key so the client can confirm the identity of the server. The complete package of information sent to the client is called a digital certificate which conforms to the X.509 standard.

This X.509 digital certificate includes, among other things, the Distinguished Name (DN) of the Server organization, the public key created by the server organization, the Distinguished Name of the organization issuing the certificate, and the issuer's signature. The organization issuing the certificate may be a well-known Certificate Authority (CA) or you may issue (create) your own certificate, called a self-signed certificate.

To create a signature, the certificate issuer first generates a message digest from the owner's DN, the owner's public key, and the issuer's DN. The message digest is the result of hashing this information down to a small size (usually 128 or 160 bits). The message digest result is unique for that information based on the hashing algorithm used. The message digest is encrypted with the issuer's private key creating the issuer's signature.

When the client receives the server certificate, the client must have the public key of the certificate signer. The public key is used to decrypt the message digest. The server certificate also contains the hashing algorithm used to create the message digest. The client uses the same algorithm to create another message digest using the Distinguished Names and public key information in the received server certificate. If this new message digest exactly matches the decrypted message digest (issuer's signature) created by the certificate issuer, the client can be assured that the certificate has not been altered. This method of authentication is dependent on the security of the private key that is used by the certificate issuer.

To conduct commercial business on the Internet, you should obtain a server certificate signed by a well-known Certificate Authority. Server certificates issued by a well-known CA gives the client high assurance that the server is authentic. Most client key rings have been primed with several well-known CA's certificates. That enables the client to authenticate a Server certificate signed by a well-known CA without having to first obtain the issuer's certificate which includes the public key. For relatively small, private networks within your own enterprise you can create your own self-signed server certificate. The only difference between a CA issued certificate and a self-signed certificate is the issuer's Distinguished Name and who's private key was used to encrypt the message digest. The client needs to

use the correct public key to decrypt the message digest. The CA certificate containing the CA's public key is probably already in the client's key ring and it can be used to decrypt the CA signature (message digest). The self-signed certificate containing the organization's public key needs to be added to the client's list of signer certificates so the client can decrypt the signature (message digest) created when the self-signed certificate was created. Some client products allow the client to add the server certificate to its list of signer certificates when the server certificate is received during SSL negotiation. If the client is confident the certificate really came from the correct server, this is an easy way to add the certificate rather than getting a copy and adding it manually.

For server authentication to work, the server must have a private key and associated server certificate in the server key database file. The gskkyman utility or RACF Common Keyring support can be used to manage the keys and certificates needed for SSL support. If the gskkyman utility was used to create the key ring, a password stash file is also required.

SSL requires a server certificate as part of its server authentication process. The server certificate and the Certificate Authority certificates are stored in a key ring (also referred to as a key database). The server's key ring can be created using the gskkyman utility provided by the System Secure Socket Layer (System SSL) element of z/OS or by using RACF's certificate management support. The key ring is associated with a server or client using server or client specific statements.

Note: Global step-up type certificates are not supported by the Telnet TN3270 server if the client application sends the handshake complete message to the server before completing the second handshake.

Client authentication

Client authentication provides additional authentication and access control by checking client certificates at the server. This support prevents a client from obtaining a connection without an installation approved certificate.

The server authenticates the client by receiving the client's certificate during the SSL handshake and verifying the certificate is valid. Validation is done by the server the same way the client validates the server's certificate. The client sends a signed certificate to the server. System SSL at the server decrypts the signature (message digest) using the public key of the client certificate issuer found in the server key database file. The server then creates a new message digest using the certificate's Distinguished Names and public key and compares the new message digest with the decrypted one. If they match, the server can be assured the client is authentic. Depending on where the client certificate is stored, up to three different levels of client authentication are available. Refer to the security information in the appropriate server or client chapters for setup details.

Level 1 authentication is performed by system SSL. The client passes an X.509 certificate to the Server as part of the SSL Handshake. To pass authentication, the Certificate Authority (CA) that signed the client certificate must be considered trusted by the server. That is, the certificate for the CA must be in the key ring used by the Server and designated as trusted. Note that the value of this option alone is based on which CAs are considered trusted. If the CA is a public CA and the certificate is in an easily obtained class, anyone can obtain such a certificate. In this case passing level 1 SSL Client Authentication does not provide much additional security unless coupled with the level 2 RACF support described below.

If the CA is controlled by the enterprise, some level of access control is provided because the client that possesses such a certificate is at least known to the organization.

Level 2 authentication requires that the client certificate be registered with RACF (or other SAF compliant security product) and mapped to a user ID. This is in addition to the checking done with the first level of client authentication support. The client certificate received during the SSL handshake is used to query the security product to verify that the certificate maps to a user ID known to the system prior to connection negotiation. This level of support provides additional access control at the server and ensures that the end user is known to have a valid user ID on the server host. Each server uses the returned user ID in a different way. Refer to the security information in the appropriate server or client chapters to see how the user ID is used for a particular server. Level 1 authentication is performed prior to level 2 authentication.

Level 3 authentication provides, in addition to level 1 and level 2 support, the capability to restrict access to the server based on the user ID returned from RACF. In some cases a certificate may be valid and mapped to a user ID but should be valid for only one of several servers. The third level of control uses the SERVAUTH RACF class to restrict access to the server based on client user ID. If the SERVAUTH class is not active or the SERVAUTH profile for the server is not defined, it is assumed level 3 authentication is not requested. If the SERVAUTH class is active and the server profile is defined, a connection is accepted only if the requestor's user ID associated with the client certificate is in the profile. Otherwise, the connection is dropped. See "Add user IDs to the SERVAUTH profile access list" on page 1184 for RACF setup details.

To enable Client Authentication for each server, use the following server-specific statements:

SERVER TYPE	**FTP**	**TN3270**	**DCAS**
STATEMENT	**SECURE_LOGIN**	**CLIENTAUTH**	**CLIENTAUTH**
AUTH LEVEL 1	REQUIRED	SSLCERT	LOCAL 1
AUTH LEVEL 2	VERIFY_USER	SAFCERT	LOCAL 2
AUTH LEVEL 3	VERIFY_USER	SAFCERT	LOCAL 3

Level 1 client authentication is done by SSL using a gskkyman key ring or a RACF key ring. If the client certificate was issued by a well-known Certificate Authority, it is likely the CA certificate is already primed in the gskkyman key ring. The CA certificate is probably also in RACF. However, all CA certificates in RACF initially have a status of NOTRUST. The CA certificate must be set to TRUST and connected to the appropriate RACF key ring. See "Update CA certificates to TRUST status" on page 1181 for detailed information. If the certificate issuer (a CA or self-signed) is not part of the list of well-known CAs, the key ring must be primed with the signer certificate of the CA or the self-signed client certificate.

After Level 1 authentication is performed by SSL using either key ring, the certificate is passed to the server which accesses the RACF database for Level 2 and Level 3 authentication.

Encryption algorithms

After authentication is performed, the client and server must agree on a symmetric encryption method and generate a single encryption key to use for data encryption. The agreed-on key is exchanged using the PKI method of encryption. Once the symmetric encryption algorithm (such as DES) and a single encryption key are chosen, all data exchanges use this algorithm and key instead of the PKI method of encryption.

In an SSL-encrypted session, all data is encrypted using the symmetric encryption algorithm immediately before it is sent to the client. Data from the client is decrypted immediately after it is received. The encryption algorithm that is used for the connection depends on a combination of the encryption algorithm list the SSL subsystem supports, the list the server wants to use, and the encryption algorithms the client requests. During the SSL handshake the client sends a list of encryption algorithms it is willing to use. The server submits its list and the SSL subsystem picks an algorithm all parties support giving preference to the order specified by the server. If the server does not support any of the encryption algorithms requested by the client, the connection is closed. The Telnet, FTP and DCAS servers and the FTP client use the SSL support provided by the System Secure Sockets Layer (System SSL) element of z/OS. The encryption algorithms supported by the servers and client are therefore dependent on the level of System SSL installed. The following encryption algorithms are supported by the base level of System SSL: NULL, RC2 export, RC4 export, DES. The System SSL Level 3 feature is required for Triple DES and RC4 non-export (128 bit) encryption algorithms. The encryption algorithm list can be customized for the servers and client to a subset of the System SSL list. Refer to the security information in the appropriate server or client chapters for specific server and client statements used for encryption list creation.

Encryption is provided either by BSafe software shipped with System SSL or by hardware. There is no TCP profile definition that controls whether the cryptographic hardware will be used for secure connections. When SSL initialization has completed, System SSL checks if ICSF is installed and active and if the hardware is enabled and loaded with the necessary Master Keys. If the hardware is not available at that time, all subsequent encryption is performed using software. If hardware is valid and ICSF is active at that time, the public key functions required during the SSL handshake and requests for encryption using DES and Triple-DES algorithms will be sent to the hardware. Otherwise, all cryptographic functions will be performed by software. Encryption requests using RC2 or RC4 algorithms are always performed by software. Also note that if ICSF subsequently becomes unavailable, System SSL will assume the hardware encryption is still wanted and encryption processing using DES or Triple-DES algorithms will fail until access to the hardware is restored. If subsequent session handshakes are attempted, they will also fail. Completion of SSL initialization is different for each server and client. Refer to the security information in the appropriate server or client chapters to understand when SSL initialization is complete and how to refresh SSL.

If hardware encryption is to be used, be sure that the RACF user ID associated with the server has read access to the RACF CSFSERV class resources. If ICSF is available but the server has not been given access to these resources, the SSL initialization may fail. The reason code is likely to be 4 (bad password) because System SSL will attempt to use the hardware encryption during processing of the key ring.

Enable CSFSERV resources

If hardware encryption and ICSF are installed, system SSL verifies that the user ID associated with the server is permitted to use CSFSERV resources. The RACF administrator should permit the RACF user ID to use the CSFSERV resources described here.

```
PERMIT service-name CLASS(CSFSERV) ID(serverid) ACCESS(READ)
```

The following CSFSERV resources (service-names) are accessed by System SSL.

1. CSFCKI Clear Key Import
2. CSFCKM Clear Key Import Multiple
3. CSFDEC DES and TripleDES Decipher
4. CSFENC DES and TripleDES Encipher
5. CSFOWH MD5 and SHA1 Hashing
6. CSFRNG Random Number Generate
7. CSFPKB RSA Key Token Build
8. CSFPKX RSA Public Key Extrace
9. CSFPKE RSA Public Key Encipher
10. CSFPKD RSA Private Key Decipher
11. CSFPKI RSA Key Import
12. CSFDSG Digital Signature Generate
13. CSFDSV Digital Signature Verify

z/OS FTP users can either permit every FTP client user ID to these general resource profiles, or they can mark these profiles as delegated and permit only the FTP daemon user ID to the profiles.

In the following example, resource CSFENC in class CSFSERV is delegated, and only the FTP daemon user ID (FTPD for this example) needs to be permitted. Make these changes before starting FTPD.

```
Permit the FTP daemon to the resource:
     PERMIT CSFENC CLASS(CSFSERV) ID(FTPD) ACCESS(READ)
Mark the resouce profile as delegated:
     RALTER CSFSERV CSFENC APPLDATA('RACF-DELEGATED')
Refresh the CSFSERV class:
     SETROPTS RACLIST(CSFSERV) REFRESH
```

For more examples, see the EZARACF sample in SEZAINST. For more information on authorizing daemons to use delegated resources, refer to *z/OS Security Server RACF Security Administrator's Guide*.

The MAXLEN installation option for hardware cryptography determines the maximum length that can be used to encrypt and decrypt data using ICSF/MVS. Set this option to 65527 or greater, because this is the maximum TCP/IP packet size.

The System SSL GSKSRVR server provides the capability to determine whether cryptographic hardware is being used through its DISPLAY CRYPTO operator command (for example, f gsksrvr,d crypto). The System SSL GSKSRVR server is not automatically started. For additional information on the SSL started task and setting up and using the GSKSRVR server, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*.

For additional information on controlling who can use cryptographic keys and services, refer to *z/OS Cryptographic Services ICSF Administrator's Guide*.

Creating and managing keys and certificates at the server

Overview

The gskkyman utility or RACF Common Keyring support can be used to manage the keys and certificates needed for SSL support.

The following table describes the steps necessary to implement the different levels of SSL security for each key ring management product.

		Key ring management product	
SSL function	**Steps**	**gskkyman**	**RACF**
Server Auth	1. Create a key ring file 2. Create a server certificate **If server certificate is self-signed** 3. Extract server certificate from server key ring 4. Add server certificate to client key ring	1. Page 1176 2. Page 1177 3. Page 1179 4. Page 1191	1. Page 1182 2. Page 1182 3. Page 1183 4. Page 1191
Client Auth Level 1	1. Set up server authentication **If client certificate is self-signed** 2. Extract client certificate from client key ring 3. Add client certificate to server key ring	1. See above 2. Page 1188 3. Page 1180	1. See above 2. Page 1188 3. Page 1180
Client Auth Level 2	1. Set up server authentication 2. Set up level 1 authentication 3. Associate certificate to RACF User ID[1]	1. See above 2. See above 3. Page 1184	1. See above 2. See above 3. Page 1184
Client Auth Level 3	1. Set up server authentication 2. Set up Level 1 authentication 3. Set up Level 2 authentication 4. Add User IDs to the server's SERVAUTH profile access list	1. See above 2. See above 3. See above 4. Page 1184	1. See above 2. See above 3. See above 4. Page 1184
Express Logon Feature	1. Set up server authentication 2. Set up level 1 authentication (optional for DCAS) 3. Set up Level 2 authentication (optional for DCAS) 4. Define passticket profiles	1. See above 2. See above 3. See above 4. Page 1185	1. See above 2. See above 3. See above 4. Page 1185

1. The gskkyman utility cannot associate a user ID with a client. If gskkyman is used for key ring and certificate management, a second security product is needed to associate the certificate to a user ID. In this example, RACF is used for the association while system SSL continues to use gskkyman.

Certificate file types

The following sections mention several certificate formats. Below is a high level summary of the differences.

- PKCS12 files are used to move the server certificate to another server key ring. Because this format contains the private key, the file is usually password protected. IBM recommends only using this format when required.
  - Commonly used file extension is .p12
  - Contains private key, public key and certificate
  - Created by
    - HOD export function

      When the HOD client specifies a client certificate to send to the server during SSL processing, it must be in this format. The private key portion is not sent to the server.
    - Netscape export function
    - gskkyman ″Export keys to a PKCS12 file″ function
- Certificate files are normally needed when a self-signed certificate is used. In this case, each self-signed certificate appears to be signed by a unique CA. Therefore, the client's key ring (if this is a self-signed server certificate) or server's key ring (if this is a self-signed client certificate) must be primed to recognize the issuer of the self-signed certificate. This format can be used to prime a key ring with the issuer's CA certificate. This format can also be used when registering a client certificate with RACF.
  - Commonly used file extensions are .crt and .der
  - Contains public key and certificate
  - Created by
    - HOD extract function
    - gskkyman's ″Create a self-signed certificate″ function

Common terminology

The following variable names are used throughout the appendix:

tnserverid
: The user ID defined to RACF given superuser status that is associated with the job started for Telnet. This can be the job name of the TCP/IP stack if Telnet is running as part of the stack, or it can be the job name of the stand-alone Telnet application.

dcasserverid
: The user ID defined to RACF given superuser status that represents the Digital Certificate Access Server.

ftpserverid
: The user ID defined to RACF given superuser status that represents the FTP daemon and all spawned FTP Servers.

serverid
: The user ID defined to RACF given superuser status that represents any of the servers above.

userid
: The user ID that is associated with a client certificate in the RACF key ring database. Or the TSO user ID that requires authority to issue certain RACF commands.

Copying z/OS UNIX files to MVS data sets

Certificate and database files are often stored in HFS file formats and sometimes need to be copied into MVS data set formats. MVS data sets can be created from HFS files by using the TSO OGET command with the BINARY option and can be protected using RACF. For example:

```
OGET '/tmp/telnet/mvs180.kdb' 'TCPCS6.MVS180.KDB' BINARY
OGET '/tmp/telnet/mvs180.sth' 'TCPCS6.MVS180.STH' BINARY
```

It is recommended that the MVS dsnames be the HFS filenames prefixed by one or more high-level qualifiers. The same high-level qualifier(s) must be used for both the key ring and the stash file. This ensures that the name relation used to generate the stash file from the key ring file is unchanged. Refer to *z/OS UNIX System Services Command Reference* for more information on the use of the OGET command.

Using the gskkyman utility

This section gives examples of how to use the gskkyman utility to:

- Create key rings
- Create a server self-signed certificate
- Extract a server certificate
- Add client certificates into a key ring

The gskkyman utility is a command-line utility. It prompts you for the information you need to perform a task. If you make an error, it issues a message and prompts you again for the information.

The gskkyman utility is documented in *z/OS Crytographic Service System Secure Sockets Layer Programming*. It is recommended that you read the gskkyman topics in this document before starting to use the gskkyman.

Additional information and examples can also be found in the following Redbooks:

- *SecureWay Communications Server for OS/390 V2R8 TCP/IP: Guide to Enhancements*
- *IBM SecureWay Host On-Demand 4.0: Enterprise Communications in the Era of Network Computing*

To run gskkyman, you must have access to the z/OS Cryptographic Services message catalogs and DLLs. The C DLL Library (that is, SYS1.SCLBDLL) must be available and APF authorized to avoid possible abends caused by trying to access a nonexisting or non-APF authorized system. For example, if the z/OS Cryptographic Service DLL library is not part of the linklist concatenation, an "export STEPLIB=SYS1.SIEALNKE" command might be needed. For additional information, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*.

The gskkyman utility is shipped with z/OS in System SSL as a part of the Cryptographic Services Base element of z/OS. It supports the generation of key sizes of 1024 or 2048 bits. Note that if you have existing keys with a size of 512, these keys are still usable. The gskkyman utility runs under the z/OS shell and can create several types of HFS files. System SSL requires the following files:

- A key ring file (also known as a key database).
- A password file (also known as a stash file) which contains the password associated with the key ring file.

The key ring file and the stash file are used by the server to obtain the server's certificate and the public/private key pair used during SSL handshake processing. The server uses the stash file as the mechanism to obtain the key ring password rather than using a configuration parameter which might be accessible to a larger number of people. The stash file is created by using gskkyman's 'Store encrypted database password' function on the main menu.

Security of these files is an installation responsibility. It is recommended to restrict the file access to users with superuser authority.

The server must have read and write access to the key database and read access to the password file.

Note: The gskkyman utility accepts only the HFS files.

The gskkyman utility allows you to enter the fully qualified path and file name when it prompts you for a key ring, certificate request, or certificate file name. However, you should change to the path where the file will be stored before you start gskkyman.

Create a key ring file

Before starting gskkyman, we suggest that you start in the directory where the key ring is to be created. Otherwise, be sure to include the full file name when specifying the key ring name.

1. Start gskkyman. This will display the *Database Menu*. Select *Create new database* (option 1).
2. Specify the key database name, password, and expiration information as requested.
3. Create a password file (also known as a stash file) by selecting *Store database password* (option 10) from the *Key Management Menu*.

Following is a sample of the gskkyman output for creating a key database. Sample user replies are shown in **bold** characters.

1. The *Database Menu* that starts the process for creating a key database is shown below. Select the *Create new database* option.

```
       Database Menu

   1 - Create new database
   2 - Open database
   3 - Change database password
   4 - Change database record length
   5 - Delete database

   0 - Exit program

Enter option number: 1
```

2. Specify the key database name, password, and expiration information as requested.

```
Enter key database name (press ENTER to return to menu): server.kdb
Enter database password (press ENTER to return to menu):
Re-enter database password:
Enter password expiration in days (press ENTER for no expiration): 365
Enter database record length (press ENTER to use 2500):

Key database /SYSTEM/usr/keyring/server.kdb created.

Press ENTER to continue.
```

3. Pressing enter brings up the *Key Management Menu* as follows. Create the password file by selecting the *Store database password* option.

```
     Key Management Menu

   Database: /SYSTEM/usr/keyring/server.kdb

    1 - Manage keys and certificates
    2 - Manage certificates
    3 - Manage certificate requests
    4 - Create new certificate request
    5 - Receive certificate issued for your request
    6 - Create a self-signed certificate
    7 - Import a certificate
    8 - Import a certificate and a private key
    9 - Show the default key
   10 - Store database password
   11 - Show database record length

    0 - Exit program

Enter option number (press ENTER to return to previous menu): 10
Database password stored in /SYSTEM/usr/keyring/server.sth.

Press ENTER to continue.
```

Pressing enter returns you to the *Key Management Menu*. You can proceed to create your server certificate or exit.

Create a server self-signed certificate

Refer to gskkyman documentation for the steps necessary to create a CA-signed server certificate. With gskkyman, you can create your own self-signed certificate for testing:

1. From the *Database Menu*, open your key ring file (Option 2–*Open database*). From the *Key Management Menu*, select *Create a self-signed certificate* (option 6).
2. Specify the certificate type to be created from one of the end user certificate options. The type of end user certificate created depends on the security requirements of your installation.
3. Specify the label, subject name, and length of time the certificate is valid as requested.
4. After the certificate is created, set the certificate as the default by selecting *Manage keys and certificates* (option 1), selecting the certificate you created, followed by selecting *Set key as default* (option 3).

The following is a sample of the gskkyman output for creating a self-signed certificate. Sample user replies are shown in **bold** characters.

1. The *Key Management Menu* that starts the process for creating a self-signed certificate is shown below. Select *Create a self-signed certificate* (option 6).

```
     Key Management Menu

     Database: /SYSTEM/usr/keyring/server.kdb

 1 - Manage keys and certificates
 2 - Manage certificates
 3 - Manage certificate requests
 4 - Create new certificate request
 5 - Receive certificate issued for your request
 6 - Create a self-signed certificate
 7 - Import a certificate
 8 - Import a certificate and a private key
 9 - Show the default key
10 - Store database password
11 - Show database record length

 0 - Exit program


Enter option number (press ENTER to return to previous menu): 6
```

2. The *Certificate Type* menu will be displayed. Select one of the server certificate types. Make sure you pick a certificate type that your client applications support.

```
       Certificate Type

   1 - CA certificate with 1024-bit RSA key
   2 - CA certificate with 2048-bit RSA key
   3 - CA certificate with 4096-bit RSA key
   4 - CA certificate with 1024-bit DSA key
   5 - User or server certificate with 1024-bit RSA key
   6 - User or server certificate with 2048-bit RSA key
   7 - User or server certificate with 4096-bit RSA key
   8 - User or server certificate with 1024-bit DSA key

Select certificate type (press ENTER to return to menu): 5
```

3. You will then be asked for the certificate data and the certificate will be created.

```
Enter label (press ENTER to return to menu): selfsignedcert
Enter subject name for certificate
  Common name (required): test server certificate
  Organizational unit (optional): dev
  Organization (required): dev
  City/Locality (optional):
  State/Province (optional):
  Country/Region (2 characters - required): US
Enter number of days certificate will be valid (default 365): 5000

Please wait .....

Certificate created.

Press ENTER to continue.
```

4. Make the self-signed certificate the default server certificate. Pressing enter returns you to the *Key Management Menu* shown in step 1 above. Select *Manage keys and certificates* (option 1). This will bring up the *Key and Certificate List* menu shown below. Select the number that corresponds to the self-signed certificate just created.

```
       Key and Certificate List

       Database: /SYSTEM/usr/keyring/server.kdb

   1 - selfsignedcert

   0 - Return to selection menu

Enter label number (ENTER to return to selection menu, p for previous list): 1
```

This will bring up the *Key and Certificate Menu*. Select *Set key as default* (option 3).

```
       Key and Certificate Menu

       Label: selfsignedcert

   1 - Show certificate information
   2 - Show key information
   3 - Set key as default
   4 - Set certificate trust status
   5 - Copy certificate and key to another database
   6 - Export certificate to a file
   7 - Export certificate and key to a file
   8 - Delete certificate and key
   9 - Change label

   0 - Exit program

Enter option number (press ENTER to return to previous menu): 3

Default key set.
```

To use the new default server certificate, the server must reinitialize its SSL environment. Refer to the security information in the appropriate server or client chapters for SSL initialization details. The self-signed server certificate will need to be added to the client's key database as a CA. For additional information, see "Extract the server certificate from the key ring."

Extract the server certificate from the key ring

If using a self-signed server certificate, the server certificate must be added to the client's key database as a CA certificate. Some clients provide the capability to extract the server's certificate when the client connects to the server. For some clients, however, this process must be done manually by exporting the server's certificate to a file, sending the server's certificate file to the client, and adding the server's certificate to the client's key database.

The server certificate can be exported to a file from the *Key and Certificate Menu*. If you have just opened your key ring file and are at the *Key Management Menu*, select *Manage keys and certificates* (option 1) and select the certificate you want to export from the list. You will then see the *Key and Certificate Menu*.

1. Select *Export certificate to a file* (option 6).
2. Specify the desired format and file name.

The following is a sample of the gskkyman output for exporting a certificate. Sample user replies are shown in **bold** characters.

1. The *Key and Certificate Menu* that starts the export process follows. Select *Export certificate to a file* (option 6).

```
       Key and Certificate Menu

       Label: selfsignedcert

   1 - Show certificate information
   2 - Show key information
   3 - Set key as default
   4 - Set certificate trust status
   5 - Copy certificate and key to another database
   6 - Export certificate to a file
   7 - Export certificate and key to a file
   8 - Delete certificate and key
   9 - Change label

   0 - Exit program

Enter option number (press ENTER to return to previous menu): 6
```

2. The *Export File Format* menu is displayed, and you will be asked to select the export format and to enter a file name.

```
       Export File Format

   1 - Binary ASN.1 DER
   2 - Base64 ASN.1 DER
   3 - Binary PKCS #7
   4 - Base64 PKCS #7

Select export format (press ENTER to return to menu): 1
Enter export file name (press ENTER to return to menu): binservercert.der

Certificate exported.
```

   If using ftp to send the export file to the client, remember to send in binary format if the file format chosen above was binary.

Add client certificates to the server key ring

If using a client self-signed certificate, the certificate must be added to the server key ring as a CA certificate. Send the client certificate to the MVS host using FTP (with the BINARY send option if the certificate was extracted in binary format). The client may be the DCAS client (DCAR), FTP client, or TN3270 client. Use gskkyman's 'Import a certificate' option to obtain the client CA from the binary DER file. If the client certificate was issued by a well-known CA, only the signer's certificate needs to be in the key ring. The gskkyman utility primes its key rings with several well-known CAs.

Using RACF's Common Keyring support

This section gives examples of how to use RACF Common Keyring support to:

- Create key rings
- Create a server self-signed certificate
- Extract a server certificate
- Add client certificates into a key ring

RACF can be used to manage the keys and certificates normally stored in the key database. All the functions that the gskkyman utility provides, are also available in

the RACF support. However, because RACF can manage multiple key rings, certificates and key rings are added independently. A certificate is then connected to one or more key rings.

All the server key rings and certificates are stored in the RACF database. There are no separate key database or stash files.

Refer to *z/OS Security Server RACF Security Administrator's Guide* for information about how to use RACF to manage your key database information.

For detailed information on the RACDCERT command and other commands that might be useful in managing your RACF key ring data, refer to *z/OS Security Server RACF Command Language Reference* for the full syntax and description of these commands.

RACF key ring names, labels, and so on are case sensitive. When adding the key ring name to the server profile, be sure that the correct case is used.

Before using RACF to store your key database information:

1. Ensure that RACDCERT is defined as an authorized TSO command in the IKJTSOxx member.
2. The well-known CA certificates are initially marked as NOTRUST in the RACF database and you will have to update the CA certificates that you plan to support to TRUST status.
3. Refresh the applicable RACF class after making changes.
4. There are various ways to register the client certificate with RACF or set up a RACF Certificate Name Filter. For a complete description of RACF management of digital certificates and options available, refer to *z/OS Security Server RACF Security Administrator's Guide*. If using RACF's Certificate Name Filtering with MultiID filters, TN3270 client authentication processing only matches filters that specify generic (*) criteria.
5. Most tasks related to certificates are managed using the RACDCERT command. The issuer of these commands must have appropriate RACF authority to the IRR.DIGTCERT.function resource in the FACILITY class with UPDATE, CONTROL, or READ ACCESS. Refer to *z/OS Security Server RACF Security Administrator's Guide* for more information on controlling the use of the RACDCERT command and for a complete description of functions needed for key ring and certificate use.
6. For best performance, consider RACLISTing the DIGTCERT and DIGTRING classes.

RACF panels also support most of the certificate and key ring functions and can be used to perform these actions.

Configuring RACF services for the servers

This section describes some commands needed for configuring RACF for the servers. These commands are in EZARACF in the SEZAINST data set.

Update CA certificates to TRUST status: Several well-known CA certificates are primed in the RACF database but are initially marked NOTRUST. To use the certificate, it must be changed to TRUST status. As an example, the commands below show how to change the Verisign Class 3 CA to trusted status and then connect it to a key ring.

```
RACDCERT CERTAUTH ALTER( LABEL('Verisign Class 3 Primary CA')) TRUST

RACDCERT ID(serverid) CONNECT (CERTAUTH    RING(SERVERKeyring)
      LABEL('Verisign Class 3 Primary CA') USAGE(CERTAUTH)  )
```

Activate the DIGTCERT, DIGTRING, and optional DIGTNMAP classes: The DIGTCERT and DIGTRING classes must be active before defining certificates or key rings to RACF by using the SETROPTS commands. For example:

```
SETROPTS CLASSACT(DIGTCERT)
SETROPTS CLASSACT(DIGTRING)
```

If using Certificate Name Filtering, ensure that the DIGTNMAP class is active. For example:

```
SETOPTS CLASSACT(DIGTNMAP)
```

Also be sure to do a refresh after any changes. For example:

```
SETROPTS RACLIST(DIGTCERT) REFRESH
SETROPTS RACLIST(DIGTRING) REFRESH
SETROPTS RACLIST(DIGTNMAP) REFRESH
```

Allow SSL key ring confirmation: System SSL verifies that the server RACF user ID does have access to the key ring. Therefore, if the server is started as an MVS started procedure, you must permit the RACF user ID associated with the server to have control access to the IRR.DIGTCERT.LISTRING. For example:

```
RDEFINE FACILITY  (IRR.DIGTCERT.LISTRING)  UACC(NONE)
PERMIT  IRR.DIGTCERT.LISTRING  CLASS(FACILITY)  ID(serverid) ACCESS(CONTROL)
```

To access a certificate authority (CA) or site certificate, the server must have control access to the IRR.DIGTCERT.LIST resource in the FACILITY class. For example:

```
RDEFINE FACILITY  (IRR.DIGTCERT.LIST)  UACC(NONE)
PERMIT  IRR.DIGTCERT.LIST  CLASS(FACILITY)  ID(userid) ACCESS(CONTROL)
```

If the server certificate was added as a site certificate, the server must have control access to the IRR.DIGTCERT.GENCERT resource in the FACILITY class. For example:

```
RDEFINE FACILITY  (IRR.DIGTCERT.GENCERT)  UACC(NONE)
PERMIT  IRR.DIGTCERT.GENCERT  CLASS(FACILITY)  ID(userid) ACCESS(CONTROL)
```

If the DCAS is started from a TSO user ID under the z/OS UNIX shell, you must also permit that ID. For example:

```
RDEFINE FACILITY  (IRR.DIGTCERT.LISTRING)  UACC(NONE)
PERMIT  IRR.DIGTCERT.LISTRING  CLASS(FACILITY)  ID(userid) ACCESS(CONTROL)
```

Create a key ring file

Use the following RACF command to add (create) a server key ring and associate it with the server RACF user ID:

```
RACDCERT ID(serverid) ADDRING(SERVERRING)
```

To delete or list a key ring:

```
racdcert ID(serverid) delring(SERVERRING)
racdcert ID(serverid) listring(SERVERRING)
```

Create a server self-signed certificate

Refer to the RACF documentation for the steps necessary to create a CA-signed server certificate. To create a self-signed server certificate called XXXDN, for user

ID serverid, use the following command, where CN is the common name and OU is the organization unit name. Additional options are available within SUBJECTSDN.

```
RACDCERT ID(serverid)  GENCERT  SUBJECTSDN(CN('UNIT1') OU('TESTING') C('US') ) TRUST
  WITHLABEL('XXXDN') SIZE(1024)
```

To connect the certificate to a key ring and make it the default certificate, use the following command. This example assumes a key ring called serverRing has already been created.

```
RACDCERT ID(serverid) CONNECT(ID(serverid)  LABEL('XXXDN') RING(SERVERRING) DEFAULT)
```

Extract a server certificate from a server key ring

If using FTP to send the server self-signed certificate to the client, use RACDCERT Export to export the server certificate to an MVS data set in DER format. The server self-signed certificate must be added to the client key ring to prime it for decrypting the server certificate. EXPORT generally implies exporting both a certificate and private key. However, the CERTDER format instructs the command to export only the certificate in DER format, which is generally considered an EXTRACT. Use the following RACF command:

```
RACDCERT ID(serverid) EXPORT(LABEL('XXXDN')) DSN('dataset name') FORMAT(CERTDER)
```

Add client certificates to the server key ring

If using a client self-signed certificate, the certificate must be added to the server key ring as a CA certificate. Send the client certificate to the MVS host using FTP (with the BINARY send option). The client may be the DCAR, FTP client, or TN3270 client. If the client certificate is issued by a well-known CA, only the signer's certificate needs to be in the key ring. The well-known CA certificates are initially marked as NOTRUST and must be updated to TRUST status.

The client self-signed certificate must be registered into the RACF database before the certificate can be associated with a key ring. Associate the certificate to a RACF user ID to register the certificate into RACF. For example, the following command uses the binary DER client certificate in an MVS data set named SSCLNTCERT.USER2.DER, associates it with RACF user ID USER2, and gives it a label of CLNTCERT_USER2:

```
RACDCERT ID(USER2) ADD('SSCLNTCERT.USER2.DER') WITHLABEL('CLNTCERT_USER2') TRUST
```

This command requires that the certificate be defined in an MVS data set. If the certificate is defined in the HFS, you can use the TSO OGET command (with the BINARY send option) to move the certificate to an MVS data set.

Use the RACDCERT CONNECT command to connect the client certificate to the RACF key ring as a CA certificate. In this example, the RACF user ID associated with the server is serverid and the key ring name used by the server is serverRing:

```
RACDCERT ID(serverid) CONNECT (ID(USER2) RING(serverRing)
  LABEL('CLNTCERT_USER2') USAGE(CERTAUTH) )
```

Be sure to refresh the DIGTCERT and DIGTRING class. For example:

```
SETROPTS RACLIST (DIGTRING) REFRESH
SETROPTS RACLIST (DIGTCERT) REFRESH
```

Reinitialize SSL: Reinitialize the server SSL to pick up the certificates that have been added to the key database. Refer to the security information in the appropriate server and client chapters to understand when SSL initialization is complete and how to refresh SSL.

Add user IDs to the SERVAUTH profile access list

With level 3 authentication, you specify the user IDs that are allowed to connect into a specific Telnet port, DCAS server, or FTP port by associating the user IDs to each server's RACF SERVAUTH profile. Refer to the security information in the appropriate server and client chapters for level 3 setup. The user ID associated with the client certificate can then be checked against the SERVAUTH class profile entry. The use of this RACF class is optional. If the SERVAUTH RACF class is active and a RACF profile for the port is defined, this level of RACF authorization will be verified prior to connection negotiation. If the SERVAUTH class is not active or there is no RACF profile, this indicates that this level of check is not required and the client is allowed to connect to the server as long as the client certificate was validated.

TN3270 server: The RACF profile name is:

```
EZB.TN3270.sysname.tcpname.PORTnnnnn
```

where nnnnn is the port number with leading zeros. The profile name can contain wildcards to the extent that the security product allows. All security product rules (for example wildcards, PROTECTALL, and so on) apply. For example, the profile name for TCP stack TCPCS running on system MVSA for port 992 would be:

```
EZB.TN3270.MVSA.TCPCS.PORT00992
```

If all systems will use the same access list, and RACF generic profile checking is active for the SERVAUTH class, the following profile name could be used:

```
EZB.TN3270.*.TCPCS.PORT00992
```

To protect all ports with a single profile, the following security product profile name could be used:

```
EZB.TN3270.MVS.TCPCS.PORT*
```

To restrict access on a port basis, the following RACF setup is needed and must be done by a user ID that has authority to issue the specified RACF commands:

- Activate the RACF SERVAUTH class, if not active:
  ```
  SETROPTS CLASSACT(SERVAUTH)
  ```
- Define the profile for the Telnet port:
  ```
  RDEFINE  SERVAUTH   EZB.TN3270.sysname.tcpname.PORTnnnnn  UACC(NONE)
  ```
- Permit the user ID associated with TCP to the port profile:
  ```
  PERMIT EZB.TN3270.sysname.tcpname.PORTnnnnn CL(SERVAUTH) ID(tcpuserid) ACCESS(READ)
  ```
- Ensure the SERVAUTH class is RACLISTed. If it is not, RACLIST it:
  ```
  SETROPTS RACLIST(SERVAUTH)
  ```
- Refresh the SERVAUTH class before using:
  ```
  SETROPTS RACLIST(SERVAUTH) REFRESH
  ```

DCAS: The RACF profile name is:

```
EZA.DCAS.cvtsysname
```

- Activate the RACF SERVAUTH class, if not active:
  ```
  SETROPTS CLASSACT(SERVAUTH)
  ```
- Define the profile:
  ```
  RDEFINE SERVAUTH EZA.DCAS.cvtsysname UACC(NONE)
  ```
- Permit the user ID associated with TCP to the port profile:
  ```
  PERMIT EZA.DCAS.cvtsysname CLASS(SERVAUTH) ACCESS(CONTROL) ID(dcasid)
  ```
- Ensure the SERVAUTH class is RACLISTed. If it is not, RACLIST it:

```
SETROPTS RACLIST(SERVAUTH)
```

- Refresh the SERVAUTH class before using:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Note: The RACF user ID associated with the certificate and the EZA.DCAS.cvtsysname can be any valid user ID.

FTP server: The RACF profile name is:

```
EZB.FTP.sysname.ftpdaemonname.PORTnnnnn
```

where nnnnn is the port number. The profile name can contain wildcards to the extent that the security product allows. All security product rules (for example wildcards, PROTECTALL, and so on) apply. For example, the profile name for FTP daemon FTPD running on system MVSA for port 992 would be:

```
EZB.FTP.MVSA.FTPD.PORT992
```

If all systems will use the same access list and RACF generic profile checking is active for the SERVAUTH class, the following profile name could be used:

```
EZB.FTP.* .FTPD.PORT992
```

To protect all ports with a single profile, the following security product profile name could be used:

```
EZB.FTP.MVS.FTPD.PORT*
```

To restrict access on a port basis, the following RACF setup is needed and must be done by a user ID that has authority to issue the specified RACF commands:

- Activate the RACF SERVAUTH class, if not active:

```
SETROPTS CLASSACT(SERVAUTH)
```

- Define the profile for the FTP port:

```
RDEFINE  SERVAUTH   EZB.FTP.sysname.ftpdaemonname.PORTxxxxx  UACC(NONE)
```

- Permit the user ID associated with the FTP daemon to the port profile:

```
PERMIT EZB.FTP.sysname.ftpdaemonname.PORTnnnnn CL(SERVAUTH)   ID(tcpuserid)  ACCESS(READ)
```

- Ensure the SERVAUTH class is RACLISTed. If it is not, RACLIST it:

```
SETROPTS RACLIST(SERVAUTH)
```

- Refresh the SERVAUTH class before using:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Define PassTicket profiles to RACF

The following recommendations apply when defining PassTicket profiles to RACF:

- Use the SETROPTS CLASSACT(PTKTDATA) command to activate the PTKTDATA class.
- Use the SETROPTS RACLIST (PTKTDATA) command to RACLIST the PTKTDATA class.
- For each application to which users want to gain access with a PassTicket, you must define a PTKTDATA class profile. For example, to give users access to TSO, define a profile for TSO using the following command:

```
RDEFINE PTKTDATA  TSO3050
                    SSIGNON(KEYMASKED(key_value))
                    UACC(NONE)
```

The DIGTNMAP and DIGTCRIT classes support profiles. The application ID used for DIGTCRIT profiles must be the same as that used on the HOD V5 Application ID popup window.

Defining profiles for applications such as TSO can be tricky because RACF has special methods for naming profiles. For more information, refer to *z/OS Security Server RACF Security Administrator's Guide*.

Add the client certificate to the RACF server key ring and associate the client certificate with a user ID.

Define a PassTicket profile for each application accessed by Express Logon. The application ID portion of the profile must match that configured on the HOD V5 workstation Application ID popup window.

```
SETR  CLASSACT(PTKTDATA)        //* Activate the PassTicket data class
SETROPTS  RACLIST(PTKTDATA)     //* Raclist the class
SETR RACLIST(PTKTDATA) REFRESH
RDEFINE   PTKTDATA  TSO3390  SSIGNON(KEYMASKED(key_value))  UACC(NONE)
```

Migrating an existing gskkyman key database to RACF

To migrate from an existing key database (kdb) created by gskkyman, each certificate that the customer has added must be individually exported and then added to the RACF database. The RACF database is already primed with some well-known Certificate Authorities (CA), so it is not necessary to migrate these CA certificates to RACF. Note however, that the well-known CAs are initially marked as NOTRUST in the RACF database and you will have to update the CA certificates that you plan to support to TRUST status.

To migrate a server certificate from your kdb created by gskkyman to RACF:

1. Use gskkyman to export the certificate and key to a PKCS12 format file:
   a. Open the key database file that you want to migrate.
   b. Select 'List/Manage keys and certificates'.
   c. Select the certificate to be exported.
   d. Select 'Export the certificate and key to a file', supply the name of the HFS file where the certificate and key will be stored (in PKCS12 format) and enter the password to protect the file when prompted.
2. Use the OGET command described in "Copying z/OS UNIX files to MVS data sets" on page 1175 to create an MVS data set.
3. Add the certificate and key to the RACF database and assign it to a user, if applicable. If this is the server certificate, assign it to the user ID associated with the server. For example:

   ```
   RACDCERT ID(serverid) ADD('Serverid.mycert.p12') WITHLABEL('ServerCert')
   PASSWORD('mypw') TRUST
   ```

You will also need to create a key ring for your server. For example:

```
RACDCERT ID(serverid) ADDRING(serverring)
```

Then connect the appropriate certificates to the key ring. For example, to connect the default server certificate that was migrated above ('ServerCertificate') to the 'serverring' that we associated with serverid:

```
RACDCERT ID(serverid) CONNECT( ID(serverid) LABEL('ServerCert')
RING(ServerRing) DEFAULT )
```

Creating and managing keys and certificates at the client

Normally, a client certificate should be obtained from a well-known Certificate Authority (CA). The Certificate Authority's root certificate needs to be included in the server's key data base as a trusted authority in order for the client's certificate to pass the SSL protocol's client authentication process. If the client certificate has been issued by a well-known CA, the client certificate need not reside in the server's key database. If an installation uses self-signed client certificates for testing purposes, each certificate appears to be issued by a unique CA. Therefore, each self-signed client certificate must be added to the server's key database as a CA.

If you also want verification that the client certificate is registered with your security product, the client certificate must reside in the server's security product database (using the RACDCERT command with the ADD option is one way to add the client certificate to RACF). See "Using RACF's Common Keyring support" on page 1180 for more information on using RACF to store certificates. If the installation is using self-signed client certificates and requests verification that the client certificate is registered with the security product, the client certificate must reside in both the server's key database (as a CA certificate) and in RACF.

Steps to create a self-signed client certificate vary depending on the source of the client certificate. See "Create a self-signed client certificate" for details on creating a client certificate using HOD's Certificate Management utility.

After the client certificate has been created and extracted as a DER data file at the client, FTP the binary DER data file to the z/OS host using FTP's binary option. If using RACF for the key ring, level 2 client authentication, or level 3 client authentication, an MVS data set will need to be created. If FTP created a z/OS UNIX file, use the OGET command described in "Copying z/OS UNIX files to MVS data sets" on page 1175 to create an MVS data set.

Create a self-signed client certificate

See HOD's online documentation for additional details. This sample uses a locally installed HOD V4 client on an NT system using HODs Key Management Utility.

1. On the HOD client, go to HOD's Certificate Management panels (go to Start, Programs, Host On Demand, Administration, Certificate Management.) and open up the key database by selecting the open icon. Usually a key database will exist. If you have never used the key database, it might have a default password (usually ncod - the help menu should contain help information that specifies the default password for your system) or you can select the new option. If new is selected, the correct path and file name will normally be filled in by HOD. Do not change this file name. The HOD key database is normally in HOD's bin subdirectory and named HODClientKeyDb.kdb.

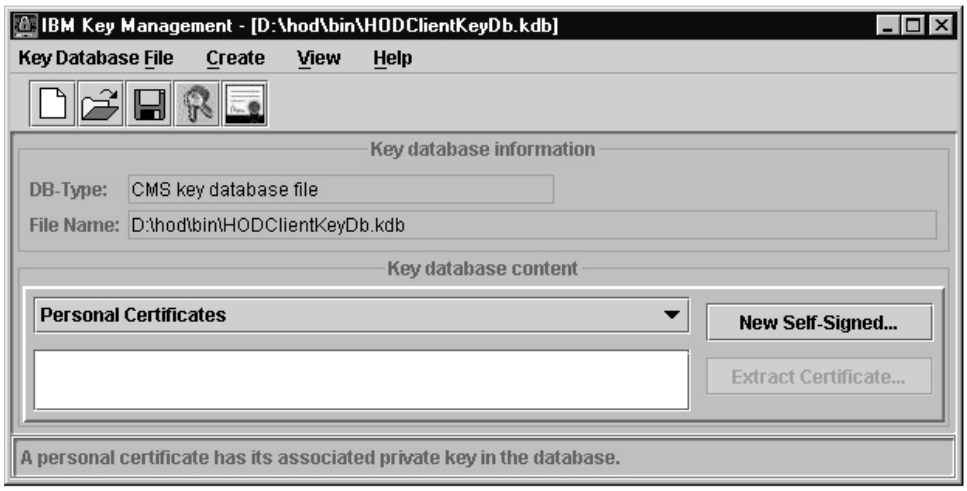


Figure 121. IBM Keys Management

If *Personal Certificates* is not displayed, click the drop-down list arrow and select *Personal Certificates* from the pull-down list.

2. Create a self-signed personal certificate by selecting the *New Self-Signed* button. The *Create New Self-Signed Certificate* screen is displayed.

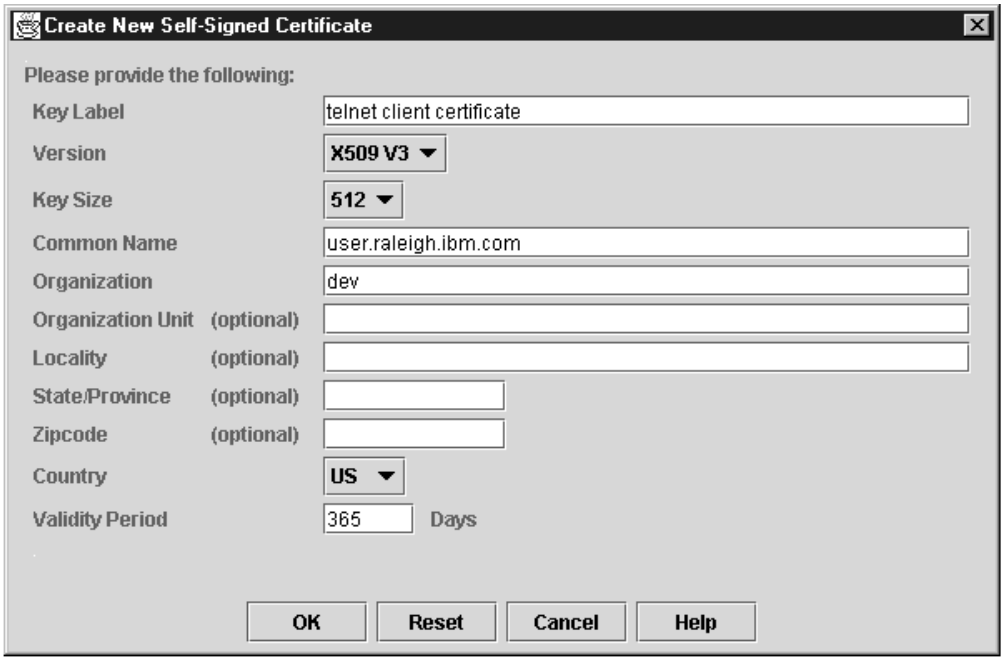


Figure 122. Create New Self-Signed Certificate

Fill in the requested information and then click OK. The new certificates will now be in the personal certificates list.

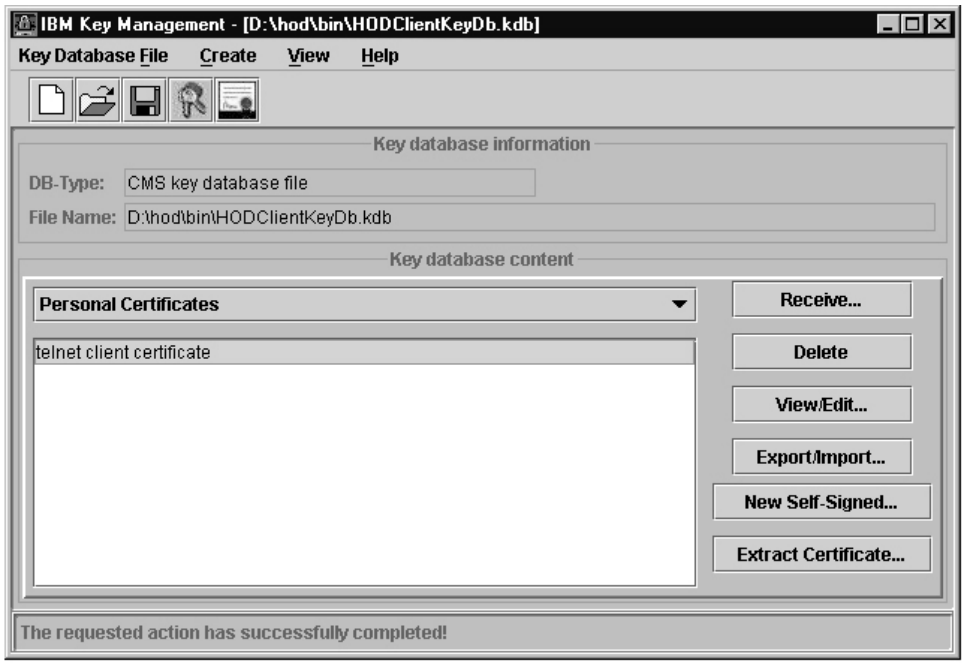


Figure 123. IBM Key Management

3. Use the export function by selecting the Export/Import button to create a PKCS12 file. This is the file that the HOD client will use.

   Specify the path and file where the exported PKCS12 file will be stored and click OK. Enter a password to protect the file when prompted.

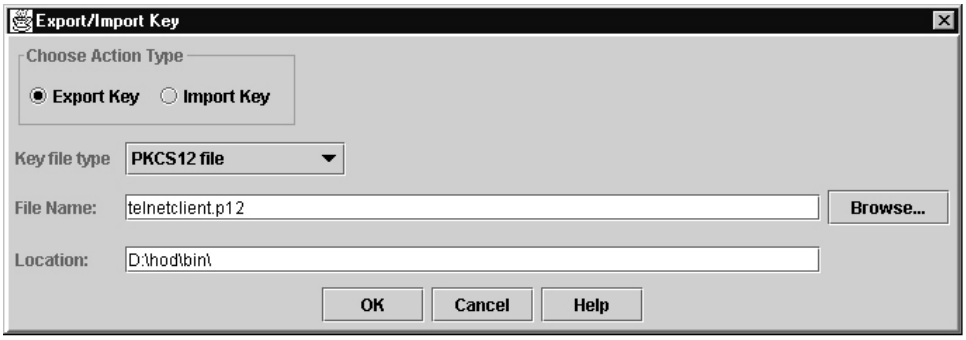


Figure 124. Export/Import Key

4. Create the certificate file that will be used to prime the server's key ring with the CA for the self-signed client certificate.

   Use the Extract Certificate function from the panel shown in Figure 123 to create a binary DER data file. This file will have the format filename.der.

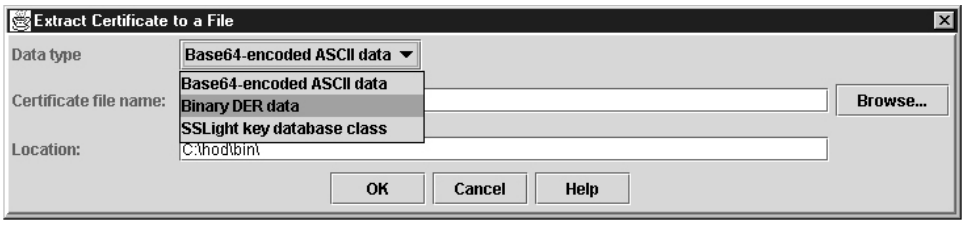


Figure 125. Extract Certificate to a File

Specify the path and file where the exported binary DER file will be stored and click on OK. This is the file you will FTP to the server host.

5. When you start a session from HOD to a port that requires a client certificate, HOD will display a panel that requests the client certificate file and password. The pkcs12 file created in step 3, should be specified.

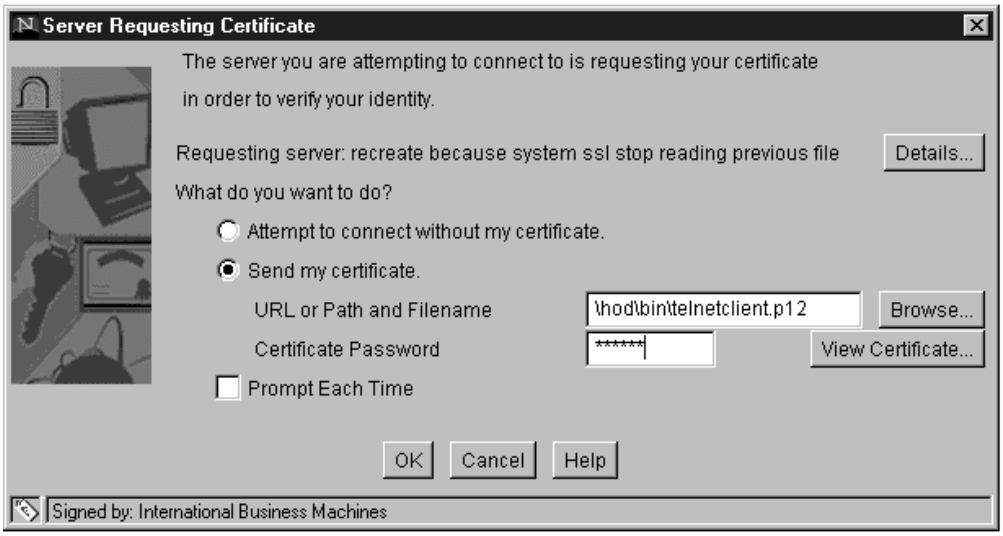


Figure 126. HOD connection using a client certificate

Note: If using HOD and you are connecting to a port that requires a client certificate, the security properties for the connection must indicate that a certificate should be sent. The following example shows the HOD Security properties screen.

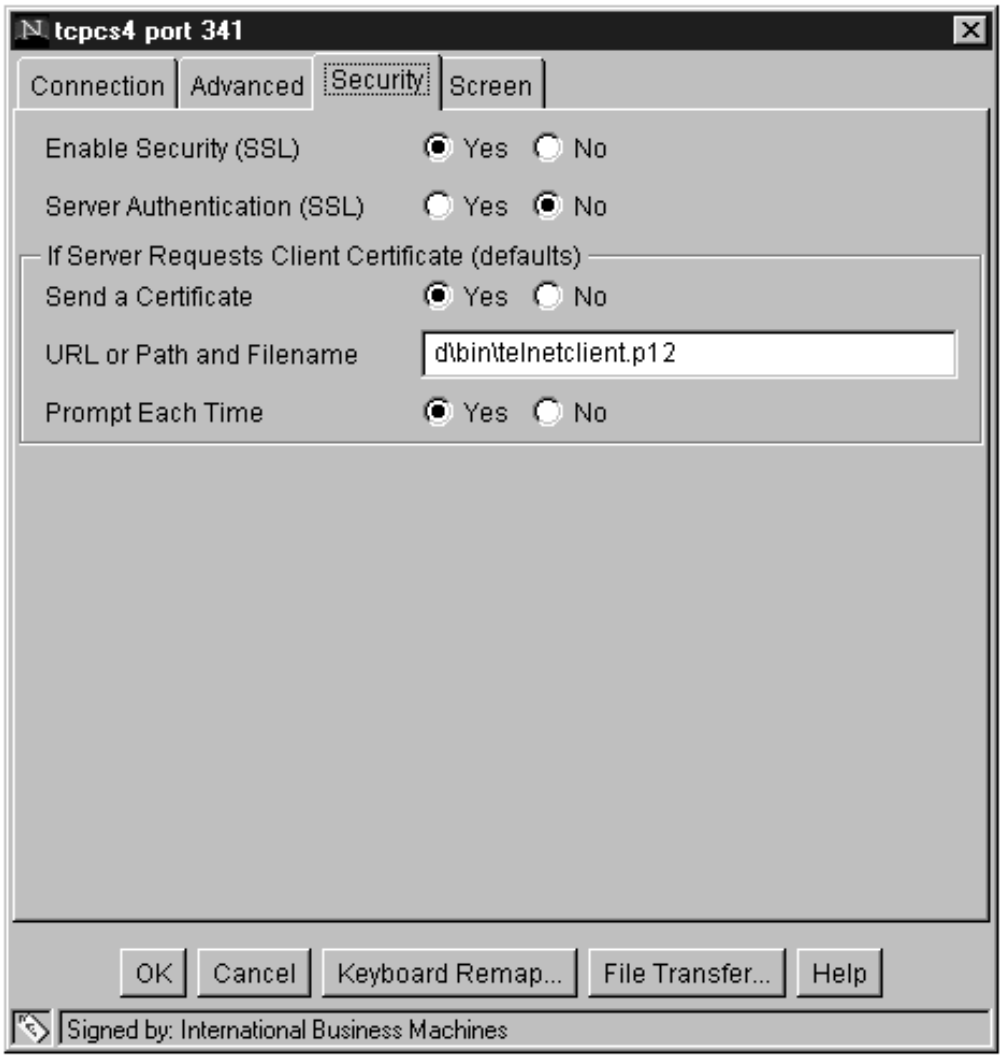


Figure 127. HOD security properties

Add server certificates to the client key ring

1. Get the server certificate information to the client machine.
   - Extract the server certificate from the server key ring. FTP can be used to ship the server certificate file to the client.
   - Newer versions of HOD allow you to extract the server information directly from the HOD client window during connection setup and eliminate the need to FTP the server certificate to the clients. The following is an example of using this method. Once the server side has been configured for the secure port and the port is active:
     - Setup your hod client to connect into the secure port and try the connection.
     - If the connection fails with a 662 (indicating the 'server presented a certificate that was not trusted'), you do not have the certificate of the CA that issued the server certificate in your client's key ring.
       - From the client window, select communication from the action bar, then select security. Information for the server certificate should be displayed:

Figure 128. Security Information

- Select extract and indicate binary format and where to store the certificate, then click OK.

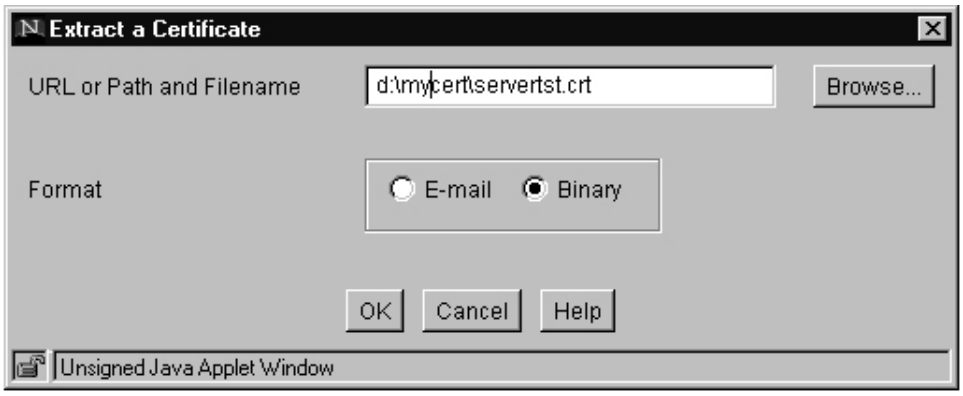


Figure 129. Extract a Certificate

If the action was successful, the following is displayed, which indicates that the server certificate information is on your client system.

Figure 130. Certificate was extracted

2. Add the self-signed server certificate to HOD's CustomizedCAs:
   - On the HOD client, go to HOD's Certificate Management panels. To do this, select Start, Programs, IBM Host On-Demand, Administration, Certificate Management.

- Open the CustomizedCAs.class file. If customized CA certificates have previously been added to HOD, or if this is the first customized CA, create a new class file by:
  - Selecting File, then New.
  - Click on the Key Database Type arrow and select SSLight key database class. This automatically fills in the required filename and path

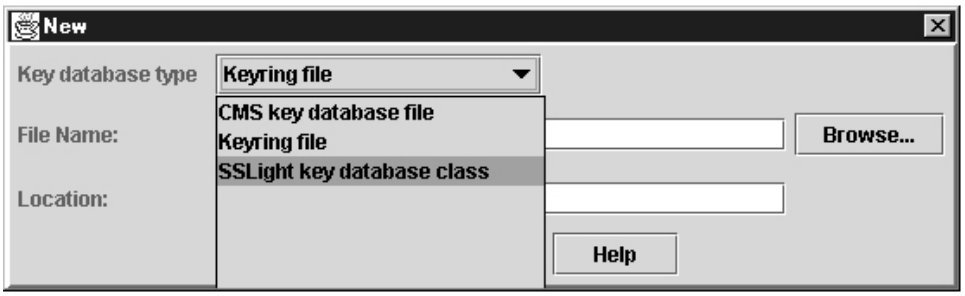


Figure 131. Creating a new CustomizedCAs.class

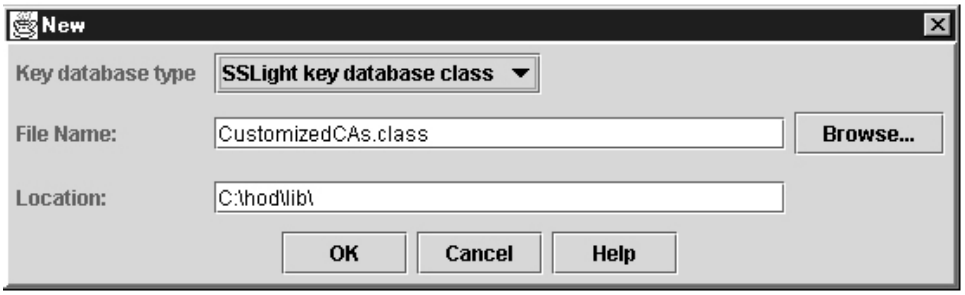


Figure 132. Default location displayed

  - Click OK. The *Signed Certificates* window is displayed.
  - Add the server certificate information:
    - Select ADD. The *Add CA's Certificate from a File* window is displayed.
    - Select data type 'Binary DER data'.

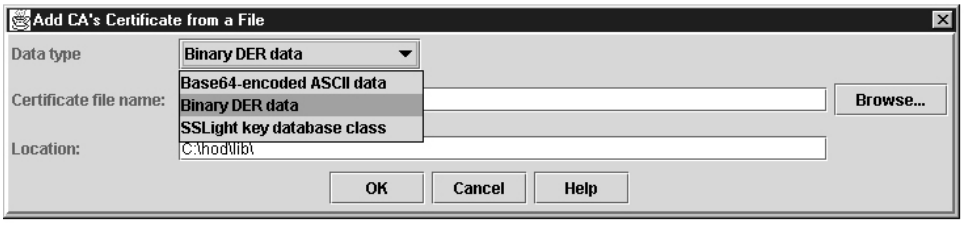


Figure 133. Add CA's Certificate From a File

  - Specify the path and name of the binary certificate file that was FTPed from the server or the file extracted using the HOD client window above. Click on the OK button to complete the add.

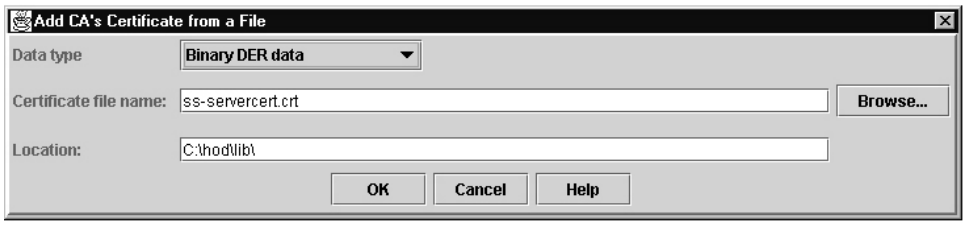


Figure 134. Add CA's Certificate From a File — continued

- Close the CustomizedCAs.class after completing the ADD.

3. Restart HOD to pick up the updated CustomizedCAs.class.

Appendix C. Express Logon Feature (ELF)

Users accessing SNA applications using TN3270 clients such as Host On-Demand (HOD) are generally required to know the user ID and password for the application they want to access. The ID-and-password authentication process creates several potential problems. For example, users may forget their IDs and passwords. If they do forget, the passwords must be reset by a system administrator, a time-consuming process. On the other hand, writing down the IDs and passwords or sharing them with someone else creates a security risk, especially since passwords are usually valid for relatively long periods of time.

IBM's solution to these problems is the Express Logon Feature (ELF), a process which allows a user on a workstation with a TN3270 client and an X.509 certificate to log on to a SNA application without entering an ID or password. The Express Logon Feature is supported on two-tier and three-tier network designs. The two-tier design utilizes the z/OS TN3270 Telnet server. The three-tier design utilizes a middle-tier TN3270 server and a Digital Certificate Access Server (DCAS).

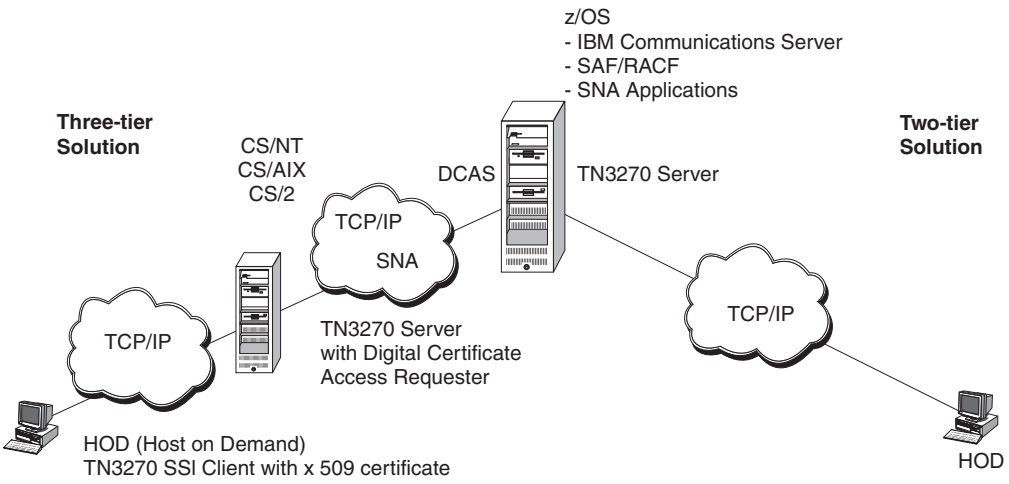


Figure 135. Express Logon network

Both network designs require a TN3270 client workstation that supports Secure Sockets Layer (SSL) connections with client authentication and an X.509 certificate. Using RACF services in z/OS, the client certificate must be associated with a valid user ID. The only client-side product that supports the Express Logon Feature is the IBM WebSphere Host On-Demand V5.0 and later releases.

The two-tier design requires the z/OS TN3270 Telnet server with SSL, client authentication, and Express Logon functions turned on. Refer to "Express Logon Feature (ELF)" on page 512 for server setup information.

The three-tier design requires a middle-tier TN3270 Telnet server and a Digital Certificate Access Server (DCAS). A middle-tier TN3270 server, so called because it does not reside on the host, but rather between the TN3270 workstation client and the host. This server includes a Digital Certificate Access Requester (DCAR). The middle-tier IBM TN3270 servers supporting Express Logon are:

- CS2 6.1

- CS/NT 6.1.1 PTF
- CS/AIX 6.0.0.1 PTF

Note: The term *DCAR* is used to describe the part of the TN3270 middle-tier server that supports the Express Logon Feature and communicates as a client with the DCAS. It is not separate from the TN3270 middle-tier server. The term *DCAR* might not be used in other documents that describe ELF but has been used here to simplify the description of this function.

A Digital Certificate Access Server (DCAS) resides on the host. DCAS uses RACF services to obtain a user ID.

The host also provides RACF Secured Signon services, which the DCAS or the MVS host Telnet server use to generate a *PassTicket*. A PassTicket is a RACF token similar to a password except that it is valid only for ten minutes.

In a typical scenario, a HOD client wants to log on to a TSO application on the host.

- In the two-tier design, the user starts an SSL connection with level 2 client authentication which passes the client certificate to the MVS host TN3270 server. The MVS host TN3270 server uses RACF Secured Signon services to obtain a user ID and PassTicket.
- In the three-tier design, the user starts the TN3270 connection to the middle-tier server. The DCAS's client is the middle-tier TN3270 server or DCAR, which attempts to log on to an SNA application for the workstation client. The DCAS receives a digital certificate from the DCAR and returns a user ID and PassTicket. SSL communication is used between the DCAS and the DCAR. The server recognizes that the client wants the Express Logon function and invokes the DCAR, which opens an SSL connection with client authentication and passes the workstation's certificate and application name to the DCAS on the host. The DCAS uses RACF Secured Signon services to obtain a user ID and PassTicket, which the DCAS returns to the DCAR. The DCAR passes this information back to the TN3270 server.

In both cases the ELF-enabled client and server now have enough information to complete the logon to TSO. This occurs without the user ever having to enter a user ID or password.

Note: You can use RACF or any other SAF-compliant security product that supports PassTickets with Express Logon. RACF APAR OW44393 is required when using the following:

- TSO with generic resources and PTKTDATA class profiles.
- Applications with shared user IDs that could access the application simultaneously. RACF requires the PTKTDATA profile to specify APPLDATA('NO REPLAY PROTECTION.')

For a final list of PTFs, refer to the HOD README file.

Configuring RACF services for Express Logon

At a minimum, you must register all workstation client certificates with RACF using the RACDCERT command. This associates the certificates with the IDs of users attempting to log on. In the two-tier solution, the certificate is passed from the client to the TN3270 server. In the three-tier solution, the certificate is passed from the client to the middle-tier TN3270 server, then to the DCAR, and then to the DCAS.

You must also create a RACF PTKTDATA profile for each application ID the end user is attempting to access. The PTKTDATA profile allows the DCAS or z/OS TN3270 server to obtain a PassTicket and user ID for the application. In the three-tier solution, the DCAS must pass the passticket and user ID back to the DCAR. For HOD, the application ID part of the profile name must be the same as that configured in the HOD Express Logon Application ID popup window. In most cases, the application name with which the user logs on will match the application ID portion of the RACF PTKTDATA class profile. However, for TSO and some other applications, the names and IDs may not match:

- If VTAM generic resources are used for TSO, define the application name portion on the RACF profile using the TCASGNAM defined in the TSOKEYxx, SYS1.PARMLIB member.
- If VTAM generic resources are not used, define the application name on the RACF profile as TSO.
- When configuring for TSO application logon, use the format TSO<SID> in the PassTicket profile, where SID is the SMF system ID defined in the SMFPRMxx member of SYS1.PARMLIB. (For example, if the SID is 3390, you would type TSO3390 in the profile.) For details, refer to the *z/OS Security Server RACF Security Administrator's Guide*.

For applications that allow shared user IDs (multiple users request access to the application simultaneously with the same user ID), you must specify the APPLDATA('NO REPLAY PROTECTION') option on the RDEFINE command in the PTKTDATA profile. This bypasses the default RACF protection against replay of PassTickets.

Configuring the Express Logon components

The following describes, in general terms, how to set up and configure currently supported Express Logon components:

- HOD TN3270 client
- z/OS TN3270 server
- Middle-tier TN3270 server (CS/2 V6.1, CS/NT 6.1.1, and CS/AIX 6.0.0.1)
- DCAS

For details on configuring each product, refer to the appropriate documentation for that product.

Configuring the HOD TN3270 client

To setup and configure the HOD client, follow these steps:

1. For each application to which the end user will log on, create a macro to record the logon screen of the application.

   The end user plays this macro when displaying the logon screen on the workstation.

2. Enter the application ID in the application ID popup window.

   The application ID must be the same name specified on the z/OS for the application ID portion of the PTKTDATA profile. The ID in the profile in most cases must be the same as the application name.
3. Use the HOD key-management utility to:
   a. Create a key database (key ring).
   b. Create a certificate request or generate a self-signed certificate and associate the certificate with the key ring.
   c. Use FTP to transmit the middle-tier TN3270 certificate to the workstation and store the server certificate in the key database of the client.
   d. Use FTP to transmit the TN3270 HOD client certificate to the middle-tier server and store in the SSL key database of the server.
4. Use FTP to transmit the workstation certificate to an MVS data set on the z/OS host. Use the RACF Certificate Services RACDCERT command to associate the certificate with a valid user ID.

Configuring the z/OS TN3270 server

Express Logon Feature requires SSL with level 2 client authentication functionality at the server. Once that level of security is working, specify the EXPRESSLOGON parameter statement to enable ELF in the z/OS TN3270 server.

Configuring the middle-tier TN3270 server (CS/2 example)

The middle-tier server is a TN3270 server, such as CS/2 V6.1, that communicates with the HOD client using an SSL connection with client authentication. The middle-tier server DCAR also communicates with the DCAS on the host. The DCAS and DCAR communicate over a TCP/IP connection using SSL with client authentication.

To configure the TN3270 server, follow these steps:

1. Configure the NDF file for the Express Logon function and communication with the DCAS using the following command:

```
DEFINE_EXPRESS_LOGON_SUPPORT
        ENABLED(YES)
        DCAS_ID(9.25.55.182)
        DCAS_ID_TYPE(IP_ADDRESS)
        DCAS_PORT(8990)
```

2. Use the local key management utility to store the workstation client certificate and the DCAS certificate in the local key ring:
   a. Create a key database file.
   b. Create a certificate request or generate a self-signed certificate and associate the certificate with the key ring.
   c. Store the workstation client certificate and the DCAS certificate in the key ring of the server.
3. Use FTP to transmit the DCAR certificate to the z/OS host and use gskkyman or RACF Certificate Services to store the DCAR certificate in the DCAS key ring.

Configuring the Digital Certificate Access Server (DCAS)

Setting up DCAS

The DCAS must run from a user ID defined with a UID=0 and from an APF authorized library. The shipped executable resides in /usr/lpp/tcpip/sbin with the

sticky bit on. The DCAS uses the z/OS SSL product, shipped with the z/OS base element. The SSL library SYS1.SIEALNKE must be in the run-time STEPLIB.

The default port used by the DCAS is port 8990. If you want to ensure that no other application uses the same port, use the following statement in the TCPIP profile data set:

```
PORT
    8990 TCP DCAS
```

If you choose to run the DCAS from the z/OS UNIX shell, use the following statement:

```
PORT
   8990 TCP OMVS
```

It is recommended that you use port access controls.

Create a configuration file for the DCAS taking into account the following:

- Define the port DCAS will listen on or use the default 8990.
- Define the KEYRING or SAFKEYRING for SSL communication.
- Decide the type of CLIENTAUTH needed.

For details on configuring the DCAS, see *z/OS Communications Server: IP Configuration Reference*.

Define a user ID as superuser to OMVS services

The server requires the user ID from which it is started be defined to use OMVS services as a superuser. The OMVS(UID(0)) could be put on the ADDUSER command. However, if the user ID already exists the ADDUSER will fail and the user ID will not be altered to superuser. ALTUSER will set the user ID to superuser whether the user ID existed before or was just created by the ADDUSER command.

```
ADDUSER DCAS
ALTUSER DCAS   DFLTGRP(OMVS)  OMVS(UID(0)HOME('/'))
```

Give the user ID access to operator commands

If the server is started from an MVS procedure, it is required that the user ID have access to the appropriate server resources in the OPERCMDS class. Use the following commands to provide access:

```
RDEFINE OPERCMDS(MVS.SERVMGR.DCAS) UACC(NONE)
PERMIT MVS.SERVMGR.DCAS  CLASS(OPERCMDS) ACCESS(CONTROL) ID(DCAS)
SETROPTS RACLIST(OPERCMDS) REFRESH
```

Provide a RACF definition for MVS startup

If the server is started as an MVS procedure, use the following RACF definitions to define the server to RACF:

```
RDEFINE STARTED  DCAS.* STDATA(USER(DCAS))
SETROPTS RACLIST(STARTED)  REFRESH
```

Starting, stopping, and toggling DCAS

DCAS can be started as either a generic server without stack affinity or as a server with affinity to a specific TCP/IP stack.

1. Start the DCAS.

   You can start the DCAS from the z/OS UNIX shell or with an MVS started procedure using optional parameters for debugging, logging, and specifying the

configuration file. Refer to the *z/OS Communications Server: IP Configuration Reference* for information on parameters used to start DCAS.

You can start the DCAS automatically when the TCP/IP address space is started or from the z/OS UNIX shell:

- To start the DCAS automatically when the TCP/IP address space is started, specify DCAS on the AUTOLOG statement in the TCPIP profile data set as shown in the following example:

```
AUTOLOG
  DCAS
ENDAUTOLOG
```

  A sample start procedure for the DCAS is provided in SEZAINST(EZADCASP).
- To start the DCAS from the UNIX shell, use the **dcas** command with optional parameters. (Refer to *z/OS Communications Server: IP Configuration Reference*) Note that the DCAS must run as a background job.

When DCAS is started, it stores its process ID (pid) in a Hierarchical File System (HFS) file. The file name under which it is stored depends upon how you configure DCAS:

- If you configure the DCAS with TCP/IP stack affinity, the pid file is named /tmp/dcas.tcpipname.pid where tcpipname is the name of the TCP/IP stack for which DCAS has affinity.
- If you configure the DCAS without stack affinity, the process ID file is named /tmp/dcas.INET.pid.

You can stop the DCAS from the UNIX shell or from MVS:

- To stop DCAS from the UNIX shell, use the following command:

```
kill  -s SIGTERM  pid
```

- To stop the DCAS from MVS, use the following command:

```
P DCAS
```

If the DCAS was started without debugging, you can toggle it (turn it on and off):

- To toggle from the z/OS UNIX shell, use the following command:

```
kill -s SIGHUP pid
```

- To toggle for an MVS started procedure, use the MODIFY command:

```
F DCAS,x
```

  where *x* is any valid character.

DCAS and system SSL

The DCAS and the DCAR use SSL V3 to communicate. The SSL protocol begins with a handshake. Then, the DCAR authenticates the DCAS and vice versa. At this time, the DCAS and the DCAR also agree on how to encrypt and decrypt the data.

You can specify the cipher level used for encryption and decryption for each connection at the time DCAS is configured, using the V3CIPHER configuration keyword. Alternatively, you can set the cipher level dynamically when DCAS starts, based on the level of cipher installed on the system. To set the cipher level dynamically, do not specify the V3CIPHER keyword.

Refer to Appendix B, "TLS/SSL security," on page 1167 for information on creating and managing key rings and certificates using either:

- The gskkyman tool
- The RACDCERT command

Authenticating the DCAS and the DCAR: The type of security and authentication required will determine the way certificates are created and managed. The DCAS, in conjunction with SSL and RACF, supports several levels of authentication.

Authenticating the DCAS: DCAS authentication is always performed by the DCAR. Authentication requires that the DCAS has a private key and an associated X.509 digital certificate defined in a key ring.

Authenticating the DCAR: The DCAR is the client that interacts with the DCAS. Authenticating the DCAR involves additional levels of control in which the client must have a key database with a certificate. Depending on the control level, the certificate is authenticated by SSL and the DCAS using RACF services.

There are three levels of client authentication from which to choose. Refer to Appendix B, "TLS/SSL security," on page 1167 for details.

To configure DCAS for level 1 authentication, specify the CLIENTAUTH LOCAL1 keyword and value in the DCAS configuration file. Use the KEYRING or the SAFKEYRING keywords in the DCAS configuration file to specify the key ring used by the DCAS.

To configure DCAS for level 2 authentication, specify the CLIENTAUTH LOCAL2 keyword and value in the DCAS configuration file.

If CLIENTAUTH LOCAL2 is coded in the DCAS configuration file, at a minimum, you must use RACF to associate the DCAR certificate with a valid user ID. You can do this using the RACDCERT ADD command. The user ID could be the one associated with the DCAS itself or it could be any valid user ID. If you want additional checking, you must activate the SERVAUTH class and define an EZA.DCAS.cvtsysname profile with the user ID associated with the DCAR certificate to access the profile.

Appendix D. Using HCD

The information in this chapter shows examples of panels that are used to define IQD channels and devices for z/OS Communications Server using HCD.

1. Select processors

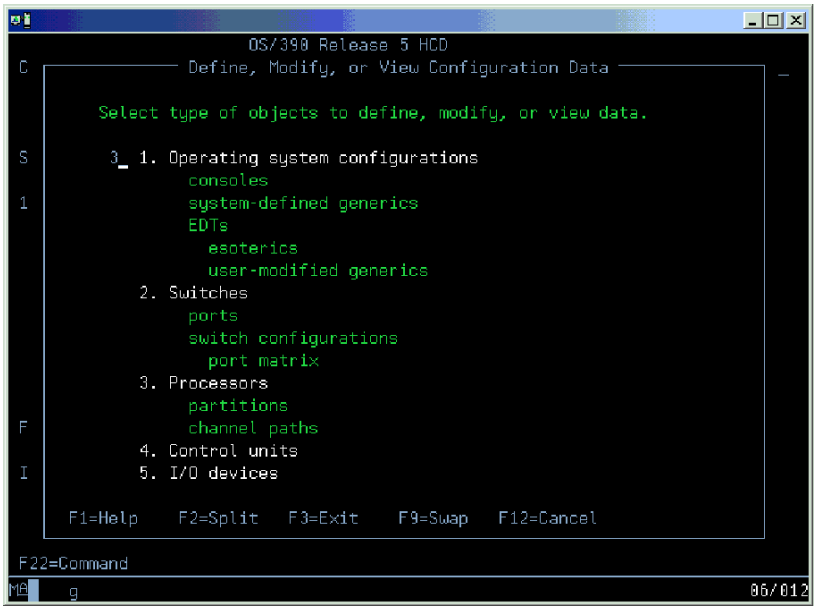


Figure 136. Select processors

2. Select 'S' Work with attached channel paths

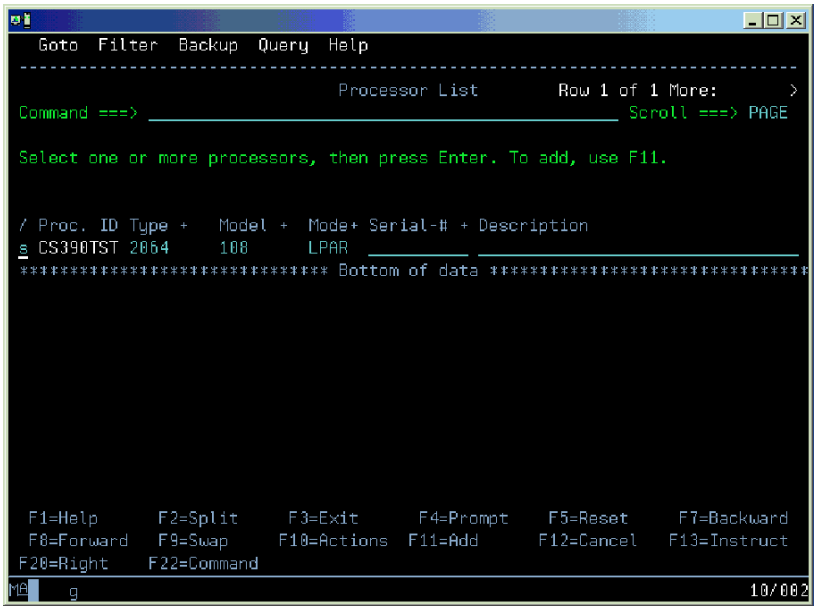


Figure 137. Work with attached channel paths

3. On the Channel Path List enter the Add command (or press F11) to initiate the Define Channel Path dialog.

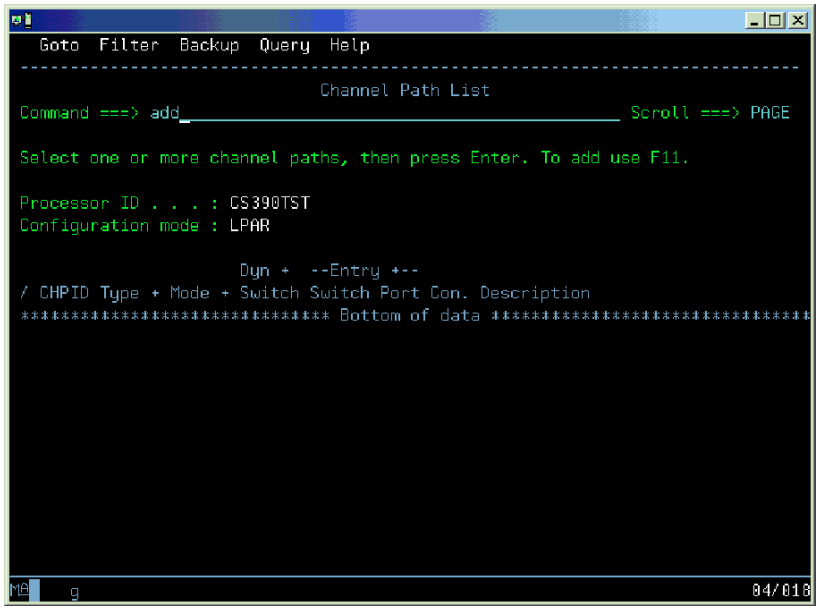


Figure 138. Initiate the Define Channel Path dialog

4. Fill in the Add Channel Path panel, then press Enter (Select SHR for Operation mode to share IQD Chpids across LPARs).

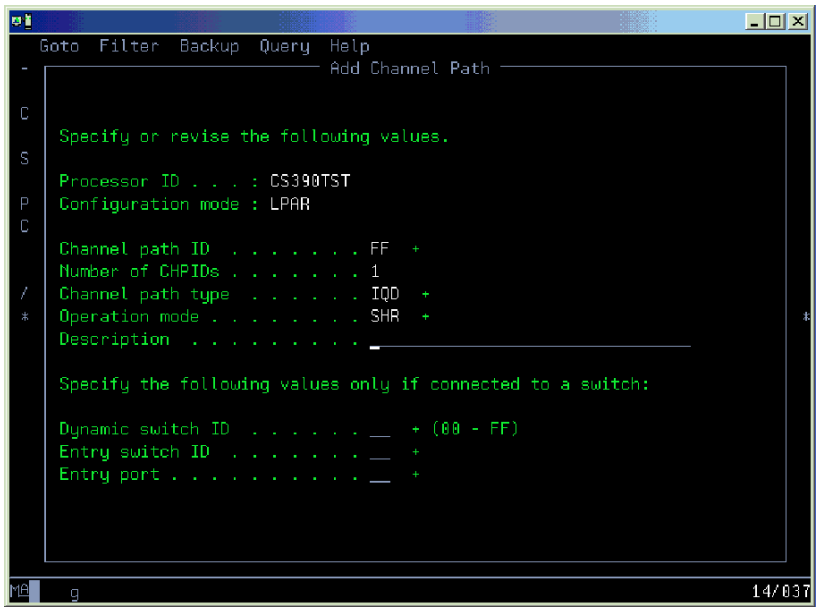


Figure 139. Add channel path

5. For an IQD channel path, the Specify Maximum Frame Size panel pops up with the default value of 16 KB.

Figure 140. Specify Maximum Frame Size

Or change the frame size to desired size:

Table 51. Frame size specification

Maximum Frame Size	TCP/IP MTU size
16K	8K
24K	16K
40K	32K
64K	56K

6. Define the channel path access list that each LPAR should have access to.

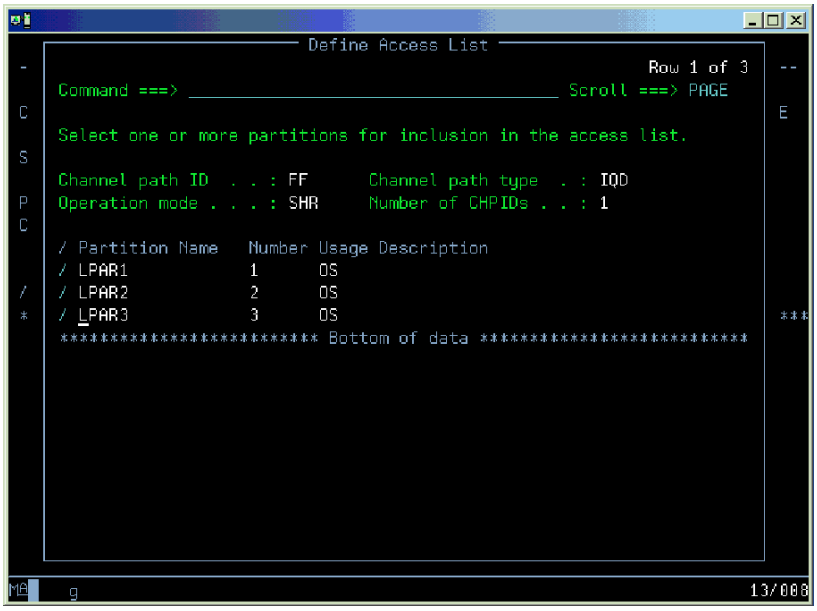


Figure 141. Define the channel path access list

7. Having pressed Enter, the Channel Path List is redisplayed with channel path number FF defined.

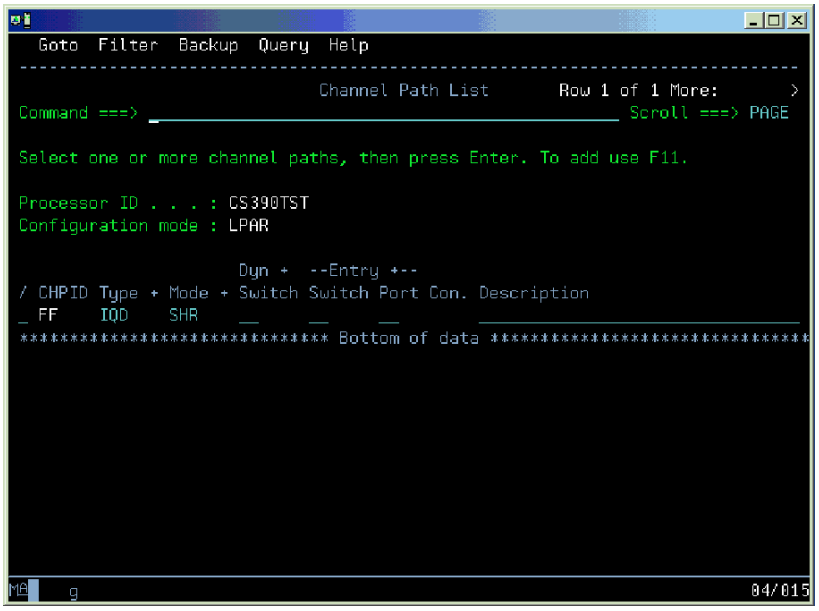


Figure 142. Channel path number FF defined

8. As the next step, add the control unit(s) to the IQD channel path. Select the defined channel path with action "Work with attached control units" (action code 'S').

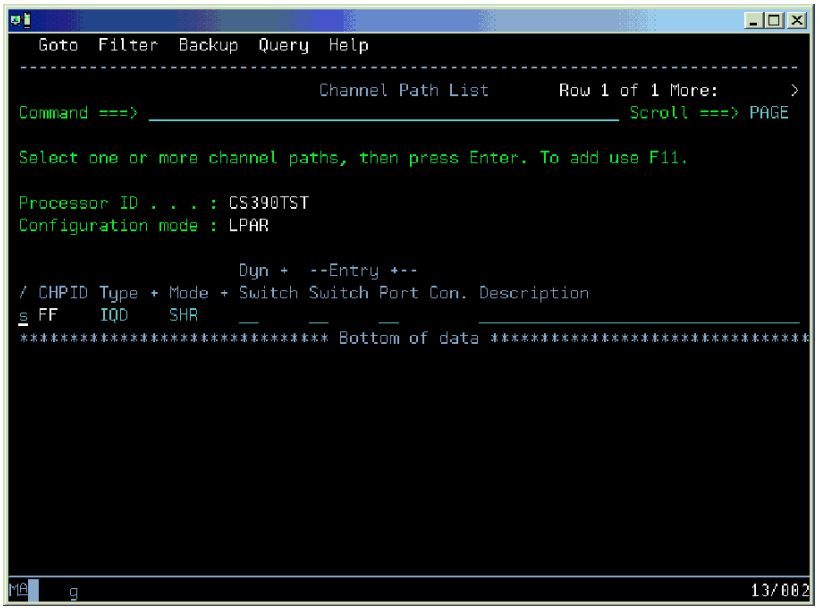


Figure 143. Work with attached control units

9. An empty control unit list is displayed. Enter the 'Add' command or F11.

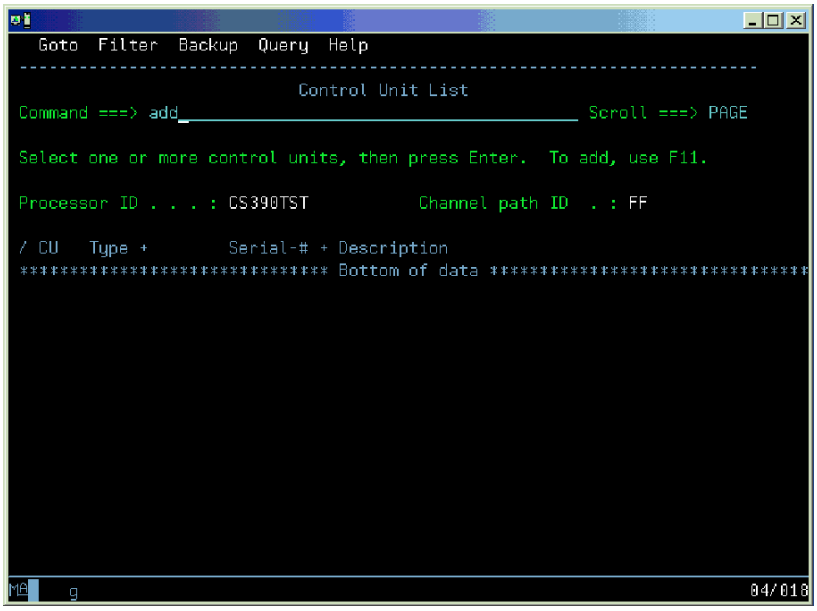


Figure 144. Add the control unit(s)

10. Define a control unit of type 'IQD' for channel path FF.

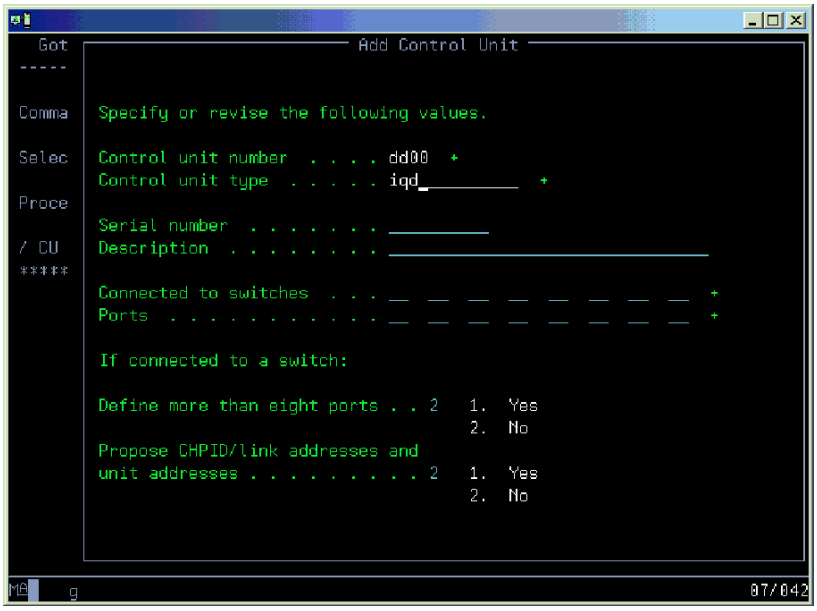


Figure 145. Define a control unit

11. Define it to the processor:

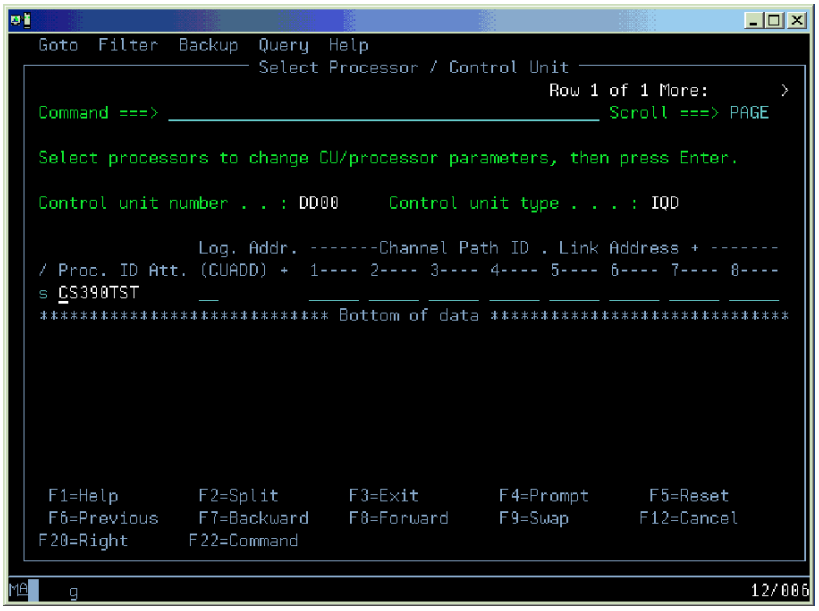


Figure 146. Define it to the processor

12. The processor settings are already preset. Pressing Enter, returns to the Select Processor/Control unit panel. Pressing Enter again, returns to the Control Unit List panel which shows the currently defined control unit.

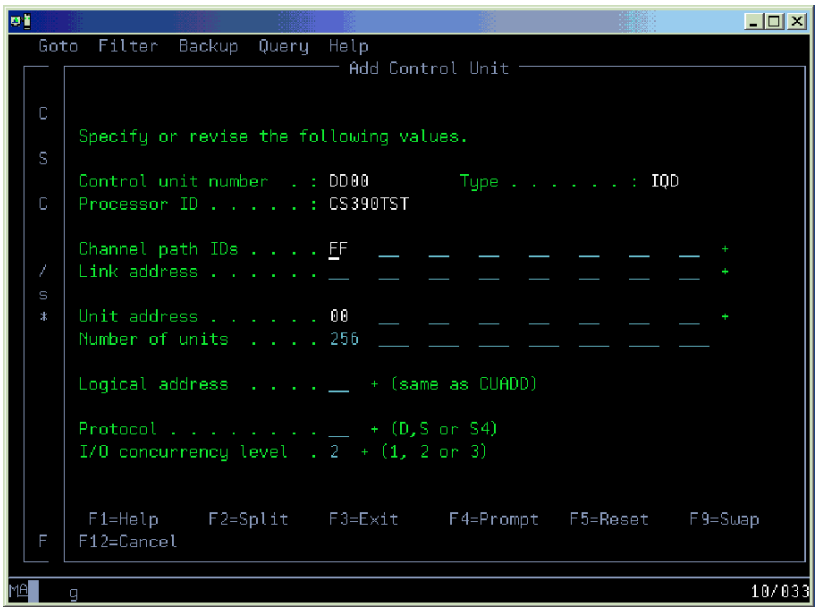


Figure 147. Currently defined control unit

13. Next, define the devices. Successively, select a control unit and perform action "Work with attached Devices".

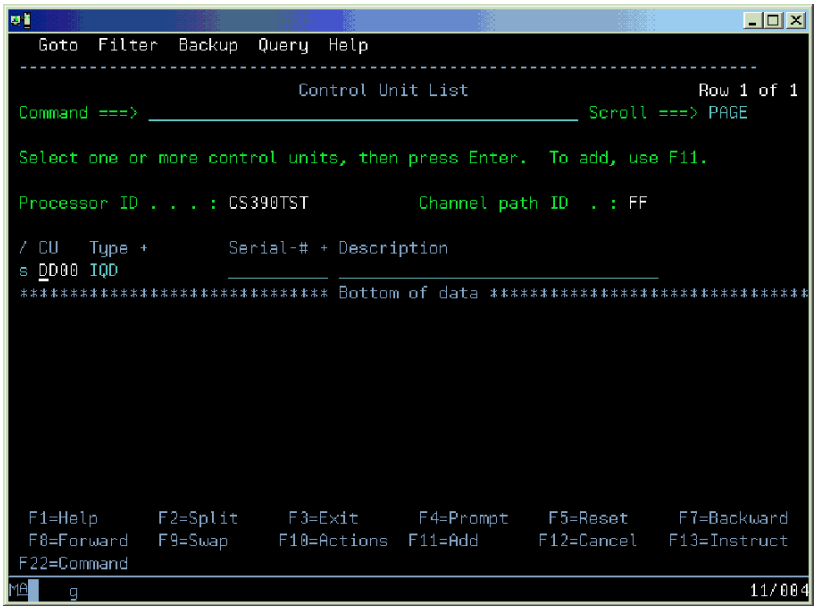


Figure 148. Define the devices

14. This leads to an empty device list.

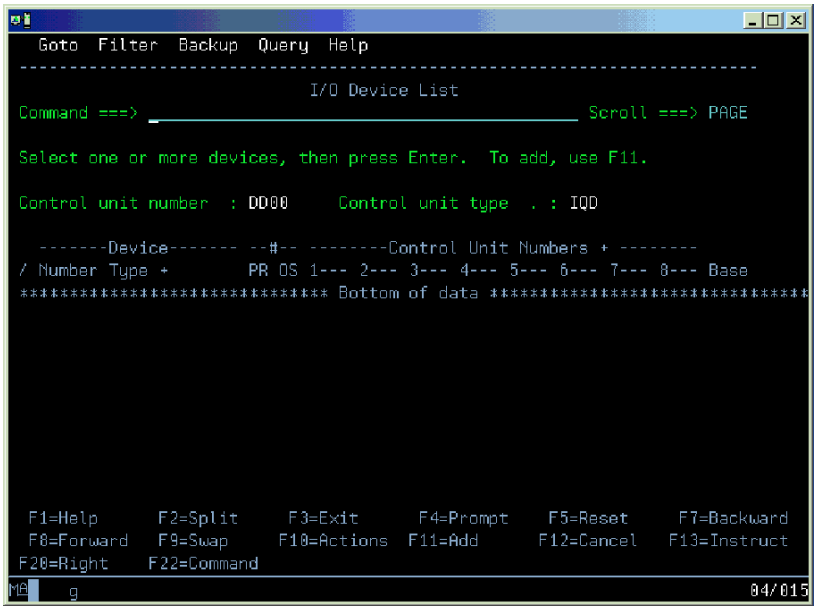


Figure 149. Empty device list

15. Perform the Add action to define the devices for the control unit selected in the previous step.

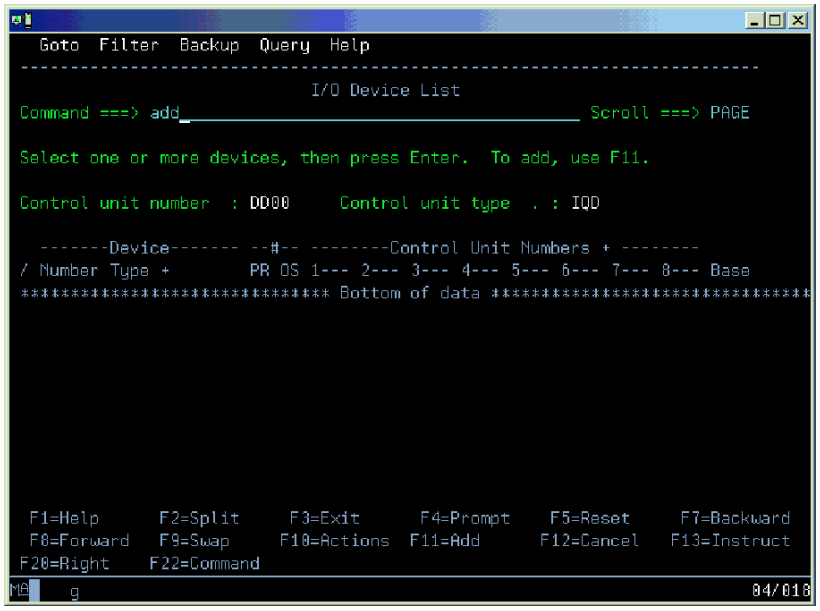


Figure 150. Define the devices for the control unit

16. Add devices of type IQD to the selected control unit.

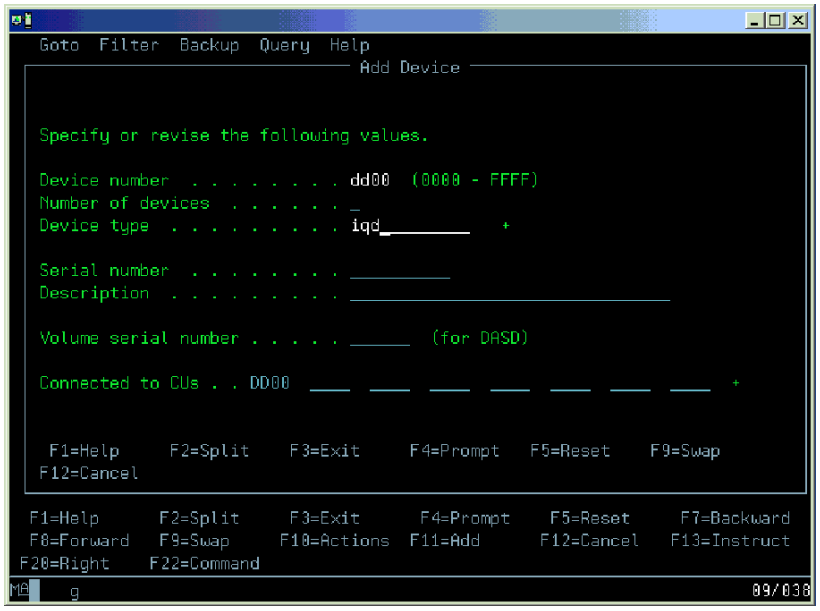


Figure 151. Add devices of type IQD

17. If the number of devices has been left unspecified (as in this example), HCD defines 10 devices.

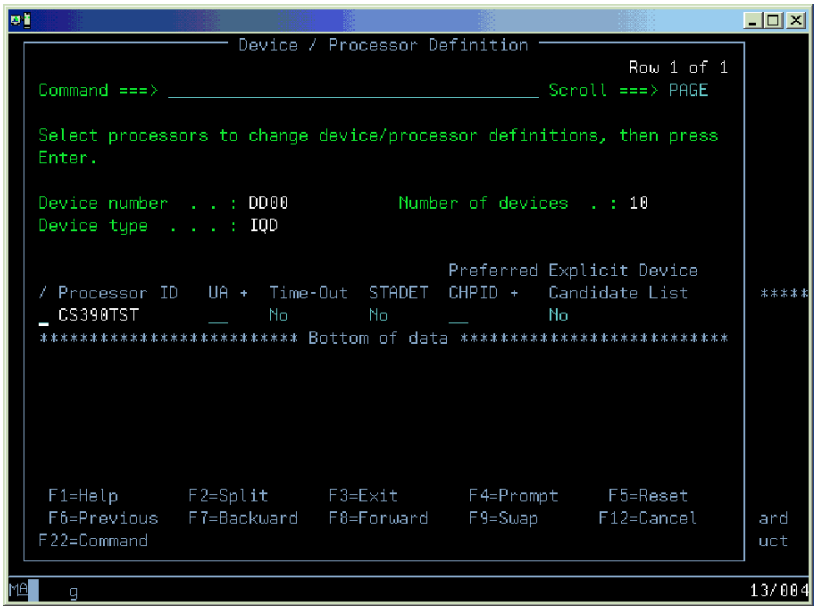


Figure 152. Define number of devices

18. Hit enter and the next panel displayed will be:

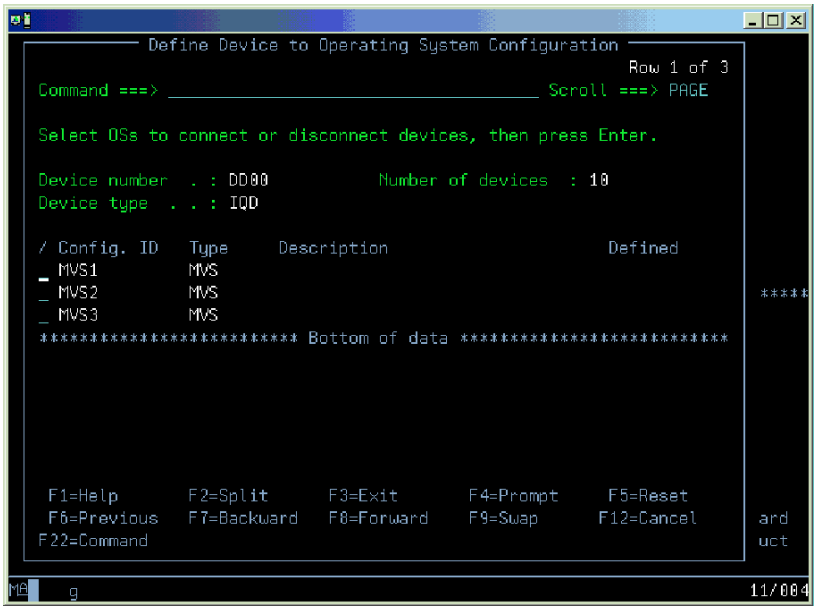


Figure 153. Define device to operating system

19. Next, define the devices to the operating system by selecting an 'S' on each system you want to have them defined on.

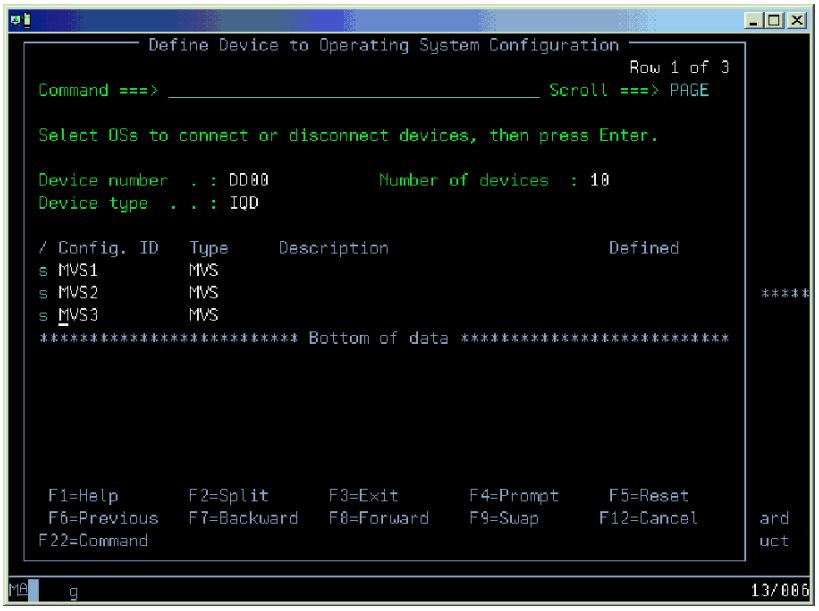


Figure 154. Select systems

20. The device parameters are shown with default values. Press Enter, to complete the definition for each system.

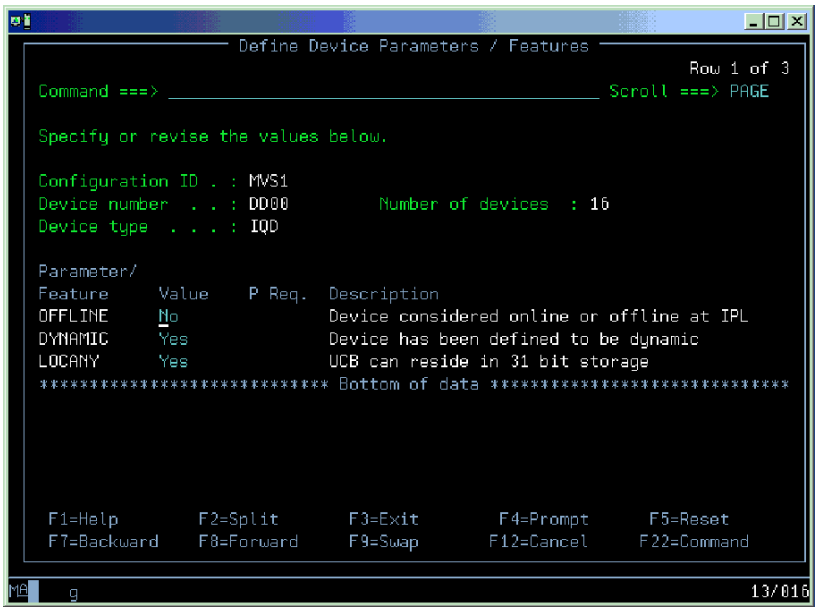


Figure 155. Complete the definition

21. Press Enter until you return to the I/O device list panel, the definition for channel path FF is complete.

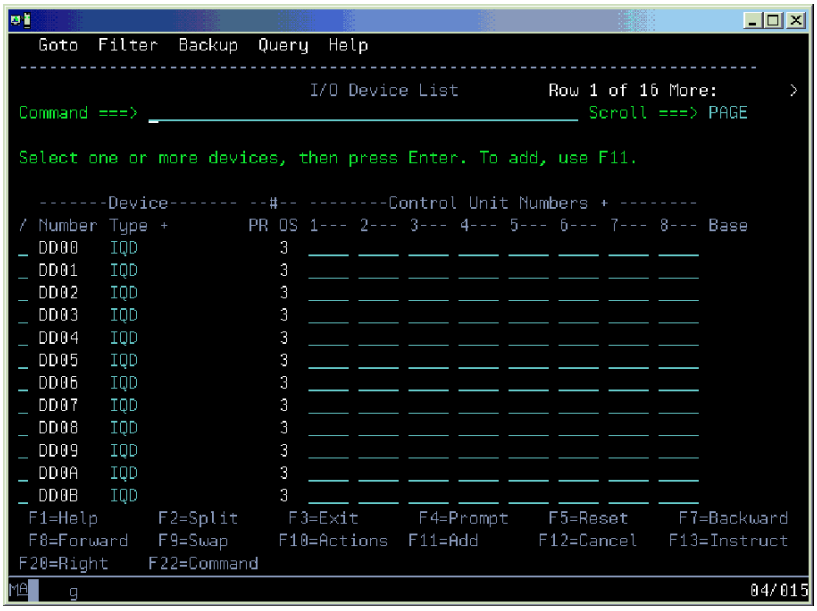


Figure 156. Definition completed

The sample IOCP input for this example would be:

```
----+----1----+----2----+----3----+----4----+----5----+----6----+----7--
         CHPID PATH=(FF),SHARED,                                        *
               PARTITION=((LPAR1,LPAR2,LPAR3),(LPAR1,LPAR2,LPAR3)),     *
               TYPE=IQD,OS=00
         CNTLUNIT CUNUMBR=DD00,PATH=(FF),UNIT=IQD
         IODEVICE ADDRESS=(DD00,010),CUNUMBR=(DD00),UNIT=IQD
```

Appendix E. Steps for preparing to run IP security

Perform the following steps to prepare to run the IKE daemon:

1. Set the appropriate UNIX System Services parameters.
2. Authorize the IKE daemon to the external security manager.
3. Authorize IP security to ICSF/MVS (optional).
4. Set up the IKE server for RSA signature mode authentication (optional).

These steps are described in detail in the following subsections.

Step 1: Setting appropriate UNIX System Services parameters

Verify that AF_UNIX and AF_INET are defined in the BPXPRMxx member of SYS1.PARMLIB. If the domains are not defined, refer to *z/OS UNIX System Services Planning* for the steps to customize the BPXPRMxx parmlib member.

Step 2: Authorizing the IKE daemon to the external security manager

To authorize the IKE daemon and the **ipsec** command to RACF, the following steps are necessary. The commands that are used are in EZARACF in the SEZAINST data set. For more information on **ipsec** command security and the SERVAUTH profile, refer to *z/OS Communications Server: IP System Administrator's Commands*.

Steps for authorizing the IKE daemon and ipsec command to RACF

Perform the following steps to authorize the IKE daemon and **ipsec** command to RACF:

1. Add user ID `IKED`, and add `IKED` to the STARTED class as follows:

```
ADDUSER  IKED     DFLTGRP(OMVSGRP)  OMVS(UID(0)  HOME('/'))
RDEFINE  STARTED  IKED.*            STDATA(USER(IKED))
PERMIT   BPX.DAEMON CLASS(FACILITY) ID(IKED)     ACCESS(READ)
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED)  REFRESH
```

2. Allow IKED to access SYS1.PARMLIB as follows:

```
PERMIT   SYS1.PARMLIB  ID(IKED)        ACCESS(READ)
```

3. Enable IKED to access certificates on an ESM key ring as follows:

```
RDEFINE  FACILITY IRR.DIGTCERT.LISTRING UACC(NONE)
RDEFINE  FACILITY IRR.DIGTCERT.LIST     UACC(NONE)
PERMIT   IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(IKED) ACCESS(READ)
PERMIT   IRR.DIGTCERT.LIST     CLASS(FACILITY) ID(IKED) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

4. Define access control for the **ipsec** command.

   The **ipsec** command uses both display and control features, and access to each feature can be controlled independently.

   - To control access to both the display and control capabilities of the **ipsec** command, issue the following commands:

```
SETROPTS GENERIC(SERVAUTH)
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.* UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.* CLASS(SERVAUTH) ID(userid) ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
```

- To specifically control access to the display capabilities of the **ipsec** command, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.DISPLAY UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.DISPLAY CLASS(SERVAUTH) ID(userid) ACCESS(READ)
```

- To specifically control access to the control capabilities of the **ipsec** command, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.CONTROL UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.CONTROL CLASS(SERVAUTH) ID(userid) ACCESS(READ)
```

To refresh the in-storage RACF profiles in the SERVAUTH class, issue the following command:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Step 3: Authorizing IP security to ICSF/MVS (optional)

IP security can take advantage of the encryption and decryption functions that are available on zSeries hardware in the following ways:

- Encrypting and decrypting TCP/IP packet data in an IPSec tunnel.
- Encrypting signature data included as part of IKE message flows. This encryption is performed only when RSA signature mode authentication is requested and the associated certificate was defined by the RACDCERT command with the ICSF keyword.

This support is provided by the combination of the Integrated Cryptographic Feature (ICRF) on the processor and the Integrated Cryptographic Service Facility/MVS (ICSF/MVS) software product. ICSF provides cryptography support through various cryptographic hardware features. The cryptographic features that are available to your applications depends on your processor or server model. For information about which features are available on your hardware, refer to the information about callable service support by hardware configuration in *z/OS Cryptographic Services ICSF Overview*.

To use this support, ICSF/MVS must be started and running. Preferably, start ICSF/MVS prior to starting TCP/IP. However, it can also be started when TCP/IP is active. For details on configuring ICSF, refer to *z/OS Cryptographic Services ICSF Administrator's Guide*.

When using a cryptographic coprocessor, the callable ICSF service names that are used by IPSec are as follows:

- CSFCKI
- CSFDEC1
- CSFENC1
- CSFRNG
- CSFCKM
- CSFOWH1

If triple DES hardware cryptographic support is not available on your zSeries processor, the CSFCKM service is not used.

When using CP Assist for Cryptographic Functions (CPACF), the following callable ICSF service names are used by IP security:

- CSNBSYE1
- CSNBSYD1

Access to some of the callable services that are provided by ICSF can be optionally controlled by RACF. This applies only to a z/OS system that is using a cryptographic coprocessor to encrypt and decrypt TCP/IP packet data in an IPSec tunnel. RACF access controls are not available for the CP Assist functions.

Requirement: If you plan to control access to the cryptographic coprocessor support, TCP/IP must be permitted to access the ICSF/MVS cryptographic services (CSFSERV).

Steps for setting up profiles in the CSFSERV resource class

Perform the following steps to set up profiles in the CSFSERV resource class:

1. Define the appropriate profiles in the CSFSERV class:

   ```
   RDEFINE  CSFSERV service-name  UACC(NONE)
   ```

2. Give TCP/IP access to the profiles:

   ```
   PERMIT service-name CLASS(CSFSERV) ID(stackname) ACCESS(READ)
   ```

3. For applications that run under a specific user's ID, such as oping, give the user's ID access to the profiles:

   ```
   PERMIT service-name CLASS(CSFSERV) ID (userid)
   ```

4. Activate the CSFSERV class and refresh the in-storage RACF profiles:

   ```
   SETROPTS CLASSACT(CSFSERV)
   SETROPTS RACLIST(CSFSERV) REFRESH
   ```

5. Set the MAXLEN ICSF/MVS installation option to 65535 or greater, because this is the maximum TCP/IP packet size. The MAXLEN installation option for hardware cryptography determines the maximum length that can be used to encrypt and decrypt data using ICSF/MVS.

Step 4: Setting up the IKE server for RSA signature mode authentication (optional)

The following steps are required if the IKE server is to use RSA signature mode for peer authentication in phase 1 negotiations.

Steps for setting up the IKE server for RSA signature mode authentication

Before you begin: Use the following references to understand the concepts that are involved in using digital certificates with RACF:

- For more information about RACF key rings and digital certificates , refer to *z/OS Security Server RACF Security Administrator's Guide*.
- For more information about controlling the use of the RACDCERT command, refer to *z/OS Security Server RACF Security Administrator's Guide*.
- For a complete description of the facilities and authorizations that are needed to create and modify digital certificates and key rings, refer to *z/OS Security Server RACF Command Language Reference*.

Perform the following steps to set up the IKE server for RSA signature mode authentication:

1. Define RACF facilities and access controls.
2. Define profiles to control access to the RACDCERT command.
3. Create a RACF key ring for the user ID under which IKED is to run.
4. Install an X509 digital certificate for the IKE server.

Step 1: Defining RACF facilities and access controls

To support RSA signature mode authentication in phase 1 negotiations, perform the following steps to give the IKE server the required access to a RACF key ring:

1. If they are not already defined, create the definitions that are required to allow certificates to be stored and accessed from the RACF database by issuing the following TSO commands:

   ```
   RDEFINE FACILITY IRR.DIGTCERT.LISTRING UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.LIST UACC(NONE)
   ```

2. To permit IKED to the facilities, issue the following TSO commands:

   ```
   PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(IKED) ACC(READ)
   PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(IKED) ACC(READ)
   ```

3. Refresh the FACILITY class:

   ```
   SETROPTS RACLIST(FACILITY) REFRESH
   ```

Step 2: Defining profiles to control access to the RACDCERT command

The RACF database provides digital certificate and key ring support through the RACDCERT command. The administrator who is responsible for managing the RACF key ring for IKED needs appropriate access to this command. Perform the following steps to define profiles to control access to this command:

1. If they are not already defined, create the definitions that are required to control access to the basic RACDCERT actions by issuing the following TSO commands:

   ```
   RDEFINE FACILITY IRR.DIGTCERT.ADD UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.ADDRING UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.CONNECT UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.GENCERT UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.GENREQ UACC(NONE)
   ```

2. Issue the following TSO commands (where *userid* is the ID of the person who will be executing the RACDCERT command to manage digital certificates):

   ```
   PERMIT IRR.DIGTCERT.ADD CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.ADDRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
   PERMIT IRR.DIGTCERT.CONNECT CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.GENCERT CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.GENREQ CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
   ```

3. Refresh the FACILITY class:

   ```
   SETROPTS RACLIST(FACILITY) REFRESH
   ```

Step 3: Creating a RACF key ring for the user ID under which IKED is to run

Digital certificates are made available to the IKE server by connecting them to a key ring that is owned by the IKE server. To create a key ring for the IKE server, issue the following TSO command:

```
RACDCERT ID(IKED) ADDRING(ikeyring)
```

The value used for *ikeyring* is case sensitive.

Step 4: Installing an X509 digital certificate for the IKE server

The IKE server requires the ability to retrieve digital certificates from a RACF key ring, each associated with a particular identity, and also to perform operations with the associated private key. IKED can own multiple certificates on its RACF key ring. You can install an X509 digital certificate in the following ways:

- Generate an X509 digital certificate for the IKE server and have it signed by a certificate authority.
- Generating a self-signed X509 digital certificate for the IKE server.
- Migrate an existing key database to a RACF key ring.

Steps for generating an X509 digital certificate for the IKE server and having it signed by a certificate authority

Before you begin: Assume that you have an X509 digital certificate that has the X500 distinguished name of `CN=IBM IKE Server,OU=Inventory,O=IBM,C=US` and a domain name of `ibm.com`. The certificate identifies the local IKE server that executes on z/OS with the user ID `IKED`.

Perform the following steps to install the X509 digital certificate:

1. Generate a self-signed certificate for the server. This certificate is associated with the user ID under which IKED is to run. (In all likelihood, this is the user `IKED`.):

   ```
   RACDCERT ID(IKED) GENCERT SUBJECTSDN(CN('IBM IKE Server') OU('Inventory') O('IBM') C('US'))
                     WITHLABEL('IKE Server') ALTNAME(DOMAIN('ibm.com'))
   ```

2. Do one of the following:
   - RACF supplies certificates for many commercial certificate authorities. If you are using one of these supplied certificate authority certificates, follow the steps for supplied digital certificates in *z/OS Security Server RACF Security Administrator's Guide*.
   - If the certificate authority you are using is not one of the supplied certificate authorities, obtain the root certificate of the certificate authority that is to sign IKED's certificate, and place it in an MVS data set (for example, USER1.EXTCA1.CERT). Add it to the RACF database as follows:

     ```
     RACDCERT ID(IKED) ADD('USER1.EXTCA1.CERT') WITHLABEL('External CA') CERTAUTH
     ```

3. Create a certificate request to send to the chosen certificate authority. The certificate request that you create is based on the certificate that was created in step 1. Place this certificate into a data set called USER1.IKED.GENREQ as follows:

   ```
   RACDCERT ID(IKED) GENREQ(LABEL('IKE Server')) DSN('USER1.IKED.GENREQ')
   ```

4. Send the certificate request to the certificate authority. The certificate request is in base 64-encoded text. Typically, the request is sent to the certificate authority by cutting and pasting the certificate request into an e-mail that is sent to the certificate authority.

The certificate authority validates the certificate. If the certificate is approved by the certificate authority, it is signed by the certificate authority and returned to the requestor.

5. Receive the returned certificate into a data set (for example, USER1.IKED.CERT). The returned certificate is in base 64-encoded text. This can be done by cutting and pasting, with FTP, or with another technique.
6. Replace the self-signed certificate with the certificate that is signed by the certificate authority. The certificate is replaced only if the user ID that is specified as the ID value on the RACDCERT ADD command is the same user ID that was specified when the certificate was created. Ensure that the user ID is the same. Otherwise, the certificate is added, rather than replacing the self-signed certificate, and does not contain the certificate's private key.

   ```
   RACDCERT ID(IKED) ADD('USER1.IKED.CERT') WITHLABEL('IKE Server')
   ```
7. Connect the certificate to IKED's existing key ring:

   ```
   RACDCERT ID(IKED) CONNECT(LABEL('IKE Server') RING(ikeyring) USAGE(PERSONAL))
   ```
8. Connect the certificate authority's certificate to the key ring:

   ```
   RACDCERT ID(IKED) CONNECT(CERTAUTH LABEL('External CA') RING(ikeyring) USAGE(CERTAUTH))
   ```

   This completes the certificate hierarchy from root to the IKED server.
9. Add the following statement to the IKE daemon configuration file, iked.conf:

   ```
   Keyring   IKED/ikeyring
   ```

You know you are done when the X509 digital certificate is available to the IKE server, and is mapped to the X500DN identity `CN=IBM IKE Server,OU=Inventory,O=IBM,C=US` from the certificate's subject name, and the FQDN identity `ibm.com` from the certificate's alternate subject name.

You can verify that the certificates that you have created are connected to the key ring associated with user ID IKED by using the RACDCERT command and examining the output of the `Ring Associations` field. Verify that the certificate authority was created and added to IKED/*ikeyring* as follows:

```
RACDCERT CERTAUTH LIST(LABEL('External CA')
```

Verify that the personal certificate for the IKE server was created and added to IKED/*ikeyring* as follows:

```
RACDCERT ID(IKED) LIST(LABEL('IKE Server'))
```

Steps for generating a self-signed X509 digital certificate for the IKE server

Before you begin: The certificate that is assigned to the secure server is a locally-signed certificate rather than one signed by a certificate authority. Assume that the local certificate authority has the distinguished name of `OU='Local Certificate Authority',O=IBM,C=US`.

Perform the following steps to implement a locally signed server certificate:

1. Generate a self-signed certificate to represent the local certificate authority:

   ```
   RACDCERT CERTAUTH GENCERT SUBJECTSDN(OU('Local Certificate Authority') O('IBM') C('US'))
   KEYUSAGE(CERTSIGN) WITHLABEL('IBM Local Certificate Authority')
   ```

   This certificate is used as the certificate authority certificate.
2. Export the certificate to a data set (in this case, USER1.LOCCERTA.CERT):

```
RACDCERT CERTAUTH EXPORT(LABEL('IBM Local Certificate Authority')) DSN('USER1.LOCCERTA.CERT')
```

3. Create a certificate for the server that is signed with the certificate authority certificate that was created in step 1:

```
RACDCERT ID(IKED) GENCERT SUBJECTSDN(CN('IBM IKE Server') OU('Inventory') O('IBM') C('US'))
WITHLABEL('IKE Server') ALTNAME(DOMAIN('ibm.com'))
SIGNWITH(CERTAUTH LABEL('IBM Local Certificate Authority'))
```

4. Connect the certificate to IKED's existing key ring:

```
RACDCERT ID(IKED) CONNECT(LABEL('IKE Server') RING(ikeyring) USAGE(PERSONAL))
```

5. Connect the local certificate authority certificate to the key ring:

```
RACDCERT ID(IKED) CONNECT(CERTAUTH LABEL('IBM Local Certificate Authority') RING(ikeyring) USAGE(CERTAUTH))
```

   This completes the certificate hierarchy from root to inventory server.

6. Add the following statement to the IKE daemon configuration file, iked.conf:

```
Keyring   IKED/ikeyring
```

You know you are done when the X509 digital certificate is available to the IKE server, and is mapped to the X500DN identity `CN=IBM IKE Server,OU=Inventory,O=IBM,C=US` from the certificate's subject name, and the FQDN identity `ibm.com` from the certificate's alternate subject name.

You can verify that the certificates that you have created are connected to the key ring associated with user ID IKED by using the RACDCERT command and examining the output of the `Ring Associations` field. Verify that the certificate authority was created and added to IKED/*ikeyring* as follows:

```
RACDCERT CERTAUTH LIST(LABEL('IBM Local Certificate Authority'))
```

Verify that the personal certificate for the IKE server was created and added to IKED/*ikeyring* as follows:

```
RACDCERT ID(IKED) LIST(LABEL('IKE Server'))
```

Steps for migrating an existing key database to a RACF key ring

Before you begin: To migrate an existing key database file to a RACF key ring, refer to the information on migrating key database files to RACF key rings in *z/OS Crytographic Service System Secure Sockets Layer Programming*.

Perform the following steps to migrate keys and certificates that are stored in an existing z/OS HFS key database into a RACF key ring:

1. Using gskkyman, export the certificate and private key to a password-protected PKCS#12 file. For details on copying a certificate with its private key, refer to *z/OS Crytographic Service System Secure Sockets Layer Programming*.
2. Copy the newly created PKCS#12 file to an MVS data set.
3. Use the RACDCERT command with the ADD operand to define a certificate and private key. The data set name that was created in step 2 contains the certificate.
4. Use the RACDCERT command with the ADDRING operand to create a new key ring in RACF.
5. Use the RACDCERT command with the CONNECT operand to add the certificate and private key to one or more existing RACF key rings.

Appendix F. Related protocol specifications (RFCs)

This appendix lists the related protocol specifications for TCP/IP. The Internet Protocol suite is still evolving through requests for comments (RFC). New protocols are being designed and implemented by researchers and are brought to the attention of the Internet community in the form of RFCs. Some of these protocols are so useful that they become recommended protocols. That is, all future implementations for TCP/IP are recommended to implement these particular functions or protocols. These become the *de facto* standards, on which the TCP/IP protocol suite is built.

You can request RFCs through electronic mail, from the automated Network Information Center (NIC) mail server, by sending a message to `service@nic.ddn.mil` with a subject line of `RFC` *`nnnn`* for text versions or a subject line of `RFC` *`nnnn`*`.PS` for PostScript versions. To request a copy of the RFC index, send a message with a subject line of `RFC INDEX`.

For more information, contact `nic@nic.ddn.mil` or at:

Government Systems, Inc.
Attn: Network Information Center
14200 Park Meadow Drive
Suite 200
Chantilly, VA 22021

Hard copies of all RFCs are available from the NIC, either individually or by subscription. Online copies are available at the following Web address: http://www.rfc-editor.org/rfc.html.

See "Internet drafts" on page 1236 for draft RFCs implemented in this and previous Communications Server releases.

Many features of TCP/IP Services are based on the following RFCs:

RFC	Title and Author
652	*Telnet output carriage-return disposition option* D. Crocker
653	*Telnet output horizontal tabstops option* D. Crocker
654	*Telnet output horizontal tab disposition option* D. Crocker
655	*Telnet output formfeed disposition option* D. Crocker
657	*Telnet output vertical tab disposition option* D. Crocker
658	*Telnet output linefeed disposition* D. Crocker
698	*Telnet extended ASCII option* T. Mock
726	*Remote Controlled Transmission and Echoing Telnet option* J. Postel, D. Crocker
727	*Telnet logout option* M.R. Crispin
732	*Telnet Data Entry Terminal option* J.D. Day
733	*Standard for the format of ARPA network text messages* D. Crocker, J. Vittal, K.T. Pogran, D.A. Henderson

734 *SUPDUP Protocol* M.R. Crispin

735 *Revised Telnet byte macro option* D. Crocker, R.H. Gumpertz

736 *Telnet SUPDUP option* M.R. Crispin

749 *Telnet SUPDUP—Output option* B. Greenberg

765 *File Transfer Protocol specification* J. Postel

768 *User Datagram Protocol* J. Postel

779 *Telnet send-location option* E. Killian

783 *TFTP Protocol (revision 2)* K.R. Sollins

791 *Internet Protocol* J. Postel

792 *Internet Control Message Protocol* J. Postel

793 *Transmission Control Protocol* J. Postel

820 *Assigned numbers* J. Postel

821 *Simple Mail Transfer Protocol* J. Postel

822 *Standard for the format of ARPA Internet text messages* D. Crocker

823 *DARPA Internet gateway* R. Hinden, A. Sheltzer

826 *Ethernet Address Resolution Protocol: Or converting network protocol addresses to 48.bit Ethernet address for transmission on Ethernet hardware* D. Plummer

854 *Telnet Protocol Specification* J. Postel, J. Reynolds

855 *Telnet Option Specification* J. Postel, J. Reynolds

856 *Telnet Binary Transmission* J. Postel, J. Reynolds

857 *Telnet Echo Option* J. Postel, J. Reynolds

858 *Telnet Suppress Go Ahead Option* J. Postel, J. Reynolds

859 *Telnet Status Option* J. Postel, J. Reynolds

860 *Telnet Timing Mark Option* J. Postel, J. Reynolds

861 *Telnet Extended Options: List Option* J. Postel, J. Reynolds

862 *Echo Protocol* J. Postel

863 *Discard Protocol* J. Postel

864 *Character Generator Protocol* J. Postel

865 *Quote of the Day Protocol* J. Postel

868 *Time Protocol* J. Postel, K. Harrenstien

877 *Standard for the transmission of IP datagrams over public data networks* J.T. Korb

883 *Domain names: Implementation specification* P.V. Mockapetris

884 *Telnet terminal type option* M. Solomon, E. Wimmers

885 *Telnet end of record option* J. Postel

894 *Standard for the transmission of IP datagrams over Ethernet networks* C. Hornig

896 *Congestion control in IP/TCP internetworks* J. Nagle

903 *Reverse Address Resolution Protocol* R. Finlayson, T. Mann, J. Mogul, M. Theimer

904 *Exterior Gateway Protocol formal specification* D. Mills

919 *Broadcasting Internet Datagrams* J. Mogul

922 *Broadcasting Internet datagrams in the presence of subnets* J. Mogul

927 *TACACS user identification Telnet option* B.A. Anderson

933 *Output marking Telnet option* S. Silverman

946 *Telnet terminal location number option* R. Nedved

950 *Internet Standard Subnetting Procedure* J. Mogul, J. Postel

951 *Bootstrap Protocol* W.J. Croft, J. Gilmore

952 *DoD Internet host table specification* K. Harrenstien, M. Stahl, E. Feinler

959 *File Transfer Protocol* J. Postel, J.K. Reynolds

961 *Official ARPA-Internet protocols* J.K. Reynolds, J. Postel

974 *Mail routing and the domain system* C. Partridge

1001 *Protocol standard for a NetBIOS service on a TCP/UDP transport: Concepts and methods* NetBios Working Group in the Defense Advanced Research Projects Agency, Internet Activities Board, End-to-End Services Task Force

1002 *Protocol Standard for a NetBIOS service on a TCP/UDP transport: Detailed specifications* NetBios Working Group in the Defense Advanced Research Projects Agency, Internet Activities Board, End-to-End Services Task Force

1006 *ISO transport services on top of the TCP: Version 3* M.T. Rose, D.E. Cass

1009 *Requirements for Internet gateways* R. Braden, J. Postel

1011 *Official Internet protocols* J. Reynolds, J. Postel

1013 *X Window System Protocol, version 11: Alpha update April 1987* R. Scheifler

1014 *XDR: External Data Representation standard* Sun Microsystems

1027 *Using ARP to implement transparent subnet gateways* S. Carl-Mitchell, J. Quarterman

1032 *Domain administrators guide* M. Stahl

1033 *Domain administrators operations guide* M. Lottor

1034 *Domain names—concepts and facilities* P.V. Mockapetris

1035 *Domain names—implementation and specification* P.V. Mockapetris

1038 *Draft revised IP security option* M. St. Johns

1041 *Telnet 3270 regime option* Y. Rekhter

1042 *Standard for the transmission of IP datagrams over IEEE 802 networks* J. Postel, J. Reynolds

1043 *Telnet Data Entry Terminal option: DODIIS implementation* A. Yasuda, T. Thompson

1044 *Internet Protocol on Network System's HYPERchannel: Protocol specification* K. Hardwick, J. Lekashman

1053 *Telnet X.3 PAD option* S. Levy, T. Jacobson

1055 *Nonstandard for transmission of IP datagrams over serial lines: SLIP* J. Romkey

1057 *RPC: Remote Procedure Call Protocol Specification: Version 2* Sun Microsystems

1058 *Routing Information Protocol* C. Hedrick

1060 *Assigned numbers* J. Reynolds, J. Postel

1067 *Simple Network Management Protocol* J.D. Case, M. Fedor, M.L. Schoffstall, J. Davin

1071 *Computing the Internet checksum* R.T. Braden, D.A. Borman, C. Partridge

1072 *TCP extensions for long-delay paths* V. Jacobson, R.T. Braden

1073 *Telnet window size option* D. Waitzman

1079 *Telnet terminal speed option* C. Hedrick

1085 *ISO presentation services on top of TCP/IP based internets* M.T. Rose

1091 *Telnet terminal-type option* J. VanBokkelen

1094 *NFS: Network File System Protocol specification* Sun Microsystems

1096 *Telnet X display location option* G. Marcy

1101 *DNS encoding of network names and other types* P. Mockapetris

1112 *Host extensions for IP multicasting* S.E. Deering

1113 *Privacy enhancement for Internet electronic mail: Part I — message encipherment and authentication procedures* J. Linn

1118 *Hitchhikers Guide to the Internet* E. Krol

1122 *Requirements for Internet Hosts—Communication Layers* R. Braden, Ed.

1123 *Requirements for Internet Hosts—Application and Support* R. Braden, Ed.

1146 *TCP alternate checksum options* J. Zweig, C. Partridge

1155 *Structure and identification of management information for TCP/IP-based internets* M. Rose, K. McCloghrie

1156 *Management Information Base for network management of TCP/IP-based internets* K. McCloghrie, M. Rose

1157 *Simple Network Management Protocol (SNMP)* J. Case, M. Fedor, M. Schoffstall, J. Davin

1158 *Management Information Base for network management of TCP/IP-based internets: MIB-II* M. Rose

1166 *Internet numbers* S. Kirkpatrick, M.K. Stahl, M. Recker

1179 *Line printer daemon protocol* L. McLaughlin

1180 *TCP/IP tutorial* T. Socolofsky, C. Kale

1183 *New DNS RR Definitions* C.F. Everhart, L.A. Mamakos, R. Ullmann, P.V. Mockapetris

1184 *Telnet Linemode Option* D. Borman

1186 *MD4 Message Digest Algorithm* R.L. Rivest

1187 *Bulk Table Retrieval with the SNMP* M. Rose, K. McCloghrie, J. Davin

1188 *Proposed Standard for the Transmission of IP Datagrams over FDDI Networks* D. Katz

1190 *Experimental Internet Stream Protocol: Version 2 (ST-II)* C. Topolcic

1191 *Path MTU discovery* J. Mogul, S. Deering

1198 *FYI on the X window system* R. Scheifler

1207 *FYI on Questions and Answers: Answers to commonly asked "experienced Internet user" questions* G. Malkin, A. Marine, J. Reynolds

1208 *Glossary of networking terms* O. Jacobsen, D. Lynch

1213 *Management Information Base for Network Management of TCP/IP-based internets: MIB-II* K. McCloghrie, M.T. Rose

1215 *Convention for defining traps for use with the SNMP* M. Rose

1227 *SNMP MUX protocol and MIB* M.T. Rose

1228 *SNMP-DPI: Simple Network Management Protocol Distributed Program Interface* G. Carpenter, B. Wijnen

1229 *Extensions to the generic-interface MIB* K. McCloghrie

1230 *IEEE 802.4 Token Bus MIB* K. McCloghrie, R. Fox

1231 *IEEE 802.5 Token Ring MIB* K. McCloghrie, R. Fox, E. Decker

1236 *IP to X.121 address mapping for DDN* L. Morales, P. Hasse

1256 *ICMP Router Discovery Messages* S. Deering, Ed.

1267 *Border Gateway Protocol 3 (BGP-3)* K. Lougheed, Y. Rekhter

1268 *Application of the Border Gateway Protocol in the Internet* Y. Rekhter, P. Gross

1269 *Definitions of Managed Objects for the Border Gateway Protocol: Version 3* S. Willis, J. Burruss

1270 *SNMP Communications Services* F. Kastenholz, ed.

1285 *FDDI Management Information Base* J. Case

1315 *Management Information Base for Frame Relay DTEs* C. Brown, F. Baker, C. Carvalho

1321 *The MD5 Message-Digest Algorithm* R. Rivest

1323 *TCP Extensions for High Performance* V. Jacobson, R. Braden, D. Borman

1325 *FYI on Questions and Answers: Answers to Commonly Asked "New Internet User" Questions* G. Malkin, A. Marine

1327 *Mapping between X.400 (1988)/ISO 10021 and RFC 822* S. Hardcastle-Kille

1340 *Assigned Numbers* J. Reynolds, J. Postel

1344 *Implications of MIME for Internet Mail Gateways* N. Bornstein

1349 *Type of Service in the Internet Protocol Suite* P. Almquist

1350 *The TFTP Protocol (Revision 2)* K.R. Sollins

1351 *SNMP Administrative Model* J. Davin, J. Galvin, K. McCloghrie

1352 *SNMP Security Protocols* J. Galvin, K. McCloghrie, J. Davin

1353 *Definitions of Managed Objects for Administration of SNMP Parties* K. McCloghrie, J. Davin, J. Galvin

1354 *IP Forwarding Table MIB* F. Baker

1356 *Multiprotocol Interconnect on X.25 and ISDN in the Packet Mode* A. Malis, D. Robinson, R. Ullmann

1358 *Charter of the Internet Architecture Board (IAB)* L. Chapin

1363 *A Proposed Flow Specification* C. Partridge

1368 *Definition of Managed Objects for IEEE 802.3 Repeater Devices* D. McMaster, K. McCloghrie

1372 *Telnet Remote Flow Control Option* C. L. Hedrick, D. Borman

1374 *IP and ARP on HIPPI* J. Renwick, A. Nicholson

1381 *SNMP MIB Extension for X.25 LAPB* D. Throop, F. Baker

1382 *SNMP MIB Extension for the X.25 Packet Layer* D. Throop

1387 *RIP Version 2 Protocol Analysis* G. Malkin

1388 *RIP Version 2 Carrying Additional Information* G. Malkin

1389 *RIP Version 2 MIB Extensions* G. Malkin, F. Baker

1390 *Transmission of IP and ARP over FDDI Networks* D. Katz

1393 *Traceroute Using an IP Option* G. Malkin

1398 *Definitions of Managed Objects for the Ethernet-Like Interface Types* F. Kastenholz

1408 *Telnet Environment Option* D. Borman, Ed.

1413 *Identification Protocol* M. St. Johns

1416 *Telnet Authentication Option* D. Borman, ed.

1420 *SNMP over IPX* S. Bostock

1428 *Transition of Internet Mail from Just-Send-8 to 8bit-SMTP/MIME* G. Vaudreuil

1442 *Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1443 *Textual Conventions for version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1445 *Administrative Model for version 2 of the Simple Network Management Protocol (SNMPv2)* J. Galvin, K. McCloghrie

1447 *Party MIB for version 2 of the Simple Network Management Protocol (SNMPv2)* K. McCloghrie, J. Galvin

1448 *Protocol Operations for version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1464 *Using the Domain Name System to Store Arbitrary String Attributes* R. Rosenbaum

1469 *IP Multicast over Token-Ring Local Area Networks* T. Pusateri

1483 *Multiprotocol Encapsulation over ATM Adaptation Layer 5* Juha Heinanen

1497 *BOOTP Vendor Information Extensions* J. Reynolds

1514 *Host Resources MIB* P. Grillo, S. Waldbusser

1516 *Definitions of Managed Objects for IEEE 802.3 Repeater Devices* D. McMaster, K. McCloghrie

1521 *MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies* N. Borenstein, N. Freed

1533 *DHCP Options and BOOTP Vendor Extensions* S. Alexander, R. Droms

1534 *Interoperation Between DHCP and BOOTP* R. Droms

1535 *A Security Problem and Proposed Correction With Widely Deployed DNS Software* E. Gavron

1536 *Common DNS Implementation Errors and Suggested Fixes* A. Kumar, J. Postel, C. Neuman, P. Danzig, S. Miller

1537 *Common DNS Data File Configuration Errors* P. Beertema

1540 *Internet Official Protocol Standards* J. Postel

1541 *Dynamic Host Configuration Protocol* R. Droms

1542 *Clarifications and Extensions for the Bootstrap Protocol* W. Wimer

1571 *Telnet Environment Option Interoperability Issues* D. Borman

1572 *Telnet Environment Option* S. Alexander

1573 *Evolution of the Interfaces Group of MIB-II* K. McCloghrie, F. Kastenholz

1577 *Classical IP and ARP over ATM* M. Laubach

1583 *OSPF Version 2* J. Moy

1591 *Domain Name System Structure and Delegation* J. Postel

1592 *Simple Network Management Protocol Distributed Protocol Interface Version 2.0* B. Wijnen, G. Carpenter, K. Curran, A. Sehgal, G. Waters

1594 *FYI on Questions and Answers— Answers to Commonly Asked "New Internet User" Questions* A. Marine, J. Reynolds, G. Malkin

1644 *T/TCP — TCP Extensions for Transactions Functional Specification* R. Braden

1646 *TN3270 Extensions for LUname and Printer Selection* C. Graves, T. Butts, M. Angel

1647 *TN3270 Enhancements* B. Kelly

1652 *SMTP Service Extension for 8bit-MIMEtransport* J. Klensin, N. Freed, M. Rose, E. Stefferud, D. Crocker

1664 *Using the Internet DNS to Distribute RFC1327 Mail Address Mapping Tables* C. Allochio, A. Bonito, B. Cole, S. Giordano, R. Hagens

1693 *An Extension to TCP: Partial Order Service* T. Connolly, P. Amer, P. Conrad

1695 *Definitions of Managed Objects for ATM Management Version 8.0 using SMIv2* M. Ahmed, K. Tesink

1701 *Generic Routing Encapsulation (GRE)* S. Hanks, T. Li, D. Farinacci, P. Traina

1702 *Generic Routing Encapsulation over IPv4 networks* S. Hanks, T. Li, D. Farinacci, P. Traina

1706 *DNS NSAP Resource Records* B. Manning, R. Colella

1712 *DNS Encoding of Geographical Location* C. Farrell, M. Schulze, S. Pleitner D. Baldoni

1713 *Tools for DNS debugging* A. Romao

1723 *RIP Version 2—Carrying Additional Information* G. Malkin

1752 *The Recommendation for the IP Next Generation Protocol* S. Bradner, A. Mankin

1766 *Tags for the Identification of Languages* H. Alvestrand

1771 *A Border Gateway Protocol 4 (BGP-4)* Y. Rekhter, T. Li

1794 *DNS Support for Load Balancing* T. Brisco

1819 *Internet Stream Protocol Version 2 (ST2) Protocol Specification—Version ST2+* L. Delgrossi, L. Berger Eds.

1826 *IP Authentication Header* R. Atkinson

1828 *IP Authentication using Keyed MD5* P. Metzger, W. Simpson

1829 *The ESP DES-CBC Transform* P. Karn, P. Metzger, W. Simpson

1830 *SMTP Service Extensions for Transmission of Large and Binary MIME Messages* G. Vaudreuil

1832 *XDR: External Data Representation Standard* R. Srinivasan

1850 *OSPF Version 2 Management Information Base* F. Baker, R. Coltun

1854 *SMTP Service Extension for Command Pipelining* N. Freed

1869 *SMTP Service Extensions* J. Klensin, N. Freed, M. Rose, E. Stefferud, D. Crocker

1870 *SMTP Service Extension for Message Size Declaration* J. Klensin, N. Freed, K. Moore

1876 *A Means for Expressing Location Information in the Domain Name System* C. Davis, P. Vixie, T. Goodwin, I. Dickinson

1883 *Internet Protocol, Version 6 (IPv6) Specification* S. Deering, R. Hinden

1884 *IP Version 6 Addressing Architecture* R. Hinden, S. Deering, Eds.

1886 *DNS Extensions to support IP version 6* S. Thomson, C. Huitema

1888 *OSI NSAPs and IPv6* J. Bound, B. Carpenter, D. Harrington, J. Houldsworth, A. Lloyd

1891 *SMTP Service Extension for Delivery Status Notifications* K. Moore

1892 *The Multipart/Report Content Type for the Reporting of Mail System Administrative Messages* G. Vaudreuil

1894 *An Extensible Message Format for Delivery Status Notifications*K. Moore, G. Vaudreuil

1901 *Introduction to Community-based SNMPv2* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1902 *Structure of Management Information for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1903 *Textual Conventions for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1904 *Conformance Statements for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1905 *Protocol Operations for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1906 *Transport Mappings for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1907 *Management Information Base for Version 2 of the Simple Network Management Protocol (SNMPv2)* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1908 *Coexistence between Version 1 and Version 2 of the Internet-standard Network Management Framework* J. Case, K. McCloghrie, M. Rose, S. Waldbusser

1912 *Common DNS Operational and Configuration Errors* D. Barr

1918 *Address Allocation for Private Internets* Y. Rekhter, B. Moskowitz, D. Karrenberg, G.J. de Groot, E. Lear

1928 *SOCKS Protocol Version 5* M. Leech, M. Ganis, Y. Lee, R. Kuris, D. Koblas, L. Jones

1930 *Guidelines for creation, selection, and registration of an Autonomous System (AS)* J. Hawkinson, T. Bates

1939 *Post Office Protocol-Version 3* J. Myers, M. Rose

1981 *Path MTU Discovery for IP version 6* J. McCann, S. Deering, J. Mogul

1982 *Serial Number Arithmetic* R. Elz, R. Bush

1985 *SMTP Service Extension for Remote Message Queue Starting* J. De Winter

1995 *Incremental Zone Transfer in DNS* M. Ohta

1996 *A Mechanism for Prompt Notification of Zone Changes (DNS NOTIFY)* P. Vixie

2010 *Operational Criteria for Root Name Servers* B. Manning, P. Vixie

2011 *SNMPv2 Management Information Base for the Internet Protocol using SMIv2* K. McCloghrie, Ed.

2012 *SNMPv2 Management Information Base for the Transmission Control Protocol using SMIv2* K. McCloghrie, Ed.

2013 *SNMPv2 Management Information Base for the User Datagram Protocol using SMIv2* K. McCloghrie, Ed.

2018 *TCP Selective Acknowledgement Options* M. Mathis, J. Mahdavi, S. Floyd, A. Romanow

2026 *The Internet Standards Process — Revision 3* S. Bradner

2030 *Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI* D. Mills

2033 *Local Mail Transfer Protocol* J. Myers

2034 *SMTP Service Extension for Returning Enhanced Error Codes*N. Freed

2040 *The RC5, RC5–CBC, RC-5–CBC-Pad, and RC5–CTS Algorithms*R. Baldwin, R. Rivest

2045 *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies* N. Freed, N. Borenstein

2052 *A DNS RR for specifying the location of services (DNS SRV)* A. Gulbrandsen, P. Vixie

2065 *Domain Name System Security Extensions* D. Eastlake 3rd, C. Kaufman

2066 *TELNET CHARSET Option* R. Gellens

2080 *RIPng for IPv6* G. Malkin, R. Minnear

2096 *IP Forwarding Table MIB* F. Baker

2104 *HMAC: Keyed-Hashing for Message Authentication* H. Krawczyk, M. Bellare, R. Canetti

2119 *Keywords for use in RFCs to Indicate Requirement Levels* S. Bradner

2132 *DHCP Options and BOOTP Vendor Extensions* S. Alexander, R. Droms

2133 *Basic Socket Interface Extensions for IPv6* R. Gilligan, S. Thomson, J. Bound, W. Stevens

2136 *Dynamic Updates in the Domain Name System (DNS UPDATE)* P. Vixie, Ed., S. Thomson, Y. Rekhter, J. Bound

2137 *Secure Domain Name System Dynamic Update* D. Eastlake 3rd

2163 *Using the Internet DNS to Distribute MIXER Conformant Global Address Mapping (MCGAM)* C. Allocchio

2168 *Resolution of Uniform Resource Identifiers using the Domain Name System* R. Daniel, M. Mealling

2178 *OSPF Version 2* J. Moy

2181 *Clarifications to the DNS Specification* R. Elz, R. Bush

2205 *Resource ReSerVation Protocol (RSVP)—Version 1 Functional Specification* R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, S. Jamin

2210 *The Use of RSVP with IETF Integrated Services* J. Wroclawski

2211 *Specification of the Controlled-Load Network Element Service* J. Wroclawski

2212 *Specification of Guaranteed Quality of Service* S. Shenker, C. Partridge, R. Guerin

2215 *General Characterization Parameters for Integrated Service Network Elements* S. Shenker, J. Wroclawski

2217 *Telnet Com Port Control Option* G. Clarke

2219 *Use of DNS Aliases for Network Services* M. Hamilton, R. Wright

2228 *FTP Security Extensions* M. Horowitz, S. Lunt

2230 *Key Exchange Delegation Record for the DNS* R. Atkinson

2233 *The Interfaces Group MIB using SMIv2* K. McCloghrie, F. Kastenholz

2240 *A Legal Basis for Domain Name Allocation* O. Vaughn

2246 *The TLS Protocol Version 1.0* T. Dierks, C. Allen

2251 *Lightweight Directory Access Protocol (v3)* M. Wahl, T. Howes, S. Kille

2253 *Lightweight Directory Access Protocol (v3): UTF-8 String Representation of Distinguished Names* M. Wahl, S. Kille, T. Howes

2254 *The String Representation of LDAP Search Filters* T. Howes

2261 *An Architecture for Describing SNMP Management Frameworks* D. Harrington, R. Presuhn, B. Wijnen

2262 *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)* J. Case, D. Harrington, R. Presuhn, B. Wijnen

2271 *An Architecture for Describing SNMP Management Frameworks* D. Harrington, R. Presuhn, B. Wijnen

2273 *SNMPv3 Applications* D. Levi, P. Meyer, B. Stewartz

2274 *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)* U. Blumenthal, B. Wijnen

2275 *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)* B. Wijnen, R. Presuhn, K. McCloghrie

2292 *Advanced Sockets API for IPv6* W. Stevens, M. Thomas

2308 *Negative Caching of DNS Queries (DNS NCACHE)* M. Andrews

2317 *Classless IN-ADDR.ARPA delegation* H. Eidnes, G. de Groot, P. Vixie

2320 *Definitions of Managed Objects for Classical IP and ARP Over ATM Using SMIv2 (IPOA-MIB)* M. Greene, J. Luciani, K. White, T. Kuo

2328 *OSPF Version 2* J. Moy

2345 *Domain Names and Company Name Retrieval* J. Klensin, T. Wolf, G. Oglesby

2352 *A Convention for Using Legal Names as Domain Names* O. Vaughn

2355 *TN3270 Enhancements* B. Kelly

2358 *Definitions of Managed Objects for the Ethernet-like Interface Types* J. Flick, J. Johnson

2373 *IP Version 6 Addressing Architecture* R. Hinden, S. Deering

2374 *An IPv6 Aggregatable Global Unicast Address Format* R. Hinden, M. O'Dell, S. Deering

2375 *IPv6 Multicast Address Assignments* R. Hinden, S. Deering

2385 *Protection of BGP Sessions via the TCP MD5 Signature Option*A. Hefferman

2389 *Feature negotiation mechanism for the File Transfer Protocol* P. Hethmon, R. Elz

2401 *Security Architecture for Internet Protocol* S. Kent, R. Atkinson

2402 *IP Authentication Header* S. Kent, R. Atkinson

2403 *The Use of HMAC-MD5–96 within ESP and AH* C. Madson, R. Glenn

2404 *The Use of HMAC-SHA–1–96 within ESP and AH* C. Madson, R. Glenn

2405 *The ESP DES-CBC Cipher Algorithm With Explicit IV* C. Madson, N. Doraswamy

2406 *IP Encapsulating Security Payload (ESP)* S. Kent, R. Atkinson

2407 *The Internet IP Security Domain of Interpretation for ISAKMP*D. Piper

2408 *Internet Security Association and Key Management Protocol (ISAKMP)* D. Maughan, M. Schertler, M. Schneider, J. Turner

2409 *The Internet Key Exchange (IKE)* D. Harkins, D. Carrel

2410 *The NULL Encryption Algorithm and Its Use With IPsec* R. Glenn, S. Kent,

2428 *FTP Extensions for IPv6 and NATs* M. Allman, S. Ostermann, C. Metz

2445 *Internet Calendaring and Scheduling Core Object Specification (iCalendar)* F. Dawson, D. Stenerson

2459 *Internet X.509 Public Key Infrastructure Certificate and CRL Profile* R. Housley, W. Ford, W. Polk, D. Solo

2460 *Internet Protocol, Version 6 (IPv6) Specification* S. Deering, R. Hinden

2461 *Neighbor Discovery for IP Version 6 (IPv6)* T. Narten, E. Nordmark, W. Simpson

2462 *IPv6 Stateless Address Autoconfiguration* S. Thomson, T. Narten

2463 *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification* A. Conta, S. Deering

2464 *Transmission of IPv6 Packets over Ethernet Networks* M. Crawford

2466 *Management Information Base for IP Version 6: ICMPv6 Group* D. Haskin, S. Onishi

2476 *Message Submission* R. Gellens, J. Klensin

2487 *SMTP Service Extension for Secure SMTP over TLS* P. Hoffman

2505 *Anti-Spam Recommendations for SMTP MTAs* G. Lindberg

2523 *Photuris: Extended Schemes and Attributes* P. Karn, W. Simpson

2535 *Domain Name System Security Extensions* D. Eastlake 3rd

2538 *Storing Certificates in the Domain Name System (DNS)* D. Eastlake 3rd, O. Gudmundsson

2539 *Storage of Diffie-Hellman Keys in the Domain Name System (DNS)* D. Eastlake 3rd

2540 *Detached Domain Name System (DNS) Information* D. Eastlake 3rd

2554 *SMTP Service Extension for Authentication* J. Myers

2570 *Introduction to Version 3 of the Internet-standard Network Management Framework* J. Case, R. Mundy, D. Partain, B. Stewart

2571 *An Architecture for Describing SNMP Management Frameworks* B. Wijnen, D. Harrington, R. Presuhn

2572 *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)* J. Case, D. Harrington, R. Presuhn, B. Wijnen

2573 *SNMP Applications* D. Levi, P. Meyer, B. Stewart

2574 *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)* U. Blumenthal, B. Wijnen

2575 *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)* B. Wijnen, R. Presuhn, K. McCloghrie

2576 *Co-Existence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework* R. Frye, D. Levi, S. Routhier, B. Wijnen

2578 *Structure of Management Information Version 2 (SMIv2)* K. McCloghrie, D. Perkins, J. Schoenwaelder

2579 *Textual Conventions for SMIv2* K. McCloghrie, D. Perkins, J. Schoenwaelder

2580 *Conformance Statements for SMIv2* K. McCloghrie, D. Perkins, J. Schoenwaelder

2581 *TCP Congestion Control* M. Allman, V. Paxson, W. Stevens

2583 *Guidelines for Next Hop Client (NHC) Developers* R. Carlson, L. Winkler

2591 *Definitions of Managed Objects for Scheduling Management Operations* D. Levi, J. Schoenwaelder

2625 *IP and ARP over Fibre Channel* M. Rajagopal, R. Bhagwat, W. Rickard

2635 *Don't SPEW A Set of Guidelines for Mass Unsolicited Mailings and Postings (spam*)* S. Hambridge, A. Lunde

2637 *Point-to-Point Tunneling Protocol* K. Hamzeh, G. Pall, W. Verthein, J. Taarud, W. Little, G. Zorn

2640 *Internationalization of the File Transfer Protocol* B. Curtin

2665 *Definitions of Managed Objects for the Ethernet-like Interface Types* J. Flick, J. Johnson

2671 *Extension Mechanisms for DNS (EDNS0)* P. Vixie

2672 *Non-Terminal DNS Name Redirection* M. Crawford

2675 *IPv6 Jumbograms* D. Borman, S. Deering, R. Hinden

2710 *Multicast Listener Discovery (MLD) for IPv6* S. Deering, W. Fenner, B. Haberman

2711 *IPv6 Router Alert Option* C. Partridge, A. Jackson

2740 *OSPF for IPv6* R. Coltun, D. Ferguson, J. Moy

2753 *A Framework for Policy-based Admission Control* R. Yavatkar, D. Pendarakis, R. Guerin

2758 *Definitions of Managed Objects for Service Level Agreements Performance Monitoring* K. White

2782 *A DNS RR for specifying the location of services (DNS SRV)* A. Gubrandsen, P. Vixix, L. Esibov

2821 *Simple Mail Transfer Protocol* J. Klensin, Ed.

2822 *Internet Message Format* P. Resnick, Ed.

2840 *TELNET KERMIT OPTION* J. Altman, F. da Cruz

2845 *Secret Key Transaction Authentication for DNS (TSIG)* P. Vixie, O. Gudmundsson, D. Eastlake 3rd, B. Wellington

2851 *Textual Conventions for Internet Network Addresses* M. Daniele, B. Haberman, S. Routhier, J. Schoenwaelder

2852 *Deliver By SMTP Service Extension* D. Newman

2874 *DNS Extensions to Support IPv6 Address Aggregation and Renumbering* M. Crawford, C. Huitema

2915 *The Naming Authority Pointer (NAPTR) DNS Resource Record* M. Mealling, R. Daniel

2920 *SMTP Service Extension for Command Pipelining* N. Freed

2930 *Secret Key Establishment for DNS (TKEY RR)* D. Eastlake, 3rd

2941 *Telnet Authentication Option* T. Ts'o, ed., J. Altman

2942 *Telnet Authentication: Kerberos Version 5* T. Ts'o

2946 *Telnet Data Encryption Option* T. Ts'o

2952 *Telnet Encryption: DES 64 bit Cipher Feedback* T. Ts'o

2953 *Telnet Encryption: DES 64 bit Output Feedback* T. Ts'o

2992 *Analysis of an Equal-Cost Multi-Path Algorithm* C. Hopps

3019 *IP Version 6 Management Information Base for The Multicast Listener Discovery Protocol* B. Haberman, R. Worzella

3060 *Policy Core Information Model—Version 1 Specification* B. Moore, E. Ellesson, J. Strassner, A. Westerinen

3152 *Delegation of IP6.ARPA* R. Bush

3291 *Textual Conventions for Internet Network Addresses* M. Daniele, B. Haberman, S. Routhier, J. Schoenwaelder

3363 *Representing Internet Protocol version 6 (IPv6) Addresses in the Domain Name System* R. Bush, A. Durand, B. Fink, O. Gudmundsson, T. Hain

3390 *Increasing TCP's Initial Window* M. Allman, S. Floyd, C. Partridge

3411 *An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks* D. Harrington, R. Presuhn, B. Wijnen

3412 *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)* J. Case, D. Harrington, R. Presuhn, B. Wijnen

3413 *Simple Network Management Protocol (SNMP) Applications* D. Levi, P. Meyer, B. Stewart

3414 *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)* U. Blumenthal, B. Wijnen

3415 *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)* B. Wijnen, R. Presuhn, K. McCloghrie

3419 *Textual Conventions for Transport Addresses* M. Daniele, J. Schoenwaelder

3484 *Default Address Selection for Internet Protocol version 6 (IPv6)* R. Draves

3493 *Basic Socket Interface Extensions for IPv6* R. Gilligan, S. Thomson, J. Bound, J. McCann, W. Stevens

3513 *Internet Protocol Version 6 (IPv6) Addressing Architecture* R. Hinden, S. Deering

3542 *Advanced Sockets Application Programming Interface (API) for IPv6* W. Richard Stevens, M. Thomas, E. Nordmark, T. Jinmei

3658 *Delegation Signer (DS) Resource Record (RR)* O. Gudmundsson

3715 *IPsec-Network Address Translation (NAT) Compatibility Requirements* B. Aboba, W. Dixon

3947 *Negotiation of NAT-Traversal in the IKE* T. Kivinen, B. Swander, A. Huttunen, V. Volpe

3948 *UDP Encapsulation of IPsec ESP Packets* A. Huttunen, B. Swander, V. Volpe, L. DiBurro, M. Stenberg

Internet drafts

Internet drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Other groups may also distribute working documents as Internet drafts. You can see Internet drafts at http://www.ietf.org/ID.html.

Several areas of IPv6 implementation include elements of the following Internet drafts and are subject to change during the RFC review process.

Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
A. Conta, S. Deering

Appendix G. Information APARs

This appendix lists information APARs for IP and SNA documents.

Notes:

1. Information APARs contain updates to previous editions of the manuals listed below. Documents updated for V1R7 are complete except for the updates contained in the information APARs that might be issued after V1R7 documents went to press.
2. Information APARs are predefined for z/OS V1R7 Communications Server and might not contain updates.
3. Information APARs for z/OS documents are in the document called *z/OS and z/OS.e DOC APAR and PTF ++HOLD Documentation*, which can be found at http://publibz.boulder.ibm.com:80/cgi-bin/bookmgr_OS390/BOOKS/ZIDOCMST/CCONTENTS.

Information APARs for IP documents

Table 52 lists information APARs for IP documents. For information APARs for V1R7, see http://www.ibm.com/support/docview.wss?uid=swg21178966.

Table 52. IP information APARs for z/OS Communications Server

Title	V1R6	V1R5	V1R4
New Function Summary (both IP and SNA)	II13824		
Quick Reference (both IP and SNA)	II13831		II13246
IP and SNA Codes	II13842		II13254
IP API Guide	II13844	II13577	II13255 II13790
IP CICS Sockets Guide		II13578	II13257
IP Configuration Guide	II13826	II13568	II13244 II13541 II13652 II13646
IP Configuration Reference	II13827	II13569 II13789	II13245 II13521 II13647 II13739
IP Diagnosis	II13836	II13571	II13249 II13493
IP Messages Volume 1	II13838	II13572	II13624 II13250
IP Messages Volume 2	II13839	II13573	II13251
IP Messages Volume 3	II13840	II13574	II13252
IP Messages Volume 4	II13841	II13575	II13253 II13628
IP Migration		II13566	II13242 II13738
IP Network and Application Design Guide	II13825	II13567	II13243

Table 52. IP information APARs for z/OS Communications Server (continued)

Title	V1R6	V1R5	V1R4
IP Network Print Facility			
IP Programmer's Reference	II13843	II13581	II13256
IP User's Guide and Commands	II13832	II13570	II13247
IP System Admin Commands	II13833	II13580	II13248 II13792

Information APARs for SNA documents

Table 53 lists information APARs for SNA documents. For information APARs for V1R7, see http://www.ibm.com/support/docview.wss?uid=swg21178966.

Table 53. SNA information APARs for z/OS Communications Server

Title	V1R6	V1R5	V1R4
New Function Summary (both IP and SNA)	II13824		
Quick Reference (both IP and SNA)	II13831		II13246
IP and SNA Codes	II13842		II13254
SNA Customization	II13857	II13560	II13240
SNA Diagnosis		II13558	II13236 II13735
SNA Diagnosis, Vol. 1: Techniques and Procedures	II13852		
SNA Diagnosis, Vol. 2: FFST Dumps and the VIT	II13853		
SNA Messages	II13854	II13559	II13238 II13736
SNA Network Implementation Guide	II13849	II13555	II13234 II13733
SNA Operation	II13851	II13557	II13237
SNA Migration		II13554	II13233 II13732
SNA Programming	II13858		II13241
SNA Resource Definition Reference	II13850	II13556	II13235 II13734
SNA Data Areas, Vol. 1 and 2			II13239
SNA Data Areas, 1	II13855		
SNA Data Areas, 2	II13856		

Other information APARs

Table 54 lists information APARs not related to documents.

Table 54. Non-document information APARs

Content	Number
Index to APARs that list recommended VTAM maintenance	II11220

Table 54. Non-document information APARs (continued)

Content	Number
Index to APARs that list trace and dump requests for VTAM problems	II13202
Index of Communication Server IP information APARs	II12028
MPC and CTC	II01501
Collecting TCPIP CTRACEs	II12014
CSM for VTAM	II13442
CSM for TCP/IP	II13951
DLUR/DLUS for z/OS V1R2, V1R4, and V1R5	II12986, II13456, and II13783
DOCUMENTATION REQUIRED FOR OSA/2, OSA EXPRESS AND OSA QDIO	II13016
DYNAMIC VIPA (BIND)	II13215
DNS — common problems and solutions	II13453
Enterprise Extender	II12223
FTPing doc to z/OS Support	II12030
FTP problems	II12079
Generic resources	II10986
HPR	II10953
iQDIO	II13142
LPR problems	II12022
MNPS	II10370
NCPROUTE problems	II12025
OMPROUTE	II12026
PASCAL API	II11814
Performance	II11710 II11711 II11712
Resolver	II13398 II13399 II13452
Socket API	II11996 II12020
SMTP problems	II12023
SNMP	II13477 II13478
SYSLOGD howto	II12021
TCPIP connection states	II12449
Telnet	II11574 II13135
TN3270 TELNET SSL common problems	II13369

Appendix H. Accessibility

Accessibility features help a user who has a physical disability, such as restricted mobility or limited vision, to use software products successfully. The major accessibility features in z/OS enable users to:

- Use assistive technologies such as screen readers and screen magnifier software
- Operate specific or equivalent features using only the keyboard
- Customize display attributes such as color, contrast, and font size

Using assistive technologies

Assistive technology products, such as screen readers, function with the user interfaces found in z/OS. Consult the assistive technology documentation for specific information when using such products to access z/OS interfaces.

Keyboard navigation of the user interface

Users can access z/OS user interfaces using TSO/E or ISPF. Refer to *z/OS TSO/E Primer*, *z/OS TSO/E User's Guide*, and *z/OS ISPF User's Guide Vol I* for information about accessing TSO/E and ISPF interfaces. These guides describe how to use TSO/E and ISPF, including the use of keyboard shortcuts or function keys (PF keys). Each guide includes the default settings for the PF keys and explains how to modify their functions.

z/OS information

z/OS information is accessible using screen readers with the BookServer/Library Server versions of z/OS books in the Internet library at:

`www.ibm.com/servers/eserver/zseries/zos/bkserv/`

Notices

IBM may not offer all of the products, services, or features discussed in this document. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:
INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

Site Counsel
IBM Corporation
P.O. Box 12195
3039 Cornwallis Road
Research Triangle Park, North Carolina 27709-2195
U.S.A

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this information and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application

programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Each copy or any portion of these sample programs or any derivative work must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

IBM is required to include the following statements in order to distribute portions of this document and the software described herein to which contributions have been made by The University of California. Portions herein © Copyright 1979, 1980, 1983, 1986, Regents of the University of California. Reproduced by permission. Portions herein were developed at the Electrical Engineering and Computer Sciences Department at the Berkeley campus of the University of California under the auspices of the Regents of the University of California.

Portions of this publication relating to RPC are Copyright © Sun Microsystems, Inc., 1988, 1989.

Some portions of this publication relating to X Window System** are Copyright © 1987, 1988 by Digital Equipment Corporation, Maynard, Massachusetts, and the Massachusetts Institute Of Technology, Cambridge, Massachusetts. All Rights Reserved.

Some portions of this publication relating to X Window System are Copyright © 1986, 1987, 1988 by Hewlett-Packard Corporation.

Permission to use, copy, modify, and distribute the M.I.T., Digital Equipment Corporation, and Hewlett-Packard Corporation portions of this software and its documentation for any purpose without fee is hereby granted, provided that the above copyright notice appears in all copies and that both that copyright notice and this permission notice appear in supporting documentation, and that the names of M.I.T., Digital, and Hewlett-Packard not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission. M.I.T., Digital, and Hewlett-Packard make no representation about the suitability of this software for any purpose. It is provided ″as is″ without express or implied warranty.

Copyright © 1983, 1995-1997 Eric P. Allman

Copyright © 1988, 1993 The Regents of the University of California. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. All advertising materials mentioning features or use of this software must display the following acknowledgement:

   ```
   This product includes software developed by the University of
   California, Berkeley and its contributors.
   ```

4. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software program contains code, and/or derivatives or modifications of code originating from the software program "Popper." Popper is Copyright ©1989-1991 The Regents of the University of California, All Rights Reserved. Popper was created by Austin Shelton, Information Systems and Technology, University of California, Berkeley.

Permission from the Regents of the University of California to use, copy, modify, and distribute the "Popper" software contained herein for any purpose, without fee, and without a written agreement is hereby granted, provided that the above copyright notice and this paragraph and the following two paragraphs appear in all copies. HOWEVER, ADDITIONAL PERMISSIONS MAY BE NECESSARY FROM OTHER PERSONS OR ENTITIES, TO USE DERIVATIVES OR MODIFICATIONS OF POPPER.

IN NO EVENT SHALL THE UNIVERSITY OF CALIFORNIA BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, INCLUDING LOST PROFITS, ARISING OUT OF THE USE OF THE POPPER SOFTWARE, OR ITS DERIVATIVES OR MODIFICATIONS, AND ITS DOCUMENTATION, EVEN IF THE UNIVERSITY OF CALIFORNIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

THE UNIVERSITY OF CALIFORNIA SPECIFICALLY DISCLAIMS ANY WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE POPPER SOFTWARE PROVIDED HEREUNDER IS ON AN "AS IS" BASIS, AND THE UNIVERSITY OF CALIFORNIA HAS NO OBLIGATIONS TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS.

Redistribution and use in source and binary forms are permitted provided that the above copyright notice and this paragraph are duplicated in all such forms and that any documentation, advertising materials, and other materials related to such distribution and use acknowledge that the software was developed by the University of California, Berkeley. The name of the University may not be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED ``AS IS'' AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. All advertising materials mentioning features or use of this software must display the following acknowledgement:

   `This product includes software developed by the University of California, Berkeley and its contributors.`
4. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Export of this software from the United States of America may require a specific license from the United States Government. It is the responsibility of any person or organization contemplating export to obtain such a license before exporting.

WITHIN THAT CONSTRAINT, permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation, and that the name of M.I.T. not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission. Furthermore

if you modify this software you must label your software as modified software and not distribute it in such a fashion that it might be confused with the original M.I.T. software. M.I.T. makes no representations about the suitability of this software for any purpose. It is provided ″as is″ without express or implied warranty.

Export of this software from the United States of America may require a specific license from the United States Government. It is the responsibility of any person or organization contemplating export to obtain such a license before exporting.

WITHIN THAT CONSTRAINT, permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation, and that the name of FundsXpress not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission. FundsXpress makes no representations about the suitability of this software for any purpose. It is provided ″as is″ without express or implied warranty.

THIS SOFTWARE IS PROVIDED ``AS IS'' AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

THE SOFTWARE IS PROVIDED ″AS IS″ AND INTERNET SOFTWARE CONSORTIUM DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS. IN NO EVENT SHALL INTERNET SOFTWARE CONSORTIUM BE LIABLE FOR ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

This package is an SSL implementation written by Eric Young (eay@cryptsoft.com). The implementation was written so as to conform with Netscape's SSL.

This library is free for commercial and non-commercial use as long as the following conditions are adhered to. The following conditions apply to all code found in this distribution, be it the RC4, RSA, lhash, DES, etc., code; not just the SSL code. The SSL documentation included with this distribution is covered by the same copyright terms except that the holder is Tim Hudson (tjh@cryptsoft.com).

Copyright remains Eric Young's, and as such any Copyright notices in the code are not to be removed. If this package is used in a product, Eric Young should be

given attribution as the author of the parts of the library used. This can be in the form of a textual message at program startup or in documentation (online or textual) provided with the package.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. All advertising materials mentioning features or use of this software must display the following acknowledgement: "This product includes cryptographic software written by Eric Young (eay@cryptsoft.com)". The word 'cryptographic' can be left out if the routines from the library being used are not cryptographic related.
4. If you include any Windows specific code (or a derivative thereof) from the apps directory (application code) you must include acknowledgement:

   "This product includes software written by Tim Hudson (tjh@cryptsoft.com)"

THIS SOFTWARE IS PROVIDED BY ERIC YOUNG ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

The license and distribution terms for any publicly available version or derivative of this code cannot be changed. i.e. this code cannot simply be copied and put under another distribution license [including the GNU Public License.]

This product includes cryptographic software written by Eric Young.

Copyright © 1999, 2000 Internet Software Consortium.

Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

THE SOFTWARE IS PROVIDED "AS IS" AND INTERNET SOFTWARE CONSORTIUM DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS. IN NO EVENT SHALL INTERNET SOFTWARE CONSORTIUM BE LIABLE FOR ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

If you are viewing this information softcopy, photographs and color illustrations may not appear.

You can obtain softcopy from the z/OS Collection (SK3T-4269), which contains BookManager and PDF formats of unlicensed books and the z/OS Licensed Product Library (LK3T-4307), which contains BookManager and PDF formats of licensed books.

Trademarks

The following terms are trademarks of the IBM Corporation in the United States or other countries or both:

Advanced Peer-to-Peer Networking
AFP
AD/Cycle
AIX
AIX/ESA
AnyNet
APL2
AS/400
AT
BookManager
BookMaster
C/370
CICS
CICS/ESA
C/MVS
Common User Access
C Set ++
CT
CUA
DB2
DFSMSdfp
DFSMShsm
DFSMS/MVS
DPI
Domino
DRDA
Enterprise Systems Architecture/370
ESCON
eServer
ES/3090
ES/9000
ES/9370
EtherStreamer
Extended Services
FFST
FFST/2
First Failure Support Technology
GDDM
IBM
IBMLink
IMS
IMS/ESA
HiperSockets
Language Environment
LANStreamer
Library Reader
LPDA
Micro Channel
Multiprise
MVS
MVS/DFP
MVS/ESA
MVS/SP
MVS/XA
NetView
Network Station
Nways
Notes
OfficeVision/MVS
OfficeVision/VM
Open Class
OS/2
OS/390
OS/400
Parallel Sysplex
PR/SM
PROFS
PS/2
RACF
Redbooks
Resource Link
RETAIN
RISC System/6000
RMF
RS/6000
S/370
S/390
S/390 Parallel Enterprise Server
SAA
SecureWay
SP
SP2
SQL/DS
System/360
System/370
System/390
SystemView
Tivoli
TURBOWAYS
VM/ESA
VSE/ESA
VTAM
WebSphere
XT
z/Architecture
z/OS
zSeries
z/VM
400
3090
3890

DB2 and NetView are registered trademarks of International Business Machines Corporation or Tivoli Systems Inc. in the U.S., other countries, or both.

The following terms are trademarks of other companies:

ATM is a trademark of Adobe Systems, Incorporated.

BSC is a trademark of BusiSoft Corporation.

CSA is a trademark of Canadian Standards Association.

DCE is a trademark of The Open Software Foundation.

HYPERchannel is a trademark of Network Systems Corporation.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

Other company, product, and service names may be trademarks or service marks of others.

Bibliography

z/OS Communications Server information

This section contains descriptions of the documents in the z/OS Communications Server library.

z/OS Communications Server documentation is available:

- Online at the z/OS Internet Library web page at http://www.ibm.com/servers/eserver/zseries/zos/bkserv
- In softcopy on CD-ROM collections. See "Softcopy information" on page xxvii.

z/OS Communications Server library

z/OS Communications Server documents are available on the CD-ROM accompanying z/OS (SK3T-4269 or SK3T-4307). Unlicensed documents can be viewed at the z/OS Internet library site.

Updates to documents are available on RETAIN® and in information APARs (info APARs). See Appendix G, "Information APARs," on page 1239 for a list of the documents and the info APARs associated with them.

Info APARs for z/OS documents are in the document called *z/OS and z/OS.e DOC APAR and PTF ++HOLD Documentation* which can be found at http://publibz.boulder.ibm.com:80/cgi-bin/bookmgr_OS390/ BOOKS/ZIDOCMST/CCONTENTS.

Planning

Title	Number	Description
z/OS Communications Server: New Function Summary	GC31-8771	This document is intended to help you plan for new IP for SNA function, whether you are migrating from a previous version or installing z/OS for the first time. It summarizes what is new in the release and identifies the suggested and required modifications needed to use the enhanced functions.
z/OS Communications Server: IPv6 Network and Application Design Guide	SC31-8885	This document is a high-level introduction to IPv6. It describes concepts of z/OS Communications Server's support of IPv6, coexistence with IPv4, and migration issues.

Resource definition, configuration, and tuning

Title	Number	Description
z/OS Communications Server: IP Configuration Guide	SC31-8775	This document describes the major concepts involved in understanding and configuring an IP network. Familiarity with the z/OS operating system, IP protocols, z/OS UNIX System Services, and IBM Time Sharing Option (TSO) is recommended. Use this document in conjunction with the *z/OS Communications Server: IP Configuration Reference*.

Title	Number	Description
z/OS Communications Server: IP Configuration Reference	SC31-8776	This document presents information for people who want to administer and maintain IP. Use this document in conjunction with the *z/OS Communications Server: IP Configuration Guide*. The information in this document includes: • TCP/IP configuration data sets • Configuration statements • Translation tables • SMF records • Protocol number and port assignments
z/OS Communications Server: SNA Network Implementation Guide	SC31-8777	This document presents the major concepts involved in implementing an SNA network. Use this document in conjunction with the *z/OS Communications Server: SNA Resource Definition Reference*.
z/OS Communications Server: SNA Resource Definition Reference	SC31-8778	This document describes each SNA definition statement, start option, and macroinstruction for user tables. It also describes NCP definition statements that affect SNA. Use this document in conjunction with the *z/OS Communications Server: SNA Network Implementation Guide*.
z/OS Communications Server: SNA Resource Definition Samples	SC31-8836	This document contains sample definitions to help you implement SNA functions in your networks, and includes sample major node definitions.
z/OS Communications Server: AnyNet SNA over TCP/IP	SC31-8832	This guide provides information to help you install, configure, use, and diagnose SNA over TCP/IP.
z/OS Communications Server: AnyNet Sockets over SNA	SC31-8831	This guide provides information to help you install, configure, use, and diagnose sockets over SNA. It also provides information to help you prepare application programs to use sockets over SNA.
z/OS Communications Server: IP Network Print Facility	SC31-8833	This document is for system programmers and network administrators who need to prepare their network to route SNA, JES2, or JES3 printer output to remote printers using TCP/IP Services.

Operation

Title	Number	Description
z/OS Communications Server: IP User's Guide and Commands	SC31-8780	This document describes how to use TCP/IP applications. It contains requests that allow a user to log on to a remote host using Telnet, transfer data sets using FTP, send and receive electronic mail, print on remote printers, and authenticate network users.
z/OS Communications Server: IP System Administrator's Commands	SC31-8781	This document describes the functions and commands helpful in configuring or monitoring your system. It contains system administrator's commands, such as TSO NETSTAT, PING, TRACERTE and their UNIX counterparts. It also includes TSO and MVS commands commonly used during the IP configuration process.
z/OS Communications Server: SNA Operation	SC31-8779	This document serves as a reference for programmers and operators requiring detailed information about specific operator commands.
z/OS Communications Server: Quick Reference	SX75-0124	This document contains essential information about SNA and IP commands.

Customization

Title	Number	Description
z/OS Communications Server: SNA Customization	SC31-6854	This document enables you to customize SNA, and includes the following: • Communication network management (CNM) routing table • Logon-interpret routine requirements • Logon manager installation-wide exit routine for the CLU search exit • TSO/SNA installation-wide exit routines • SNA installation-wide exit routines

Writing application programs

Title	Number	Description
z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference	SC31-8788	This document describes the syntax and semantics of program source code necessary to write your own application programming interface (API) into TCP/IP. You can use this interface as the communication base for writing your own client or server application. You can also use this document to adapt your existing applications to communicate with each other using sockets over TCP/IP.
z/OS Communications Server: IP CICS Sockets Guide	SC31-8807	This document is for programmers who want to set up, write application programs for, and diagnose problems with the socket interface for CICS using z/OS TCP/IP.
z/OS Communications Server: IP IMS Sockets Guide	SC31-8830	This document is for programmers who want application programs that use the IMS TCP/IP application development services provided by IBM's TCP/IP Services.
z/OS Communications Server: IP Programmer's Guide and Reference	SC31-8787	This document describes the syntax and semantics of a set of high-level application functions that you can use to program your own applications in a TCP/IP environment. These functions provide support for application facilities, such as user authentication, distributed databases, distributed processing, network management, and device sharing. Familiarity with the z/OS operating system, TCP/IP protocols, and IBM Time Sharing Option (TSO) is recommended.
z/OS Communications Server: SNA Programming	SC31-8829	This document describes how to use SNA macroinstructions to send data to and receive data from (1) a terminal in either the same or a different domain, or (2) another application program in either the same or a different domain.
z/OS Communications Server: SNA Programmer's LU 6.2 Guide	SC31-8811	This document describes how to use the SNA LU 6.2 application programming interface for host application programs. This document applies to programs that use only LU 6.2 sessions or that use LU 6.2 sessions along with other session types. (Only LU 6.2 sessions are covered in this document.)
z/OS Communications Server: SNA Programmer's LU 6.2 Reference	SC31-8810	This document provides reference material for the SNA LU 6.2 programming interface for host application programs.
z/OS Communications Server: CSM Guide	SC31-8808	This document describes how applications use the communications storage manager.

Title	Number	Description
z/OS Communications Server: CMIP Services and Topology Agent Guide	SC31-8828	This document describes the Common Management Information Protocol (CMIP) programming interface for application programmers to use in coding CMIP application programs. The document provides guide and reference information about CMIP services and the SNA topology agent.

Diagnosis

Title	Number	Description
z/OS Communications Server: IP Diagnosis Guide	GC31-8782	This document explains how to diagnose TCP/IP problems and how to determine whether a specific problem is in the TCP/IP product code. It explains how to gather information for and describe problems to the IBM Software Support Center.
z/OS Communications Server: SNA Diagnosis Vol 1, Techniques and Procedures and *z/OS Communications Server: SNA Diagnosis Vol 2, FFST Dumps and the VIT*	GC31-6850 GC31-6851	These documents help you identify an SNA problem, classify it, and collect information about it before you call the IBM Support Center. The information collected includes traces, dumps, and other problem documentation.
z/OS Communications Server: SNA Data Areas Volume 1 and *z/OS Communications Server: SNA Data Areas Volume 2*	GC31-6852 GC31-6853	These documents describe SNA data areas and can be used to read an SNA dump. They are intended for IBM programming service representatives and customer personnel who are diagnosing problems with SNA.

Messages and codes

Title	Number	Description
z/OS Communications Server: SNA Messages	SC31-8790	This document describes the ELM, IKT, IST, ISU, IUT, IVT, and USS messages. Other information in this document includes: • Command and RU types in SNA messages • Node and ID types in SNA messages • Supplemental message-related information
z/OS Communications Server: IP Messages Volume 1 (EZA)	SC31-8783	This volume contains TCP/IP messages beginning with EZA.
z/OS Communications Server: IP Messages Volume 2 (EZB, EZD)	SC31-8784	This volume contains TCP/IP messages beginning with EZB or EZD.
z/OS Communications Server: IP Messages Volume 3 (EZY)	SC31-8785	This volume contains TCP/IP messages beginning with EZY.
z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM)	SC31-8786	This volume contains TCP/IP messages beginning with EZZ and SNM.
z/OS Communications Server: IP and SNA Codes	SC31-8791	This document describes codes and other information that appear in z/OS Communications Server messages.

APPC Application Suite

Title	Number	Description
z/OS Communications Server: APPC Application Suite User's Guide	SC31-8809	This documents the end-user interface (concepts, commands, and messages) for the AFTP, ANAME, and APING facilities of the APPC application suite. Although its primary audience is the end user, administrators and application programmers may also find it useful.

Title	Number	Description
z/OS Communications Server: APPC Application Suite Administration	SC31-8835	This document contains the information that administrators need to configure the APPC application suite and to manage the APING, ANAME, AFTP, and A3270 servers.
z/OS Communications Server: APPC Application Suite Programming	SC31-8834	This document provides the information application programmers need to add the functions of the AFTP and ANAME APIs to their application programs.

Index

Special characters

Numerics

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z

Communicating Your Comments to IBM

If you especially like or dislike anything about this document, please use one of the methods listed below to send your comments to IBM. Whichever method you choose, make sure you send your name, address, and telephone number if you would like a reply.

Feel free to comment on specific errors or omissions, accuracy, organization, subject matter, or completeness of this document. However, the comments you send should pertain to only the information in this manual and the way in which the information is presented. To request additional publications, or to ask questions or make comments about the functions of IBM products or systems, you should talk to your IBM representative or to your IBM authorized remarketer.

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

Please send your comments to us in either of the following ways:

- If you prefer to send comments by FAX, use this number: 1+919-254-4028
- If you prefer to send comments electronically, use this address:
  - comsvrcf@us.ibm.com.
- If you prefer to send comments by post, use this address:

```
International Business Machines Corporation
Attn: z/OS Communications Server Information Development
P.O. Box 12195, 3039 Cornwallis Road
Department AKCA, Building 501
Research Triangle Park, North Carolina 27709-2195
```

Make sure to include the following in your note:

- Title and publication number of this document
- Page number or topic to which your comment applies.

Program Number: 5694–A01 and 5655–G52

Printed in USA

SC31-8775-08

Spine information:

z/OS Communications Server

z/OS V1R7.0 Comm Svr: IP Configuration Guide

Version 1
Release 7