H3C SR6600 路由器

用户手册

杭州华三通信技术有限公司

http://www.h3c.com.cn

资料版本：20101124-C-1.11

声明

Copyright © 2007-2010 杭州华三通信技术有限公司及其许可者 版权所有，保留一切权利。

未经本公司书面许可，任何单位和个人不得擅自摘抄、复制本书内容的部分或全部，并不得以任何形式传播。

H3C、H3C、Aolynk、  、H^{3}Care、  、TOP G、  、IRF、NetPilot、Neocean、NeoVTL、SecPro、SecPoint、SecEngine、SecPath、Comware、Secware、Storware、NQA、VVG、V^{2}G、V^{n}G、PSPT、XGbus、N-Bus、TiGem、InnoVision、HUASAN、华三均为杭州华三通信技术有限公司的商标。对于本手册中出现的其它公司的商标、产品标识及商品名称，由各自权利人拥有。

由于产品版本升级或其他原因，本手册内容有可能变更。H3C 保留在没有任何通知或者提示的情况下对本手册的内容进行修改的权利。本手册仅作为使用指导，H3C 尽全力在本手册中提供准确的信息，但是 H3C 并不确保手册内容完全没有错误，本手册中的所有陈述、信息和建议也不构成任何明示或暗示的担保。

技术支持

用户支持邮箱：customer_service@h3c.com

技术支持热线电话：800-810-0504（固话拨打）

400-810-0504（手机、固话均可拨打）

网址：http://www.h3c.com.cn

相关资料及其获取方式

相关资料

手册名称	用途
《H3C SR6602 路由器 安装手册》	介绍了 H3C SR6602 路由器的简介、软件维护、硬件维护、故障处理、安装及安装前的准备工作。
《H3C SR6604/6608 路由器 安装手册》	介绍了 H3C SR6604、6608 路由器的简介、软件维护、硬件维护、故障处理、安装及安装前的准备工作。
《H3C SR6616 路由器 安装手册》	介绍了 H3C SR6616 路由器的简介、软件维护、硬件维护、故障处理、安装及安装前的准备工作。
《H3C SR6600 路由器 接口卡及接口模块手册》	介绍了 H3C SR6600 路由器支持的各种接口卡的功能、面板、接口属性、指示灯、线连接方法。
《中低端系列路由器 电缆手册》	介绍了中低端路由器涉及到的所有电缆的管脚连接关系，供用户进行参考。

资料获取方式

您可以通过 H3C 网站（www.h3c.com.cn）获取最新的产品资料：

H3C 网站与产品资料相关的主要栏目介绍如下：

- [产品技术]：可以获取产品介绍和技术介绍的文档。
- [解决方案]：可以获取解决方案类资料。
- [服务支持/文档中心]：可以获取安装类、配置类或维护类等产品资料。
- [服务支持/软件下载]：可以获取与软件版本配套的资料。

资料意见反馈

如果您在使用过程中发现产品资料的任何问题，可以通过以下方式反馈：

E-mail：info@h3c.com

感谢您的反馈，让我们做得更好！

前 言

本书简介

本手册包含了 H3C SR6600 路由器的操作和命令手册。其中总述部分各章节内容如下：

- **1 简介。**介绍本手册的内容。
- **2 业务特性明细。**介绍了本书所涉及的各个模块的主要业务特性，可以链接到各个模块的操作、命令手册，帮助用户了解各模块的主要特性功能。
- **3 登录方法及用户界面介绍。**介绍了 SR6600 路由器的登录方法，帮助用户了解 SR6600 路由器所支持的用户界面以及如何通过 Console 口和以太网口进行本地或远程登录。
- **4 路由器的软件维护。**介绍了 SR6600 路由器的软件升级、维护方法及口令丢失的处理，帮助用户了解各种模式下的软件升级和配置文件的维护。
- **5 槽位顺序及接口编号。**介绍了 SR6600 路由器的设备槽位的分布及接口的编号规则。

本书约定

1. 命令行格式约定

格　式	意　义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从两个或多个选项中选取一个。
[x | y | ...]	表示从两个或多个选项中选取一个或者不选。
{ x | y | ... } *	表示从两个或多个选项中选取多个，最少选取一个，最多选取所有选项。
[x | y | ...] *	表示从两个或多个选项中选取多个或者不选。
&<1-n>	表示符号&前面的参数可以重复输入 1～n 次。
#	由“#”号开始的行表示为注释行。

2. 图形界面格式约定

格　式	意　义
< >	带尖括号“< >”表示按钮名，如“单击<确定>按钮”。
[ ]	带方括号“[ ]”表示窗口名、菜单名和数据表，如“弹出[新建用户]窗口”。

格　式	意　义
/	多级菜单用“/”隔开。如[文件/新建/文件夹]多级菜单表示[文件]菜单下的[新建]子菜单下的[文件夹]菜单项。

3. 各类标志

本书还采用各种醒目标志来表示在操作过程中应该特别注意的地方，这些标志的意义如下：

 警告	该标志后的注释需给予格外关注，不当的操作可能会对人身造成伤害。
 注意	提醒操作中应注意的事项，不当的操作可能会导致数据丢失或者设备损坏。
 提示	为确保设备配置成功或者正常工作而需要特别关注的操作或信息。
 说明	对操作内容的描述进行必要的补充和说明。
 窍门	配置、操作、或使用设备的技巧、小窍门。

目 录

1 简介 1-1
1.1 本书简介 1-1
1.2 相关手册 1-1
1.3 分册简介 1-1
2 业务特性明细 2-1
2.1 概述 2-1
2.2 特性功能索引 2-1
2.3 特性功能明细 2-2
2.3.1 接入分册 2-2
2.3.2 IP业务分册 2-3
2.3.3 IP路由分册 2-5
2.3.4 MPLS&VPN分册 2-7
2.3.5 QoS分册 2-8
2.3.6 安全分册 2-8
2.3.7 系统分册 2-10
2.3.8 IP组播分册 2-12
2.3.9 OAA分册 2-13
3 登录方法及用户界面介绍 3-1
3.1 通过Console口进行本地登录 3-1
3.2 通过以太网端口利用Telnet进行远程登录 3-2
3.3 用户界面简介 3-4
3.3.1 SR6600 路由器支持的用户界面 3-4
3.3.2 路由器用户界面编号 3-4
4 路由器的软件维护 4-1
4.1 路由器的软件组成及作用 4-1
4.1.1 路由器管理的文件 4-1
4.1.2 BootWare程序文件 4-1
4.1.3 应用程序文件 4-1
4.1.4 配置文件 4-2
4.1.5 路由器的软件维护的几种方法 4-3
4.2 BootWare菜单 4-4
4.2.1 BootWare主菜单 4-4
4.2.2 串口子菜单 4-7
4.2.3 以太网子菜单 4-7
4.2.4 文件控制子菜单 4-8
4.2.5 BootWare操作子菜单 4-8
4.2.6 Storage Device操作子菜单 4-9
4.3 通过串口升级应用程序和BootWare 4-9
4.3.1 XModem协议简介 4-9
4.3.2 串口参数的修改 4-10

4.3.3 升级应用程序 4-11
4.3.4 升级BootWare 4-13
4.4 通过TFTP升级应用程序和BootWare 4-15
4.4.1 在BootWare菜单中通过TFTP升级应用程序 4-16
4.4.2 在命令行模式通过TFTP升级和备份应用程序 4-18
4.5 通过FTP升级应用程序和BootWare 4-20
4.5.1 通过BootWare菜单中通过FTP升级应用程序 4-21
4.5.2 在命令行模式下通过FTP升级和备份应用程序 4-21
4.6 升级备用主控单元的应用程序（以SR6616 为例） 4-27
4.6.1 方法一 4-27
4.6.2 方法二 4-27
4.6.3 方法三 4-27
4.6.4 方法四 4-29
4.7 应用程序以及配置文件的维护 4-31
4.7.1 显示所有文件 4-31
4.7.2 设置应用程序文件的属性 4-32
4.7.3 设置配置文件的属性 4-33
4.7.4 删除文件 4-35
4.8 口令丢失的处理 4-36
4.8.1 BootWare口令丢失 4-36
4.8.2 用户口令丢失 4-36
4.8.3 Super Password口令丢失 4-37
4.9 备份和恢复BootWare 4-38
4.9.1 备份完整BootWare 4-38
4.9.2 恢复完整BootWare 4-38
5 槽位顺序及接口编号 5-1
5.1 SR6602 槽位排列顺序及接口编号 5-1
5.1.1 槽排列顺序 5-1
5.1.2 接口编号方法 5-1
5.1.3 举例说明 5-1
5.2 SR6604/SR6608 路由器的槽位及编号 5-2
5.2.1 插槽排列顺序 5-2
5.2.2 接口编号方法 5-2
5.2.3 举例说明 5-3
5.3 SR6616 路由器的槽位及编号 5-4
5.3.1 插槽排列顺序 5-4
5.3.2 主控单元和灵活接口平台的槽位分布 5-6
5.3.3 接口编号方法 5-6
5.3.4 举例说明 5-6

1 简介



说明

- 本手册当前适用于 H3C SR6602、SR6604、SR6608 和 SR6616 路由器产品，其中 SR6602 是集中式设备，SR6604、SR6608 和 SR6616 是分布式设备。本书对分布式和集中式设备有差异的命令分别进行了介绍，请用户注意。
- 本手册的组网及配置举例均以 SR6602 为例进行介绍。

1.1 本书简介

您可以通过以下条目了解到本书所涉及的主要内容。

- 业务特性明细：通过对本书所涉及的各个模块的主要业务特性进行简要介绍，帮助用户了解各模块的主要特性功能。
- 登录方法及用户界面介绍：通过介绍 H3C SR6600 路由器的基本登录方法，帮助用户了解 SR6600 路由器所支持的用户界面以及如何通过 Console 口和以太网口进行本地或远程登录。
- 路由器的软件维护：通过介绍 SR6600 路由器的软件升级、维护方法及口令丢失的处理，帮助用户了解各种模式下的软件升级和配置文件的维护。
- 槽位顺序及接口编号：方便用户在设备维护时，通过直观的示意图查找设备插槽的分布信息。

1.2 相关手册

- 如果需要了解路由器的安装、启动与配置、软件维护、硬件维护、电缆连接、安装故障处理等请分别参考《H3C SR6602 路由器 安装手册》、《H3C SR6608/6604 路由器 安装手册》和《H3C SR6616 路由器 安装手册》。
- 如果需要了解 SR6600 路由器支持的接口卡及接口模块的功能、接口卡线缆连接，请您参考《H3C SR6600 路由器 接口卡及接口模块手册》。
- 如果需要了解 SR6600 路由器涉及到的所有电缆的管脚连接关系等，请您参考《中低端系列路由器 电缆手册》。

可以通过 www.h3c.com.cn 获取最新版本配套的产品资料，方法如下：

- 单击主页的［服务支持］/［文档中心］，选择［IP 网络产品］；
- 在［IP 网络产品］页面中，选择［路由器产品］相应的设备类型；
- 点击相应的设备类型，即可查询该设备相关的手册。

1.3 分册简介

H3C SR6600 路由器支持的软件特性可以划分为 10 个分册，各分册的主要内容介绍如下。

- 接入分册：路由器支持的不同接口的配置及链路层协议的配置。

- IP 业务分册：路由器支持的如 IP 地址、ARP、DNS、DHCP、IP 性能、IP 单播策略路由、ACL、IPv6 等 IP 相关特性的配置。
- IP 路由分册：路由器支持的如静态路由、RIP、RIPng、OSPF、OSPFv3、IS-IS、BGP、BGP4+、路由策略等路由协议的配置。
- MPLS&VPN 分册：路由器支持的如 MPLS、MPLS TE、GRE、MPLS L3VPN、MPLS L2VPN 等 MPLS 和 VPN 相关协议的配置。
- QoS 分册：路由器支持的如流分类、流量监管、流量整形、QoS 策略、拥塞管理、拥塞避免、MPLS QoS、帧中继 QoS 等 QoS 相关协议和特性的配置。
- 安全分册：路由器支持的如 AAA 及用户管理、防火墙、NAT、IPSec、IKE、RADIUS、HWTACACS 等安全协议的配置。
- 系统分册：路由器支持的如搭建配置环境、基本配置、用户登录、文件管理、系统维护、NTP、SNMP、RMON、VRRP 等系统相关的协议和特性的配置。
- IP 组播分册：路由器支持的如 IGMP、PIM、MSDP、组播 VPN 等 IP 组播相关协议的配置。
- OAA 分册：路由器 OAA 架构支持的相关协议如 ACFP 及 OAP 单板的配置。

2 业务特性明细

2.1 概述

H3C SR6600 路由器当前共有三款产品：SR6602、SR6604、SR6608 和 SR6616。SR6602 是集中式设备，定位于高性能业务网关设备及行业网络的汇聚/接入；SR6604、SR6608 和 SR6616 为分布式设备，定位于大型网络的高性能核心、汇聚、接入设备，中小型企业网络的核心设备。

2.2 特性功能索引

表2-1 SR6600 路由器特性功能索引

功能模块	业务特性			
接入分册	CPOS 接口	POS 接口	以太网接口	WAN 接口
	帧中继	HDLC 接口	PPP	逻辑接口
	Modem 管理	ATM 接口	ATM	E-CPOS 接口
	VLAN 终结配置	镜像		
IP 业务分册	ARP	DHCP	域名解析	IP 地址
	IP 性能优化	IP 单播策略路由	URPF	IPv6 基础
	邻接表	流分类	IP 虚拟分片重组	双协议栈
	隧道技术	UDP Helper	DHCPv6	IPv6 单播策略路由
	NAT-PT 配置			
IP 路由分册	IP 路由概述	BGP	IS-IS	OSPF
	RIP	路由策略	静态路由	IPv6 BGP
	IPv6 IS-IS	OSPFv3	RIPng	IPv6 静态路由
	BFD			
MPLS&VPN 分册	MPLS 基本配置	MPLS TE	MPLS L2VPN	MPLS L3VPN
	GRE	L2TP		
QoS 分册	QoS	帧中继 QoS		
安全分册	AAA	防火墙	ALG	PKI
	会话管理	NAT	ACL	IPSec
	SSH2.0	Portal	公钥管理	Web 过滤配置
	连接限制	Rsh		

功能模块	业务特性			
系统分册	GR	VRRP	HA	设备管理
	NQA	NetStream	NTP	RMON
	SNMP	文件系统管理	系统基本配置	系统维护与调试
	信息中心	用户界面	Track	IPv6 NetStream
	Sampler 配置	热补丁	Telnet	Flow 日志
	User Profile			
IP 组播分册	组播概述	组播路由与转发	IGMP	MSDP
	PIM	组播 VPN	MBGP	IPv6 组播路由与转发
	MLD	IPv6 PIM	IPv6 MBGP	
OAA 分册	OAP 单板	ACFP	ACSEI	

2.3 特性功能明细



说明

在本手册的 PDF 版本中，通过点击“操作、命令手册”栏目中的链接访问各特性功能的操作、命令手册，如果想返回特性功能明晰页面，可以使用快捷键“Alt + ←”。

2.3.1 接入分册

表2-2 接入分册业务特性

业务特性	操作、命令手册	特性说明
CPOS 接口	CPOS 接口配置 CPOS 接口命令	• CPOS 接口介绍 • 配置 CPOS 接口 • 配置 E1 通道 • 配置 T1 通道 • 配置 E3 通道 • 配置 T3 通道
POS 接口	POS 接口配置 POS 接口命令	• POS 接口的配置和介绍。
以太网接口	以太网接口配置 以太网接口命令	• Combo 及三层以太网接口介绍。
WAN 接口	WAN 接口配置 WAN 接口命令	• 同步串口、AUX、CE1 等广域网接口介绍。
帧中继	帧中继配置 帧中继命令	• 帧中继 DCE/DTE 配置 • 多链路帧中继配置

业务特性	操作、命令手册	特性说明
HDLC	HDLC 配置 HDLC 命令	• HDLC 配置
PPP	PPP 配置 PPP 命令	• PPP 配置 • MP 配置 • PPP 链路效率机制配置 • PPPoE 的配置
逻辑接口	逻辑接口配置 逻辑接口命令	• LoopBack 接口配置 • NULL 接口配置 • 子接口配置 • MP-group 接口配置 • MFR 接口配置 • 虚拟模板接口和虚拟访问接口配置
Modem 管理	Modem 管理配置 Modem 管理命令	• Modem 管理配置
ATM 接口	ATM 接口配置 ATM 接口命令	• ATM OC-3c/STM-1 等接口的介绍。
ATM	ATM 配置 ATM 命令	• ATM、ATM 子接口、PVC 的配置和 VP 监管 • ATM 接口支持的 IPoA、PPPoA、IPoEoA 和 PPPoEoA 的应用介绍。
E-CPOS 接口	E-CPOS 接口配置 E-CPOS 接口命令	• E-CPOS 接口介绍 • 配置 E-CPOS 接口 • E-CPOS 接口典型配置举例
VLAN 终结配置	VLAN 终结配置 VLAN 终结命令	• VLAN 终结简介 • 配置报文的 TPID • 配置 VLAN 终结支持广播/组播 • 配置 Dot1q 终结
镜像	镜像配置 镜像命令	SR6602 本地镜像配置
返回特性功能索引		

2.3.2 IP 业务分册

表2-3 IP 业务分册业务特性

业务特性	操作、命令手册	特性说明
ARP	ARP 配置 ARP 命令	ARP 是将 IP 地址解析为 MAC 地址的协议。 • ARP 配置 • 免费 ARP 配置 • 代理 ARP 的配置

业务特性	操作、命令手册	特性说明
DHCP	DHCP 配置 DHCP 命令	DHCP 采用客户端/服务器通信模式，由客户端向服务器提出配置申请，服务器返回 IP 地址等相应的配置信息，以实现 IP 地址等信息的动态配置。 ● DHCP 服务器配置 ● DHCP 中继配置 ● DHCP 客户端配置
域名解析	域名解析配置 域名解析命令	域名系统（DNS，Domain Name System）是一种用于 TCP/IP 应用程序的分布式数据库，提供域名与 IP 地址之间的转换。 ● 静态域名解析配置 ● 动态域名解析配置
IP 地址	IP 地址配置 IP 地址命令	● IP 地址配置 ● 接口借用 IP 地址配置
IP 性能优化	IP 性能优化配置 IP 性能优化命令	在一些特定的网络环境里，需要调整 IP 的参数，以便网络性能达到最佳。 ● 允许接收和发送直连网段广播报文的配置 ● TCP 定时器、缓冲区大小、报文长度、ICMP 差错发送等功能属性配置 ● 接口的 TCP 最大报文段长度配置 ● ICMP 差错报文发送功能配置
IP 单播策略路由	IP 单播策略路由配置 IP 单播策略路由命令	策略路由是一种依据用户制定的策略进行路由选择的机制。 ● 单播策略路由的配置
URPF	URPF 配置 URPF 命令	URPF 的主要功能是用于防止基于源地址欺骗的网络攻击行为。 URPF 的配置
IPv6 基础	IPv6 基础配置 IPv6 基础命令	IPv6 是网络层协议的第二代标准协议，它是 IETF 设计的一套规范，是 IPv4 的升级版本。 ● IPv6 的基本功能配置 ● IPv6 邻居发现协议配置 ● PMTU 发现配置 ● TCP6 配置 ● IPv6 FIB 转发功能配置 ● 指定时间内发送 ICMPv6 差错报文的最大个数配置 ● IPv6 DNS 配置
邻接表	邻接表配置 邻接表命令	邻接表用于管理相连且处于激活态的邻居的信息。
流分类	流分类配置 流分类命令	流分类采用一定的规则识别符合某类特征的报文，是有区别地进行服务的前提和基础。
IP 虚拟分片重组	IP 虚拟分片重组配置 IP 虚拟分片重组命令	为了避免每个业务模块单独处理后片先到这种情况而导致复杂度过高，设备需要收到 IP 报文后就对分片报文进行虚拟分片重组。 IP 虚拟分片重组功能可以对分片报文进行检验、排序和缓存，保证后续业务模块处理的都是顺序正确的分片报文。

业务特性	操作、命令手册	特性说明
双协议栈	双协议栈配置	同时支持 IPv4 协议和 IPv6 协议的网络节点即成为双协议栈节点。当双栈节点配置 IPv4 地址和 IPv6 地址后，就可以在相应接口上转发 IPv4 和 IPv6 报文。 ● 双协议栈配置 ● IPv4 到 IPv6 的过渡技术
隧道技术	隧道技术配置 隧道技术命令	隧道（tunnel）是指一种协议封装到另外一种协议中的技术。隧道技术只要求隧道两端的设备支持两种协议。 ● IPv4 兼容 IPv6 手动/自动隧道配置 ● 6to4 隧道配置 ● ISATAP 隧道配置 ● IPv4 in IPv4 隧道配置 ● IPv6 in IPv6 隧道配置 ● IPv4 in IPv6 隧道配置 ● 6PE 配置
UDP Helper	UDP Helper 配置 UDP Helper 命令	UDP Helper 是实现对指定 UDP 端口广播报文的中继转发，即将指定 UDP 端口的广播报文转换为单播报文发送给指定的服务器，起到中继的作用。 ● UDP Helper 的工作原理 ● 配置 UDP Helper
DHCPv6	DHCPv6 配置 DHCPv6 命令	DHCPv6 是针对 IPv6 编址方案设计的、为主机分配 IPv6 地址和其他网络配置参数的协议。 ● DHCPv6 的典型组网应用 ● DHCPv6 的工作过程 ● 配置 DHCPv6 中继
IPv6 单播策略路由	IPv6 单播策略路由配置 IPv6 单播策略路由命令	策略路由是一种依据用户制定的策略进行路由选择的机制。 ● IPv6 单播策略路由简介 ● 配置 IPv6 单播策略路由
NAT-PT	NAT-PT 配置 NAT-PT 命令	NAT-PT 协议提供 IPv4 和 IPv6 地址之间的相互转换功能。 ● NAT-PT 的工作原理 ● 配置 NAT-PT
返回特性功能索引		

2.3.3 IP 路由分册

表2-4 IP 路由分册业务特性

业务特性	操作、命令手册	特性说明
IP 路由概述	IP 路由概述 IP 路由表显示命令	IP 路由基本概述。 ● 路由协议概述 ● IP 路由和路由表 ● 通过路由表选路
BGP	BGP 配置 BGP 命令	是一种自治系统之间的动态路由协议。 ● BGP 的基本功能配置 ● 路由属性配置 ● 大型 BGP 网络配置

业务特性	操作、命令手册	特性说明
IS-IS	IS-IS 配置 IS-IS 命令	用于自治系统内部，是一种链路状态协议，使用最短路径优先算法进行路由计算。 ● IS-IS 的基本功能配置 ● 路由信息控制配置
OSPF	OSPF 配置 OSPF 命令	是基于链路状态的内部网关协议。 ● OSPF 基本功能配置 ● OSPF 的区域特性配置 ● OSPF 的网络类型配置 ● OSPF 的路由信息控制配置
RIP	RIP 配置 RIP 命令	一种内部网关协议，主要用于规模较小的网络中。 ● RIP 基本功能配置 ● RIP 路由特性配置
路由策略	路由策略配置 路由策略命令	为了改变网络流量所经过的途径而修改路由信息的技术，主要通过改变路由属性来实现。 ● 过滤列表配置 ● 路由策略配置
静态路由	静态路由配置 静态路由命令	由管理员手工配置，恰当地设置和使用静态路由可以改进网络的性能，并可为重要的网络应用保证带宽。 ● 静态路由配置 ● 静态路由应用
IPv6 BGP	IPv6 BGP 配置 IPv6 BGP 命令	利用 BGP 的多协议扩展属性，达到在 IPv6 网络中应用的目的，BGP 协议原有的消息机制和路由机制并没有改变。 ● BGP4+基本功能配置 ● 控制路由信息的发布与接收配置 ● BGP4+路由属性配置
IPv6 IS-IS	IPv6 IS-IS 配置 IPv6 IS-IS 命令	支持多种网络层协议，包括 IPv6 协议，支持新添加的两个 TLVs 和一个新的 NLPID。 ● IPv6-ISIS 基本特性配置 ● IPv6-ISIS 路由信息控制配置
OSPFv3	OSPFv3 配置 OSPFv3 命令	提供对 IPv6 的支持。 ● OSPFv3 基本功能配置 ● OSPFv3 区域属性配置 ● OSPFv3 路由信息控制配置
RIPng	RIPng 配置 RIPng 命令	原来的 IPv4 网络中 RIP-2 协议的扩展。 ● RIPng 基本功能配置 ● RIPng 路由特性配置
IPv6 静态路由	IPv6 静态路由配置 IPv6 静态路由命令	IPv6 静态路由与 IPv4 静态路由类似，适合于一些结构比较简单的 IPv6 网络。 ● IPv6 静态路由配置
BFD	BFD 配置 BFD 命令	BFD（Bidirectional Forwarding Detection，双向转发检测）是一套全网统一的检测机制，用于快速检测、监控网络中链路连通状况。 ● BFD 基本功能配置 ● 各协议与 BFD 联动配置
返回特性功能索引		

2.3.4 MPLS&VPN 分册

表2-5 MPLS&VPN 分册业务特性

业务特性	操作、命令手册	特性说明
MPLS 基本	MPLS 基本配置 MPLS 基本配置命令	• MPLS 基本配置 • LDP 协议简介 • MPLS 基本能力配置 • PHP 特性配置 • 静态 LSP 配置 • MPLS LDP 配置 • LDP 实例配置 • MPLS 的 TTL 处理配置
MPLS TE	MPLS TE 配置 MPLS TE 命令	• MPLS 流量工程 • MPLS TE 基本能力配置 • 使用静态 CR-LSP 配置 MPLS TE 隧道 • 使用动态信令协议配置 MPLS TE 隧道 • RSVP-TE 高级特性配置 • 调整 CR-LSP 的建立 • 调整 MPLS TE 隧道的建立 • 流量转发配置 • 影响流量转发的参数配置 • 自动带宽调整过程配置 • CR-LSP 备份配置 • MPLS TE 快速重路由配置
MPLS L2VPN	MPLS L2VPN 配置 MPLS L2VPN 命令	提供基于 MPLS 网络的二层 VPN 服务，使运营商可以在统一的 MPLS 网络上提供基于不同介质的二层 VPN。 • MPLS L2VPN 配置 • CCC 方式 MPLS L2VPN 配置 • SVC 方式 MPLS L2VPN 配置 • Martini 方式的 MPLS L2VPN 配置 • Kompella 方式的 MPLS L2VPN 配置
MPLS L3VPN	MPLS L3VPN 配置 MPLS L3VPN 命令	BGP/MPLS VPN 是提供商 VPN 解决方案中一种基于 PE 的 L3VPN 技术。 • VPN 实例配置 • 基本 BGP/MPLS VPN 配置 • 跨域 VPN 配置 • 多角色主机配置 • HoVPN 配置 • OSPF 伪连接配置 • Multi-VPN-Instance CE 配置 • BGP 的 AS 号替换配置
GRE	GRE 配置 GRE 命令	通用路由封装，对某些网络层协议的数据报进行封装，使这些被封装的数据报能够在另一个网络层协议中传输。 • GRE 简介 • GRE in IPv4 隧道配置
L2TP	L2TP 配置 L2TP 命令	L2TP 协议提供了对 PPP 链路层数据包的隧道传输支持。 • LAC 配置 • LNS 配置
返回特性功能索引		

2.3.5 QoS 分册

表2-6 QoS 分册业务特性

业务特性	操作、命令手册	特性说明
QoS	QoS 配置 QoS 命令	服务质量，针对延迟、抖动、丢包率所做的差别服务。不支持峰值速率；不支持硬件队列。 ● 流分类、流量监管和流量整形配置 ● QoS 策略配置 ● 拥塞管理配置 ● 拥塞避免 ● MPLS QoS 配置
帧中继 QoS	帧中继 QoS 配置 帧中继 QoS 命令	在帧中继接口上通用的 QoS 服务为用户提供接口上的流量监管、流量整形、拥塞管理、拥塞避免等服务。
返回特性功能索引		

2.3.6 安全分册

表2-7 安全分册业务特性

业务特性	操作、命令手册	特性说明
AAA	AAA 配置 AAA 命令	供了一个用来对认证、授权和计费这三种安全功能进行配置的一致性框架。 ● AAA 配置 ● RADIUS 协议配置 ● HWTACACS 协议配置
防火墙	防火墙配置 防火墙命令	阻止来自因特网的对受保护网络的未授权的访问，另一方面允许内部网络的用户对因特网进行 Web 访问或收发 E-mail 等。 ● 包过滤防火墙配置 ● ASPF 状态防火墙配置
ALG	ALG 配置 ALG 命令	ALG（Application Level Gateway，应用层网关）特性主要完成对应用层报文的处理
PKI	PKI 配置 PKI 命令	公钥基础设施，通过使用公开密钥技术和数字证书来确保系统信息安全，并负责验证数字证书持有者身份的一种体系。 ● 为 PKI 生成 RSA 密钥 ● 证书注册信息配置 ● 证书申请方法配置 ● 证书验证配置 ● 证书属性访问控制策略配置
会话管理	会话管理配置 会话管理命令	会话管理是为了实现 NAT、ASPF、攻击防范等基于会话进行处理的业务而抽象出来的公共功能。 ● 配置会话管理 ● 配置长连接会话

业务特性	操作、命令手册	特性说明
NAT	NAT 配置 NAT 命令	网络地址转换，是将 IP 数据报报头中的 IP 地址转换为另一个 IP 地址的过程，主要用于实现私有网络访问外部网络的功能。 ● EASY IP 配置 ● 静态 NAT 配置 ● 多对多转换配置 ● NAPT（多对一）转换配置 ● NAT 内部服务器配置 ● NAT 日志配置 ● NAT 连接数限制配置
ACL	ACL 配置 ACL 命令	访问控制列表，用来实现流识别功能，不支持流模板。 ● 时间段配置 ● IPv4 基本 ACL 配置 ● IPv4 高级 ACL 配置 ● IPv6 基本 ACL 配置 ● IPv6 高级 ACL 配置
IPsec	IPsec 配置 IPsec 命令	是 IETF 制定的三层隧道加密协议，它为 Internet 上传输的数据提供了安全保证。 ● IPsec 安全提议配置 ● IPsec 安全策略配置 ● IPsec 安全策略模板配置 ● IPsec 安全策略应用 ● IKE 安全提议配置 ● IKE 对等体配置 ● IKE 存活机制配置
SSH2.0	SSH2.0 配置 SSH2.0 命令	安全外壳，用户通过一个不能保证安全的网络环境远程登录到设备时，SSH 特性可以提供安全保障和强大的认证功能。 ● SSH 服务器配置 ● 配置路由器做为 SSH 客户端 ● SSH 客户端软件配置
Portal	Portal 配置 Portal 命令	Portal 认证通常也称为 Web 认证，一般将 Portal 认证网站称为门户网站
公钥管理	公钥管理配置 公钥管理命令	● 本地的非对称密钥对配置 ● 远端主机的公钥配置
Web 过滤	Web 过滤配置 Web 过滤命令	● 网页地址过滤 ● URL 参数过滤 ● Java 阻断 ● ActiveX 阻断 ● 配置信息的保存和恢复
连接限制	连接限制配置 连接限制命令	为了保护内部网络资源（网络或主机或服务器）以及合理分配设备系统资源，可以通过制定相应的连接限制策略来对设备上建立的连接进行统计和限制。 ● 创建连接限制策略 ● 配置连接限制策略 ● 应用连接限制策略

业务特性	操作、命令手册	特性说明
Rsh	Rsh 配置 Rsh 命令	Rsh 用来在远程主机上执行特定的命令。远程主机需要运行 Rsh 守护程序以支持 Rsh 服务。Rsh 客户端与远程主机的守护进程之间进行通信。
返回特性功能索引		

2.3.7 系统分册

表2-8 系统分册业务特性

业务特性	操作、命令手册	特性说明
GR	GR 概述	完美重启，路由协议重启时，能够保证转发业务不中断。 ● 只支持 FIB6、ISIS 和 BGP 的协议级 GR 的配置
VRRP	VRRP 配置 VRRP 命令	基于虚拟路由冗余协议，能在某台设备出现故障时仍然提供高可靠的缺省链路 IPv4 的 VRRP 配置。
HA	HA 配置 HA 命令	HA（High Availability，高可靠性）特性用于实现系统的高可靠性
设备管理	设备管理配置 设备管理命令	设备管理功能，使用户能够查看设备当前的工作状态，配置设备运行的相关参数，实现对设备的日常维护和管理。不支持 bootROM 合法性检查。 ● 设备管理的配置
NQA	NQA 配置 NQA 命令	NQA 通过各种探测方式对网络或服务进行质量分析，并提供测试结果。 ● NQA 测试的配置 ● NQA 可选参数的配置 ● 对 Ping 功能的增强
NetStream	NetStream 配置 NetStream 命令	NetStream 提供报文统计功能，它根据报文的目的 IP 地址、源 IP 地址、目的端口号、源端口号、协议号、ToS、输入接口和输出接口来区分流，并针对不同的流进行独立的数据统计。 ● NetStream 配置 ● NetStream 统计功能的配置 ● NetStream 输出的 UDP 报文属性配置 ● NetStream 流老化配置
NTP	NTP 配置 NTP 命令	网络时间协议，用来在分布式时间服务器和客户端之间进行时间同步。 ● NTP 工作模式配置 ● 配置本地时钟作为参考时钟 ● NTP 可选参数配置 ● 访问控制权限配置 ● NTP 验证功能配置
RMON	RMON 配置 RMON 命令	远程网络监视，使 SNMP 更有效、更积极主动地监测远程网络设备能够减少网管站同代理间的通讯流量，达到简便而有力地管理大型互连网络的目的。 ● RMON 配置

业务特性	操作、命令手册	特性说明
SNMP	SNMP 配置 SNMP 命令	简单网络管理协议是使用 TCP/IP 协议族对互联网上的设备进行管理的一个框架，它提供一组基本的操作来监视和维护互联网。 ● SNMP 基本功能配置 ● Trap 配置
文件系统管理	文件系统管理配置 文件系统管理命令	设备运行过程中所需要的文件保存在设备的存储设备中，为了方便用户对存储设备进行有效的管理，设备以文件系统的方式对这些文件进行管理。 ● 文件系统的管理配置 ● FTP/TFTP 配置 ● TFTP 配置
系统基本配置	系统基本配置操作 系统基本配置命令	● 系统基本配置
系统维护与调试	系统维护与调试配置 系统维护与调试命令	对于设备所支持的各种协议和特性，系统基本上都提供了相应的调试功能，帮助用户对错误进行诊断和定位。 ● 系统调试配置 ● ping、tracert 等命令配置
信息中心	信息中心配置 信息中心命令	信息中心是系统的信息枢纽，它能够对所有的系统信息进行分类、管理。 ● 信息中心的配置
用户界面	用户界面配置 用户界面命令	用户界面视图是系统提供的一种视图，主要用来管理工作在流方式下的异步串口。通过在用户界面视图下的各种操作，可以达到统一管理各种用户配置的目的。 ● 异步串口配置用户属性的配置 ● 终端属性的配置 ● Modem 属性的配置 ● 自动执行命令的配置 ● 用户界面所能访问的命令级别的配置 ● VTY 用户界面访问限制\支持协议的配置 ● 异步串口重定向功能的配置 ● 登陆用户界面认证方式的配置
Track	Track 配置 Track 命令	联动功能是指通过建立联动项，实现不同模块之间的联动。联动功能由应用模块、Track 模块和监测功能模块三部分组成 ● Track 简介 ● 配置 Track 与 NQA 联动 ● 配置 Track 与应用模块联动
IPv6 NetStream	IPv6 NetStream 配置 IPv6 NetStream 命令	IPv6 NetStream 提供报文统计功能，它根据 IPv6 报文的目的 IP 地址、源 IP 地址、目的端口号、源端口号、协议号、流标签、输入接口和输出接口来区分流，并针对不同的流进行独立的数据统计。 ● IPv6 NetStream 统计功能的配置 ● IPv6 NetStream 聚合统计功能的配置 ● Pv6 NetStream 统计输出报文的属性配置 ● IPv6 NetStream 的流老化的配置
Sampler	Sampler 配置 Sampler 命令	Sampler 提供报文采样功能，从一组固定数量的报文中选出一个报文，送交后续业务模块处理

业务特性	操作、命令手册	特性说明
热补丁	热补丁配置 热补丁命令	热补丁是一种快速、低成本修复产品软件版本缺陷的方式。和升级软件版本相比，热补丁的主要优势是不会使设备当前正在运行的业务中断，即在不重启设备的情况下，可以对设备当前软件版本的缺陷进行修复。 该模块介绍了热补丁的配置和典型的应用
Telnet	Telnet 命令	该模块介绍了 telnet 过程中用到命令行。
Flow 日志	Flow 日志配置 Flow 日志命令	Flow 日志是指用户访问外部网络流信息的相关记录。该模块介绍了。 • Flow 日志的版本配置 • Flow 日志报文的源地址配置 • Flow 日志的输出
User Profile	User Profile 配置 User Profile 命令	User Profile（用户配置文件）提供一个配置模板，能够保存预设配置（一系列配置的集合）。用户可以根据不同的应用场景为 User Profile 配置不同的内容。 • 创建 User Profile • 配置 User Profile • 激活 User Profile
返回特性功能索引		

2.3.8 IP 组播分册

表2-9 IP 组播分册业务特性

业务特性	操作、命令手册	特性说明
组播概述	组播概述	组播技术概述，不支持二层组播特性。 • 组播模型分类 • 组播的框架结构 • 组播的转发机制
组播路由与转发	组播路由与转发配置 组播路由与转发命令	组播策略就是为了对 RPF 路由选择信息进行过滤而实施的一些策略。 • 组播策略简介 • 组播策略配置
IGMP	IGMP 配置 IGMP 命令	IGMP，互联网组管理协议，是 TCP/IP 协议族中负责 IP 组播成员管理的协议。 • IGMP 基本功能配置 • 调整 IGMP 的性能
MSDP	MSDP 配置 MSDP 命令	组播源发现协议，基于多个 PIM-SM 域的互连而开发的一种域间组播解决方案。 • MSDP 基本功能配置 • MSDP 对等体配置 • SA 消息配置
PIM	PIM 配置 PIM 命令	协议无关组播，可以利用任何单播路由协议生成的单播路由表为 IP 组播提供路由。 • PIM-DM 配置 • PIM-SM 配置 • PIM-SSM 配置 • PIM 公共信息配置

业务特性	操作、命令手册	特性说明
组播 VPN	组播 VPN 配置 组播 VPN 命令	• 组播 VPN 简介 • MD VPN 实现原理 • MD VPN 配置任务简介 • 配置 MD VPN
MBGP	MBGP 配置 MBGP 命令	为了提供对多种网络层协议的支持，IETF 对 BGP-4 进行了扩展，形成 MP-BGP，使 BGP 能够为多种路由应用提供路由信息。 • 配置 MBGP 基本功能 • 配置 MBGP 的路由属性 • 调整和优化 MBGP 网络 • 配置大型 MBGP 网络
IPv6 组播路由与转发	IPv6 组播路由与转发配置 IPv6 组播路由与转发命令	• IPv6 组播路由与转发简介 • 使能 IPv6 组播路由 • 配置 IPv6 组播路由与转发
MLD	MLD 配置 MLD 命令	MLD 用于 IPv6 路由器在其直连网段上发现组播侦听者。组播侦听者是那些希望接收组播数据的主机节点。 • 配置 MLD 基本功能 • 调整 MLD 性能 • 配置 MLD SSM Mapping • 配置 MLD Proxying
IPv6 PIM	IPv6 PIM 配置 IPv6 PIM 命令	IPv6 PIM 可以利用静态路由或者任意 IPv6 单播路由协议所生成的 IPv6 单播路由表为 IPv6 组播提供路由。 • 配置 IPv6 PIM-DM • 配置 IPv6 PIM-SM • 配置 IPv6 PIM-SSM • 配置 IPv6 PIM 公共特性
IPv6 MBGP	IPv6 MBGP 配置 IPv6 MBGP 命令	• IPv6 MBGP 简介 • IPv6 MBGP 配置任务简介 • 配置 IPv6 MBGP 的基本功能 • 控制路由信息的发布与接收 • 配置 IPv6 MBGP 的路由属性 • 调整和优化 IPv6 MBGP 网络 • 配置大型 IPv6 MBGP 网络
返回特性功能索引		

2.3.9 OAA 分册

表2-10 OAA 分册业务特性

业务特性	操作、命令手册	特性说明
OAP 单板	OAP 单板配置 OAP 单板命令	OAP 是针对新兴业务提供的一个开放式应用平台。OAP 单板根据客户需要可加载安全、语音等业务软件，为客户提供多样化的服务
ACFP	ACFP 配置 ACFP 命令	ACFP 是基于 OAA 架构设计的应用控制转发协议

业务特性	操作、命令手册	特性说明
ACSEI	ACSEI 配置 ACSEI 命令	ACSEI 为 ACFP server 和 client 提供了一种交互信息的方法，可以为 ACFP 联动提供很好的支撑，保障 ACFP client 与 ACFP server 之间有效交互信息、协作运行某种业务。
返回特性功能索引		

3 登录方法及用户界面介绍

SR6600 路由器可以通过以下方式登录：

- 通过 Console 口进行本地登录
- 通过以太网口利用 Telnet 进行本地或远程登录
- 通过 AUX 口利用 Modem 拨号进行远程登录
- 通过 NMS 登录

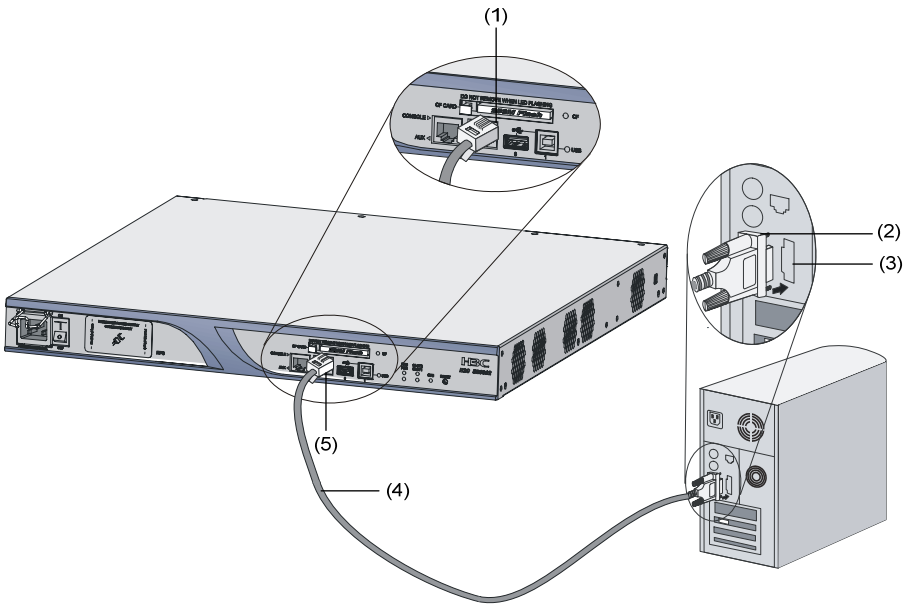
在实际应用中通过Console口进行的本地登录和通过以太网口利用Telnet进行本地或远程登录是最常用的登录方法。在此，我们介绍一下这两种登录方法。

3.1 通过 Console 口进行本地登录

1. 建立本地配置环境。

如下图所示，只需将配置口电缆的 RJ45 一端与路由器的配置口相连，DB9 一端与终端（微机上则是串口）相连；

图3-1 通过 CON 口进行本地配置

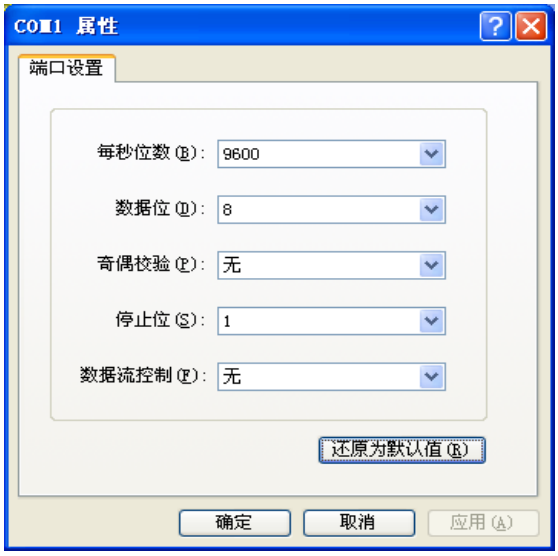


(1) 配置口（CONSOLE）	(2) DB9（母）连接器
(3) 配置终端的串口	(4) 配置口电缆
(5) RJ45 连接器	

2. 配置终端参数

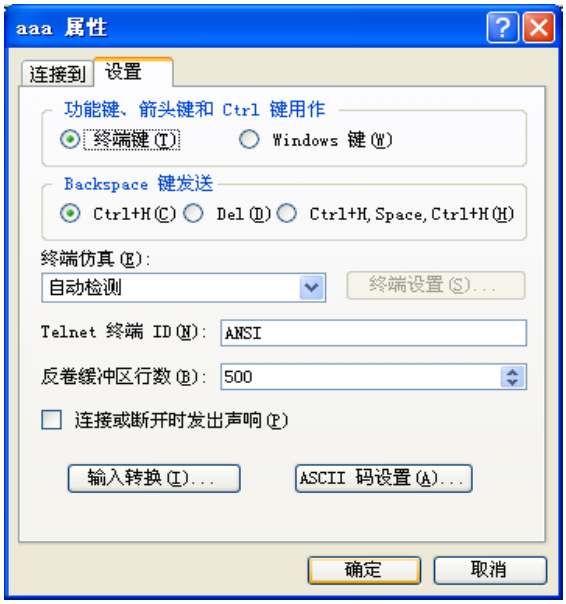
配置终端的参数，以 Windows 自带的超级终端为例。在串口的属性对话框中设置波特率为 9600，数据位为 8，奇偶校验为无，停止位为 1，流量控制为无。如下图所示：

图3-2 串口参数设置



终端类型为 VT100 或自动检测。

图3-3 终端类型设置



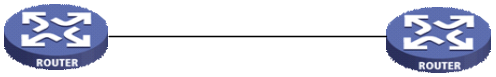
设置好各参数后，按[确定]即可对设备进行配置了。

3.2 通过以太网端口利用 Telnet 进行远程登录

1. 环境配置

将路由器通过以太网口连入网络。我们以两台路由器直连的网络为例（客户端为 Telnet 终端），我们可以通过客户端的路由器 Telnet 到服务器端，如下图：

图3-4 远程登录 Telnet 服务器



Server Client

2. 配置服务器端路由器

在对服务器端路由器进行远程登录前，必须要对设备进行配置。第一次对设备做配置时，必须通过 Console 口进行本地配置。

在服务器端路由器上做以下配置：

进入系统视图

```
Press ENTER to get started.
<H3C>system-view
```

使能 Telnet 服务器端功能

```
[H3C]telnet server enable
% Start Telnet server
```

配置与网络相连端口的 IP 地址，用户可更换为实际使用的 IP 地址

```
[H3C]interface GigabitEthernet0/0
[H3C-GigabitEthernet0/0]ip address 192.168.0.1 24
```

#进入 vty 接口视图

```
[H3C]user-interface vty 0
```

#设置登录密码及优先级：登录密码为 123456，优先级为 2

```
[H3C-ui-vty0]set authentication password cipher 123456
[H3C-ui-vty0]user privilege level 2
```

3. 登录路由器

确保终端到路由器之间网络的畅通，然后就可以通过终端远程登录路由器了（本例中，远程终端为客户路由器）。

```
<H3C>telnet 1.1.1.1
Trying 1.1.1.1 ...
Press CTRL+K to abort
Connected to 1.1.1.1 ...
******************************************************************************
* All rights reserved (2004-2006)                                            *
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************

Login authentication


Password:
<H3C> system-view
```

```
[H3C]
```

在提示输入密码时，输入密码：123456。正确输入密码后即可对路由器进行操作。



说明

IPv6 Telnet 的命令与 IPv4 的相同。命令行如下：

telnet ipv6 { *ipv6-address* | *hostname* } [**-i** *interface-type interface-number*] [*port-number*]

3.3 用户界面简介

3.3.1 SR6600 路由器支持的用户界面

表3-1 用户界面介绍

用户界面	对应用户	使用的路由器端口类型	说明
Console 用户界面	通过 Console 口登录的用户	Console 口	每台路由器只有 1 个 Console 用户
AUX 用户界面	通过 AUX 口登录的用户	AUX 口	每台路由器只有 1 个 AUX 用户
VTY 用户界面	Telnet 用户、SSH 用户	以太网端口	每台路由器最多可以有 5 个 VTY 用户

3.3.2 路由器用户界面编号

系统的用户界面共分 3 类，并按照一定的先后顺序排列：

分别是控制台（CON）、辅助接口（AUX）和虚拟接口（VTY）3 种类型。控制台和辅助接口分别只有一个；VTY 类型的用户界面可能有多个，而每种类型的多个用户界面内部又按照顺序排列。

用户界面的编号有两种方式：绝对编号方式和相对编号方式。

(1) 绝对编号

绝对编号的起始编号是 0（即 CON 口），其它接口依次类推。CON、AUX 口各占一个编号；VTY 接口不同产品可以支持的数量不同，请使用 **display user-interface** 查看即可得知。绝对编号可以唯一的指定一个用户界面或一组用户界面。

- CON 用户界面编号为 0；
- AUX 用户界面编号为 1；
- VTY 用户界面排在 AUX 用户界面之后，第一个 VTY 的绝对编号是 2（即 VTY 0），第（n+2）个 VTY 的绝对编号为（n+1）（即 VTY n-1），依此类推。

(2) 相对编号

相对编号的形式是：用户界面类型＋编号。此编号是每种类型的用户界面的编号。遵守的规则如下：

- CON 用户界面编号为：CON 0
- AUX 用户界面编号为：AUX 0；
- VTY 用户界面编号为：第一个为 VTY 0，第二个为 VTY 1，依此类推。

4 路由器的软件维护

说明

本手册以 H3C SR6608 路由器为例，SR6602、SR6604 和 SR6616 的 BootWare 菜单与 SR6608 相同。

4.1 路由器的软件组成及作用

4.1.1 路由器管理的文件

SR6608 有三类文件需要管理，分别是：

- BootWare 程序文件
- 应用程序文件
- 配置文件

4.1.2 BootWare 程序文件

BootWare 程序文件是路由器启动时用来引导应用程序的文件，BootWare 存放在 Flash 中。完整的 BootWare 包含 BootWare 基本段和 BootWare 扩展段。

- BootWare 基本段是指完成系统基本初始化的 BootWare。
- BootWare 扩展段具有丰富的人机交互功能，用于接口的初始化，可以实现升级应用程序和引导系统。
- BootWare 基本段启动后，可以在 BootWare 基本段菜单下加载和升级 BootWare 扩展段。

4.1.3 应用程序文件

SR6608 提供 Dual Image 功能，即系统缺省定义了三种用于启动的应用程序文件：

- 主用程序文件（简称“主文件”，属性标识为“M”）
- 备用程序文件（简称“备用文件”，属性标识为“B”）
- 安全程序文件（简称“安全文件”，属性标识为“S”）

这三种应用程序文件存放在 CF 卡中。

通常，设备出厂时应用程序文件缺省存储在内置的 CF 卡中。

当用户在CF卡中加载了三个属性分别为M、B、S的应用程序文件时，系统将按以上顺序选择这三个文件来启动路由器。如果希望改变这个选择顺序或者改变应用程序文件的属性，可以参考 4.7 应用程序以及配置文件的维护。

主文件、备用文件、安全文件的缺省文件名及启动时的选择顺序如下：

- 主文件：缺省文件名为 main.bin，是系统启动缺省使用的文件。
- 备份文件：缺省文件名为 backup.bin，当主文件启动失败时，系统使用备份文件启动。

- 安全文件：缺省文件名为 secure.bin，当主文件和备份文件均启动失败时，系统使用安全文件启动；如果安全文件启动失败，系统将提示启动失败信息。

需要注意的是：

- 仅有属性为 M、B、S 的应用程序文件可以用于系统启动，普通属性（即属性标识为“N/A”的应用程序文件，该类文件未指定属性）应用程序文件不会被用于系统启动。
- 应用程序文件的文件名可以在应用程序启动后通过命令修改；
- 应用程序文件的属性可以在 BootWare 菜单中修改，也可以在应用程序启动后通过命令修改；需要注意的是，由于安全文件是保证系统正常启动的最后一项措施，故安全文件的属性不允许修改，安全文件也不能由其它属性的文件修改而来，只能由用户在 BootWare 菜单中下载。
- 同一应用程序文件可以同时具有 M、B 和 S 属性，但同一时刻、同一属性的文件，设备上只能有一个。例如：CF 卡中有一个应用程序文件既具有 M 属性又具有 B 属性，那么就不可能存在其它的属性为 M 或者属性为 B 的文件；若另一个应用程序文件的属性由 N/A 改成了 B，那么以前既具有 M 属性，又具有 B 属性的文件只剩下 M 属性了。

4.1.4 配置文件

配置文件是保存路由器配置信息的文件，用于系统启动时的加载配置，配置文件支持 main/backup 属性，系统缺省定义了三种配置文件。

- 主用配置文件（属性标识为“M”）
- 备用配置文件（属性标识为“B”）
- 安全配置文件（属性标识为“S”）

主、备用属性的配置文件主要有以下应用：

- 当主用配置文件损坏或者丢失的时候，可以使用备用配置文件来加载配置信息。这种双重保护的方式增强了文件系统的安全性和可靠性。
- 保存当前配置时，可以指定配置文件的主用/备用/普通属性（即属性标识为“N/A”的配置文件，该类配置文件未指定属性）。
- 擦除设备中的配置文件时，可以指定是擦除主用配置文件还是备用配置文件；对于具有主、备用双重属性的配置文件，可以指定是擦除配置文件的主用属性还是备用属性。
- 设置下次启动使用的配置文件时，可以指定配置文件的主、备用属性。

安全配置文件用于主、备用配置文件加载配置均失败的情况。但需要注意的是：系统目前暂不支持安全配置文件。

系统在启动过程中将按如下顺序选择配置文件来加载配置。

- 如果主用配置文件存在，则以主用配置文件初始化配置。
- 如果主用配置文件不存在，备用配置文件存在，则以备用配置文件初始化配置。
- 如果备用配置文件也不存在，则以空配置启动（即系统将采用缺省参数进行初始化）。

如果希望改变上述选择顺序或者改变配置文件的属性，可以参考 4.7.3 设置配置文件的属性。

需要注意的是：

- 配置文件的文件名可以在应用程序启动后通过命令修改。

- 配置文件属性可以在BootWare菜单中修改，也可以在应用程序启动后通过命令修改。关于修改配置文件属性的说明，请参见 4.7.3 设置配置文件的属性。
- 同一配置文件可以同时具有主、备用属性，但同一时刻、同一属性的文件，设备上只能有一个。


说明

- 配置文件的文件名最长不能超过 64 个字符（含盘符和一个字符串结束符）。例如：如果盘符为：“CFa0:/”，则文件名除了盘符和一个字符串结束符外，剩余部部分最长可以为[64-1-6]=57 个字符。
- 如果超过 57 个字符就会出现文件操作错误。一般建议文件名不超过 16 个字符。
- 在 BooWare 阶段会显示配置文件，但配置文件的属性不会显示，仍是 N/A。
- 命令 **startup saved-configuration** *cfgfile* 的详细说明，请参见《H3C SR6600 路由器 用户手册》。

4.1.5 路由器的软件维护的几种方法

可以采用如下两种方法来维护设备的软件。

- 通过串口采用 XModem 协议完成 BootWare 及应用程序的升级。
- 通过以太网使用 FTP/TFTP 协议完成 BootWare 及应用程序的升级。

其中，第二种方式又可以分别通过 BootWare 菜单和命令行两种方式来实现。

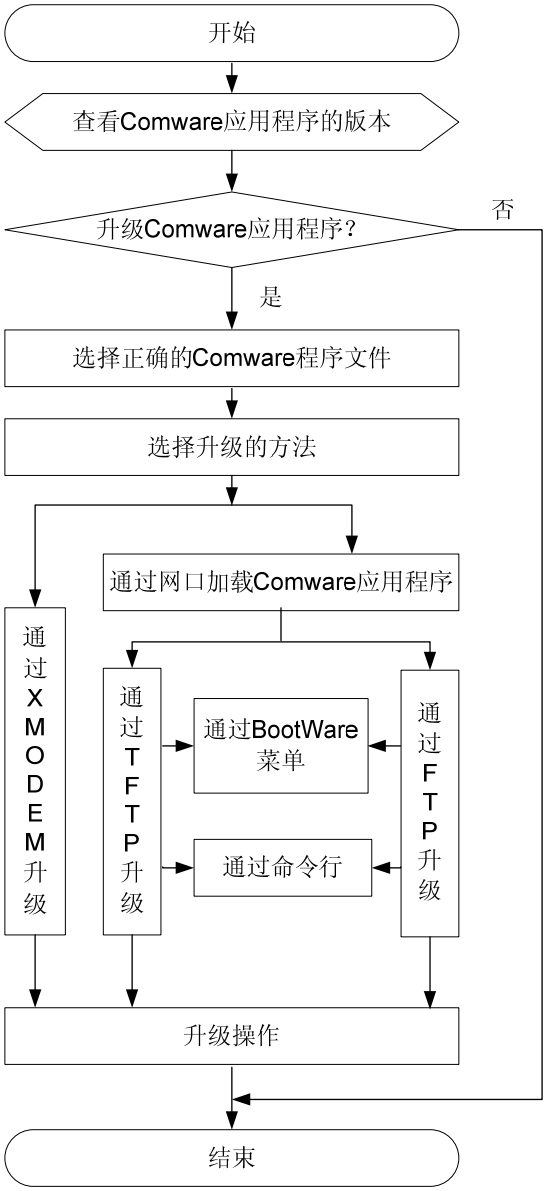
需要注意的是：

- 在升级应用程序的过程中，如果采用覆盖原有应用程序的方式升级，请确保在升级过程中不掉电，否则不仅会导致升级失败，而且原有应用程序也会被破坏。此时，如果设备原来只有一个应用程序用于启动，那么下次启动将无法正常进行。
- 如果升级的是 BootWare 基本段，请确保升级过程中不掉电，否则会破坏 BootWare 基本段，导致下次启动不能够正常进行。
- 如果升级的是 BootWare 扩展段，那么只是升级了 BootWare 的一部分，一旦出现错误可以重新升级。


说明

- BootWare 程序同 Comware 应用程序捆绑升级，即用户不需要单独升级 BootWare 程序，在升级 Comware 应用程序到最新版本并重新启动后，系统将检测当前的 BootWare 版本和主机应用程序内包含的 BootWare 版本是否一致，如果检测到不一致，系统就会自动刷新当前的 BootWare。
- 系统启动时，灵活接口平台会检测当前运行的 BootWare 版本是否与 Comware 应用程序捆绑的 BootWare 版本一致，如果不一致，将自动为用户升级 BootWare，并且在用户终端上输出提示信息。升级过程中，请不要进行重启整机、热插拔灵活接口平台以及下电等操作，否则容易导致灵活接口平台损坏。
- 进行软件升级前应确认当前的 BootWare 版本及 Comware 应用程序版本，以便使用正确的升级文件。Comware 应用程序和 BootWare 程序版本配套关系请参见《版本说明书》中的版本配套表。

图4-1 BootWare 程序及 Comware 应用程序的升级流程



4.2 BootWare 菜单

4.2.1 BootWare 主菜单

路由器上电开机后，首先运行 BootWare 基本段，然后再运行 BootWare 扩展段，终端屏幕上显示如下系统信息：

```
System is starting.....
Booting Normal Extend BootWare....
************************************************************
*                                                          *
*            H3C SR66 BootWare, Version 1.19               *
*                                                          *
************************************************************
Copyright (c) 2004-2008 Hangzhou H3C Technologies Co., Ltd.
Compiled Date       : Apr  15 2008
CPU Type            : MPC8548E
```

```
CPU L1 Cache        : 32KB
CPU Clock Speed     : 1000MHz
Memory Type         : DDR2 SDRAM
Memory Size         : 1024MB
Memory Speed        : 400MHz
BootWare Size       : 1024KB
Flash Size          : 4MB
cfa0 Size           : 495MB
NVRAM Size          : 128KB
BASIC CPLD Version  : 134.0
EXTEND CPLD Version : 133.0
PCB Version         : Ver.B

BootWare Validating...
Press Ctrl+B to enter extended boot menu...
Starting to get the main application file--cfa0:/main.bin!.............
................................................................................
.............................................................
The main application file is self-decompressing
```

键入<Ctrl+B>，将进入 Boot 扩展菜单；否则，系统将进入主应用程序文件的读入和自解压过程。



说明

- 必须在出现“Press Ctrl+B to enter extended boot menu”的 4 秒钟之内，键入<Ctrl+B>，系统方能进入 BootWare 扩展菜单；否则，系统将进入主应用程序文件的自解压过程。
- 若系统进入主应用程序的自解压过程后还希望进入 Boot 扩展菜单，则需要重新启动路由器。
- 为了便于阅读和便于理解，如果不做特殊说明，Boot 扩展菜单都将称为 BootWare 主菜单。

当出现“Press Ctrl+B to enter extended boot menu...”时，键入<Ctrl+B>，系统提示：

```
Please input BootWare password:
```

要求输入 BootWare 密码，输入正确的密码后，系统会提示：

- 当前操作设备的名称；
- 可以通过菜单<9>Storage Device Operation 来选择从哪种存储介质中加载应用程序。

需要注意的是：

- BootWare 的初始密码为空。
- 若连续三次输入不正确的 BootWare 密码，系统将挂死。此时，需要重启设备才能够继续输入密码
- 若输入的密码正确，才能够进入 BootWare 主菜单：

```
Note: The current operating device is cfa0
Enter < Storage Device Operation > to select device.

==================<EXTEND-BOOTWARE MENU>=====================
| <1> Boot System                                          |
| <2> Enter Serial SubMenu                                 |
| <3> Enter Ethernet SubMenu                               |
```

```
| <4> File Control                                          |
| <5> Modify BootWare Password                              |
| <6> Skip Current System Configuration                     |
| <7> BootWare Operation Menu                               |
| <8> Clear Super Password                                  |
| <9> Storage Device Operation                              |
| <0> Reboot                                                |
============================================================
Enter your choice(0-9):
```

该菜单含义如 表 4-1所示：

表4-1 BootWare 主菜单

菜单项	说明
<1> Boot System	从存储介质 CF 卡加载和引导应用程序
<2> Enter Serial SubMenu	进入串口子菜单 串口子菜单功能的详细描述请参见 4.2.2 串口子菜单
<3> Enter Ethernet SubMenu	进入以太网子菜单 以太网子菜单功能的详细描述请参见 4.2.3 以太网子菜单
<4> File Control	文件控制子菜单 文件控制子菜单功能的详细描述请参见 4.2.4 文件控制子菜单
<5> Modify BootWare Password	修改 BootWare 密码
<6> Skip Current System Configuration	忽略系统当前的配置 采用这种方式启动时，只是本次生效，下次启动时还需要重新设置 该功能一般在用户口令丢失之后使用
<7> BootWare Operation Menu	BootWare 操作子菜单 BootWare操作子菜单功能的详细描述请参见 4.2.5 BootWare操作子菜单
<8> Clear Super Password	清除超级用户密码 超级用户密码用来设置切换用户级别时的密码 缺省情况下没有设置超级用户密码 设置该选项后，仅在第一次重启路由器时生效，第二次重启路由器后超级用户密码将恢复
<9> Storage Device Operation	存储设备操作子菜单 用于存储介质的选择，以便选择存放在相应存储介质中的应用程序来启动系统
<a> Reboot	重新启动路由器



- H3C SR6608 路由器将内置 CF 卡标识为 cfa0，外置 CF 卡标识为 cfb0。
- 若应用程序存放在外置 CF 卡中，则需要通过 BootWare 主菜单中的选项<9>，来选择从 cfb0 加载应用程序。

4.2.2 串口子菜单

在 BootWare 主菜单下选择<2>，可以进入串口子菜单。

在串口子菜单中，可以通过 Xmodem 协议升级应用程序和 BootWare 程序。

系统显示如下：

```
======================<Enter Serial SubMenu >================
|Note:the operating device is cfa0                            |
| <1> Download Application Program To SDRAM And Run           |
| <2> Update Main Application File                            |
| <3> Update Backup Application File                          |
| <4> Update Secure Application File                          |
| <5> Modify Serial Interface Parameter                       |
| <0> Exit To Main Menu                                       |
============================================================
Enter your choice(0-5):
```

串口子菜单中各选项含义如 表 4-2所示：

表4-2 BootWare 串口子菜单

菜单项	说明
<1> Download Application Program To SDRAM And Run	通过串口下载应用程序到内存并启动
<2> Update Main Application File	升级主应用程序文件
<3> Update Backup Application File	升级备份应用程序文件
<4> Update Secure Application File	升级安全应用程序文件
<5> Modify Serial Interface Parameter	修改串口参数
<0> Exit To Main Menu	返回 BootWare 主菜单

4.2.3 以太网子菜单

在 BootWare 主菜单下键入<3>，可以进入以太网子菜单。

在以太网子菜单中，可以通过 FTP/TFTP 协议升级应用程序和 BootWare 程序。

系统显示如下：

```
====================<EEnter Ethernet SubMenu>===============
|Note:the operating device is cfa0                            |
| <1> Download Application Program To SDRAM And Run           |
| <2> Update Main Application File                            |
| <3> Update Backup Application File                          |
| <4> Update Secure Application File                          |
| <5> Modify Ethernet Parameter                               |
| <0> Exit To Main Menu                                       |
| < Ensure The Parameter Be Modified Before Downloading! >    |
============================================================
Enter your choice(0-5):
```

以太网口子菜单中各选项含义如 表 4-3所示：

表4-3 以太网口子菜单

菜单项	说明
<1> Download Application Program To SDRAM And Run	下载应用程序到内存并启动
<2> Update Main Application File	升级主应用程序
<3> Update Backup Application File	升级备份应用程序
<4> Update Secure Application File	升级安全应用程序
<5> Modify Ethernet Parameter	修改以太网参数
<0> Exit To Main Menu	返回 BootWare 主菜单

4.2.4 文件控制子菜单

在 BootWare 主菜单中键入**<4>**，系统将进入文件控制子菜单。

通过文件控制子菜单，可以实现对存储介质中保存的应用程序文件的显示、属性修改和删除等操作。

系统显示如下：

```
=======================<File CONTROL>=======================
|Note:the operating device is cfa0                          |
| <1> Display All File(s)                                   |
| <2> Set Application File type                             |
| <3> Set Configuration File type                           |
| <4> Delete File                                           |
| <0> Exit To Main Menu                                     |
=============================================================
Enter your choice(0-4):
```

文件控制子菜单中各选项含义如 表 **4-4**所示：

表4-4 文件控制子菜单

菜单项	说明
<1> Display All File	显示所有文件
<2> Set Application File type	设置应用程序文件属性
<3> Set Configuration File type	设置配置文件属性
<4> Delete File	删除文件
<0> Exit To Main Menu	返回 BootWare 主菜单

4.2.5 BootWare 操作子菜单

在 BootWare 主菜单下，键入**<7>**，进入 BootWare 操作菜单：

```
====================<BOOTWARE Operation Menu >==============
|Note:the operating device is cfa0                          |
| <1> Backup Full BootWare                                  |
| <2> Restore Full BootWare                                 |
| <3> Update BootWare By Serial                             |
| <4> Update BootWare By Ethernet                           |
| <0> Exit To Main Menu                                     |
```

```
============================================================
Enter your choice(0-4):
```

BootWare操作子菜单中各选项含义如 表 4-5所示：

表4-5 BootWare 操作子菜单

菜单项	说明
<1> Backup Full BootWare	备份完整 BootWare
<2> Restore Full BootWare	恢复完整 BootWare
<3> Update BootWare By Serial	通过串口升级 BootWare
<4> Update BootWare By Ethernet	通过以太网升级 BootWare
<0> Exit To Main Menu	返回 BootWare 主菜单

4.2.6 Storage Device 操作子菜单

在 BootWare 主菜单下，键入<9>，进入存储设备操作子菜单：

```
====================<DEVICE CONTROL>========================
| <1> Display All Available Nonvolatile Storage Device(s)   |
| <2> Set The Operating Device                              |
| <3> Set The Default Boot Device                           |
| <0> Exit To Main Menu                                     |
============================================================
Enter your choice(0-3):
```

存储设备操作子菜单中各选项含义如 表 4-6所示：

表4-6 Storage Device 操作子菜单

菜单项	说明
<1> Display All Available Nonvolatile Storage Device(s)	显示所有可用的存储设备
<2> Set The Operating Device	设置当前的操作设备
<3> Set The Default Boot Device	设置默认的启动设备
<0> Exit To Main Menu	返回 BootWare 主菜单

4.3 通过串口升级应用程序和 BootWare

4.3.1 XModem 协议简介

通过串口升级应用程序和 BootWare 请使用 XModem 协议。

XModem 协议是一种文件传输协议，因其简单性和较好的性能而被广泛应用。XModem 协议通过串口传输文件，支持 128 字节和 1K 字节两种类型的数据包，并且支持一般校验和、CRC 两种校验方式，在出现数据包错误的情况下支持多次重传（一般为 10 次）。

XModem 协议传输由接收程序和发送程序完成。先由接收程序发送协商字符，协商校验方式，协商通过之后发送程序就开始发送数据包，接收程序接收到一个完整的数据包之后按照协商的方式对数据包进行校验：

- 如果校验通过，则发送确认字符，然后发送程序继续发送下一个数据包。
- 如果校验失败，则发送否认字符，然后发送程序重传此数据包。

4.3.2 串口参数的修改

在实际应用中，有时为了节省升级软件的时间，需要提高串口的传输速率；有时为了提高传输的可靠性，还需要降低串口的传输速率。本节介绍如何调整串口的传输速率。

首先进入 BootWare 主菜单，键入<2>进入到串口子菜单，然后选择<5>，进入修改串口参数菜单，系统会提示修改串口波特率：

```
=======================<BAUDRATE SET>=======================
|Note:'*'indicates the current baudrate                     |
|     Change The HyperTerminal's Baudrate Accordingly       |
|--------------------<Baudrate Available>------------------|
| <1> 9600(Default) *                                       |
| <2> 19200                                                 |
| <3> 38400                                                 |
| <4> 57600                                                 |
| <5> 115200                                                |
| <0> Exit                                                  |
=============================================================
Enter your Choice(0-5):
```

选择合适的传输速率，以 115200bps 为例：键入<5>，系统将提示如下信息：

```
Baudrate has been changed to 115200 bps.
Please change the terminal's baudrate to 115200 bps, press ENTER when ready.
```

因为设备串口的波特率已经修改为 115200bps，而终端的波特率还为 9600bps，双方是无法通信的。因此，根据系统的提示，需要改变终端设置的波特率，使其与设备选择的下载波特率一致。

在终端上做如下操作：

图4-2 断开终端连接



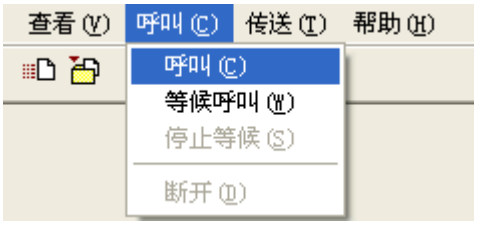
点击[文件/属性]。在属性栏点击< 配置(F)... >按钮，修改波特率为 115200：

图4-3 修改终端的波特率



点击[呼叫/呼叫]，重新建立呼叫连接：

图4-4 重新呼叫连接



然后按<Enter>键，此时系统会提示当前设置的波特率并返回上级菜单。

系统提示如下：

```
The current baudrate is 115200 bps
```

说明

如果通过改变串口速率下载文件来升级应用程序，那么完成升级后应及时将超级终端的连接速率恢复为 9600bps，以防止启动或重新启动时无法显示屏幕打印信息。

4.3.3 升级应用程序

通过串口升级应用程序，是在串口子菜单下实现的。

首先进入BootWare主菜单，键入<2>，进入串口子菜单。对该菜单的详细解释，请参见 4.2.2 串口子菜单。

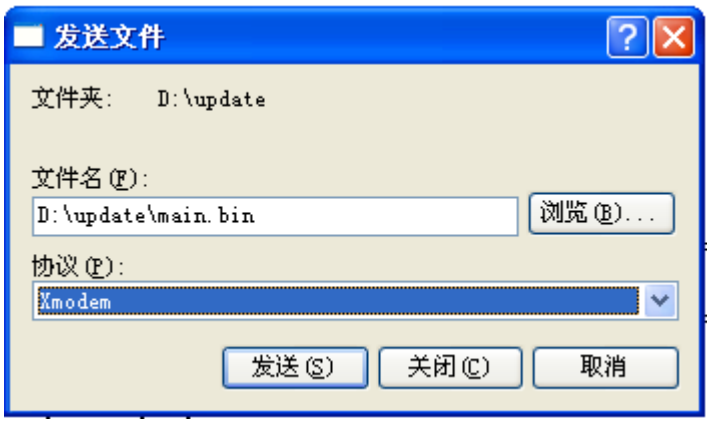
以升级主应用程序 main.bin 为例：

为了提高升级速度，可以在升级前先修改串口的传输速率（详细步骤请参见 4.3.2 串口参数的修改）。在串口子菜单下键入<2>，系统提示：

```
Please Start To Transfer File, Press <Ctrl+C> To Exit.
Waiting...CCCCCCCCCCCCCCCCCCC
```

此时，从终端窗口选择[传送/发送文件]，弹出如 图 4-5所示的对话框：

图4-5 发送文件对话框



点击<浏览>按扭，选择需要下载的应用程序文件，并将协议设置为XModem，然后点击<发送>按钮，系统弹出如 图 4-6所示的界面：

图4-6 正在发送文件界面



下载完成后，终端界面出现如下显示信息，表明下载并升级应用程序成功：

```
Download successfully!
24803072 bytes downloaded!
```

系统接着提示为下载到当前存储介质的应用程序命名：

```
Input the File Name:
```

(1) 如果输入的文件名没有与既有文件同名，那么下载的应用程序将以此名字命名，例如：

```
Input the File Name:main.bin
Updating File cfa0:/main.bin.............  ... ......................
.....................................................................................
```

```
.....................................................................................
.....................................Done!
```

需要注意的是，此时还需要返回BootWare主菜单，并进入文件控制子菜单，将此应用程序设置为默认启动程序，系统在下次启动时将采用该程序来完成系统的启动。设置默认启动程序的说明，请参见“4.7.2 设置应用程序文件的属性”。

(2) 如果输入的文件名与既有文件同名，则提示：

```
The file is exist,will you recover it? [Y/N]
```

- 如果选择“Y”，系统将覆盖原有应用程序，并且升级操作成功。设备在下次启动时，将采用新的应用程序来完成系统的启动。
- 如果选择“N”，则提示：

```
Cancel to overwrite the file.
Failed!
```

此时，系统将取消覆盖原有应用程序，并且升级操作失败。



说明

应用程序一般都比较大，多在 10M 以上。速率调整为 115200bps 之后，升级一般也需要 30 分钟左右。所以，对应用程序的升级推荐通过以太网来实现。

4.3.4 升级 BootWare

首先进入BootWare主菜单（请参见 4.2.1 BootWare主菜单），然后在BootWare主菜单下键入<7>，系统将进入BootWare操作子菜单，对BootWare的操作都是在该菜单下完成的。对该菜单的详细解释请参见 4.2.5 BootWare操作子菜单。

以升级完整的 BootWare 为例：

升级完整的 BootWare，需要先升级 BootWare 基本段，然后再升级 BootWare 扩展段。

在 BootWare 操作子菜单下键入<3>，系统提示：

```
==================<BOOTWARE OPERATION SERIAL SUB-MENU>================
|<1> Update Full BootWare                                             |
|<2> Update Extend BootWare                                           |
|<3> Update Basic BootWare                                            |
|<4> Modify Serial Interface Parameter                                |
|<0> Exit To Main Menu                                                |
======================================================================
Enter your choice(0-4):
```

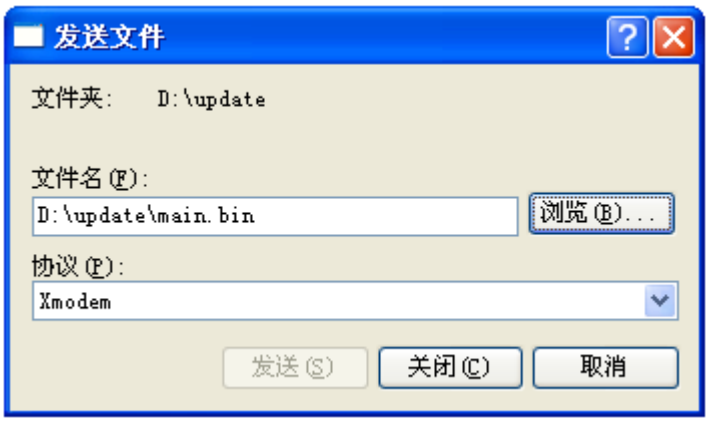
为了提高升级速度，需要先修改串口波特率，因此选择<4>，将串口的速率设置为 115200bps。此时，还需要将终端的速率设置成和串口的速率一致，否则二者无法进行通信。修改终端速率的方法请参见“4.3.2 串口参数的修改”。

串口和终端的速率设置完毕后，返回 BootWare 操作子菜单，并键入<1>，系统接着显示：

```
Waiting ...CCCCCCCCCCCCCCCCCCCCCCCC...
```

此时，从终端窗口选择［传送/发送文件］，弹出如 图 4-7所示的对话框：

图4-7 发送文件对话框



点击<浏览>按扭，选择需要下载的应用程序文件，并将协议设置为XModem，然后点击<发送>按钮，系统弹出如 图 4-8所示的界面：

图4-8 正在发送文件界面



下载完成后，终端界面出现如下信息，表明下载升级成功：

```
Download successfully!
24803072 bytes downloaded!
Updating Basic BootWare? [Y/N]
```

情况一：如果键入[N]，系统则提示：

```
Not update the Basic!
Updating Extend BootWare? [Y/N]
```

此时，系统将停止升级 BootWare 基本段，并提示是否升级 BootWare 扩展段。

情况二：如果键入[Y]，系统接下来会升级 BootWare 基本段：

```
Updating Basic BootWare..............Done!
Updating Extend BootWare? [Y/N]Y
```

此时，升级 BootWare 基本段成功，并且系统提示是否继续升级 BootWare 扩展段：

- 如果键入[N]，系统将放弃升级 BootWare 扩展段。此时，系统只完成了 BootWare 基本段的升级；
- 如果键入[Y]，系统将继续升级 BootWare 扩展段：

```
Updating Extend BootWare.................Done!
```

此时，BootWare 扩展段升级成功。

至此，完整的 BootWare 升级成功。



说明

- BootWare 程序同 Comware 应用程序捆绑升级，即用户不需要单独升级 BootWare 程序，在升级最新版本的 Comware 应用程序之后，下次启动运行时系统即会更新 Bootware。
- 系统启动时，灵活接口平台会检测当前运行的 BootWare 版本是否与 Comware 应用程序捆绑的 BootWare 版本一致，如果不一致，将自动为用户升级 BootWare，并且在用户终端上输出提示信息。升级过程中，请不要进行重启整机、热插拔灵活接口平台以及下电等操作，否则容易导致灵活接口平台损坏。图中所示文件名、文件大小、文件路径等参数会因具体情况而不同，进行升级前应确认当前的 BootWare 版本及应用程序版本，以便使用正确的文件。

4.4 通过 TFTP 升级应用程序和 BootWare

当应用程序文件比较大时，可以通过 TFTP 来进行升级，这样可以大大缩短升级维护的时间。

TFTP（Trivial File Transfer Protocol，简单文件传输协议）是 TCP/IP 协议族中的一个用来在客户机与服务器之间进行简单文件传输的协议，提供不复杂、开销不大的文件传输服务。TFTP 承载在 UDP 上，提供不可靠的数据流传输服务，不提供存取授权与认证机制，使用超时重传方式来保证数据的到达。

SR6608 提供的 TFTP 服务为 TFTP Client，即路由器作为 TFTP Client，文件服务器作为 TFTP Server，用户通过在终端输入相应的命令，可将本路由器的应用程序上传到文件服务器上，或者从文件服务器下载应用程序到路由器中。

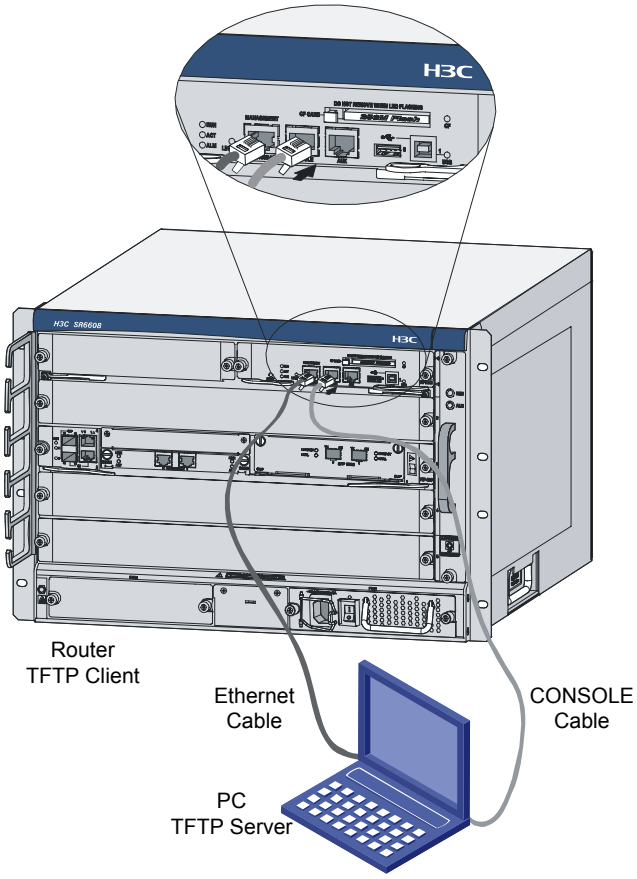
通过 TFTP 升级应用程序，有两种方式：

- 在 BootWare 菜单中通过 TFTP 升级应用程序
- 在命令行模式下通过 TFTP 升级应用程序

4.4.1 在 BootWare 菜单中通过 TFTP 升级应用程序

(1) 搭建升级环境

图4-9 搭建 TFTP 升级环境



- 路由器作为 TFTP Client，PC 作为 TFTP Server。
- 用交叉以太网线将路由器的管理以太网口 MANAGEMENT 和 PC 相连，并且保证路由器与 PC 的网络互通。本例设置管理以太网口的 IP 地址为 192.168.80.10，PC 的 IP 地址为 192.168.80.200。
- 在 PC 上启动 TFTP Server 程序，并设置好应用程序存放的路径。

注意

- TFTP Server 由用户自己购买和安装，H3C SR6608 路由器不附带此软件。
- H3C SR6608 路由器可以通过配置口和管理以太网口来升级应用程序。

---

(2) 在 BootWare 菜单中配置以太网口参数。

首先进入 BootWare 主菜单，接着键入<3>，进入以太网子菜单，然后键入<5>，进入以太网口参数设置菜单来配置以太网参数：

```
==================<ETHERNET PARAMETER SET>==================
 Note:       '.' = Clear field.
             '-' = Go to previous field.
          Ctrl+D = Quit.
============================================================
```

```
Protocol (FTP or TFTP):tftp
Load File Name      :main.bin
                    : main.bin
Target File Name    :main.bin
                    : main.bin
Server IP Address   :192.168.80.200
Local IP Address    :192.168.80.10
Gateway IP Address  :
FTP User Name       :
FTP User Password   :
```

表4-7 配置以太网口参数显示信息描述表

显示信息	说明
'.' = Clear field	快捷键：“.”表示清除当前输入
'-' = Go to previous field	快捷键：“-”表示返回到前一个参数域
Ctrl+D = Quit	快捷键：表示退出参数配置界面
Protocol (FTP or TFTP)	选择通过 FTP/TFTP 协议升级应用程序
Load File Name	下载文件名，要与下载的实际文件名一致 同时，需要在 FTP/TFTP 客户端设置下载路径 说明： ● 第一个文件名“main.bin”是系统自动记忆上次设置的文件名； ● 第二个文件名“main.bin”是用户本次设置的文件名 ● 应用程序文件名的长度最长为 50 个字符（包括后缀，如.bin）
Target File Name	下载到路由器后的目标文件名 文件的后缀需要和下载文件的后缀保持一致 说明： ● 第一个文件名“main.bin”是系统自动记忆上次设置的文件名； ● 第二个文件名“main.bin”是用户本次设置的文件名 ● 应用程序文件名的长度最长为 50 个字符（包括后缀，如.bin）
Server IP Address	FTP/TFTP 服务器的 IP 地址
Local IP Address	和 FTP/TFTP 服务器相连的接口 IP 地址
Gateway IP Address	网关 IP 地址，无需设置
FTP User Name	FTP 用户名 FTP 下载用到该设置，TFTP 下载无需设置
FTP User Password	FTP 用户的密码 TFTP 下载无需考虑

说明

- 如果需要使用冒号后面的默认参数，则直接按<ENTER>。
- 当网口参数检验 CRC 错误时，系统会将网口参数修改成默认值，并且打印：Check net params crc error, use the default value。

(3) 以上配置完成后，系统将自动返回到以太网子菜单，然后选择<2>，系统开始升级主应用程序：

```
Loading.........
.................
...... .... done
24802996 bytes downloaded!
Updating File cfa0:/main.bin
.........Done!
```

(4) 升级成功后，首先选择<0>，返回到 BootWare 主菜单，然后选择<1>，从 CF 卡重新引导系统。

注意

- 如果输入的应用程序文件名与 CF 卡中原有文件的文件名一样，系统将提示：The file is exist,will you recover it? [Y/N]，选择[y]，则直接覆盖 CF 卡中的应用程序文件。升级后的应用程序文件将直接替换原来该类型文件，成为唯一的应用程序。
- 请注意存储介质的存储空间是否足够，否则系统将提示空间不足：The free space isn't enough!。
- 各种文件类型的详细说明，请参见 4.1 路由器的软件组成及作用。

4.4.2 在命令行模式通过 TFTP 升级和备份应用程序

(1) 搭建升级环境

- 路由器作为 TFTP Client，PC 作为 TFTP Server。
- 升级环境的搭建方法同 4.4.1 在BootWare菜单中通过TFTP升级应用程序。
- 在 PC 上启动终端仿真程序，然后配置 Client 和 Server 的 IP 地址为同一网段：以保证路由器与 PC 的网络互通。本例设置路由器管理以太网口的 IP 地址为 192.168.80.10，PC 的 IP 地址为 192.168.80.200。
- 可以采用 **ping** 命令来检验两端是否连接成功。

注意

H3C SR6608 路由器可以通过配置口和管理以太网口来升级应用程序。

---

(2) 查看存储介质中的文件和剩余空间

在配置终端上用 **dir** 命令查询当前文件系统所包含的文件，以及存储器剩余的空间，为升级应用程序做准备。

```
<H3C>dir
Directory of cfa0:/

   0     drw-         -  Oct 28 2000 04:09:30   logfile
   1     -rw-  24802996  Oct 21 2007 17:03:26   sr6608.bin
   2     -rw-      1355  Oct 04 2007 17:22:12   startup.cfg
   3     -rw-  24802996  Sep 13 2037 13:21:20   main.bin

505480 KB total (456576 KB free)
```

```
File system type of cfa0: FAT16
```

表4-8 dir 命令显示信息描述表

显示信息	说明
Directory of cfa0:/	当前目录名称
505480 KB total (456576 KB free)	已用 CF 卡空间（CF 卡剩余空间） 如果 CF 卡剩余空间不够，将导致升级失败
File system type of cfa0: FAT16	CF 卡的文件系统类型

(3) 升级应用程序

通过 TFTP，可以将服务器上的应用程序文件下载到路由器，并覆盖原有的主应用程序来实现应用程序的更新。所作的更改在下次启动时生效。

从 PC 上下载应用程序文件 main.bin 到路由器上。

```
<H3C> tftp 192.168.80.200 get main.bin main.bin
The file main.bin exists. Overwrite it? [Y/N]:y
  Verifying server file...
  Deleting the old file, please wait...

  File will be transferred in binary mode
  Downloading file from remote TFTP server, please wait...|
  TFTP: 24802996  bytes received in 907 second(s)
  File downloaded successfully.
```



说明

- 下载应用程序文件的时候，如果路由器上已经有一个同名文件，那么系统将会提示是否覆盖，选择[Y/N]进行确认。
- 命令 **tftp** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方式来实现配置文件的更新。配置文件可以由文本编辑器修改，通过下载修改过的配置文件可以实现对配置的更改，所做更改会在下次启动后生效。

(4) 备份应用程序文件

通过 TFTP，可以将路由器上的应用程序文件上传到服务器上，实现应用程序文件的备份。

把路由器上的 main.bin 文件上传到 PC 上，并保存为 main.bin。

```
<H3C> tftp 192.168.80.200 put main.bin main.bin

  File will be transferred in binary mode
  Sending file to remote TFTP server. Please wait...
  TFTP:  24802996 bytes sent in 32 second(s).
  File uploaded successfully.
```

说明

- 备份应用程序文件的时候，如果有同名文件，系统不会出现提示信息，而是直接覆盖服务器上的文件。
- 命令 **tftp** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方式来实现配置文件的备份。

表4-9 路由器作为 TFTP Client 时升级和备份应用程序文件的命令和显示信息说明

显示信息	说明
tftp 192.168.80.200 get main.bin main.bin	从服务器下载需要升级的文件
The file main.bin exists. Overwrite it? [Y/N]:	提示是否要覆盖重名文件
TFTP: 24802996 bytes received in 907 second(s)	下载完毕并升级成功
tftp 192.168.80.200 put main.bin main.bin	从路由器上传文件到服务器
TFTP: 24802996 bytes sent in 32 second(s)	上传完毕并备份成功

4.5 通过 FTP 升级应用程序和 BootWare

当应用程序文件比较大时，也可以通过 FTP 来进行升级，这样可以大大缩短升级维护的时间。

FTP（File Transfer Protocol，文件传输协议）在 TCP/IP 协议族中属于应用层协议，主要向用户提供远程主机之间的文件传输。FTP 承载于 TCP 上，提供可靠的、面向连接的数据流传输服务。与 TFTP 相比，FTP 软件要大得多。

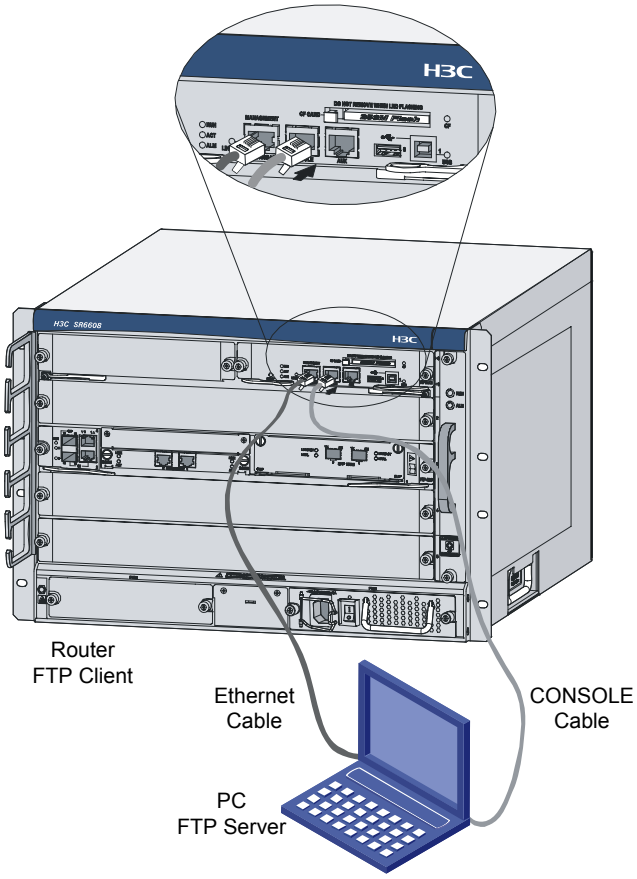
通过 FTP 升级应用程序有两种方式：

- 在 BootWare 菜单中通过 FTP 升级应用程序：在这种方式下，路由器只能够充作为 FTP Client，不能够作为 FTP Server。
- 在命令行模式下升级应用程序：在这种方式下，路由器既可以作为 FTP Server，也可以作为 FTP Client。

4.5.1 通过 BootWare 菜单中通过 FTP 升级应用程序

(1) 搭建升级环境

图4-10 搭建 FTP 升级环境



- 路由器作为 FTP Client，PC 作为 FTP Server。
- 用交叉以太网线将路由器的管理以太网口与 PC 相连，并且保证路由器与 PC 的网络互通。本例设置路由器管理以太网口的 IP 地址为 192.168.80.10，PC 的 IP 地址为 192.168.80.200。
- 在 PC 上启动 FTP Server 程序，设置应用程序所在的路径以及设置 FTP 用户名和密码。

注意

- FTP Server 由用户自己购买和安装，H3C SR6608 路由器不附带此软件。
- H3C SR6608 路由器可以通过配置口和管理以太网口来升级应用程序。

升级过程中的第（2）步至第（4）步请参考 TFTP 升级步骤中对应的描述，这里不再赘述。

4.5.2 在命令行模式下通过 FTP 升级和备份应用程序

1. 路由器作为 FTP Client，PC 作为 FTP Server

(1) 搭建升级环境，搭建过程请参见 4.5.1 通过BootWare菜单中通过FTP升级应用程序。

(2) 在配置终端上用**dir**命令查询当前文件系统所包含的文件，以及存储器剩余的空间，为升级应用程序文件做准备。具体的查询方法，请参见 4.4.2 (2)查看存储介质中的文件和剩余空间。

(3) 在 PC 上启动 FTP Server 程序，设置应用程序所在的路径，以及 FTP 用户名和密码。本例设置用户名为 guest，密码为 123456。

(4) 登录 FTP Server。

```
<H3C>ftp 192.168.80.200
Trying 192.168.80.200 ...
Press CTRL+K to abort
Connected to 192.168.80.200.
220 3Com 3CDaemon FTP Server Version 2.0
User(192.168.80.200:(none)):guest
331 User name ok, need password
Password:
230 User logged in
[ftp]
```

连接成功后，通过命令行完成对应用程序文件的升级和备份。

表4-10 路由器作为 FTP Client 时登录 FTP Server 的显示信息说明

显示信息	说明
ftp 192.168.80.200	登录 FTP Server，并进入 FTP 客户端视图
User(192.168.80.200:(none))	输入 FTP Server 上配置的用户名
331 User name ok, need password	用户名正确，需要输入该用户的密码
230 User logged in	成功登录 FTP Server

(5) 升级应用程序文件

通过 FTP，可以将服务器上的应用程序下载到路由器上，并覆盖原有的主应用程序来实现应用程序的更新。所作的更改在下次启动时生效。

从 PC 上下载应用程序文件 main.bin 到路由器上，并保存成 main.bin。。

```
[ftp]get main.bin main.bin
 cfa0:/main.bin has been existing. Overwrite it? [Y/N]:y
227 Entering passive mode (192,168,80,200,7,237)
125 Using existing data connection
226 Closing data connection; File transfer successful.
FTP: 24802996 byte(s) received in 467.570 second(s), 53.00K byte(s)/sec.
```



说明

- 下载应用程序文件的时候，如果路由器上已经有一个同名文件，那么系统将会提示是否覆盖，选择[Y/N]进行确认。
- 命令 **get** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方法来实现配置文件的更新。配置文件可以由文本编辑器修改，通过下载修改过的配置文件可以实现对配置的更改，所做更改会在下次启动后生效。

(6) 备份应用程序文件

通过 FTP，可以将路由器上的应用程序文件上传到服务器上，以实现应用程序文件的备份。

把路由器上的 main.bin 文件上传到 PC 上，并保存为 main.bin。

```
[ftp]put main.bin main.bin
227 Entering passive mode (192,168,80,200,5,34)
125 Using existing data connection
226 Closing data connection; File transfer successful.
FTP: 24802996 byte(s) sent in 14.400 second(s), 1722.00Kbyte(s)/sec
[ftp]quit
221 Service closing control connection
```



- 备份应用程序文件的时候，如果有同名文件，系统将直接覆盖服务器上的文件。
- 命令 **put** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方式来实现配置文件的备份。

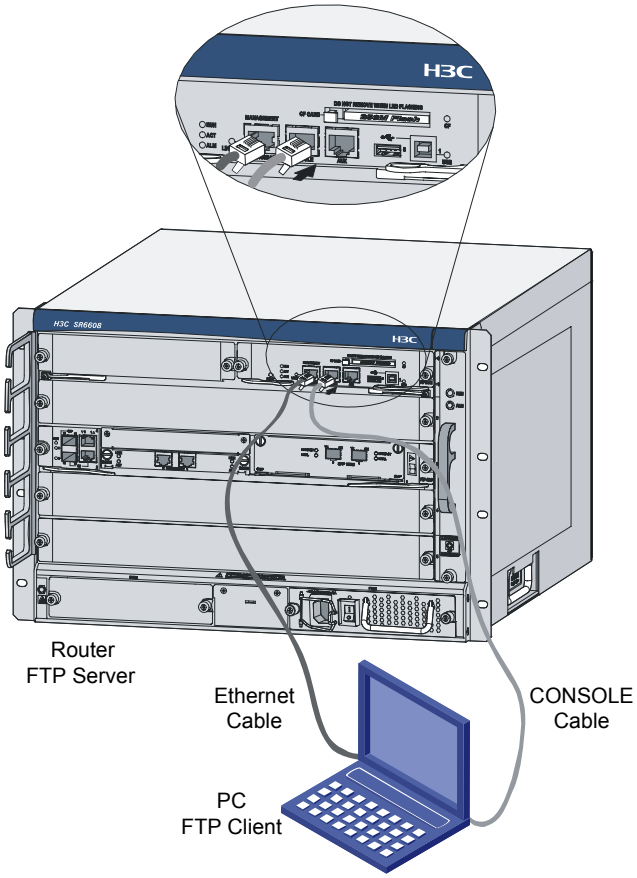
表4-11 路由器作为 FTP Client 时的升级和备份应用程序文件的命令和显示信息说明

显示信息	说明
[ftp]get main.bin main.bin	从服务器下载需要升级的文件
cfa0:/main.bin has been existing. Overwrite it? [Y/N]:y	提示是否要覆盖重名文件
FTP: 24802996 byte(s) received in 467.570 second(s), 53.00K byte(s)/sec	下载完毕并升级成功
[ftp]put main.bin main.bin	从路由器上传文件到服务器进行备份
FTP: 24802996 byte(s) sent in 14.400 second(s), 1722.00Kbyte(s)/sec	上传完毕并备份成功
[ftp]quit	退出 FTP 客户端视图
221 Service closing control connection	关闭服务控制连接

2. 路由器为 FTP Server，PC 为 FTP Client

(1) 搭建升级环境

图4-11 搭建 FTP 升级环境



- 用交叉以太网线将路由器的管理以太网口与 PC 相连。
- 配置路由器和 PC 两端的 IP 地址为同一网段，以保证路由器与 PC 的网络互通。本例设置路由器的管理以太网口的 IP 地址为 192.168.80.10，PC 的 IP 地址为 192.168.80.200。
- 可以采用 **ping** 命令来检验两端是否连接成功。

注意

H3C SR6608 路由器可以通过配置口和管理以太网口来升级应用程序。

---

(2) 启动 FTP 服务

启动 FTP Server 功能。

```
<H3C>system-view
[H3C] ftp server enable
```

添加 FTP 授权用户名和密码。

```
[H3C] local-user guest
New local user added.
[H3C-luser- guest] service-type ftp
[H3C-luser- guest] password simple 123456
[H3C-luser-guest] level 3 authorization-attribute level 3
```

表4-12 启动 FTP Server 命令说明

显示信息	说明
ftp server enable	启动 FTP Server 功能
local-user guest	创建本地用户，并进入该用户视图
service-type ftp	设置当前本地可以使用的服务类型为 FTP
password simple 123456	设置当前本地用户的密码
authorization-attribute level 3	设置用户的级别为 3



- 配置 FTP Server 的验证和授权之后，就可以启动 FTP 服务了。FTP Server 可同时支持多个用户访问。远端 FTP 用户向 FTP Server 发送请求，FTP Server 执行相应的动作，并向用户返回执行的结果。
- 配置 FTP Server 的验证和授权之后，还需要设置用户的级别为 3，否则，将备份的文件恢复到路由器时，系统会提示：You have no rights to store files。有关用户级别的设置，请参考《H3C SR6600 路由器 用户手册》。

(3) 在 PC 端启动 FTP Client 程序

以 Windows XP 自带的 FTP 客户端为例。

在 DOS 窗口下输入<ftp>：

```
C:\Documents and Settings\Administrator>ftp
ftp>
ftp> open 192.168.80.10
Connected to 192.168.80.10.
220 FTP service ready.
User (192.168.80.10:(none)): guest
331 Password required for guest
Password:
230 User logged in.
```

表4-13 启动 FTP Server 命令和显示信息说明

显示信息	说明
C:\Documents and Settings\Administrator>ftp	在 PC 上启动 FTP Client 程序
ftp> open 192.168.80.10	在 FTP 客户端视图下，登录 IPv4 FTP 服务器
User (192.168.80.10:(none))	输入 FTP Server 上配置的用户名
331 Password required for guest	输入该用户的密码
230 User logged in	登陆 FTP Server 成功

(4) 升级应用程序文件

通过 FTP，将 Client 上的应用程序文件上传到 Server 上，并覆盖原有的主应用程序文件，来实现应用程序的升级。所作的更改在下次启动时生效。

将 PC 上的应用程序文件 main.bin 上传到路由器上，并命名为 main.bin。

```
ftp> binary
200 Type set to I.
ftp> lcd d:\update
Local directory now D:\update.
ftp> put main.bin main.bin
200 Port command okay.
150 Opening BINARY mode data connection for main.bin.
226 Transfer complete.
```

说明

- 升级应用程序文件的时候，如果有同名文件，系统将直接覆盖服务器上的文件。
- 命令 **put** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方式来实现配置文件的更新。配置文件可以由文本编辑器修改，通过上传修改过的配置文件可以实现对配置的更改，所做更改会在下次启动后生效。

(5) 备份应用程序文件

通过 FTP，可以将 Server 上的应用程序文件下载到 Client，来实现应用程序的备份。

从路由器下载应用程序文件 main.bin 到 PC 上，并保存成 main.bin。

```
ftp> get main.bin main_bac.bin
200 Port command okay.
150 Opening BINARY mode data connection for main.bin.
226 Transfer complete.
```

说明

- 下载应用程序文件的时候，如果有同名文件，系统将直接覆盖服务器上的文件。
- 命令 **get** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。
- 可以通过相同的方式来实现配置文件的备份。

表4-14 启动 FTP Server 操作和显示信息说明

显示信息	说明
ftp> binary	更改传输模式为二进制模式 缺省情况下为 AscII 模式
ftp> lcd d:\update	改本地路径
ftp> put main.bin main.bin	从 PC 上传文件到路由器
ftp> get main.bin main.bin	从路由器下载文件到 PC
ftp> quit	退出 FTP 服务器
221 Server closing.	关闭服务控制连接

4.6 升级备用主控单元的应用程序（以 SR6616 为例）

当设备采用双主控启动时，需要保证主/备用主控单元上的主启动文件版本一致，否则备用主控单元就不能正常启动。在这种情况下，无法实现主/备用主控单元之间的相互备份以及文件操作，也不能保证设备的高可靠性。

通过下列方法，可以将备用主控单元的主启动文件更新到与主用主控单元的主启动文件版本一致。需要注意的是，采用下列方法更新备用主控单元的主启动文件前，需要保证主用主控单元的 CF 卡中已经存在主启动文件，否则需要重新下载主启动文件到主用主控单元，具体方法请参见本文的前半部分。

4.6.1 方法一

通过在主用主控单元上执行 表 4-15中的命令，即可完成备用主控单元的快速升级，该命令相当于分步执行了以下操作：

(1) 将主用主控单元当前使用的启动文件拷贝到备用主控单元；

(2) 将拷贝后的文件指定为备用主控单元的下次启动文件；

(3) 重启备用主控单元。

无论主/备用主控单元上的主启动文件版本是否一致，该方法均可以适用。

表4-15 同步升级备用主控单元的启动文件

操作	命令	说明
同步升级备用主控单元的启动文件	**boot-loader update slot** *slot-number*	必选 该命令在用户视图下执行



说明

命令 **boot-loader update slot** *slot-number* 的详细说明，请参见《H3C SR6600 路由器 用户手册》中的“系统分册/设备管理”。

4.6.2 方法二

重启备用主控单元，在备用主控单元的Bootware主菜单中更新备用主控单元的主启动文件，更新方法，请参见 4.3 通过串口升级应用程序和BootWare、4.4 通过TFTP升级应用程序和BootWare、4.5 通过FTP升级应用程序和BootWare（推荐使用 4.4 或者 4.5 这两种方式，速度比较快。）

需要注意的是，无论主/备用主控单元的主启动文件版本是否一致，必须明确哪一个文件是当前的主启动文件，如果不明确，可以通过 TFTP/FTP 方式从主用主控单元的 CF 卡将当前使用的主启动文件上传至文件服务器，然后再进行备用主控单元主启动文件的更新。

4.6.3 方法三

使用命令 **tftp** 将当前主启动文件下载到备用主控单元，并设置为主启动文件，然后重启备用主控单元。需要注意的是：

- 主/备用主控单元的主启动文件版本必须一致。

- 必须明确哪一个文件是当前的主启动文件，如果不明确，可以通过 TFTP/FTP 方式从主用主控单元的 CF 卡将当前使用的主启动文件上传至文件服务器，然后再进行备用主控单元主启动文件的更新。

具体方法如下：

(1) 通过 **dir** 命令查看主用主控单元上是否已经存在待升级的应用程序文件，如果存在，则继续第(2)步操作；如果不存在，则需要先通过 XModem/FTP/TFPT 等传输协议把待升级的应用程序文件下载到主用主控单元上（具体方法请参见本章前半部分内容）；

(2) 使用 **dir** 命令查看备用主控单元 CF 卡的剩余空间，确保备用主控单元 CF 卡的剩余空间能够容纳待升级的应用程序文件；

(3) 使用命令 **tftp** 将服务器上的应用程序文件下载到备用主控单元：

- 当设备配置的主控单元为 RPE-X1 时：

```
<H3C>tftp 192.168.80.200 get main.bin slot4#cfa0：/main.bin

 File will be transferred in binary mode
 Downloading file from remote TFTP server, please wait...|
 TFTP: 24384324 bytes received in 101 second(s)
 File downloaded successfully.
```

- 当设备配置的主控单元为 RSE-X1 时：

```
<H3C>tftp 192.168.80.200 get main.bin slot6#cfa0：/main.bin

 File will be transferred in binary mode
 Downloading file from remote TFTP server, please wait...|
 TFTP: 24384324 bytes received in 101 second(s)
 File downloaded successfully.
```

(4) 指定备用主控单元下次启动时使用的应用程序文件：

- 当设备配置的主控单元为 RPE-X1 时：

```
<H3C>boot-loader file slot4#cfa0:/main.bin slot 4 main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 4!
```

- 当设备配置的主控单元为 RSE-X1 时：

```
<H3C>boot-loader file slot6#cfa0:/main.bin slot 6 main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 6!
```

(5) 使用命令 **display boot-loader** [**slot** *slot-number*] 查看主用主控单元和备用主控单元用于启动的应用程序文件信息，确保二者一致。

(6) 在系统视图下使用命令 **slave restart** 重启备用主控单元之后，先用命令 **display version** 查看设备当前的版本，如果版本号是要升级的版本就表明升级成功。

(7) 通过 **display device** 命令查看主用主控单元和备用主控单元的状态，如果主用主控单元的状态和备用主控单元的状态都是 Normal，就表明主用主控单元和备用主控单元的状态正常：

- 当设备配置的主控单元为 RPE-X1 时：

```
<H3C> display device
Slot No.  Board type    Status       Primary    SubSlots
```

```
---------------------------------------------------------------------
0         N/A           Absent       N/A        N/A
1         FIP-200       Normal       N/A        2
2         N/A           Absent       N/A        N/A
3         N/A           Absent       N/A        N/A
4         RPE-X1        Normal       Slave      0
5         RPE-X1        Normal       Master     0
6         N/A           Absent       N/A        0
7         N/A           Absent       N/A        N/A
8         FIP-200       Normal       N/A        2
9         N/A           Absent       N/A        N/A
```

- 当设备配置的主控单元为 RSE-X1 时：

```
<H3C> display device
 Slot No.  Board type    Status       Primary    SubSlots
 ---------------------------------------------------------------------
 0         N/A           Absent       N/A        N/A
 1         FIP-210       Normal       N/A        2
 2         N/A           Absent       N/A        N/A
 3         N/A           Absent       N/A        N/A
 5         RSE-X1        Normal       Master     0
 6         RSE-X1        Normal       Slave      0
 7         N/A           Absent       N/A        N/A
 8         FIP-210       Normal       N/A        2
 9         N/A           Absent       N/A        N/A
```

4.6.4 方法四

使用命令 **copy** 将当前主用主控单元上的主启动文件拷贝到备用主控单元，并设置为备用主控单元的主启动文件，然后重启备用主控单元。需要注意的是，采用该方法时，主/备用主控单元的主启动文件版本必须一致。

具体方法如下：

(1) 通过 **dir** 命令查看主用主控单元上是否已经存在待升级的应用程序文件，如果存在，则继续第(2)步操作；如果不存在，则需要先通过 XModem/FTP/TFPT 等传输协议把待升级的应用程序文件下载到主用主控单元上（具体方法请参见本章前半部分内容）；

(2) 使用 **dir** 命令查看备用主控单元 CF 卡的剩余空间，确保备用主控单元 CF 卡的剩余空间能够容纳待升级的应用程序文件；

(3) 使用 **copy** 命令把主用主控单元上的应用程序文件复制到备用主控单元上：

- 当设备配置的主控单元为 RPE-X1 时：

```
<H3C> copy b54D002main.bin slot41#cfa0:/
Copy cfa0:/mainb54d002.bin to slot14#cfa0:/mainb54d002.bin?[Y/N]:y

%Copy file cfa0:/mainb54d002.bin to slot14#cfa0:/mainb54d002.bin...Done.
```

- 当设备配置的主控单元为 RSE-X1 时：

```
<H3C> copy main.bin slot6#cfa0:/
Copy cfa0:/main.bin to slot6#cfa0:/main.bin?[Y/N]:y
```

```
%Copy file cfa0:/main.bin to slot6#cfa0:/main.bin...Done.
```

(4) 指定备用主控单元下次启动时使用的应用程序文件，例如：

- 当设备配置的主控单元为 RPE-X1 时：

```
<H3C> boot-loader file slot14#cfa0:/mainb54d002.bin slot 14 main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 14!
```

- 当设备配置的主控单元为 RSE-X1 时：

```
<H3C> boot-loader file slot6#cfa0:/main.bin slot 6 main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 6!
```

(5) 使用命令 **display boot-loader** [**slot** *slot-number*] 查看主用主控单元和备用主控单元用于启动的应用程序文件信息，确保二者一致。

(6) 在系统视图下使用命令 **slave restart** 重启备用主控单元之后，先用命令 **display version** 查看设备当前的版本，如果版本号是要升级的版本就表明升级成功。

(7) 通过 **display device** 命令查看主用主控单元和备用主控单元的状态：

- 当设备配置的主控单元为 RPE-X1 时：如果主用主控单元的状态是 master，备用主控单元的状态是 ok，就表明主用主控单元和备用主控单元的状态正常（缺省情况下，Board5 为主用主控单元，Board4 为备用主控单元）。

```
<H3C> display device
Slot No.  Board type    Status       Primary    SubSlots
 ------------------------------------------------------------------------
 0         N/A           Absent       N/A        N/A
 1         FIP-200       Normal       N/A        2
 2         N/A           Absent       N/A        N/A
 3         N/A           Absent       N/A        N/A
 4         RPE-X1        Normal       Slave      0
 5         RPE-X1        Normal       Master     0
 6         N/A           Absent       N/A        0
 7         N/A           Absent       N/A        N/A
 8         FIP-200       Normal       N/A        2
 9         N/A           Absent       N/A        N/A
```

- 当设备配置的主控单元为 RSE-X1 时：如果主/备用主控单元的状态均是 Normal，就表明主/备用主控单元的状态正常（缺省情况下，Board5 为主用主控单元，Board6 为备用主控单元）。

```
<H3C> display device
 Slot No.  Board type    Status       Primary    SubSlots
 ------------------------------------------------------------------------
 0         N/A           Absent       N/A        N/A
 1         FIP-210       Normal       N/A        2
 2         N/A           Absent       N/A        N/A
 3         N/A           Absent       N/A        N/A
 5         RSE-X1        Normal       Master     0
 6         RSE-X1        Normal       Slave      0
```

```
 7         N/A           Absent       N/A        N/A
 8         FIP-210       Normal       N/A        2
 9         N/A           Absent       N/A        N/A
```

4.7 应用程序以及配置文件的维护

对文件类型的修改、显示等操作可以在文件控制子菜单下完成。

在 BootWare 主菜单下键入<4>，系统会进入文件控制子菜单。系统显示如下：

```
========================<File CONTROL>=======================
|Note:the operating device is cfa0                           |
| <1> Display All File(s)                                    |
| <2> Set Application File type                              |
| <3> Set Configuration File type                            |
| <4> Delete File                                            |
| <0> Exit To Main Menu                                      |
============================================================
Enter your choice(0-3):
```

4.7.1 显示所有文件

1. 在 BootWare 菜单中显示所有文件

在文件控制子菜单下键入<1>，系统提示：

```
Display all file(s) in cfa0:
 'M' = MAIN      'B' = BACKUP      'S' = SECURE    'N/A' = NOT ASSIGNED
============================================================================
NO. Size(B)    Time                 Type    Name
1   941352     Nov/26/2007 12:02:42 N/A     cfa0:/logfile/logfile.log
2   24802996   Nov/04/2007 17:03:26 N/A     cfa0:/sr6608.bin
3   1355       Nov/04/2007 17:22:12 N/A     cfa0:/startup.cfg
4   1597       Nov/04/2007 10:07:10 N/A     cfa0:/~/startup.cfg
5   24802996   Nov/11/2007 15:11:42 M       cfa0:/main.bin
============================================================================
```

2. 在命令行中显示所有文件

```
<H3C> dir
Directory of cfa0:/

   0     drw-          -  Nov 28 2000 04:09:30   logfile
   1     -rw-   24802996  Nov 04 2007 17:03:26   sr6608.bin
   2     -rw-       1355  Nov 04 2007 17:22:12   startup.cfg
   3     -rw-   24802996  Nov 13 2037 13:21:20   main.bin

505480 KB total (456576 KB free)

File system type of cfa0: FAT16
```

表4-16 命令 dir 的显示信息说明

显示信息	说明
Directory of cfa0:/	当前目录名
505480 KB total (456576 KB free)	CF 卡已用空间（剩余空间）
File system type of cfa0: FAT16	CF 卡的文件系统类型



说明

命令 **dir** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。

4.7.2 设置应用程序文件的属性

1. 在 BootWare 菜单中设置应用程序文件的属性

应用程序文件的属性可以在BootWare菜单中修改，也可以在应用程序启动后通过命令修改；安全文件的属性不允许修改。关于应用程序文件各种属性的详细说明，请参见 4.1.3 应用程序文件。

在文件控制子菜单下键入<2>，系统提示：

```
'M' = MAIN      'B' = BACKUP      'S' = SECURE    'N/A' = NOT ASSIGNED
============================================================================
NO. Size(B)   Time                 Type   Name
|1   24802996  Nov/04/2007 17:03:26 B      cfa0:/main.bin           |
|2   24802996  Nov/11/2007 15:11:42 M      cfa0:/main_bak.bin       |
|0   Exit
============================================================================
Enter file no:
```

输入文件号，这里选择 2，系统将提示如下：

```
Modify the file attribute:
============================================================================
| <1> +Main                                                              |
| <2> -Main                                                              |
| <3> +Backup                                                            |
| <4> -Backup                                                            |
| <0> Exit                                                               |
Enter your choice(0-4):
```

键入 1～4 可以对操作的文件设置为M属性、取消M属性、设置为B属性和取消B属性，文件属性的详细说明请参见 4.1 路由器的软件组成及作用。

本例选择<1>，系统将属性为 B 的文件 main.bin 设置成属性为 M+B 的文件。

```
Set the file attribute success!
```

此时，原来属性为 M 的文件 main_bak.bin 会自动变成属性为 N/A 的文件。

2. 在命令行中设置应用程序文件的属性

以上一小节中的文件 main.bin 和 main_bak.bin 为例。

将文件 main.bin 的属性由 B 设置成 M+B。

```
<H3C> boot-loader file main.bin slot 1 main
  This command will set the boot file of the specified board. Continue? [Y/N]:
  The specified file will be used as the main boot file at the next reboot!
```

此时，文件 main.bin 的属性变成了 M+B，并且成为设备下次启动时的主启动文件。

此外，CF 卡中的另一个应用程序文件 main_bak.bin 的属性也由 M 自动变成了 N/A。

通过文件控制子菜单查看所有文件，可以验证这两个文件属性的改变。

在文件控制子菜单下键入<2>，系统提示：

```
'M' = MAIN      'B' = BACKUP      'S' = SECURE    'N/A' = NOT ASSIGNED
======================================================================
|NO. Size(B)    Time                 Type    Name                      |
|1   24802996   Nov/04/2007 17:03:26 M+B     cfa0:/main.bin            |
|2   24802996   Nov/11/2007 15:11:42 N/A     cfa0:/main_bak.bin        |
|0   Exit                                                              |
======================================================================
Enter file no:
```



说明

命令 **boot-loader** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。

4.7.3 设置配置文件的属性

1. 在 BootWare 菜单中设置配置文件的属性

在BootWare菜单中可以完成配置文件属性的设置。关于配置文件的main/backup属性的详细说明，请参见 4.1.4 配置文件。

在文件控制子菜单下键入<3>，系统将提示：

```
'M' = MAIN      'B' = BACKUP      'S' = SECURE      'N/A' = NOT ASSIGNED
========================================================================
|NO. Size(B)    Time                 Type    Name                      |
|1   734        Dec/17/2000 11:47:20 N/A     cfa0:/startup.cfg         |
|2   760        Dec/17/2000 18:49:50 N/A     cfa0:/startup_bak2.cfg    |
|3   810        Dec/17/2000 18:49:26 M       cfa0:/startup_bak1.cfg    |
|4   850        Dec/17/2000 18:50:08 N/A     cfa0:/startup_bak3.cfg    |
|0   Exit                                                              |
========================================================================
Enter file No:
```

输入文件号，这里选择 1，系统接着提示：

```
Modify the file attribute:
========================================================================
|<1> +Main                                                             |
|<2> -Main                                                             |
|<3> +Backup                                                           |
|<4> -Backup                                                           |
|<0> Exit                                                              |
========================================================================
```

```
Enter your choice(0-4):
```

键入 1～4，可以将选定的文件设置为 M 属性、取消 M 属性、设置为 B 属性和取消 B 属性。

本例选择<1>，系统将属性标识为 N/A 的文件 startup.cfg 设置成属性为 M 的文件。

```
Set the file attribute success!
```

此时，原来属性为 M 的文件 startup_bak1.cfg 会自动变成属性标识为 N/A 的文件。

 说明

如果在文件控制子菜单中键入<3> Set Configuration File type，直接键入<0> Exit，那么系统将进入主、备配置文件的默认模式，即：主、备配置文件分别为 cfa0:/main.cfg、cfa0:/backup.cfg，此时如果 cfa0 中恰好有上述同名文件，将自动拥有对应属性，否则对应属性的文件不存在。

---

2. 在命令行中设置配置文件的属性

可以使用命令 **startup saved-configuration** *cfgfile* [**backup** | **main**]来配置系统下次启动时使用的配置文件及其属性。其中，参数 *cfgfile* 表示配置文件的文件名。

需要注意的是：

- 对属性标识为 N/A 的配置文件，在执行 **startup saved-configuration** *cfgfile* **main** 命令之后，它具有主用属性；
- 对属性标识为 N/A 的配置文件，在执行 **startup saved-configuration** *cfgfile* **backup** 命令之后，它具有备用属性；
- 对主用配置文件，在执行 **startup saved-configuration** *cfgfile* **backup** 命令之后，它具有主用和备用两种属性；
- 对备用配置文件，在执行 **startup saved-configuration** *cfgfile* **main** 命令之后，它具有主用和备用两种属性；
- 不指定 main/backup 属性的情况下，设置的是主用配置文件；
- 在设置配置文件的主用属性时，如果系统中已有主用配置文件，则将原主用配置文件的主用属性擦除，使系统中只有唯一的主用配置文件；
- 在设置配置文件的备用属性时，如果系统中已有备用配置文件，则将原备用配置文件的备用属性擦除，使系统中只有唯一的备用配置文件；
- 在执行 **undo startup saved-configuration** 命令之后，系统还会擦除配置文件的主、备用属性（即执行该命令后，系统中不再有主用配置文件和备用配置文件）。

配置系统下次启动时使用的备用配置文件。

```
<H3C> startup saved-configuration startup.cfg backup
Please wait......Done!
```

 说明

命令 **startup saved-configuration** *cfgfile* 的详细说明，请参见《H3C SR6600 路由器 用户手册》。

4.7.4 删除文件

1. 在 BootWare 菜单中删除文件

在文件控制子菜单下键入<3>，系统将提示：

```
Deleting the file in cfa0:
 'M' = MAIN      'B' = BACKUP      'S' = SECURE   'N/A' = NOT ASSIGNED
========================================================================
NO. Size(B)   Time                 Type   Name
1   941352    Nov/26/2000 12:02:42 N/A    cfa0:/logfile/logfile.log
2   24802996  Nov/04/2007 17:03:26 N/A    cfa0:/sr6608.bin
3   1355      Nov/04/2007 17:22:12 N/A    cfa0:/startup.cfg
4   1597      Nov/04/2007 10:07:10 N/A    cfa0:/~/startup.cfg
5   24802996  Nov/11/2055 15:11:42 M      cfa0:/main.bin
0   Exit
========================================================================
Enter file no:
```

输入文件号，这里选择 4，系统接着提示：

```
The file you selected is cfa0:/~/startup.cfg,Delete it? [Y/N]
```

键入[Y]，

系统如果出现如下提示信息，则表明文件删除成功。

```
Deleting.........Done!
```



注意

删除的时候系统不会提示确认，请在进行删除操作前确认无误！

---

2. 在命令行中删除文件

在用户视图下键入命令 **delete** [**/unreserved**] *file-url*，可以删除文件。其中，

- **/unreserved** 表示彻底删除该文件；
- file-url 表示要删除的文件名。

删除根目录下的文件 test.txt。

```
<H3C> delete test.txt
Delete cfa0:/test.txt?[Y/N]:y
%Delete file cfa0:/test.txt...Done.
```

此时，文件test.txt暂时被删除到了回收站中，如果想恢复回收站里的文件，则可以通过命令**undelete**来实现。

从回收站中恢复被删除的文件 test.txt。

```
<H3C> undelete test.txt
Undelete cfa0:/test.txt? [Y/N]:y
% Undeleted file cfa0:/test.txt.
```

说明

命令 **delete** 和 **undelete** 的详细说明，请参见《H3C SR6600 路由器 用户手册》。

4.8 口令丢失的处理

如果路由器的 BootWare 口令、用户口令和 Super Password 丢失，可以采用如下方法解决。

4.8.1 BootWare 口令丢失

如果路由器的 BootWare 口令丢失，请与代理商联系。

BootWare 口令修改是在 BootWare 主菜单下实现的。在 BootWare 主菜单下键入<5>，按提示更改密码。

配置终端显示如下：

```
please input old password:
```

输入旧密码：

```
please input old password: ******
```

说明

- 如果旧密码输入错误，那么系统会提示再次输入旧密码: Wrong password,Please input password again:。
- 如果连续输入 3 次均错误，系统会提示：Wrong password,system halt，即系统挂死。

旧密码输入完毕后，系统会提示输入两遍新密码，以保证输入的新密码前后一致：

```
Please input new password: ******
Please input new password again: ******
```

如果系统出现如下提示信息，则表示新密码设置成功：

```
Password Set Successfully.
```

说明

- 输入的 BootWare 口令以“**”形式回显。
- 口令长度支持 32 个字符，如果输入的口令超过 32 个字符，系统将自动截取前 32 个字符作为口令。

4.8.2 用户口令丢失

用户口令的丢失会使用户无法进入系统，这时可以采用忽略系统配置的方式启动。请按如下步骤操作：

(1) 进入 BootWare 主菜单，选择<6>，即以忽略系统配置方式启动。

系统出现如下提示：

```
Flag Set Success.
```

表明系统已经设置成功。

(2) 当再次出现 BootWare 主菜单时，选择<0>，系统开始重新启动。

```
System is rebooting now.

System start booting...
Booting Normal Extend BootWare....
```

(3) 重启后在系统视图下设置新的用户口令。

```
<H3C> system-view
[H3C] user-interface console 0
[H3C-ui-console0] authentication-mode password
[H3C-ui-console0] set authentication password simple 123456
```

以上就表示设置 CONSOLE 口验证方式为密码验证，且为 CONSOLE 口设置密码为 123456，密码采用明文存储。

用命令 **set authentication password { cipher | simple }** *password* 设置密码时：

- 如果使用参数 **cipher**，则密码以密文方式存储，用命令 **display current-configuration** 是不可以从当前配置中看到密码的；
- 如果使用参数 simple，则密码以明文方式存储，用命令 display current-configuration 可以从当前配置中看到密码。



说明

- 重启后，系统按初始缺省配置运行，但原配置文件仍保存在存储介质中。为了恢复原配置，可使用 **display saved-configuration** 命令将原配置显示出来，然后拷贝并执行这些配置。
- 密码采用明文存储时用命令 **display current-configuration** 可以从当前配置中看到密码，设置密码时用命令设置 **set authentication password cipher** 123456 后，密码会加密存储。

---

(4) 保存新配置。

```
[H3C] save
```



说明

- 修改用户口令后应执行 **save** 命令，以保存修改。
- 建议用户将所作的修改保存到默认的配置文件。

4.8.3 Super Password 口令丢失

Super Password 可以使用户在四个 Super 等级的权限中切换，Super Password 丢失会使用户无法进行一些权限较高的操作。

在 BootWare 主菜单下键入<8>，可以清除 Super Password：

```
==================<EXTEND-BOOTWARE MENU>=====================
| <1> Boot System                                           |
| <2> Enter Serial SubMenu                                  |
| <3> Enter Ethernet SubMenu                                |
| <4> File Control                                          |
| <5> Modify BootWare Password                              |
| <6> Skip Current System Configuration                     |
```

```
| <7> BootWare Operation Menu                                |
| <8> Clear Super Password                                   |
| <9> Storage Device Operation                               |
| <0> Reboot                                                 |
============================================================
Enter your choice(0-9):8
```

如果系统出现如下提示信息，则表明清除 Super Password 成功。

```
Clear Application Password Success
```



说明

- 选择清除 Super Password 后再退出并重启路由器时，用户可以直接进入系统视图。
- 设置该选项后，仅在第一次重启路由器时生效，第二次重启路由器后超级用户口令将恢复。

4.9 备份和恢复 BootWare

在 BootWare 操作子菜单中可以完成 BootWare 的备份和恢复。

在BootWare主菜单下键入<7>，可以进入BootWare操作子菜单（BootWare操作子菜单的详细说明，请参见 4.2.5 BootWare操作子菜单。）

4.9.1 备份完整 BootWare

备份完整的 BootWare 需要先备份 BootWare 基本段，然后再备份 BootWare 扩展段。

在 BootWare 操作子菜单中键入<1>，系统会提示如下：

```
Will you backup the Basic BootWare? [Y/N]
```

这里选择[Y]：

```
Begin to backup the Basic BootWare.....................Done!
```

此时，BootWare 基本段备份完毕。

然后，系统接着提示：

```
Will you backup the Extend BootWare? [Y/N]
```

这里选择[Y]：

```
Begin to backup the Extend BootWare.....................Done!
```

这时，BootWare 扩展段也备份完毕。



说明

- BootWare 基本段和扩展段均备份到 Flash 中。
- 通过 BootWare 子菜单备份完整的 BootWare，备份的是槽位 0 或者 1 中的当前主控板的 BootWare。

4.9.2 恢复完整 BootWare

在 BootWare 操作子菜单中键入<2>，会把备份到 Flash 中的 BootWare 恢复到系统中。恢复完整的 BootWare 时，需要先恢复 BootWare 基本段，然后再恢复 BootWare 扩展段。

```
Will you restore the Basic BootWare? [Y/N]
```

这里选择[Y]：

```
Begin to restore Normal Basic BootWare.....................Done!
```

此时，BootWare 基本段恢复完毕。

然后，系统接着提示：

```
Will you restore the Extend BootWare? [Y/N]
```

这里选择[Y]：

```
Begin to restore Normal Extend BootWare.....................Done!
```

这时 BootWare 扩展段也恢复完毕。

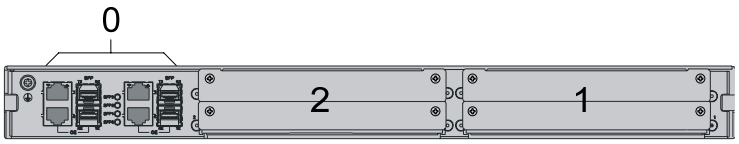
5 槽位顺序及接口编号

5.1 SR6602 槽位排列顺序及接口编号

5.1.1 槽排列顺序

SR6602 支持多种接口，包括 Console 口、AUX 口、GE 接口、Serial 口（同步串口）和 POS 接口等，每种接口在配置时按编号排列。

图5-1 SR6602 的插槽排列顺序



说明

图 5-1中的数字“0～2”分别代表设备的槽位编号Slot0～Slot2，实物上并没有这六个丝印标志。

5.1.2 接口编号方法

SR6602 的接口采用“二维”编号方法，原则如下：

(1) 接口编号为 *interface-type X/Y*，其中：

- *interface-type*：表示接口类型，如 GigabitEthernet 和 Serial 等。
- *X*：表示槽位号，即模块所处的插槽号，取值范围为 0～2，其中 0 为固定 GE 口的槽位号，1 和 2 是 HIM/MIM 所在的槽位号如：1 或者 2。
- *Y*：表示接口序号，即接口在模块上的编号。

(2) 同一模块上的不同接口槽位号 *X* 相同。

(3) 每种接口的序号 *Y* 均从 0 开始，并按照模块上的编号排列（从左至右）。

5.1.3 举例说明

(1) SR6602 固定 GigabitEthernet 接口的编号为：

- GigabitEthernet 0/0
- GigabitEthernet 0/1
- GigabitEthernet 0/2
- GigabitEthernet 0/3

(2) 如果 SR6602 的 Slot1 和 Slot2 分别安装了一块 2GBE、2SAE 模块，那么 GigabitEthernet 接口和 Serial 接口的编号分别为：

- Slot1：GigabitEthernet 1/0、GigabitEthernet 1/1
- Slot2：Serial 2/0、Serial2/1。

5.2 SR6604/SR6608 路由器的槽位及编号

5.2.1 插槽排列顺序

SR6604/6608 支持多种接口，包括 CONSOLE 口、AUX 口、GE 接口、Serial 口（同步串口）、POS 接口和 E1 口等，每种接口在配置时按编号排列。

图5-2 SR6604 的插槽排列顺序

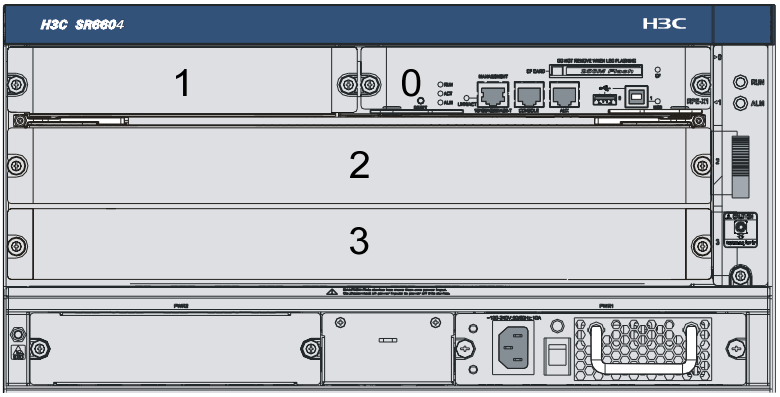
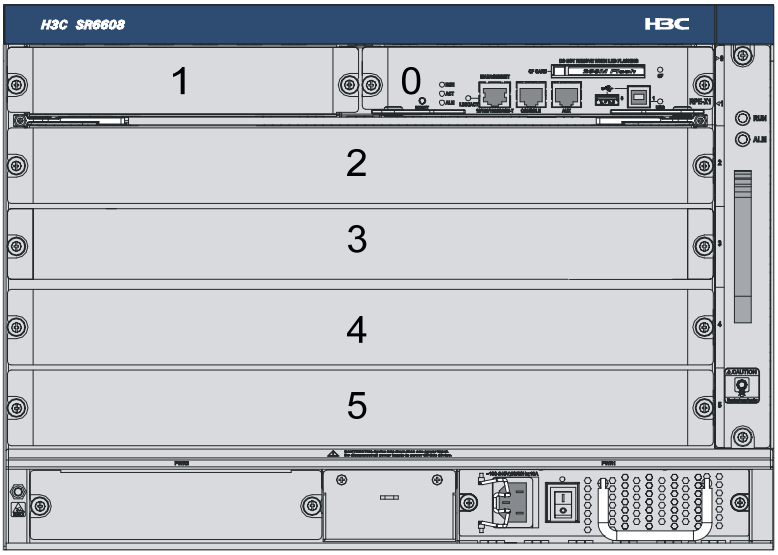


图5-3 SR6608 的插槽排列顺序



说明

图 5-3中的数字 “0～5” 分别代表SR6608 的槽位编号Slot0～Slot5，图 5-2上的数字 “0～3” 分别代表SR6604 的槽位编号Slot0～Slot3，实物上并没有这些丝印标志。

5.2.2 接口编号方法

SR6604 的 Slot2～Slot3、SR6608 的 Slot2～Slot5 需要先安装一个 FIP-100/FIP-200，才能够安装 HIM/MIM。其中，FIP-200 既支持 HIM，也支持 MIM，而 FIP-100 只支持 MIM。

SR6604/6608 的接口采用 “三维” 编号方法，原则如下：

接口编号为 *interface-type* X/Y/Z，其中：

- *interface-type*：表示接口类型，如 GigabitEthernet 和 Serial 等。
- *X*：表示槽位号，即业务板 FIP-100/FIP-200 在设备上的槽位号，如：1 或者 2。
- *Y*：表示子槽位，即 HIM/MIM 在业务板上的槽位号。
- *Z*：表示接口序号，即接口在 HIM/MIM 上的编号。

需要注意的是：

- 对于同一块业务板上的不同接口卡，槽位号 *X* 相同。
- 对于同一块 HIM/MIM 上的不同接口，子槽位号 *Y* 相同。
- 每种接口的序号 *Z* 均从 0 开始，并按照模块上的编号排列（从左至右）。
- 主控板无论是安装在 Slot0 还是 Slot1，或者同时安装在 Slot0 和 Slot1 上，因为同一时间只有一块主控板起作用，所以管理以太网口的编号是固定的： M-GigabitEthernet 0/0/0。

5.2.3 举例说明

1. 槽位 3 上的 FIP-100 安装了一块 2GBE 模块

(1) FIP-100 的固定 GigabitEthernet 接口的编号

- GigabitEthernet 3/0/0
- GigabitEthernet 3/0/1


说明

SR6604/6608 业务板上的固定以太网口子槽位号 *Y* 为 0。

---

(2) 2GBE 模块的 GigabitEthernet 接口的编号

若 2GBE 模块安装在 FIP-100 的 Slot1，那么 2GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/1/0
- GigabitEthernet 3/1/1

若 2GBE 模块安装在 FIP-100 的 Slot2，那么 2GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/2/0
- GigabitEthernet 3/2/1

2. 槽位 3 上的 FIP-200 安装了一块 4GBE 模块

(1) FIP-200 的固定 GigabitEthernet 接口编号

- GigabitEthernet 3/0/0
- GigabitEthernet 3/0/1

(2) 4GBE 的 GigabitEthernet 接口编号

若 4GBE 模块安装在 FIP-200 的 Slot1，那么 4GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/1/0
- GigabitEthernet 3/1/1
- GigabitEthernet 3/1/2
- GigabitEthernet 3/1/3

若 4GBE 模块安装在 FIP-200 的 Slot2，那么 4GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/2/0
- GigabitEthernet 3/2/1
- GigabitEthernet 3/2/2
- GigabitEthernet 3/2/3

3. 槽位 2 上的 FIP-200 安装了一块 4SAE 模块

(1) FIP-200 的固定 GigabitEthernet 接口编号

- GigabitEthernet 2/0/0
- GigabitEthernet 2/0/1

(2) 4SAE 的串口编号

若 4SAE 模块安装在 FIP-200 的 Slot1，那么 4SAE 模块的串口编号为：

- Serial 2/1/0
- Serial 2/1/1
- Serial 2/1/2
- Serial 2/1/3

若 4SAE 模块安装在 FIP-200 的 Slot2，那么 4SAE 模块的串口编号为：

- Serial 2/2/0
- Serial 2/2/1
- Serial 2/2/2
- Serial 2/2/3

5.3 SR6616 路由器的槽位及编号

5.3.1 插槽排列顺序

设备支持多种接口，包括 CONSOLE 口、AUX 口、GE 接口、Serial 口（同步串口）、POS 接口和 E1 口等，每种接口在配置时按编号排列。

图5-4 SR6616 的插槽排列顺序（配置了 RPE-X1 主控单元）

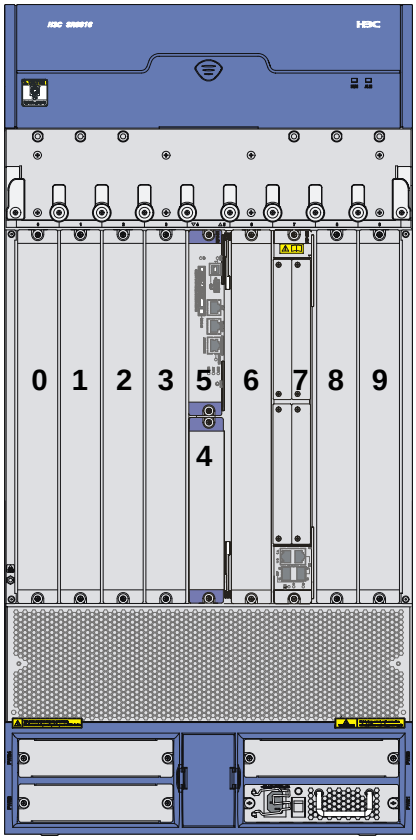
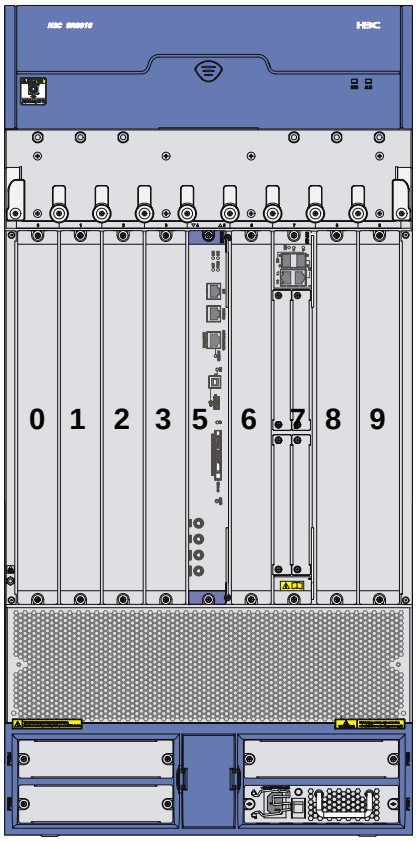


图5-5 SR6616 的插槽排列顺序（配置了 RSE-X1 主控单元）



说明

- 图 5-4中的数字“0～9”分别代表SR6616（配置RPE-X1 时）的槽位编号Slot0～Slot9，实物上并没有这些丝印标志。
- 图 5-5中的数字“0～3、5～9”分别代表SR6616（配置RSE-X1 时）的槽位编号Slot0～Slot3、Slot5～Slot9，实物上并没有这些丝印标志。需要注意的是：SR6616 配置主控单元RSE-X1 时，没有Slot4。

5.3.2 主控单元和灵活接口平台的槽位分布

表5-1 主控单元和灵活接口平台的槽位分布

主控单元	槽位分布
RPE-X1（支持 1+1 冗余备份）	Slot4、Slot5
RSE-X1（支持 1+1 冗余备份）	Slot5、Slot6 （当设备配置的主控单元为 RSE-X1 时，没有 Slot4）
灵活接口平台	**槽位分布**
FIP-100	Slot0～Slot3、Slot6～Slot9
FIP-200	Slot0～Slot3、Slot6～Slot9
FIP-210	Slot0～Slot3、Slot6～Slot9 （其中 Slot6 还可以安装 RSE-X1）

5.3.3 接口编号方法

在安装 HIM/MIM 前，需要先安装灵活接口平台。需要注意的是，FIP-200/FIP-210/既支持 HIM，也支持 MIM，而 FIP-100 只支持 MIM。

设备的接口采用“三维”编号方法，原则如下：

接口编号为 *interface-type* X/Y/Z，其中：

- *interface-type*：表示接口类型，如 GigabitEthernet 和 Serial 等。
- *X*：表示槽位号，即灵活接口平台在设备上的槽位号。
- *Y*：表示子槽位，即 HIM/MIM 在灵活接口平台上的槽位号。
- *Z*：表示接口序号，即接口在 HIM/MIM 上的编号。

需要注意的是：

- 对于同一灵活接口平台上的不同接口卡，槽位号 *X* 相同。
- 对于同一块 HIM/MIM 上的不同接口，子槽位号 *Y* 相同。
- 每种接口的序号 *Z* 均从 0 开始，并按照模块上的编号排列。

5.3.4 举例说明

1. 槽位 3 上的 FIP-100 安装了一块 MIM-2GBE 模块

(1) FIP-100 的固定 GigabitEthernet 接口的编号

- GigabitEthernet 3/0/0
- GigabitEthernet 3/0/1

说明

灵活接口平台上的固定以太网口子槽位号 *Y* 为 0。

(2) MIM-2GBE 模块的 GigabitEthernet 接口的编号

若 MIM-2GBE 模块安装在 FIP-100 的 Slot1，那么 MIM-2GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/1/0
- GigabitEthernet 3/1/1

若 MIM-2GBE 模块安装在 FIP-100 的 Slot2，那么 MIM-2GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/2/0
- GigabitEthernet 3/2/1

2. 槽位 3 上的 FIP-200 安装了一块 MIM-4GBE 模块

(1) FIP-200 的固定 GigabitEthernet 接口编号

- GigabitEthernet 3/0/0
- GigabitEthernet 3/0/1

(2) MIM-4GBE 的 GigabitEthernet 接口编号

若 MIM-4GBE 模块安装在 FIP-200 的 Slot1，那么 MIM-4GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/1/0
- GigabitEthernet 3/1/1
- GigabitEthernet 3/1/2
- GigabitEthernet 3/1/3

若 MIM-4GBE 模块安装在 FIP-200 的 Slot2，那么 MIM-4GBE 模块的 GigabitEthernet 接口编号为：

- GigabitEthernet 3/2/0
- GigabitEthernet 3/2/1
- GigabitEthernet 3/2/2
- GigabitEthernet 3/2/3

3. 槽位 2 上的 FIP-210 安装了一块 MIM-4SAE 模块

(1) FIP-210 的固定 GigabitEthernet 接口编号

- GigabitEthernet 2/0/0
- GigabitEthernet 2/0/1

(2) MIM-4SAE 的串口编号

若 MIM-4SAE 模块安装在 FIP-210 的 Slot1，那么 MIM-4SAE 模块的串口编号为：

- Serial 2/1/0
- Serial 2/1/1
- Serial 2/1/2
- Serial 2/1/3

若 MIM-4SAE 模块安装在 FIP-210 的 Slot2，那么 MIM-4SAE 模块的串口编号为：

- Serial 2/2/0
- Serial 2/2/1
- Serial 2/2/2
- Serial 2/2/3

目 录

1 CPOS接口配置 1-1
1.1 CPOS接口介绍 1-1
1.1.1 SONET/SDH 1-1
1.1.2 CPOS 1-1
1.1.3 SDH的帧结构 1-1
1.1.4 相关术语 1-2
1.1.5 E1/T1/E3/T3 向STM-1 的复用 1-2
1.1.6 E1/T1/E3/T3 通道编号的计算 1-3
1.1.7 开销字节 1-4
1.1.8 CPOS接口应用举例 1-5
1.2 配置CPOS接口 1-6
1.3 配置E1 通道 1-7
1.4 配置T1 通道 1-7
1.5 配置E3 通道 1-8
1.6 配置T3 通道 1-9
1.7 配置接口卡的工作模式 1-10
1.8 CPOS接口显示和维护 1-10
1.9 CPOS接口典型配置举例 1-11
1.9.1 CPOS-E1 典型举例 1-11
1.9.2 CPOS-E3 典型举例 1-12
1.10 CPOS接口故障的诊断与排除 1-13
1.10.1 接口物理状态为up，链路协议状态为down，并且检测到线路环回 1-13

1 CPOS 接口配置

1.1 CPOS 接口介绍

1.1.1 SONET/SDH

SONET（Synchronous Optical Network）是 ANSI 定义的同步传输体制，是一种全球化的标准传输协议，采用光传输。

SDH（Synchronous Digital Hierarchy）是 CCITT（现在的 ITU-T）定义的，采用同步复用方式和灵活的映射结构，可以从 SDH 信号中直接分插出低速的支路信号，而不需要使用大量的复接/分接设备，从而能够减少信号损耗和设备投资。

1.1.2 CPOS

当把 SDH 信号看成由低速信号复用而成时，这些低速支路信号就称为通道。CPOS，即通道化的 POS 接口，它充分利用了 SDH 体制的特点，提供对带宽精细划分的能力，可减少组网中对设备低速物理端口的数量要求，增强设备的低速端口汇聚能力，并提高设备的专线接入能力。

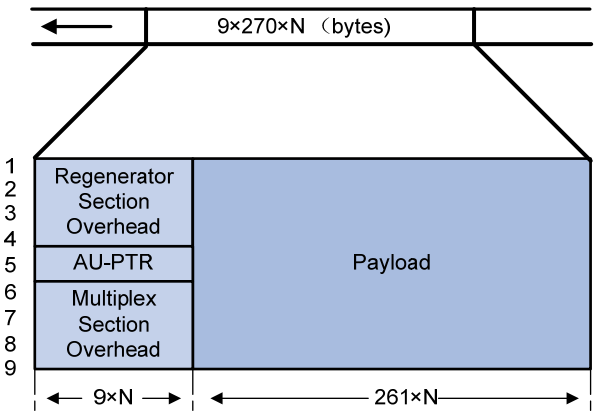
CPOS 接口主要用于提高设备对低速接入的汇聚能力，主要有两种速率 STM-1 和 STM-16。

1.1.3 SDH 的帧结构

下面简单介绍一下 SDH 信号－STM-N 的帧结构。

为方便地从高速信号中直接分/插低速支路信号，应尽可能使低速支路信号在一帧内均匀地、有规律地分布。ITU-T 规定 STM-N 的帧采用以字节为单位的矩形块状结构，如下图所示：

图1-1 STM-N 帧结构



STM-N 是 9 行×270×N 列的块状帧结构，此处的 N 与 STM-N 的 N 相一致，取值范围：1，4，16，……，表示此信号由 N 个 STM-1 信号复用而成。

STM-N 的帧结构由 3 部分组成：段开销（SOH，Section Overhead），包括再生段开销（RSOH，Regenerator Section Overhead）和复用段开销（MSOH，Multiplex Section Overhead）；管理单元指针（AU-PTR）；信息净负荷（payload）。AU-PTR 是指示信息净负荷的第一个字节在 STM-N 帧内位置的指针，以便接收端能根据这个指针正确分离信息净负荷。

1.1.4 相关术语

- 复用单元：SDH 的基本复用单元包括若干容器（C-n）、虚容器（VC-n）、支路单元（TU-n）、支路单元组（TUG-n）、管理单元（AU-n）和管理单元组（AUG-n），其中 n 为单元等级序号。
- 容器（Container）：用来装载各种速率的业务信号的信息结构单元，G.709 定义了 C-11、C-12、C-2、C-3 和 C-4 五种标准容器的规范。
- 虚容器（VC，Virtual Container）：用来支持 SDH 的通道层连接的信息结构单元，是 SDH 通道的信息终端。虚容器分为低阶虚容器和高阶虚容器两类，VC-4 和 AU-3 中的 VC-3 都属于高阶虚容器。
- 支路单元（TU，Tributary Unit）和支路单元组（TUG，Tributary Unit Group）：TU 是提供低阶通道层和高阶通道层之间适配的信息结构。在高阶 VC 净负荷中，固定地占有规定位置的一个或多个 TU 的集合称为 TUG。
- 管理单元（AU，Administration Unit）和管理单元组（AUG，Administration Unit Group）：AU 是提供高阶通道层和复用段层之间适配的信息结构。在 STM-N 的净负荷中，固定地占有规定位置的一个或多个 AU 的集合称为 AUG。

1.1.5 E1/T1/E3/T3 向 STM-1 的复用

在 G.709 建议的 SDH 复用过程中，从一个有效载荷到 STM-N 的复用线路并不唯一，E1、T1、E3 和 T3 向 STM-1 的复用过程分别如下图所示：

图1-2 E1 向 STM-1 的复用过程

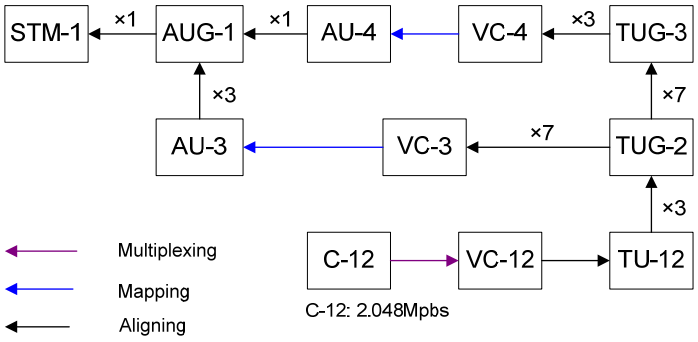


图1-3 T1 向 STM-1 的复用过程

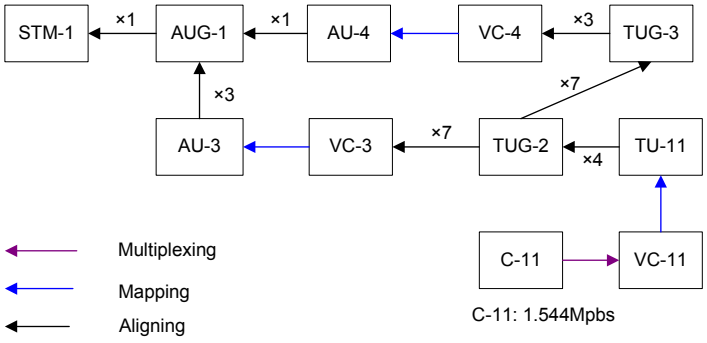


图1-4 E3 向 STM-1 的复用过程

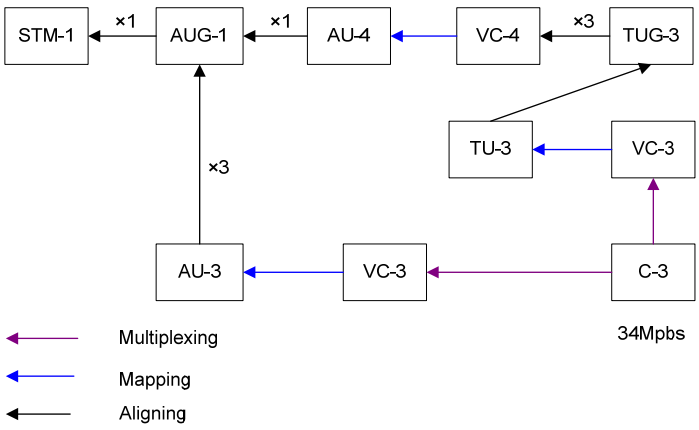
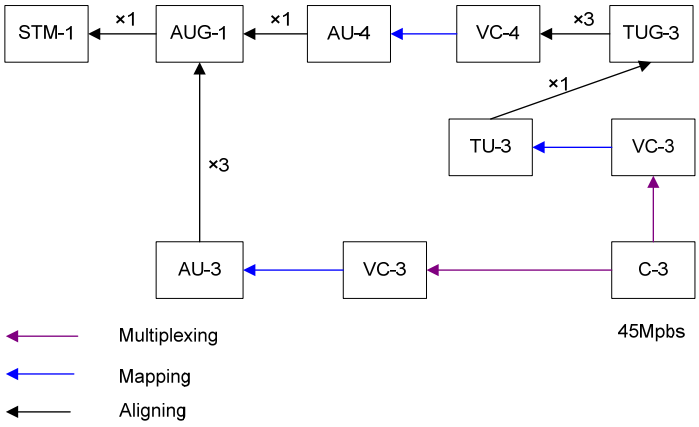


图1-5 T3 向 STM-1 的复用过程



实际应用中，不同的国家和地区可能采用不同的复用路径，为保证互通，系统在 CPOS 接口上提供 **multiplex mode** 命令，使用户可以选择 AU-3 还是 AU-4 复用路径（我国光同步传输网技术体制选用的是 AU-4 的复用路径）。

1.1.6 E1/T1/E3/T3 通道编号的计算

对于 CPOS 接口，由于采用字节间插的复用方式，所以，一个高阶虚容器中的低阶虚容器并不是顺序排列的。为方便用户配置，下面以 CPOS 的 E1 采用 AU-4 复用路径时为例，介绍支路单元 TU 的序号计算方法。

当采用AU-4 的复用路径时，从 图 1-2的复用过程可以看出：2Mbps的复用结构是 3－7－3 结构。计算同一个VC-4 中不同位置TU-12 序号的公式如下：

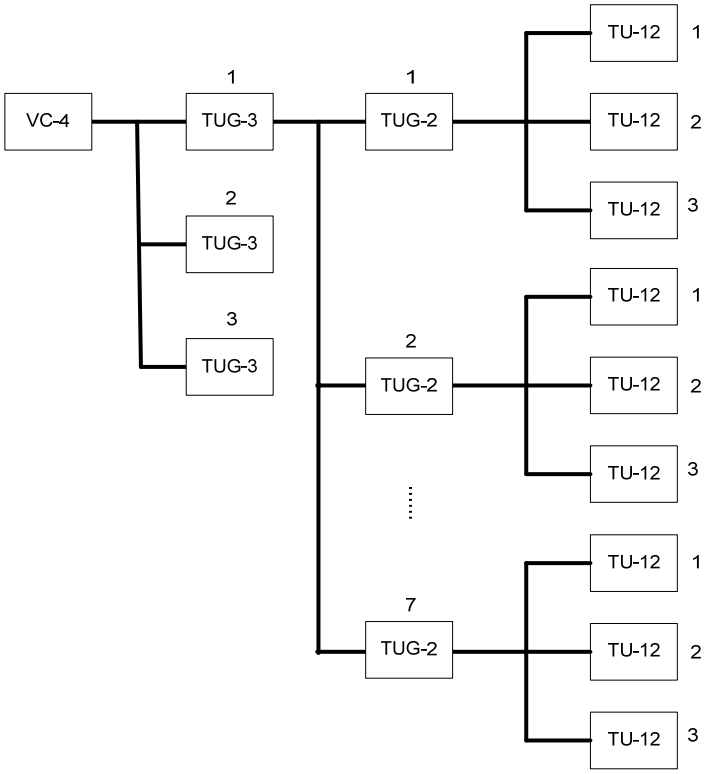
TU-12 序号＝TUG-3 编号＋（TUG-2 编号－1）×3＋（TU-12 编号－1）×21。

在 VC-4 中 TUG-3 编号相同，TUG-2 编号相同，而 TU-12 编号相差为 1 的两个 TU-12 称为位置相邻。

说明

上述公式中的编号是指 VC4 帧中的位置编号，TUG-3 编号范围：1～3；TUG-2 编号范围：1～7；TU-12 编号范围：1～3。TU-12 序号是指本 TU-12 是 VC-4 帧 63 个 TU-12 的按复用先后顺序的第几个 TU-12，也即是第几个 E1 通道。

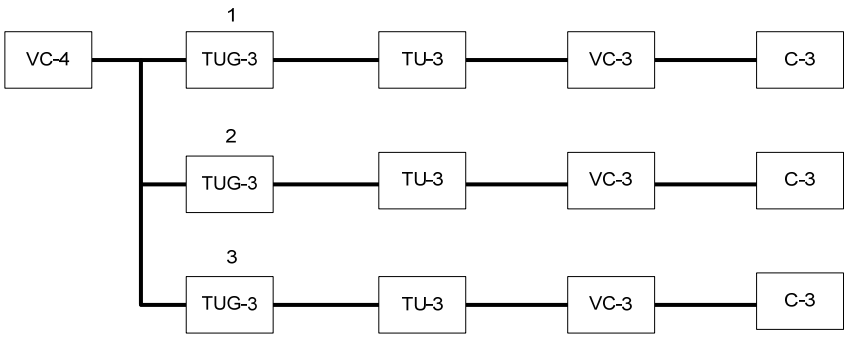
图1-6 VC-4 中 TUG-3、TUG-2、TU-12 的排放顺序



当采用 AU-3 的复用路径时，可同理推导出 TU-12 的序号计算方法。

CPOS 接口配置为 63 个 E1 通道或 84 个 T1 通道时，直接使用编号 1～63 或 1～84 来引用，当本公司设备与其它公司设备的通道化 STM-1 接口共同使用时，需要注意由于对通道的引用方法不同而导致的编号上的差异。

图1-7 VC-4 中 TUG-3、VC-3、C-3 的排放顺序



对于E3/T3 的通道编号计算较它的E1/T1 通道编号的计算简单很多。当采用AU-4 复用路径时，如 图1-7所示，由于TUG-3 和VC-3 中不再进一步分解成低阶通道，所以E3/T3 通道编号等于TUG-3 编号。当采用AU-3 复用路径时，可同理推出E3/T3 的通道编号。

1.1.7 开销字节

SDH 提供层层细化的监控管理功能。具体来讲，监控分为段监控和通道监控，段层次的监控分为再生段和复用段的监控，通道层次的监控分为高阶通道和低阶通道的监控。这些监控功能是通过不同的开销字节实现的。

说明

SDH 的开销字节非常丰富，本节只对在 CPOS 配置过程中用到的几种做简单介绍，如果希望更多了解 SDH 的开销字节，请查阅相关的专业书籍。

- 段开销

段开销（SOH）包括再生段开销（RSOH）和复用段开销（MSOH）。

再生段踪迹字节 J0（Regeneration Section Trace Message）包含在 RSOH 中，该字节被用来重复地发送段接入点标识符（Section Access Point Identifier），以便接收端能据此确认与指定的发送端处于持续连接状态。在同一个运营者的网络内该字节可为任意字符，而在不同两个运营者的网络边界处要使设备收、发两端的 J0 字节匹配。通过 J0 字节可使运营者提前发现和解决故障，缩短网络恢复时间。

- 通道开销

STM-N 帧的 payload 部分包含对低速支路信号进行监控的开销字节 POH（Path Overhead，通道开销字节）。

段开销负责段层的监控功能，而通道开销负责的是通道层的监控功能。通道开销又分为高阶通道开销（Higher-Order Path Overhead）和低阶通道开销（Lower-Order Path Overhead）。

高阶通道开销对 VC-4/VC-3 级别的通道进行监测。

通道踪迹字节 J1（Higher-Order VC-N path trace byte）包含在高阶通道开销中，该字节的作用与 J0 字节类似，被用来重复发送高阶通道接入点标识符，使通道接收端能据此确认与指定的发送端处于持续连接（该通道处于持续连接）状态。要求收发两端 J1 字节匹配。

信号标记字节 C2（Path signal label byte）也包含在高阶通道开销中，C2 用来指示 VC 帧的复接结构和信息净负荷的性质，例如通道是否已装载、所载业务种类和它们的映射方式。要求收发两端 C2 字节匹配。

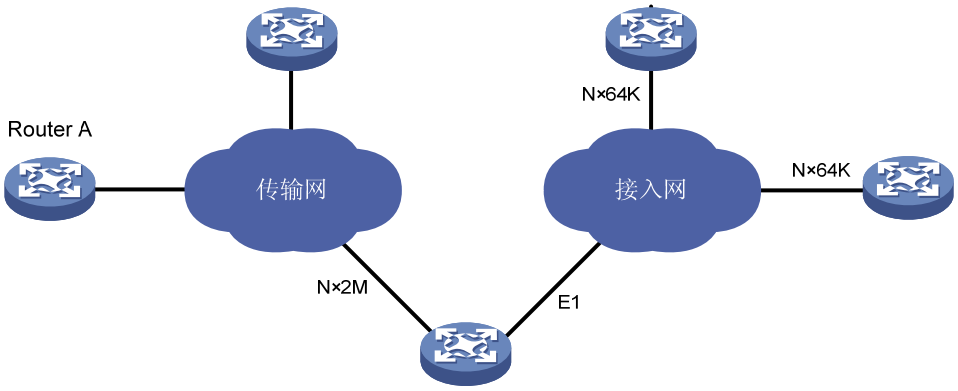
1.1.8 CPOS 接口应用举例

目前，一些政府机关和企事业单位使用中低端设备通过 E1/T1 租用线接入到传输网；而带宽需求介于 E1 和 T3（44M）之间的用户，例如一些数据中心，则同时租用几个 E1/T1。

所有这些用户的带宽经过传输网会聚到一个或者几个 CPOS 接口，再接入到高端设备，高端设备通过时隙唯一识别各低端设备。

实际情况中，CPOS接口与各低端设备之间可能经过不止一级传输网，各低端设备与传输网之间可能还需要其它的传输手段进行中继。这种应用在逻辑上等同于各低端设备分别通过E1/T1 或者n×E1/T1 的专线接入设备Router A，如 图 1-8所示。

图1-8 CPOS 典型应用组网示意图



1.2 配置 CPOS 接口

表1-1 配置 CPOS 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CPOS 的接口视图	**controller cpos** *cpos-number*	必选
配置 CPOS 帧格式为 SDH 类型	**frame-format { sdh | sonet }**	可选 缺省情况下，CPOS 的帧格式为 **sdh**，即应用在 SDH 模式下
配置 CPOS 的时钟模式	**clock { master | slave }**	可选 缺省情况下，CPOS 的时钟模式为从时钟（**slave**）
配置 CPOS 的环回方式	**loopback { local | remote }**	可选 缺省情况下，禁止任何形式的环回
配置 CPOS 的 AUG 复用路径	**multiplex mode { au-3 | au-4 }**	可选 该命令只在 SDH 模式下可用 缺省情况下，AUG 使用 AU-4 复用
配置段开销字节和高阶通道开销字节	**flag { j0** *j0-string* **| j1** *path-number j1-string* **| c2** *path-number c2-value* **| s1** *s1-value* **| s1s0** *path-number s1s0-value* **}**	可选 目前，不支持配置 **s1** 参数
关闭 CPOS 物理端口	**shutdown**	可选 缺省情况下，CPOS 物理接口为打开状态
配置 CPOS 接口的 SD（信号劣化）和 SF（信号失败）的门限值	**threshold { sd | sf }** *value*	可选 缺省情况下，SD 门限值为 10e-6（即 X 取值为 6），SF 门限值为 10e-3（即 X 取值为 3）
配置 E1/E3 通道或者 T1/T3 通道属性	请参见 1.3 配置E1 通道	可选
	请参见 1.4 配置T1 通道	
	请参见 1.5 配置E3 通道	
	请参见 1.6 配置T3 通道	

说明

- 当设备的某物理接口闲置，没有连接电缆时，请使用 **shutdown** 命令禁止该接口，以防止由于干扰导致接口异常。
- 使用 **shutdown** 命令关闭接口会导致接口停止工作，请慎用此命令。

1.3 配置 E1 通道

表1-2 配置 E1 通道

配置步骤		命令	说明
进入系统视图		**system-view**	-
进入指定 CPOS 的接口视图		**controller cpos** *cpos-number*	必选
配置 E1 通道的帧格式		**e1** *e1-number* **set frame-format** { **crc4** | **no-crc4** }	可选 缺省情况下，E1 通道的帧格式为 **no-crc4**
配置 E1 通道的时钟模式		**e1** *e1-number* **set clock** { **master** | **slave** }	可选 缺省情况下，E1 通道的时钟模式为从时钟（**slave**）
配置 E1 通道的环回方式		**e1** *e1-number* **set loopback** { **local** | **payload** | **remote** }	可选 缺省情况下，E1 通道不进行任何形式的环回
配置 E1 通道开销		**e1** *e1-number* **set flag c2** *c2-value* **e1** *e1-number* **set flag j2** { **sdh** | **sonet** } *j2-string*	可选 缺省情况下，**c2** 取值为 02（十六进制），**j2** 循环发送空字符“”
配置E1通道的工作模式（两者必选其一）	配置 E1 工作在非成帧模式	**e1** *e1-number* **unframed**	必选 缺省情况下，E1 不进行通道化
	配置 E1 工作在通道模式并对对 E1 通道的时隙进行捆绑	**undo e1** *e1-number* **unframed**	可选 缺省情况下，E1 工作在通道化模式
		e1 *e1-number* **channel-set** *set-number* **timeslot-list** *range*	必选 缺省情况下，E1 不进行通道化
关闭指定的 E1 通道		**e1** *e1-number* **shutdown**	可选 缺省情况下，E1 通道为打开状态
配置 E1 通道捆绑生成的串口的捆绑模式		**ppp mp soft-binding** **undo ppp mp soft-binding**	可选 缺省情况下，采用硬件捆绑模式

1.4 配置 T1 通道

表1-3 配置 T1 通道

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入指定 CPOS 的接口视图	**controller cpos** *cpos-number*	必选

配置步骤		命令	说明
配置 T1 通道的帧格式		**t1** *t1-number* **set frame-format** { **esf** | **sf** }	可选 缺省情况下， T1 通道的帧格式为 **esf**
配置 T1 通道的时钟模式		**t1** *t1-number* **set clock** { **master** | **slave** }	可选 缺省情况下，T1 通道的时钟模式为从时钟（**slave**）
配置 T1 通道的环回方式		**t1** *t1-number* **set loopback** { **local** | **payload** | **remote** }	可选 缺省情况下，T1 通道不进行任何形式的环回
配置 T1 通道开销		**t1** *t1-number* **set flag c2** *c2-value* **t1** *t1-number* **set flag j2** { **sdh** | **sonet** } *j2-string*	可选 缺省情况下，**c2** 取值为 02（十六进制），**j2** 循环发送空字符“”
配置 T1 通道的工作模式（两者必选其一）	配置 T1 工作在非成帧模式	**t1** *t1-number* **unframed**	必选 缺省情况下，T1 不进行通道化
	配置 T1 工作在通道模式	**undo t1** *t1-number* **unframed**	可选 缺省情况下，T1 工作在通道化模式
		t1 *t1-number* **channel-set** *set-number* **timeslot-list** *range* [**speed** { **56k** | **64k** }]	必选 缺省情况下，T1 不进行通道化
关闭指定的 T1 通道		**t1** *t1-number* **shutdown**	可选 缺省情况下，T1 通道为打开状态
配置 T1 通道捆绑生成的串口的捆绑模式		**ppp mp soft-binding** **undo ppp mp soft-binding**	可选 缺省情况下，采用硬件捆绑模式

1.5 配置 E3 通道

表1-4 配置 E3 通道

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入指定 CPOS 的接口视图	**controller cpos** *cpos-number*	必选
创建非成帧模式的 E3 通道对应的串口	**using e3** *e3-number*	可选 缺省情况下，未创建串口
配置 vc-3 帧的开销字节	**flag vc-3** *path-number* { **c2** *c2-value* | **j1** { **sdh** *sdh-string* | **sonet** *sonet-string* } | **s1s0** *s1s0-value* }	可选 本命令中 c2、j1 和 s1s0 参数的支持情况与设备的型号有关，请以设备的实际情况为准
配置 vc-4 帧的开销字节	**flag vc-4** *path-number* { **c2** *c2-value* | **j1** { **sdh** *sdh-string* | **sonet** *sonet-string* } | **s1s0** *s1s0-value* }	可选 本命令中 c2、j1 和 s1s0 参数的支持情况与设备的型号有关，请以设备的实际情况为准
配置指定的 E3 通道工作在 FE3 模式，并配置 DSU 模式或子速率	**fe3** *e3-number* { **dsu-mode** { **0** | **1** } | **subrate** *sub-number* }	可选 缺省情况下，DSU 模式为 1，即 Kentrox 模式；子速率为 34010kbps
配置 E3 通道的时钟模式	**e3** *e3-number* **set clock** { **master** | **slave** }	缺省情况下，E3 通道的时钟模式为从时钟（slave）

配置步骤	命令	说明
配置 E3 通道的环回模式	**e3** *e3-number* **set loopback** { **local** | **payload** | **remote** }	缺省情况下，未进行任何形式的环回
配置 E3 通道的 national bit 通信码	**e3** *e3-number* **set national-bit** { **0** | **1** }	缺省情况下，E3 通道 national bit 为 1
关闭 E3 通道	**e3** *e3-number* **shutdown**	缺省情况下，E3 通道处于开启状态



说明

E3/T3 通道支持部分（Fractional）化也称之为 FE3/FT3。这种非标准的应用，可以根据带宽需求将其速率变化为不同等级，称之为子速率。

1.6 配置 T3 通道

表1-5 配置 T3 通道

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入指定 CPOS 的接口视图	**controller cpos** *cpos-number*	必选
创建非成帧模式的T3通道对应的串口	**using t3** *t3-number*	可选 缺省情况下，未创建串口
配置T3通道的告警信号检测与发送功能	**t3** *t3-number* **set alarm** { **detect | generate** { **ais** | **febe** | **idle** | **rai** } }	可选 缺省情况下，T3 通道的告警信号检测与发送功能都未使能。
配置进行线路位（Bit）错误率的测试	**t3** *t3-number* **set bert pattern** { **2^7** | **2^11** | **2^15** | **qrss** } **time** *number*	可选 缺省情况下，不进行线路位错误率的测试
配置T3通道的时钟模式	**t3** *t3-number* **set clock** { **master** | **slave** }	可选 缺省情况下，T3 通道的时钟模式为从时钟（**slave**）
配置T3接口的FEAC链路信号的检测和传输功能	**t3** *t3-number* **set feac** { **detect** | **generate** { **ds3-los** | **ds3-ais** | **ds3-oof** | **ds3-idle** | **ds3-eqptfail** | **loopback** { **ds3-line** | **ds3-payload** } } }	可选 缺省情况下，T3 接口的 FEAC 链路信号的检测和传输功能都未使能
配置T3接口所使用的帧格式	**t3** *t3-number* **set frame-format** { **c-bit | m23** }	可选 缺省情况下，T3 接口的帧格式为 C-bit Parity
配置T3通道的环回模式	**t3** *t3-number* **set loopback** { **local** | **payload** | **remote** }	可选 缺省情况下，T3 通道不进行任何形式的环回
配置 T3 通道的 MDL 链路消息检测与传输功能	**t3** *t3-number* **set mdl** { **data** { **eic** *string* | **fic** *string* | | **gen-no** *string* | **lic** *string* | **pfi** *string* | **port-no** *string* | **unit** *string* } | **detect** | **generate** { **idle-signal** | **path** | **test-signal** } }	可选 本命令中 T3 通道的 MDL 链路消息检测与传输功能支持情况与设备的型号有关，请以设备的实际情况为准
配置 vc-3 帧的开销字节	**flag vc-3** *path-number* { **c2** *c2-value* | **j1** { **sdh** *sdh-string* | **sonet** *sonet-string* } | **s1s0** *s1s0-value* }	可选 本命令中 c2、j1 和 s1s0 参数的支持情况与设备的型号有关，请以设备的实际情况为准

配置步骤	命令	说明
配置 vc-4 帧的开销字节	**flag vc-4** *path-number* { **c2** *c2-value* | **j1** { **sdh** *sdh-string* | **sonet** *sonet-string* } | **s1s0** *s1s0-value* }	可选 本命令中 c2、j1 和 s1s0 参数的支持情况与设备的型号有关，请以设备的实际情况为准
配置 E3 通道工作在 FT3 模式，并配置 DSU 模式或子速率	**ft3** *t3-number* { **dsu-mode** { **0** | **1** | **2** | **3** | **4** } | **subrate** *sub-number* }	缺省情况下，DSU 模式为 0，即 Digital Link 模式；子速率为 44210kbps

说明

E3/T3 通道支持部分（Fractional）化也称之为 FE3/FT3。这种非标准的应用，可以根据带宽需求将其速率变化为不同等级，称之为子速率。

1.7 配置接口卡的工作模式

部分 CPOS 卡支持一卡多用，通过配置卡的工作模式，可以完成整个卡的工作模式的切换。

支持工作模式切换的 CPOS 接口卡包括 HIM-CL1P、HIM-CL2P、HIM-CLS1P 和 HIM-CLS2P 四种。

说明

有关接口卡的工作模式的介绍和配置，请参见“系统分册”中的“设备管理”。

1.8 CPOS 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 CPOS 接口的运行情况，通过查看显示信息验证配置的效果。

表1-6 CPOS 接口显示和维护

操作	命令
显示 CPOS 的所有通道信息	**display controller cpos** [*cpos-number*]
显示 CPOS 的指定 E1 通道信息	**display controller cpos** *cpos-number* **e1** *e1-number*
显示 CPOS 的指定 T1 通道信息	**display controller cpos** *cpos-number* **t1** *t1-number*
显示 CPOS 的指定 E3 通道信息	**display controller cpos** *cpos-number* **e3** *e3-number*
显示 CPOS 的指定 T3 通道信息	**display controller cpos** *cpos-number* **t3** *t3-number*
显示 E1/T1 串口信息	**display interface serial** *interface-number/channel-number:set-number*
清除 CPOS 接口的计数器	**reset counters controller cpos** *interface-number*

说明

display interface serial 命令请参见“接入分册”中的“WAN 接口命令”。

1.9 CPOS 接口典型配置举例

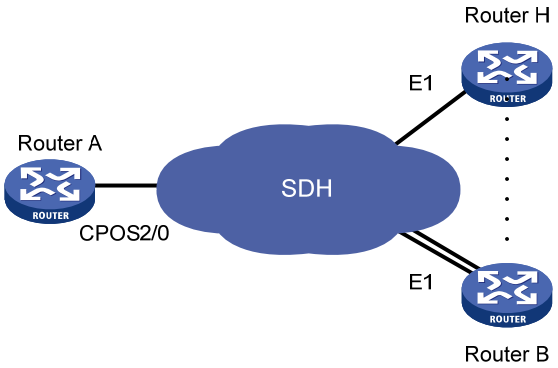
1.9.1 CPOS-E1 典型举例

1. 组网需求

Router A 中心节点下属有 B～H 七个分支节点，每个分支节点设备通过 E1 链路上行连接到中心节点设备；设备通过 CPOS 接口汇聚上行的链路。由于 Router B 分支节点进行了扩容后，一条 E1 链路满足不了需求，因此添加了一条 E1 链路。要求用 MP-group 接口的方式，对这两条 E1 链路进行捆绑。

2. 组网图

图1-9 CPOS-E1 配置组网图



3. 配置步骤

(1) 配置 Router A

此处只给出 CPOS 接口和 E1 接口的关键配置步骤，关于其他一些业务的部署，在此不再赘述。

将 CPOS 接口和其通道化出来的 E1 接口配置为主时钟模式

```
<Sysname> system-view
[Sysname] controller cpos 2/0
[Sysname-Cpos2/0] clock master
[Sysname-Cpos2/0] e1 1 unframed
[Sysname-Cpos2/0] e1 1 set clock master
[Sysname-Cpos2/0] e1 2 unframed
[Sysname-Cpos2/0] e1 2 set clock master
```

创建 Mp-group 接口，配置相应的 IP 地址。

```
[Sysname] interface mp-group 1
[Sysname-Mp-group1] ip address 10.1.1.1 30
[Sysname-Mp-group1] quit
```

配置串口 Serial2/0/1:0

```
[Sysname] interface serial 2/0/1:0
[Sysname-Serial2/0/1:0] ppp mp mp-group 1
[Sysname-Serial2/0/1:0] quit
[Sysname] interface serial 2/0/0/2:0
[Sysname-Serial2/0/2:0] ppp mp mp-group 1
```

```
[Sysname-Serial2/0/2:0] quit
```

(2)　配置 Router B

各个分支节点配置与 Router B 配置类似，配置过程省略。

```
<Sysname> system-view
[Sysname] controller e1 2/1
[Sysname-E2/1] using e1
[Sysname-E2/1] quit
[Sysname] controller e1 2/2
[Sysname-E2/2] using e1
[Sysname-E2/2] quit
```

创建 Mp-group 接口。

```
[Sysname] interface mp-group 1
[Sysname-Mp-group1] ip address 10.1.1.2 30
[Sysname-Mp-group1] quit
```

配置串口 Serial2/0/1:0 和 Serial2/0/2:0

```
[Sysname] interface serial 2/0/1:0
[Sysname-Serial2/0/1:0] ppp mp mp-group 1
[Sysname-Serial2/0/1:0] quit
[Sysname] interface serial 2/0/2:0
[Sysname-Serial2/0/2:0] ppp mp mp-group 1
[Sysname-Serial2/0/2:0] quit
```

可以通过 **display interface serial 2/0/1**、**display interface mp-group 1** 和 **display ppp mp** 命令，可以查看连通状态，用 **ping** 命令检查网络是否配通。

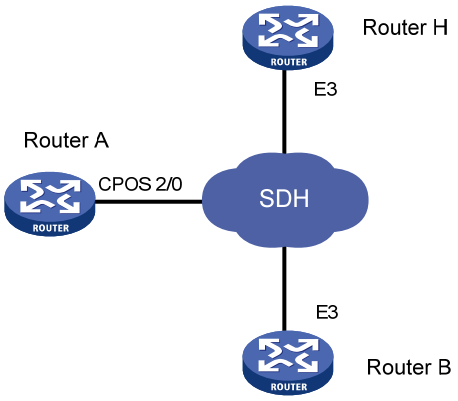
1.9.2 CPOS-E3 典型举例

1. 组网需求

Router H 和 Router B 需要从 2Mbps 扩容到 32Mbps，一条 E1 链路满足不了需求，至少需要 16 条 E1 链路，考虑到 16 条 E1 链路的复杂性，因此直接替换为一条 E3 链路。

2. 组网图

图1-10 CPOS-E3 配置组网图



3. 配置步骤

(1) 配置 Router A

此处只给出 CPOS 接口和 E3 接口的关键配置步骤，关于其他一些业务的部署，在此不再赘述。

将 CPOS 接口和其通道化出来的 E3 接口配置为主时钟模式

```
<RouterA> system-view
[RouterA] controller cpos 2/0
[RouterA-Cpos2/0] clock master
[RouterA-Cpos2/0] using e3 1
[RouterA-Cpos2/0] using e3 2
[RouterA-Cpos2/0] e3 1 set clock master
[RouterA-Cpos2/0] e3 2 set clock master
```

配置串口 Serial2/0/1:0

```
[RouterA] interface serial2/0/1:0
[RouterA-Serial2/0/1:0] ip address 11.1.1.2 24
[RouterA-Serial2/0/1:0] quit
[RouterA] interface serial2/0/2:0
[RouterA-Serial2/0/2:0] ip address 12.1.1.2 24
[RouterA-Serial2/0/2:0] quit
```

(2) 配置 Router B

各个分支节点配置与 Router B 配置类似，配置过程省略。

```
<RouterB> system-view
[RouterB] controller e3 2/1
[RouterB-E2/1] using e3
[RouterB-E2/1] quit
```

配置串口 Serial2/0/1:0

```
[RouterB] interface serial2/0/1:0
[RouterB-Serial2/0/1:0] ip address 11.1.1.8 24
[RouterB-Serial2/0/1:0] quit
```

可以通过 **display interface serial 2/0/1:0** 命令，可以查看连通状态，用 **ping** 命令检查网络是否配通。

1.10 CPOS 接口故障的诊断与排除

1.10.1 接口物理状态为 up，链路协议状态为 down，并且检测到线路环回

1. 故障现象

设备的 CPOS 接口与其他厂商的 CPOS 接口通过 SDH 传输设备互连，对 CPOS E1 通道捆绑出的串口封装 PPP，使用 **display interface serial** 命令查看接口状态，发现接口物理状态为 up，链路协议状态为 down；并且，虽然没有配置环回，但有部分接口显示检测到线路环回（“loopback is detected”）。

2. 故障排除

这种故障很可能是 SDH 传输设备的各级复用单元配置与设备 CPOS 的 E1 通道序号对应关系不正确，导致信号在传输设备中对应的时隙错误。由于两端时隙对不上，使得 PPP 的链路协商不成功，LCP 不能正常工作。并且，如果时隙对应到了传输设备一个空闲的时隙，而传输设备这个时隙的串口又处于环回状态，就会在设备上提示检测到链路发生环回，也可以打开 PPP 调试开关（使用 **debugging ppp lcp error** 命令）查看环回。

请按以下步骤处理：

(1) 使用**display controller cpos e1** 命令确定E1 通道的复用路径；或根据 1.1.6 节的内容得出 E1 通道的复用路径；

(2) 检查在传输设备上的配置，是否符合上一步计算出的 E1 复用路径，对于不一致的项进行更改。

目 录

1 POS接口配置 1-1
1.1 POS接口介绍 1-1
1.1.1 SONET/SDH 1-1
1.1.2 POS 1-1
1.2 配置POS接口 1-1
1.3 配置接口卡的工作模式 1-2
1.4 POS接口显示和维护 1-2
1.5 POS接口典型配置举例 1-3
1.5.1 路由器通过POS接口光纤直连 1-3
1.5.2 路由器通过POS接口经帧中继网互连 1-4
1.6 POS接口故障的诊断与排除 1-5
1.6.1 POS接口物理状态为down 1-5
1.6.2 物理层up，链路down 1-5
1.6.3 IP丢包严重 1-6

1 POS 接口配置

1.1 POS 接口介绍

1.1.1 SONET/SDH

SONET（Synchronous Optical Network，同步光网络）是 ANSI 定义的同步传输体制，是一种全球化的标准传输协议，采用光传输。

SDH（Synchronous Digital Hierarchy，同步数字系列）是 CCITT（现在的 ITU-T）定义的，采用同步复用方式和灵活的映射结构，可以从 SDH 信号中直接分插出低速的支路信号，而不需要使用大量的复接/分接设备，从而能够减少信号损耗和设备投资。

1.1.2 POS

POS（Packet Over SONET/SDH，SONET/SDH 上的分组）是一种应用在城域网及广域网中的技术，它具有支持分组数据的优点，如支持 IP 数据分组。

POS 将长度可变的数据包直接映射进 SONET 同步载荷中，使用 SONET 物理层传输标准，提供了一种高速、可靠、点到点的数据连接。

POS 接口在数据链路层可以使用 PPP、帧中继和 HDLC 协议，在网络层使用 IP 协议。针对不同的设备，接口传输速率会有所不同，例如 STM-1、STM-4c 和 STM-16，每一级速率都是较低一级的 4 倍。

1.2 配置 POS 接口

在进行链路协议和网络协议等配置前，需要根据对端设备的配置或应用进行如下 POS 接口物理参数的配置。此外，如果作为备份接口，则需配置接口备份工作参数；如果要在接口上建立防火墙，则需配置包过滤规则。

表1-1 配置 POS 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 POS 接口视图	**interface pos** *interface-number*	必选
配置 POS 接口的时钟模式	**clock { master | slave }**	可选 缺省情况下，POS 接口的时钟模式为从时钟模式（**slave**）
配置 POS 接口的 CRC 校验字长度	**crc { 16 | 32 }**	可选 缺省情况下，CRC 校验字长度为 32 比特
配置 POS 接口的环回方式	**loopback { local | remote }**	可选 缺省情况下，禁止任何形式的环回

操作	命令	说明
配置 POS 接口的开销字节	**flag c2** *flag-value*	可选 缺省情况下，**c2** 的缺省值为 16（十六进制）
	flag { **j0** | **j1** } { **sdh** | **sonet** } *flag-value*	可选 缺省情况下，系统使用 SDH 帧格式的缺省值 SDH 帧格式下 **j0** 和 **j1** 的缺省值都为空
配置 POS 接口的帧格式	**frame-format** { **sdh** | **sonet** }	可选 缺省情况下，POS 接口的帧格式为 SDH
配置 POS 接口的加扰功能	**scramble**	可选 缺省情况下，使能对载荷的加扰功能
配置 POS 接口的链路协议类型	**link-protocol** { **ppp** | **fr** [**nonstandard** | **ietf** | **mfr** *interface-number*] | **hdlc** }	可选 缺省情况下，POS 接口的链路协议类型为 PPP
配置 POS 接口的 MTU	**mtu** *size*	可选 取值范围为 128～1500，缺省值为 1500
配置 POS 接口的 SD（信号劣化）和 SF（信号失败）的门限值	**threshold** { **sd** | **sf** } *value*	可选 缺省情况下，SD 门限值为 10e-6（即 X 取值为 6），SF 门限值为 10e-3（即 X 取值为 3）
配置子接口速率统计功能	**sub-interface rate-statistic**	可选 缺省情况下，接口的子接口速率统计功能处于关闭状态
关闭 POS 物理端口	**shutdown**	可选 缺省情况下，POS 物理接口为打开状态



说明

当路由器的某物理接口闲置，没有连接电缆时，请使用 **shutdown** 命令关闭该接口，以防止由于干扰导致接口异常。关闭接口会导致接口停止工作，请慎用此命令。

1.3 配置接口卡的工作模式

部分 POS 卡支持一卡多用，通过配置卡的工作模式，可以完成整个卡的工作模式的切换。

支持工作模式切换的 POS 接口卡包括 HIM-MSP2P 和 HIM-MSP4P 两种。



说明

有关接口卡的工作模式的介绍和配置，请参见“系统分册”中的“设备管理”。

1.4 POS 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 POS 接口的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除接口统计信息。

表1-2 POS 接口显示和维护

操作	命令
显示 POS 接口的配置和状态信息	**display interface pos** [*interface-number*]
显示指定 POS 接口的 IP 相关配置和状态信息	**display ip interface pos** *interface-number*
显示指定 POS 接口的 IPv6 相关配置和状态信息	**display ipv6 interface pos** [*interface-number*]
清除指定接口的统计信息	**reset counters interface** [**pos** [*interface-number*]]

1.5 POS 接口典型配置举例

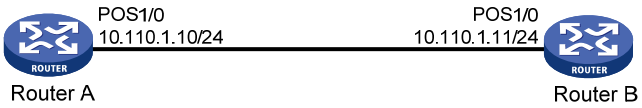
1.5.1 路由器通过 POS 接口光纤直连

1. 组网需求

用一对（接收、发送）单模光纤直接连接路由器 Router A 和 Router B 的 POS 接口，通过 PPP 互连。

2. 组网图

图1-1 路由器通过 POS 接口光纤直连组网图



3. 配置步骤

(1) 配置 Router A

配置 POS 接口 1/0，物理参数全部采用缺省配置。

```
<RouterA> system-view
[RouterA] interface pos 1/0
[RouterA-Pos1/0] ip address 10.110.1.10 255.255.255.0
[RouterA-Pos1/0] link-protocol ppp
[RouterA-Pos1/0] mtu 1500
[RouterA-Pos1/0] shutdown
[RouterA-Pos1/0] undo shutdown
```

(2) 配置 Router B

配置 POS 接口 1/0。

```
<RouterB> system-view
[RouterB] interface pos 1/0
```

时钟模式为主时钟模式，其它物理参数采用缺省配置。

```
[RouterB-Pos1/0] clock master
[RouterB-Pos1/0] ip address 10.110.1.11 255.255.255.0
[RouterB-Pos1/0] link-protocol ppp
[RouterB-Pos1/0] mtu 1500
```

```
[RouterB-Pos1/0] shutdown
[RouterB-Pos1/0] undo shutdown
```

可以通过 **display interface pos** 查看 POS 接口连通状态，用 **ping** 命令检查网络是否配通。

1.5.2 路由器通过 POS 接口经帧中继网互连

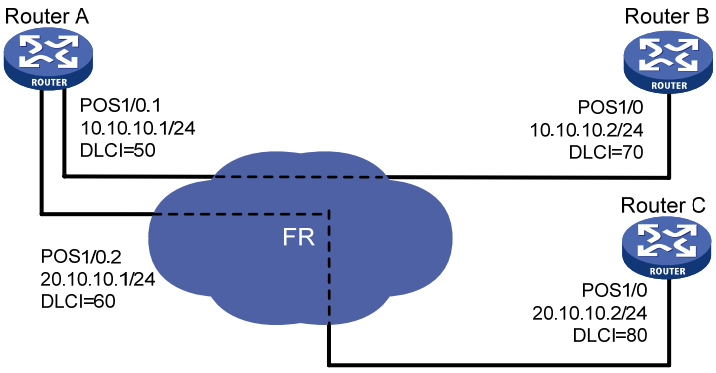
1. 组网需求

路由器可以使用 POS 接口通过公用帧中继网络互连，路由器作为用户设备，工作在帧中继的 DTE 方式。

Router A 利用帧中继子接口连接处于不同网段的 Router B 和 Router C。

2. 组网图

图1-2 路由器通过 POS 接口经帧中继网互连组网图



3. 配置步骤

(1) 配置 Router A

配置 Router A 的 POS 接口 1/0。

```
<RouterA> system-view
[RouterA] interface pos 1/0
[RouterA-Pos1/0] clock slave
```

封装帧中继链路协议。

```
[RouterA-Pos1/0] link-protocol fr
[RouterA-Pos1/0] fr interface-type dte
[RouterA-Pos1/0] quit
```

配置 POS 接口 1/0 的子接口 1。

```
[RouterA] interface pos 1/0.1
[RouterA-Pos1/0.1] ip address 10.10.10.1 255.255.255.0
[RouterA-Pos1/0.1] fr dlci 50
[RouterA-Pos1/0.1] fr map ip 10.10.10.2 50
[RouterA-Pos1/0.1] mtu 1500
[RouterA-Pos1/0.1] quit
```

配置 POS 接口 1/0 的子接口 2。

```
[RouterA] interface pos 1/0.2
[RouterA-Pos1/0.2] ip address 20.10.10.1 255.255.255.0
```

```
[RouterA-Pos1/0.2] fr dlci 60
[RouterA-Pos1/0.2] fr map ip 20.10.10.2 60
[RouterA-Pos1/0.2] mtu 1500
[RouterA-Pos1/0.2] quit
```

(2) 配置 Router B

配置 Router B 的 POS 接口 1/0。

```
[RouterB] interface pos 1/0
[RouterB-Pos1/0] clock slave
```

封装帧中继链路协议。

```
[RouterB-Pos1/0] link-protocol fr
[RouterB-Pos1/0] fr interface-type dte
[RouterB-Pos1/0] ip address 10.10.10.2 255.255.255.0
[RouterB-Pos1/0] fr dlci 70
[RouterB-Pos1/0] fr map ip 10.10.10.1 70
[RouterB-Pos1/0] mtu 1500
```

Router C 的配置方法与 Router B 的配置方法相同。

在各路由器上可以通过 **display interface pos** 查看 POS 接口连通状态，用 **ping** 命令检查网络是否配通。

1.6 POS 接口故障的诊断与排除

1.6.1 POS 接口物理状态为 down

1. 故障现象

POS 接口物理状态为 down

2. 故障排除

- 请检查插接在 POS 接口的光纤是否接错。正常情况下应该有两根光纤，分别负责接收和发送，并且不能接反。另外，如果将一根光纤的两端接在了同一个 POS 接口的接收端和发送端上，即使没有启用环回，使用 **display interface** 也会看到“loopback detected”的信息。
- 如果设备采取 POS 接口直连相连时（即所谓的“背靠背”连接），POS 接口应配置一端使用主时钟模式，另一端使用从时钟模式。

1.6.2 物理层 up，链路 down

1. 故障现象

物理层 up，链路层 down。

2. 故障排除

- POS 接口时钟、扰码等物理参数配置与对端不匹配；
- 链路层协议配置与对端不匹配；
- 本端或对端 IP 地址没有配置。

1.6.3 IP 丢包严重

1. 故障现象

IP 丢包严重。

2. 故障排除

- POS 接口时钟配置不正确（产生大量的 CRC 错误）；
- 最大传输单元 MTU 配置不匹配。

目 录

1 以太网接口配置 1-1
1.1 以太网接口通用配置 1-1
1.1.1 Combo接口配置 1-1
1.1.2 配置 10GE接口的LAN/WAN模式 1-1
1.1.3 以太网接口/子接口基本配置 1-2
1.1.4 配置以太网接口的流量控制功能 1-4
1.1.5 配置以太网接口环回测试功能 1-4
1.2 三层以太网接口/子接口的配置 1-5
1.2.1 三层以太网接口/子接口配置任务简介 1-5
1.2.2 配置以太网接口/子接口的MTU 1-5
1.2.3 配置以太网接口链路层连接状态up/down抑制时间 1-5
1.2.4 配置子接口速率统计功能 1-6
1.3 虚拟以太网接口 1-6
1.3.1 虚拟以太网接口简介 1-6
1.3.2 配置虚拟以太网接口 1-7
1.4 以太网接口显示和维护 1-7

1 以太网接口配置

为了适应网络需求，设备上定义了三种以太网接口：

- 三层以太网接口。该类接口是设备上的一种物理接口，工作在网络层，可以配置 IP 地址，处理三层协议，提供路由功能。
- 三层以太网子接口。该类接口是设备上的一种逻辑接口，工作在网络层，可以配置 IP 地址，处理三层协议。主要用来实现在三层以太网接口上支持收发 VLAN tagged 报文。
- 虚拟以太网接口。虚拟以太网接口是一种逻辑接口，在接口板上实现，主要用于以太网协议承载其它数据链路层协议。

1.1 以太网接口通用配置

该部分介绍了三层以太网接口的共有属性及其配置，各自的特有属性请参见下文中的相关章节。

1.1.1 Combo 接口配置

1. Combo 接口介绍

Combo 接口是指设备面板上的两个以太网接口（通常一个是光口一个是电口），而在设备内部只有一个转发接口。Combo 电口与其对应的光口在逻辑上是光电复用的，用户可根据实际组网情况选择其中的一个使用，但两者不能同时工作，当激活其中的一个接口时，另一个接口就自动处于禁用状态。两个以太网接口只对应一个 interface 视图，用户在同一个 interface 视图完成对两个接口的状态切换操作。

2. 配置 Combo 接口的状态

表1-1 配置 Combo 接口的状态

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	-
激活指定的单 Combo 接口	**combo enable** { **copper** | **fiber** }	可选 缺省情况下，电口处于激活状态

1.1.2 配置 10GE 接口的 LAN/WAN 模式

1. 10GE 接口 LAN/WAN 模式简介

10GE（Ten-GigabitEthernet，万兆以太网）接口根据其物理特性，可以分为两种工作模式：

- LAN 模式。工作在该模式下的 10GE 接口传输以太网报文，用于连接以太网。
- WAN 模式。工作在该模式下的 10GE 接口传输 SDH（Synchronous Digital Hierarchy，同步数字系列）报文，用于连接 SDH 网络。接口工作在 WAN 模式下仅支持点到点的报文传输。

注意

工作在 WAN 模式下的 10GE 接口将以太网报文封装成 SDH 帧；10G POS（Packet over SDH）接口将 PPP 报文封装成 SDH 帧，二者帧格式不同，因此工作在 WAN 模式下的 10GE 接口和 10G POS 接口不能互通。

2. J0/J1 开销字节简介

SDH 帧具有丰富的开销字节，可完成对传输网的分层管理等运行维护功能。J0、J1 主要用于在不同国家、不同地区、或不同厂商的设备之间提供互通支持。

再生段踪迹字节 J0 通常被设置为段接入点标识符，发送端通过重复发送该字节来保持与接收端的连接。

通道踪迹字节 J1 包含在高阶通道开销中，通常被设置为高阶通道接入点标识符，它的作用与 J0 字节类似，被用来保持与通道接收端的连接。

为了保证通信的畅通，通常要求发送端和接收端的 J0 和 J1 字节能够分别匹配。关于 SDH 及 SDH 的开销字节，请查阅相关的专业书籍。

3. 配置 10GE 接口 LAN/WAN 模式

表1-2 配置 10GE 接口 LAN/WAN 模式

操作	命令	说明
进入系统视图	**system-view**	-
进入 10GE 接口视图	**interface ten-gigabitethernet** *interface-number*	-
配置 10GE 接口的 LAN/WAN 模式	**port-mode { lan | wan }**	可选 缺省情况下，10GE 接口工作在 LAN 模式下
在 WAN 模式下，配置 10GE 接口的 J0/J1 字节的值	**flag { j0 | j1 } sdh** *flag-value*	可选 缺省情况下，J0/J1 字节的值为全 0

说明

flag 命令只有 10GE 接口工作在 WAN 模式下时配置才有效。

1.1.3 以太网接口/子接口基本配置

1. 以太网接口基本配置

设置以太网接口的双工模式时存在三种情况：

- 当希望接口在发送数据包的同时可以接收数据包，可以将接口设置为全双工（**full**）属性；
- 当希望接口同一时刻只能发送数据包或接收数据包时，可以将接口设置为半双工（**half**）属性；
- 当设置接口为自协商（**auto**）状态时，接口的双工状态由本接口和对端接口自动协商而定。

设置以太网接口的速率时，当设置接口速率为自协商（**auto**）状态时，接口的速率由本接口和对端接口双方自动协商而定。

表1-3 以太网接口基本配置

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	-
设置以太网接口的描述字符串	**description** *text*	可选 缺省情况下，描述字符串为“*该接口的接口名* **interface**”
设置以太网接口的双工模式	**duplex** { **auto** | **full** | **half** }	可选 对于 Combo 口中的光口，不支持配置 **half** 参数 缺省情况下，接口的双工模式为 **auto**（自协商）状态
设置以太网接口的速率	**speed** { **10** | **100** | **1000** | **auto** }	可选 对于 Combo 口中的光口，不支持配置 **10** 和 **100** 参数 缺省情况下，以太网接口的速率为 **auto**（自协商）状态
关闭以太网接口	**shutdown**	可选 缺省情况下，接口处于打开状态 如果想打开端口，可以使用 **undo shutdown** 命令

说明

FE 电接口和 GE 电接口的缺省速率及双工模式均为自动协商模式，用户也可强制更改速率及双工模式，但应确保速率及双工模式与连接对端相同。

2. 以太网子接口基本配置

三层以太网子接口可以解决三层以太网接口不能识别 VLAN 报文的问题。用户可以在一个以太网接口上配置多个子接口，这样，来自不同 VLAN 的报文可以从不同的子接口进行转发，为用户提供了很高的灵活性。

表1-4 以太网子接口基本配置

操作	命令	说明
进入系统视图	**system-view**	-
创建以太网子接口并进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	必选
设置当前接口的描述信息	**description** *text*	可选 缺省情况下，描述字符串为“*接口名* **interface**”，比如：GigabitEthernet1/0.1 Interface
关闭以太网子接口	**shutdown**	可选 缺省情况下，子接口处于打开状态

说明

- 以太网子接口下必须配置当前子接口允许收发的 VLAN 报文，否则，子接口不能收发报文。
- 以太网子接口只有在关联了 VLAN 后才能被激活，正常的收发报文。相关配置请参见“接入分册”中的“VLAN 终结配置”
- 以太网子接口还可以配置 IP 参数，相关配置请参见“IP 业务分册”。
- 本端设备以太网子接口号、VLAN ID 需要分别和相连的对端设备的以太网子接口号、VLAN ID 一致，否则报文将不能正确传输。
- 管理以太网口不支持子接口。

1.1.4 配置以太网接口的流量控制功能

当本端和对端设备都开启了流量控制功能后，如果本端设备发生拥塞，它将向对端设备发送消息，通知对端设备暂时停止发送报文；而对端设备在接收到该消息后将暂时停止向本端发送报文；反之亦然。从而避免了报文丢失现象的发生。只有本端和对端设备都开启了流量控制功能，才能实现对本端以太网接口的流量控制。

表1-5 开启以太网接口的流量控制功能

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	-
开启以太网接口的流量控制功能	**flow-control**	必选 缺省情况下，以太网接口的流量控制功能处于关闭状态

1.1.5 配置以太网接口环回测试功能

用户可以开启以太网接口环回测试功能，检验以太网接口能否正常工作。测试时接口将不能正常转发数据包。以太网接口环回测试功能包括内部环回测试和外部环回测试。

- 内部环回测试。该测试在交换芯片内部建立自环，用以定位端口内部的部分硬件功能是否出现故障。
- 外部环回测试。该测试需要在以太网接口上接一个自环头，从接口发出的报文通过自环头又环回到该接口，并被该接口接收。用以定位该端口内部的所有硬件功能及线缆是否出现故障。

表1-6 配置以太网接口环回测试功能

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	-
配置以太网接口进行环回测试	**loopback { external | internal }**	可选 缺省情况下，以太网接口环回测试功能处于关闭状态

说明

- 端口关闭状态(down)下可以进行内部环回测试，但不能进行外部环回测试。手工关闭(shutdown)时，则不能进行内部和外部环回测试。
- 在进行环回测试时系统将禁止在接口上进行 **speed**、**duplex** 和 **shutdown** 命令的配置；
- 以太网接口开启环回测试功能时将工作在全双工状态；关闭环回测试功能后恢复原有配置。

1.2 三层以太网接口/子接口的配置

1.2.1 三层以太网接口/子接口配置任务简介

表1-7 三层以太网接口配置任务简介

配置任务	说明	详细配置
配置以太网接口的 MTU	可选 三层以太网接口/子接口都支持	1.2.2
配置以太网接口链路层连接状态 up/down 抑制时间	可选 三层以太网接口支持	1.2.3
配置子接口速率统计功能	可选 三层以太网接口支持	1.2.4

1.2.2 配置以太网接口/子接口的 MTU

MTU（Maximum Transmission Unit，最大传输单元）参数影响 IP 报文的分片与重组。

表1-8 配置以太网接口的 MTU

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type* { *interface-number* | *interface-number.subnumber* }	-
设置 MTU	**mtu** *size*	可选 缺省情况下，以太网接口的 MTU 为 1500Bytes

说明

由于 QoS 队列长度的限制（如 FIFO 队列的缺省长度为 75），MTU 太小会造成分片太多，从而被 QoS 队列丢弃。此时，可适当增大 MTU 值或 QoS 队列的长度。以太网接口视图下的命令 **qos fifo queue-length** 可以改变 QoS 队列长度（具体配置请参见“QoS 分册”中的“QoS 配置”）。

1.2.3 配置以太网接口链路层连接状态 up/down 抑制时间

以太网接口在三层模式下有两种链路层连接状态：up 和 down。在抑制时间内，接口链路层状态之间的切换，不会被系统感知；而是等待经过抑制时间后，链路层再向系统报告链路连接状态的变化。该功能用于避免因接口在短时间内频繁改变接口链路层状态，而给系统带来的额外的开销。

表1-9 配置以太网接口链路层连接状态 up/down 抑制时间

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
设置以太网接口链路层连接状态 up/down 抑制时间	**timer hold** *seconds*	可选 缺省情况下，以太网接口链路层连接状态 up/down 抑制时间为 10 秒



说明

如果网络的延迟比较大，或拥塞程度较高，可以适当加大轮询时间间隔，以减少网络震荡的发生。

1.2.4 配置子接口速率统计功能

当接口使能了子接口速率统计功能后，设备会定时刷新子接口速率统计信息。用户可以通过 **display interface** 命令查看统计结果。

表1-10 配置子接口速率统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置子接口速率统计功能	**sub-interface rate-statistic**	可选 缺省情况下，接口的子接口速率统计功能处于关闭状态



说明

- 开启本功能可能需要耗费大量系统资源，请谨慎使用。
- 所有支持子接口的接口类型均支持该功能。

1.3 虚拟以太网接口

1.3.1 虚拟以太网接口简介

虚拟以太网接口是一种逻辑接口，在接口板上实现，主要应用于 PPPoEoA 和 IPoEoA。

- PPPoEoA 是 Point to Point Protocol over Ethernet over ATM 的简称。它有三层结构，最上层封装了 PPP 协议；中间为 PPPoE，即以太网承载 PPP 协议报文；最下一层为 ATM 上承载 PPPoE 报文。其中，以太网承载的 PPP 协议报文的各项参数是通过接入设备接口板上的三层 VE 接口来配置的。详细内容请参见“接入分册”中的“ATM 配置”。
- IPoEoA 是通过 PVC 上绑定 VE 接口，从而实现在 ATM 上承载以太网报文。详细内容请参见“接入分册”中的“ATM 配置”。

1.3.2 配置虚拟以太网接口

配置 PVC（Permanent Virtual Channel，永久虚通路）承载 PPPoEoA 和 IPoEoA 时，必须指定一个虚拟以太网接口与之对应，如果对应的虚拟以太网接口没有创建，则不能配置 PVC。

表1-11 配置子接口速率统计功能

操作	命令	说明
进入系统视图	**system-view**	-
创建虚拟以太网接口并进入虚拟以太网接口视图	**interface virtual-ethernet** *interface-number*	必选 如果指定的虚拟以太网接口不存在，则该命令先完成虚拟以太网接口的创建，然后再进入该接口的视图 用户最多可以创建 1024 个虚拟以太网接口



说明

有关 PPPoEoA 的配置，请参见“接入分册”中的“ATM 配置”。

1.4 以太网接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后接口的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除接口统计信息。

表1-12 以太网接口显示和维护

操作	命令
显示指定接口当前的运行状态和相关信息	**display interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]]
显示指定接口的接口概要信息	**display brief interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示虚拟以太网接口的状态信息	**display interface virtual-ethernet** [*interface-number*]
清除指定接口的统计信息	**reset counters interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]]

目 录

1 WAN接口配置 1-1
1.1 同步串口 1-1
1.1.1 同步串口介绍 1-1
1.1.2 配置同步串口 1-1
1.2 CE1 接口 1-2
1.2.1 CE1 接口介绍 1-2
1.2.2 配置CE1 接口（工作在E1 方式） 1-3
1.2.3 配置CE1 接口（工作在CE1 方式） 1-3
1.2.4 配置CE1 接口其他参数 1-4
1.2.5 配置错包扩散抑制功能 1-5
1.3 CT1 接口 1-5
1.3.1 CT1 接口介绍 1-5
1.3.2 配置CT1 接口 1-5
1.3.3 配置CT1 接口其他参数 1-6
1.3.4 配置CT1 接口进行线路位（Bit）错误率的测试 1-7
1.3.5 CT1 接口显示和维护 1-7
1.4 E1-F接口 1-8
1.4.1 E1-F接口介绍 1-8
1.4.2 配置E1-F接口（工作在成帧方式） 1-8
1.4.3 配置E1-F接口（工作在非成帧方式） 1-8
1.4.4 配置E1-F接口的其他参数 1-9
1.4.5 E1-F接口显示和维护 1-10
1.5 T1-F接口 1-10
1.5.1 T1-F接口介绍 1-10
1.5.2 配置T1-F接口 1-10
1.5.3 配置T1-F接口进行线路位（Bit）错误率的测试 1-12
1.5.4 T1-F接口显示和维护 1-12
1.6 CE3 接口 1-12
1.6.1 CE3 接口介绍 1-12
1.6.2 配置CE3 接口（工作在E3 方式） 1-13
1.6.3 配置CE3 接口（工作在CE3 方式） 1-13
1.6.4 配置CE3 接口其他参数 1-14
1.6.5 CE3 接口显示和维护 1-14
1.7 CT3 接口 1-15
1.7.1 CT3 接口介绍 1-15
1.7.2 配置CT3 接口（工作在T3 模式） 1-16
1.7.3 配置CT3 接口（工作在CT3 模式） 1-16
1.7.4 配置CT3 接口其他参数 1-17
1.7.5 CT3 接口显示和维护 1-18

1 WAN 接口配置

WAN（Wide Area Network，广域网）按照线路类型来分有 X.25 网、帧中继网、ATM 网、ISDN 网等类型。路由器因此也相应地有异步串口、同步串口、ATM 接口、ISDN BRI 接口、CE1 接口等等。

目前系统支持的 WAN 接口包括同步串口和 CE1 接口。

1.1 同步串口

1.1.1 同步串口介绍

同步串口特性：

- 可以工作在 DTE 和 DCE 两种方式，一般情况下，同步串口作为 DTE 设备，接受 DCE 设备提供的时钟。
- 同步串口可以外接多种类型电缆，如 V.24、V.35、X.21、RS449、RS530 等。设备可以自动检测同步串口外接电缆类型，并完成电气特性的选择，一般情况下，无需手工配置。
- 同步串口支持的链路层协议包括 PPP、帧中继等。
- 支持 IP 网络层协议。
- 可以通过执行 **display interface serial** 命令，查看同步串口的当前外接电缆类型以及工作方式（DTE/DCE）等信息。

1.1.2 配置同步串口

表1-1 配置同步串口

操作	命令	说明
进入系统视图	**system-view**	-
进入指定同步串口的视图	**interface serial** *interface-number*	必选
配置链路层协议	**link-protocol { fr | hdlc | ppp }**	可选 缺省情况下，链路层协议为 PPP
配置数字信号编码格式	**code { nrz | nrzi }**	可选 缺省情况下，同步串口使用 NRZ 编码格式
配置波特率	**baudrate** *baudrate* **virtualbaudrate** *virtualbaudrate*	可选 缺省情况下，同步串口的波特率为 64000bps
选择 DTE 侧工作时钟	**clock { dteclk1 | dteclk2 | dteclk3 | dteclk4 | dteclkauto }**	可选 DCE 侧同步串口缺省时钟为 **dceclk**，不需要配置；DTE 侧同步串口缺省时钟选择为 **dteclk1** 当使用 X.21 的 DTE 线缆时，该命令无效，时钟始终为 **dteclk3**
配置同步串口 DTE 侧的发送时钟或者接收时钟翻转	**invert { transmit-clock | receive-clock }**	可选 缺省情况下，禁止翻转

操作	命令	说明
配置 MTU 值	**mtu** *size*	可选 缺省情况下，MTU 值为 1500
配置同步串口的 CRC 校验模式	**crc { 16 | 32 | none }**	可选 缺省情况下，使用 16 位 CRC 校验
使能电平检测功能	**detect dsr-dtr**	可选 缺省情况下，同步串口电平检测功能是使能的
允许进行载波检测	**detect dcd**	可选 缺省情况下，允许同步串口进行载波检测
使能对内自环功能	**loopback**	可选 缺省情况下，禁止对内自环
配置链路状态轮询时间间隔	**timer hold** *seconds*	可选 缺省情况下，链路状态轮询时间间隔为 10 秒
配置同步串口的线路空闲码为“0xFF”	**idle-mark**	可选 缺省情况下，同步串口的线路空闲码为“0x7E”
配置翻转 RTS 信号	**reverse-rts**	可选 缺省情况下，不翻转 RTS 信号
配置子接口速率统计功能	**sub-interface rate-statistic**	可选 缺省情况下，接口的子接口速率统计功能处于关闭状态

说明

根据需要，同步串口还可能要配置 PPP/帧中继参数、IP 地址等，具体内容请参考相关章节。

1.2 CE1 接口

1.2.1 CE1 接口介绍

20 世纪 60 年代，随着 PCM（Pulse Code Modulation，脉冲编码调制）技术的出现，TDM 技术（Time Division Multiplexing，时分复用）在数字通信系统中逐渐得到广泛的应用。目前，在数字通信系统中存在两种时分复用系统，一种是 ITU-T 推荐的 E1 系统，广泛应用于欧洲以及中国；一种是由 ANSI 推荐的 T1 系统，主要应用于北美和日本（日本采用的 J1，与 T1 基本相似，可以算作 T1 系统）。

CE1 接口拥有两种工作方式：E1 工作方式（也称为非通道化工作方式）和 CE1 工作方式（也称为通道化工作方式）。

(1) 当 CE1 接口使用 E1 工作方式时，它相当于一个不分时隙、数据带宽为 2.048Mbps 的接口，其逻辑特性与同步串口相同，支持 PPP、帧中继等数据链路层协议，支持 IP 等网络协议。

(2) 当 CE1 接口使用 CE1 工作方式时，它在物理上分为 32 个时隙，对应编号为 0～31，其中 0 时隙用于传输同步信息，可以将除 0 时隙外的全部时隙任意分成若干组（channel set），每组时隙捆绑以后，作为一个接口使用，其逻辑特性与同步串口相同，支持 PPP、HDLC、帧中继等数据链路层协议，支持 IP 等网络协议。

1.2.2 配置 CE1 接口（工作在 E1 方式）

表1-2 配置 CE1 接口（工作在 E1 方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CE1 接口的视图	**controller e1** *number*	必选
配置接口工作在 E1 方式	**using e1**	必选 缺省情况下，CE1 接口工作在 CE1 工作方式
配置 CE1 接口的其他参数	请参见 1.2.4 配置CE1 接口其他参数	可选

CE1 接口工作在 E1 方式时，系统会自动创建一个 Serial 接口，接口的编号是 **serial** *interface-number*:**0**。此接口的逻辑特性与同步串口相同，可以视其为同步串口进行进一步的配置。主要的配置内容包括：

- PPP、帧中继等数据链路层协议工作参数
- IP 地址

详细内容请参考本手册的相关部分，此处不再赘述。

1.2.3 配置 CE1 接口（工作在 CE1 方式）

表1-3 配置 CE1 接口（工作在 CE1 方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CE1 接口的视图	**controller e1** *number*	必选
配置接口工作在 CE1 工作方式	**using ce1**	可选 缺省情况下，CE1 接口工作在 CE1 工作方式
将接口捆绑为 channel set	**channel-set** *set-number* **timeslot-list** *list*	必选
配置检测远端告警信号	**alarm detect rai**	可选 缺省情况下，检测远端告警信号
配置 CE1 接口的其他参数	请参见 1.2.4 配置CE1 接口其他参数	可选

当接口工作在 CE1 方式时，可以把该接口作为 CE1 接口使用。用户需要配置 channel set，才会产生相应的串口，在一个 CE1 接口上可以捆绑出多达 31 个 channel set。

在将接口时隙捆绑为 channel set 之后，系统会自动创建一个 Serial 接口，接口的编号是 **serial** *interface-number*:*set-number*。此接口的逻辑特性与同步串口相同，可以视其为同步串口进行进一步的配置，主要的配置内容包括：

- PPP、帧中继等数据链路层协议工作参数
- IP 地址

详细内容请参考本手册的相关部分，不再赘述。

1.2.4 配置 CE1 接口其他参数

表1-4 配置 CE1 接口其他参数

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CE1 接口的视图	**controller e1** *number*	必选
配置线路编解码格式	**code { ami | hdb3 }**	可选 缺省情况下，CE1 接口的线路编解码格式为 **hdb3**
配置接口进行 AIS 检测	**detect-ais**	可选 缺省情况下，进行 AIS 检测
配置接口匹配的传输线路类型	**cable { long | short }**	可选 缺省情况下，CE1 接口匹配的传输线路类型为 **long**
配置线路时钟	**clock { master | slave }**	可选 CE1 接口的时钟模式缺省为 **slave** 时钟
配置时钟自动切换功能	**clock-change auto**	可选 缺省情况下，禁止使能时钟自动切换。
配置对用户数据进行翻转	**data-coding { normal | inverted }**	可选 缺省情况下，CE1 接口不对用户数据进行翻转
配置帧格式	**frame-format { crc4 | no-crc4 }**	可选 CE1 接口的缺省帧格式为 **no-crc4**
配置接口的线路空闲码类型	**idlecode { 7e | ff }**	可选 缺省情况下，CE1 接口的线路空闲码为 **7e**
配置接口的帧间填充符类型	**itf type { 7e | ff }**	可选 缺省情况下，CE1 接口的帧间填充符为 **7e**
配置帧间填充字节的个数	**itf number** *number*	可选 缺省情况下，CE1 接口的帧间填充符的字节个数为 4 个
使能环回检测功能并配置检测方式	**loopback { local | payload | remote }**	可选 缺省情况下，环回检测功能是禁止的
退回系统视图	**quit**	-
进入 CE1 接口形成的同步串口视图	**interface serial** *interface-number*:*set-number*	必选
配置 CRC 校验模式	**crc { 16 | 32 | none }**	可选 缺省情况下，使用 16 位 CRC 校验

警告

当在接口下配置了 **code ami** 和 **data-coding inverted** 两条命令时（即接口的线路编解码格式为 AMI，且对用户数据进行翻转时），需要执行 **undo detect-ais** 命令关闭 AIS 检测，否则将导致接口异常。

1.2.5 配置错包扩散抑制功能

当一个时隙收到特定错包的时候，会影响其它时隙的正常收包，导致其它时隙也会断续地收到错包，这种情况称之为错包扩散现象。

通过配置错包扩散抑制功能，可以降低错包扩散的概率。错包扩散抑制功能提供了三个参数：*detect-timer*、*renew-timer*、*threshold*。

这几个参数的作用为：在 *detect-timer* 时间内如果接口的错包率大于 *threshold* 的值时，就认为接口出现故障并强行关闭该接口。当接口由于错包过多而被关闭后，经过 *renew-timer* 秒恢复为正常状态。

表1-5 配置错包扩散抑制功能

操作	命令	说明
进入系统视图	**system-view**	-
使能错包扩散抑制功能	**error-diffusion restraint enable**	必选
配置错包扩散抑制功能的参数	**error-diffusion restraint config** *detect-timer renew-timer threshold*	可选 缺省情况下，*detect-timer* 为 30 秒，*renew-timer* 为 600 秒，*threshold* 为 20%

1.3 CT1 接口

1.3.1 CT1 接口介绍

CT1 接口只能工作在通道化工作方式，可以将全部时隙（时隙 1～24）任意地分成若干组，每组时隙捆绑为一个 channel set。每组时隙捆绑后系统自动生成一个接口，其逻辑上等同于同步串口，支持 PPP、HDLC、帧中继等数据链路层协议，支持 IP 等网络协议。

1.3.2 配置 CT1 接口

表1-6 配置 CT1 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CT1 接口的视图	**controller t1** *number*	-
将接口捆绑为 channel set	**channel-set** *set-number* **timeslot-list** *list* [**speed** { **56k** | **64k** }]	必选 在一个 CT1 接口上可以捆绑出多达 24 个 channel set；时隙的缺省速率为 64kbps
配置 CT1 接口的其他参数	请参见 1.3.3 配置CT1 接口其他参数	可选

在将接口时隙捆绑为 channel set 之后，系统会自动创建一个 Serial 接口，接口的编号是 **serial** *number*:*set-number*。此接口的逻辑特性与同步串口相同，可以视其为同步串口进行配置。主要的配置内容包括：

- PPP、帧中继等数据链路层协议工作参数
- IP 地址

- 如果要在其上建立防火墙，则需配置地址转换和报文过滤规则

详细内容请参考本手册的相关部分，不再赘述。

1.3.3 配置 CT1 接口其他参数

表1-7 配置 CT1 接口其他参数

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CT1 接口的视图	**controller t1** *number*	-
配置 CT1 接口的描述字符串	**description** *text*	可选 缺省情况下，CT1 接口的描述字符串为“*该接口的接口名* interface”
配置传输线路衰减	**cable long { 0db | -7.5db | -15db | -22.5db }** **cable short { 133ft | 266ft | 399ft | 533ft | 655ft }**	可选 缺省情况下，CT1 接口匹配的传输线路衰减为 **long 0db**
配置线路编解码格式	**code { ami | b8zs }**	可选 缺省情况下，CT1 接口的线路编解码格式为 **b8zs**
配置线路时钟模式	**clock { master | slave }**	可选 缺省情况下，线路时钟模式为 **slave** 时钟
配置帧格式	**frame-format { sf | esf }**	可选 缺省情况下，CT1 接口的帧格式为 **esf**
配置检测远端告警信号	**alarm detect rai**	可选 在接口帧格式采用 **esf** 的情况下，可以使用该命令
配置对用户数据进行翻转	**data-coding { normal | inverted }**	可选 缺省情况下，CT1 接口不对用户数据进行翻转
配置接口的线路空闲码类型	**idlecode { 7e | ff }**	可选 缺省情况下，CT1 接口的线路空闲码为 **7e**
配置接口的帧间填充符类型	**itf type { 7e | ff }**	可选 缺省情况下，CT1 接口的帧间填充符为 **7e**
配置帧间填充字节的个数	**itf number** *number*	可选 缺省情况下，CT1 接口的帧间填充符的字节个数为 4 个
配置接口的告警门限	**alarm-threshold { ais { level-1 | level-2 } | lfa { level-1 | level-2 | level-3 | level-4 } | los { pulse-detection | pulse-recovery }** *value* **}**	可选 在缺省情况，对于 LOS 告警，**pulse-detection** 参数的值为 176，**pulse-recovery** 的值为 22，即默认的情况下，如果在 176 个脉冲周期内检测到的脉冲数小于 22 个则认为载波丢失 LOS 告警产生；对于 AIS 告警，缺省值为 level-1；对于 LFA 告警，缺省值为 level-1

操作	命令	说明
配置接口的 FDL 格式	**fdl { ansi | att | both | none }**	可选 缺省情况下，接口的 FDL 格式为 **none**
使能环回检测功能并配置检测方式	**loopback { local | payload | remote }**	可选 缺省情况下，环回检测功能是禁止的
发送远程环回控制码	**sendloopcode { fdl-ansi-llb-down | fdl-ansi-llb-up | fdl-ansi-plb-down | fdl-ansi-plb-up | fdl-att-plb-down | fdl-att-plb-up | inband-llb-down | inband-llb-up }**	可选 缺省情况下，不发送远程环回控制码
关闭 CT1 接口	**shutdown**	可选 缺省情况下，CT1 接口为打开状态
退回系统视图	**quit**	-
进入 CT1 接口通过各种方式形成的同步串口视图	**interface serial** *interface-number*:*set-number* 或 **interface serial** *interface-number*:**23**	必选
配置 CRC 校验模式	**crc { 16 | 32 | none }**	可选 缺省情况下，使用 16 位 CRC 校验

1.3.4 配置 CT1 接口进行线路位（Bit）错误率的测试

BERT 测试方式为，本端发出测试数据流，经过线路某处环回来，检测收到的测试数据流与发出的测试数据流是否一致以及位错误率达到多少，从而为用户判断线路状态提供依据。因此，要求线路中某处能环回发出的数据流，如将对方配置远端环回等。利用 **bert** 命令配置好测试模式，指定测试时间，开始测试后，可以查看接口状态中的 BERT 测试状态和测试结果。

表1-8 配置 CT1 接口进行线路位（Bit）错误率的测试

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CT1 接口的视图	**controller t1** *number*	-
配置 CT1 接口进行线路位（Bit）错误率的测试	**bert pattern { 2^20 | 2^15 } time** *minutes* **[unframed]**	必选 缺省情况下，不进行线路位错误率的测试

1.3.5 CT1 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 CT1 接口的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令清除 CT1 接口的计数器信息。

表1-9 CT1 接口显示和维护

配置	命令
显示 CT1 接口的工作状态	**display controller t1 [** *interface-number* **]**
显示 channel set 的工作状态	**display interface serial** *interface-number:set-number*

配置	命令
清除 CT1 接口的 controller 计数器	**reset counters controller t1** *interface-number*

1.4 E1-F 接口

1.4.1 E1-F 接口介绍

E1-F 接口是指部分（Fractional）化 E1 接口，它是 CE1 接口的简化版本。在 E1 接入应用中，如果不需要划分出多个通道组（channel set），使用 CE1 接口就显得浪费。此时，可以利用 E1-F 接口来满足这些简单的 E1 接入需求。相对 CE1 接口而言，使用 E1-F 接口是一种低价位的 E1 接入方案。

E1-F 接口工作在成帧方式时，只能将时隙捆绑为一个通道组，而 CE1 接口可以将时隙任意分组，捆绑出多个通道组。

E1-F 接口拥有两种工作方式：成帧方式和非成帧方式。缺省情况下，E1-F 接口工作在成帧方式。

当 E1-F 接口工作于非成帧方式时，它相当于一个不分时隙、数据带宽为 2048kbps 的接口，其逻辑特性与同步串口相同，支持 PPP、HDLC、帧中继等数据链路层协议，支持 IP 等网络协议。

当 E1-F 接口工作于成帧方式时，它在物理上分为 32 个时隙，对应编号为 0～31。其中 0 时隙用于传输同步信息，其余时隙可以被任意捆绑成一个通道组（channel set），E1-F 接口的速率为 n×64kbps，其逻辑特性与同步串口相同，支持 PPP、帧中继等数据链路层协议，支持 IP 和 IPX 等网络协议。

1.4.2 配置 E1-F 接口（工作在成帧方式）

表1-10 配置 E1-F 接口（工作在成帧方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入 E1-F 接口的视图	**interface serial** *interface-number*	-
配置接口工作方式为成帧方式	**undo fe1 unframed**	可选 E1-F 接口缺省工作在成帧方式
对 E1-F 接口进行时隙捆绑	**fe1 timeslot-list** *range*	可选 缺省情况下，E1-F 接口对所有时隙进行捆绑
配置检测远端告警信号	**alarm detect rai**	可选 缺省情况下，检测远端告警信号
配置 E1-F 接口的其他参数	请参见 1.4.4 配置E1-F接口的其他参数	可选

1.4.3 配置 E1-F 接口（工作在非成帧方式）

表1-11 配置 E1-F 接口（工作在非成帧方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入 E1-F 接口的视图	**interface serial** *interface-number*	-

操作	命令	说明
配置接口工作方式为非成帧方式	**fe1 unframed**	必选 E1-F 接口缺省工作在成帧方式
配置 E1-F 接口的其他参数	请参见 1.4.4 配置E1-F接口的其他参数	可选

1.4.4 配置 E1-F 接口的其他参数

表1-12 配置 E1-F 接口的其他参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 E1-F 接口的视图	**interface serial** *serial-number*	-
配置 E1-F 接口的描述字符串	**description** *text*	可选 缺省情况下，E1-F 接口的描述字符串为“*该接口的接口名* interface”
配置线路编解码格式	**fe1 code { ami | hdb3 }**	可选 缺省情况下，E1-F 接口的线路编解码格式为 **hdb3**
配置线路时钟模式	**fe1 clock { master | slave }**	可选 缺省情况下，E1-F 接口的线路时钟模式为 **slave** 时钟
配置接口支持的电缆类型	**fe1 cable { long | short }**	可选 缺省情况下，E1-F 接口支持长电缆类型
配置接口的 CRC 校验模式	**crc { 16 | 32 | none }**	可选 缺省情况下，使用 16 位 CRC 校验
设置对用户数据进行翻转	**fe1 data-coding { inverted | normal }**	可选 缺省情况下，不对用户数据进行翻转
配置接口进行 AIS 检测	**fe1 detect-ais**	可选 缺省情况下，进行 AIS 检测
配置接口的帧格式	**fe1 frame-format { crc4 | no-crc4 }**	可选 缺省情况下，E1-F 接口的帧格式为 **no-crc4**
配置接口的线路空闲码类型	**fe1 idlecode { 7e | ff }**	可选 缺省情况下，E1-F 接口的线路空闲码为 **7e**
配置接口的帧间填充符类型	**fe1 itf type { 7e | ff }**	可选 缺省情况下，E1-F 接口的帧间填充符为 **7e**
配置接口的帧间填充字节的个数	**fe1 itf number** *number*	可选 缺省情况下，E1-F 接口的帧间填充字节的个数为 4 个
使能环回检测功能，并配置检测方式	**fe1 loopback { local | payload | remote }**	可选 缺省情况下，禁止环回检测功能
关闭 E1-F 接口	**shutdown**	可选 缺省情况下，E1-F 接口为打开状态

1.4.5 E1-F 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 E1-F 接口的运行情况，通过查看显示信息验证配置的效果。

表1-13 E1-F 接口显示和维护

配置	命令
显示 E1-F 接口的配置和状态信息	**display fe1** [**serial** *interface-number*]
显示 E1-F 接口的工作状态	**display interface serial** *interface-number*

1.5 T1-F 接口

1.5.1 T1-F 接口介绍

T1-F 接口是指部分（Fractional）化 T1 接口，它是 CT1 接口的简化版本。在 T1 接入应用中，如果不需要划分出多个通道组（channel set），使用 CT1 接口就显得浪费。此时，可以利用 T1-F 接口来满足这些简单的 T1 接入需求。相对 CT1 接口而言，使用 T1-F 接口是一种低价位的 T1 接入方案。

T1-F 接口工作在成帧方式时，只能将时隙捆绑为一个通道组，而 CT1 接口可以将时隙任意分组，捆绑出多个通道组。

T1 线路由 24 个多路复用信道组成，即一个 T1 基群帧 DS1 包含 24 个 DS0（64kbps）时隙和 1 bit 帧同步位（framing bit），每个时隙有 8 个 bit 位，故每个基群帧共 24 X 8＋1＝193bit。由于每秒钟可以传送 8000 帧，故 DS1 的传送速率为 193 X 8k=1544kbps。

T1-F 接口只能工作在成帧工作方式，它可以将全部时隙（时隙 1～24）任意地捆绑成一个组（channel set），T1-F 接口的速率为 n×64kbps 或 n×56kbps，其逻辑上等同于同步串口，支持 PPP、HDLC、帧中继等数据链路层协议，支持 IP 等网络协议。

1.5.2 配置 T1-F 接口

表1-14 配置 T1-F 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入 T1-F 接口的视图	**interface serial** *interface-number*	-
配置 T1-F 接口的描述字符串	**description** *text*	可选 缺省情况下，T1-F 接口的描述字符串为“*该接口的接口名* interface”
配置接口捆绑时隙和速率	**ft1 timeslot-list** *range* [**speed** { **56k** | **64k** }]	可选 缺省情况下，T1-F 接口对所有时隙进行捆绑，时隙的缺省速率为 64kbps，即 T1-F 接口的缺省速率为 1536kbps
配置传输线路衰减	**ft1 cable** { **long** *decibel* | **short** *length* }	可选 缺省情况下，T1-F 接口的传输线路衰减为 **long 0db**

操作	命令	说明
配置线路编解码格式	**ft1 code** { **ami** | **b8zs** }	可选 缺省情况下，T1-F 接口的线路编解码格式为 **b8zs**
配置线路时钟的工作模式	**ft1 clock** { **master** | **slave** }	可选 缺省情况下，T1-F 接口的时钟模式为 **slave** 时钟
配置对用户数据进行翻转	**ft1 data-coding** { **inverted** | **normal** }	可选 缺省情况下，不对用户数据进行翻转
配置在 ESF 格式时 FDL 比特位的使用模式	**ft1 fdl** { **ansi** | **att** | **both** | **none** }	可选 缺省情况下，FDL 为禁用状态
配置接口的 CRC 校验模式	**crc** { **16** | **32** | **none** }	可选 缺省情况下，使用 16 位 CRC 校验
配置接口的帧格式	**ft1 frame-format** { **esf** | **sf** }	可选 缺省情况下，接口的帧格式为 **esf**
配置检测远端告警信号	**ft1 alarm detect rai**	可选 缺省情况下，检测远端告警信号 在接口帧格式采用 esf 的情况下，可以使用该命令
配置接口的告警门限	**ft1 alarm-threshold** { **ais** { **level-1** | **level-2** } | **lfa** { **level-1** | **level-2** | **level-3** | **level-4** } | **los** { **pulse-detection** | **pulse-recovery** } *value* }	可选 缺省情况，下对于 LOS 告警，**pulse-detection** 参数的值为 176，**pulse-recovery** 的值为 22，即默认的情况下，如果在 176 个脉冲周期内检测到的脉冲数小于 22 个则认为载波丢失 LOS 告警产生；对于 AIS 告警，缺省值为 level-1；对于 LFA 告警，缺省值为 level-1
配置接口的线路空闲码类型	**ft1 idlecode** { **7e** | **ff** }	可选 缺省情况下，T1-F 接口的线路空闲码为 **7e**
配置接口的帧间填充符类型	**ft1 itf type** { **7e** | **ff** }	可选 缺省情况下，T1-F 接口的帧间填充符为 **7e**
配置接口的帧间填充字节的个数	**ft1 itf number** *number*	可选 缺省情况下，T1-F 接口的帧间填充字节的个数为 4 个
使能环回检测功能，并配置检测方式	**ft1 loopback** { **local** | **payload** | **remote** }	可选 缺省情况下，环回检测功能是禁止的
配置发送远程环回控制码	**ft1 sendloopcode** { **fdl-ansi-llb-down** | **fdl-ansi-llb-up** | **fdl-ansi-plb-down** | **fdl-ansi-plb-up** | **fdl-att-plb-down** | **fdl-att-plb-up** | **inband-llb-down** | **inband-llb-up** }	可选 缺省情况下，不发送远程环回控制码
关闭 T1-F 接口	**shutdown**	可选 缺省情况下，T1-F 接口为打开状态

1.5.3 配置 T1-F 接口进行线路位（Bit）错误率的测试

BERT 测试方式为，本端发出测试数据流，经过线路某处环回来，检测收到的测试数据流与发出的测试数据流是否一致，位错误率达到多少，从而为用户判断线路状态提供依据。因此，要求线路中某处能环回发出的数据流，如将对方配置远端环回等。利用 **bert** 命令配置好测试模式，指定测试时间，开始测试后，可以查看接口状态中的 BERT 测试状态和测试结果。

表1-15 配置 T1-F 接口进行线路位（Bit）错误率的测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 T1-F 接口的视图	**interface serial** *interface-number*	-
配置 T1-F 接口进行线路位（Bit）错误率的测试	**ft1 bert pattern** { **2^20** | **2^15** } **time** *minutes* [**unframed**]	必选 缺省情况下，不进行线路位错误率的测试

1.5.4 T1-F 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 T1-F 接口的运行情况，通过查看显示信息验证配置的效果。

表1-16 T1-F 接口显示和维护

操作	命令
显示 T1-F 接口的配置和状态信息	**display ft1** [**serial** *interface-number*]
显示 T1-F 接口的工作状态	**display interface serial** *interface-number*

1.6 CE3 接口

1.6.1 CE3 接口介绍

E3 与 E1 同属于 ITU-T 的数字载波体系，数据传输速率为 34.368Mbps，线路编解码方式采用 HDB3。

CE3 接口有两种工作模式，E3 和 CE3。缺省情况下，CE3 接口工作在通道化模式（CE3）。

- E3 模式

当工作在 E3 模式时，它相当于一个不分时隙，数据带宽为 34.368Mbps 的接口。系统会自动创建一个 Serial 接口，编号为 **serial** *number/line-number*/**0:0**。此接口的速率为 34.368Mbps，其逻辑特性与同步串口相同，可以视其为同步串口进行配置。

- CE3 模式

当工作在 CE3 模式时，它可以解复用 16 路 E1 信号，E3 到 E1 的解复用符合 ITU-T G.751 和 G.742 规范。每个 E1 又可以分为 32 个时隙，对应编号为 0～31，其中 1～31 时隙可任意捆绑为 N×64kbps 的逻辑通道（时隙 0 用于传送帧同步信号，不能被捆绑）。因此，CE3 支持通道化到 E1 方式和通道化 CE1 方式。

当 E1 通道工作在非成帧方式（E1 方式）时，系统会自动创建一个 Serial 接口，编号为 **serial** *number/line-number*:**0**。此接口的速率为 2048kbps，其逻辑特性与同步串口相同，可以视其为同步串口进行的配置。

当 E1 通道工作在成帧方式（CE1 方式）时，可以对其进行时隙捆绑。系统会自动创建一个 Serial 接口，编号为 **serial** *number/line-number*:*set-number*。此接口的速率为 N×64kbps，其逻辑特性与同步串口相同，可以视其为同步串口进行配置。

CE3 接口支持 PPP、HDLC、帧中继等链路层协议，支持 IP 等网络协议。

1.6.2 配置 CE3 接口（工作在 E3 方式）

表1-17 配置 CE3 接口（工作在 E3 方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CE3 接口视图	**controller e3** *interface-number*	-
配置 CE3 接口工作在 E3 方式	**using e3**	必选 缺省情况下，CE3 接口在 CE3 方式
配置 CE3 接口工作在 FE3 模式，并配置 DSU 模式或子速率	**fe3** { **dsu-mode** { **0** | **1** } | **subrate** *number* }	可选 缺省情况下，DSU 模式为 1，即 Kentrox 模式；子速率为 34010kbps
配置 CE3 接口的其他参数	请参见 1.6.4 配置CE3 接口其他参数	可选

根据组网需要，可能还要对 CE3 接口配置 PPP、帧中继参数，以及 IP 地址等，具体内容请参考本手册相关章节。

1.6.3 配置 CE3 接口（工作在 CE3 方式）

表1-18 配置 CE3 接口（工作在 CE3 方式）

操作		命令	说明
进入系统视图		**system-view**	-
进入指定 CE3 接口视图		**controller e3** *interface-number*	-
配置 CE3 接口工作在 CE3 方式		**using ce3**	可选 缺省情况下，CE3 接口在 CE3 方式
配置 CE3 接口的 E1 通道工作方式（非成帧方式或者成帧方式，二者选择其一）	配置 CE3 接口的 E1 通道工作在 E1 方式（非成帧方式）	**e1** *line-number* **unframed**	必选 缺省情况下，E1 通道的方式是 CE1 方式
	配置 CE3 接口的 E1 通道工作在 CE1 方式（成帧方式），并进行 CE1 的时隙捆绑	**undo e1** *line-number* **unframed** **e1** *line-number* **channel-set** *set-number* **timeslot-list** *list*	可选 缺省情况下，E1 通道的方式是 CE1 方式（成帧方式） 必选 缺省情况下，不捆绑任何 channel set
配置 CE3 接口的其他参数		请参见 1.6.4 配置CE3 接口其他参数	可选

根据组网需要，可能还要对 CE3 接口配置 PPP、帧中继参数，以及 IP 地址等，具体内容请参考本手册相关章节。

1.6.4 配置 CE3 接口其他参数

表1-19 配置 CE3 接口其他参数

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CE3 接口视图	**controller e3** *interface-number*	-
配置 CE3 接口的描述字符串	**description** *text*	可选 缺省情况下，CE3 接口的描述字符串为“*该接口的接口名* interface”
配置 CE3 接口进行线路位（Bit）错误率的测试 配置 CE3 接口下某 E1 通道的线路位（Bit）错误率的测试	**bert pattern** { **2^7** | **2^11** | **2^15** | **qrss** } **time** *number* [**unframed**] **e1** *line-number* **bert pattern** { **2^11** | **2^15** | **2^20** | **2^23** | **qrss** } **time** *number* [**unframed**]	可选 缺省情况下，不进行线路位错误率的测试
配置 CE3 接口的时钟模式 配置 E1 通道的时钟模式（工作在 CE3 方式）	**clock** { **master** | **slave** } **e1** *line-number* **set clock** { **master** | **slave** }	可选 缺省情况下，CE3 使用线路时钟 **slave** 可选 缺省情况下，E1 通道使用线路时钟 **slave**
配置 CE3 接口的 national bit	**national-bit** { **0** | **1** }	可选 缺省情况下，CE3 接口的 national bit 为 1
使能环回检测功能，并配置检测方式	**loopback** { **local** | **payload** | **remote** } **e1** *line-number* **set loopback** { **local** | **remote** | **payload** }	可选 缺省情况下，环回检测功能是禁止的
配置 E1 帧格式	**e1** *line-number* **set frame-format** { **crc4** | **no-crc4** }	可选 缺省情况下，E1 通道的帧格式为 **no-crc4**
退回系统视图	**quit**	-
进入 CE3 接口通过各种方式形成的同步串口视图	**interface serial** *number*/*line-number*:**0** 或 **interface serial** *number*/*line-number*:*set-number*	必选
配置 CRC 校验模式	**crc** { **16** | **32** | **none** }	可选 缺省情况下，使用 16 位 CRC 校验

1.6.5 CE3 接口显示和维护

CE3 接口的显示和调试操作包括关闭接口和显示接口信息。



注意

关闭接口会导致接口停止工作，请慎用此命令。

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 CE3 接口的运行情况，通过查看显示信息验证配置的效果。

表1-20 CE3 接口显示和维护

配置	命令	说明
显示 CE3 接口的状态信息	**display controller e3** [*interface-number*]	**display** 命令可以在任意视图执行
查看 CE3 接口形成的串口的配置和状态信息	**display interface serial** *interface-number*	
清除 CE3 接口的 controller 计数器	**reset counters controller e3** *interface-number*	**reset** 命令可以在用户视图执行
关闭 CE3 接口	**shutdown**	请在 CE3 接口视图下执行关闭或打开接口的命令
打开 CE3 接口	**undo shutdown**	
关闭 E1 通道	**e1** *line-number* **shutdown**	
打开 E1 通道	**undo e1** *line-number* **shutdown**	

说明

- 关闭或启动 CE3 接口，对于 E3 形成的串口、CE3 解复用出的 E1 通道、E1 通道形成的串口及 E1 通过时隙捆绑形成的串口都有效。
- 关闭或启动 E1 通道，对于 E1 通道形成的串口及 E1 通过时隙捆绑形成的串口都有效。
- 如果只关闭或启动 E3 形成的串口、E1 通道形成的串口、以及 E1 通道时隙捆绑形成的串口，可在相应的串口视图下执行 **shutdown/undo shutdown** 命令。

1.7 CT3 接口

1.7.1 CT3 接口介绍

T3 和 T1 同属于美国国家标准协会 ANSI 制定的 T-载波体系，T3 对应的数字信号级别为 DS-3，数据传输速率为 44.736Mbps。

CT3 接口可支持两种工作模式：T3 模式（也称为非通道化工作模式）和 CT3 模式（也称为通道化工作模式）。

- T3 模式

当工作在 T3 模式时，相当于一个不分时隙，数据带宽为 44.736Mbps 的同步串口。系统会自动创建一个 Serial 接口，编号为 serial number/line-number/0:0。此接口的速率为 44.736Mbps，其逻辑特性与同步串口相同，可以视其为同步串口进行进一步的配置。

- CT3 模式

当工作在 CT3 模式时，可以分解为 28 路 T1 信号。每个 T1 又可分为 24 个时隙，对应编号为 1～24。和 E1 不同，T1 的每个数据通道速率可配置为 64k 和 56k 两种。因此，在 CT3 模式下，可任意捆绑 M×1.536Mbps（M=1～28）的逻辑通道或 N×56kbps 及 N×64kbps（N＝1～300）的逻辑通道。

当 T1 通道工作在非成帧方式（T1 方式）时，系统会自动创建一个 Serial 接口，编号为 **serial** *number/line-number*:**0**。此接口的速率为 1544kbps，其逻辑特性与同步串口相同，可以视其为同步串口进行进一步的配置。

当 T1 通道工作在成帧方式（CT1 方式）时，可以对其进行时隙捆绑。系统会自动创建一个 Serial 接口，编号为 **serial** *number/line-number*:*set-number*。此接口的速率为 N×64kbps（或 N×56kbps），其逻辑特性与同步串口相同，可以视其为同步串口进行进一步的配置。

1.7.2 配置 CT3 接口（工作在 T3 模式）

表1-21 配置 CT3 接口（工作在 T3 模式）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CT3 接口视图	**controller t3** *interface-number*	-
配置 CT3 接口的工作模式为 T3	**using t3**	必选 缺省情况下，CT3 接口工作在 CT3 模式
配置 CT3 接口工作在 FT3 模式，并配置 DSU 模式或子速率	**ft3** { **dsu-mode** { **0** | **1** | **2** | **3** | **4** } | **subrate** *number* }	可选 缺省情况下，DSU 模式为 0，即 Digital Link 模式；子速率为 44210kbps
配置 CT3 接口的其他参数	请参见 1.7.4 配置CT3 接口其他参数	可选

CT3 接口封装成的串口可支持 PPP、HDLC、帧中继等链路层协议，支持 IP 等网络协议。根据组网需要，可能还要对 CT3 接口配置链路层协议、IP 地址等，具体内容请参考本手册相关章节。

1.7.3 配置 CT3 接口（工作在 CT3 模式）

表1-22 配置 CT3 接口（工作在 CT3 模式）

操作		命令	说明
进入系统视图		**system-view**	-
进入指定 CT3 接口视图		**controller t3** *interface-number*	-
配置 CT3 接口的工作模式为 CT3		**using ct3**	可选 缺省情况下，CT3 接口工作在 CT3 模式
配置 CT3 接口的 T1 通道的工作方式（非成帧方式或者成帧方式，二者选择其一）	配置 CT3 接口的 T1 通道工作在 T1 方式（非成帧方式）	**t1** *line-number* **unframed**	必选 缺省情况下，T1 接口工作在 CT1 模式
	配置 CT3 接口的 T1 通道工作在 CT1 方式（成帧方式），进行 CT1 的时隙捆绑	**undo t1** *line-number* **unframed** **t1** *line-number* **channel-set** *set-number* **timeslot-list** *range* [**speed** { **56k** | **64k** }]	可选， 缺省情况下，T1 的工作方式是 CT1 方式 必选 缺省情况下，不捆绑任何 channel set。时隙的缺省速率为 64kbps
配置 CT3 接口的其他参数		请参见 1.7.4 配置CT3 接口其他参数	可选

CT3 接口封装成的串口可支持 PPP、MP、HDLC、帧中继等链路层协议，支持 IP 等网络协议。根据组网需要，可能还要对 CT3 接口配置链路层协议、IP 地址等，具体内容请参考本手册相关章节。

1.7.4 配置 CT3 接口其他参数

表1-23 配置 CT3 接口其他参数

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 CT3 接口视图	**controller t3** *interface-number*	-
配置 CT3 接口的描述字符串	**description** *text*	可选 缺省情况下，CT3 接口的描述字符串为“*该接口的接口名* interface”
配置 CT3 接口的时钟模式 配置 T1 通道的时钟模式（工作在 CT3 模式）	**clock** { **master** | **slave** } **t1** *line-number* **set clock** { **master** | **slave** }	可选 缺省情况下，CT3 接口使用线路时钟 **slave**
配置 CT3 接口的电缆长度	**cable** *feet*	可选 缺省情况下，电缆长度为 49 英尺
配置 CT3 接口的环回模式 配置 T1 通道的环回方式（工作在 CT3 模式）	**loopback** { **local** | **payload** | **remote** } **t1** *line-number* **set loopback** { **local** | **payload** | **remote** }	可选 缺省情况下，CT3 接口不进行任何形式的环回
配置 CT3 接口的帧格式	**frame-format** { **c-bit** | **m23** }	可选 缺省情况下，CT3 接口使用 C-bit 帧格式
配置 T1 通道的帧格式	**t1** *line-number* **set frame-format** { **esf** | **sf** }	可选 缺省情况下，T1 通道使用扩展超帧格式 **esf**
配置 CT3 接口的告警信号检测与发送功能 配置 T1 通道的告警信号检测与发送功能	**alarm** { **detect** | **generate** { **ais** | **febe** | **idle** | **rai** } } **t1** *line-number* **alarm** { **detect** | **generate** { **ais** | **rai** } }	可选 缺省情况下，检测功能处于打开状态
配置 CT3 接口进行线路位错误率的测试，并配置测试方式 进行 CT3 接口下某 T1 通道的线路位错误率的测试，并配置测试方式	**bert pattern** { **2^7** | **2^11** | **2^15** | **qrss** } **time** *number* [**unframed**] **t1** *line-number* **bert pattern** { **2^11** | **2^15** | **2^20** | **2^23** | **qrss** } **time** *number* [**unframed**]	可选 缺省情况下，不进行线路位错误率的测试
配置 CT3 接口的 FEAC 链路数据的检测和传输功能	**feac detect** **feac generate loopback** { **ds3-line** | **ds3-payload** } **feac generate** { **ds3-los** | **ds3-ais** | **ds3-oof** | **ds3-idle** | **ds3-eqptfail** }	可选 缺省情况下，CT3 接口的 FEAC 定时检测功能是打开的，但不发送任何 FEAC 信号
配置 CT3 接口的 MDL 链路消息检测与传输功能	**mdl** { **data** { **eic** *string* | **fic** *string* | | **gen-no** *string* | **lic** *string* | **pfi** *string* | **port-no** *string* | **unit** *string* } | **detect** | **generate** { **idle-signal** | **path** | **test-signal** } }	可选 缺省情况下，CT3 接口的 MDL 定时检测功能是禁止的，消息参数为缺省值，不发送任何消息
配置对端 CT3 接口的某个 T1 通道为某种环回状态	**t1** *line-number* **sendloopcode** { **fdl-ansi-line-up** | **fdl-ansi-payload-up** | **fdl-att-payload-up** | **inband-line-up** }	可选

操作	命令	说明
配置 T1 通道的 FDL 链路格式	**t1** *line-number* **set fdl** { **ansi** | **att** | **both** | **none** }	可选 缺省情况下，FDL 为禁止状态。 需要注意的是：只有在 CT3 接口下的 T1 通道工作在通道化模式下，且 T1 帧格式为 ESF 时候，该配置才有效
退回系统视图	**quit**	-
进入 CT3 接口通过各种方式形成的同步串口视图	**interface serial** *number/line-number*:**0** 或 **interface serial** *number/line-number*:*set-number*	必选
配置 CRC 校验模式	**crc** { **16** | **32** | **none** }	可选 缺省情况下，使用 16 位 CRC 校验

1.7.5 CT3 接口显示和维护

CT3 接口的显示和调试操作包括关闭接口和显示接口信息。

注意

关闭接口会导致接口停止工作，请慎用此命令。

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 CT3 接口的运行情况，通过查看显示信息验证配置的效果

表1-24 CT3 接口显示和维护

配置	命令	说明
查看 CT3 接口的工作状态	**display controller t3** [*interface-number*]	**display** 命令可以在任意视图执行
查看 CT3 接口形成的串口的配置和状态信息	**display interface serial** *interface-number*	
清除 CT3 接口的 controller 计数器	**reset counters controller t3** *interface-number*	**reset** 命令可以在用户视图执行
快捷显示 T1 通道线路状态	**t1** *line-number* **show**	请在 CT3 接口视图下执行该命令
关闭 CT3 接口	**shutdown**	请在 CT3 接口视图下执行关闭或打开接口的命令
启动 CT3 接口	**undo shutdown**	
关闭 T1 通道	**t1** *t1-number* **shutdown**	
启动 T1 通道	**undo t1** *t1-number* **shutdown**	

说明

- 关闭或启动 CT3 接口，对于 T3 形成的串口、CT3 解复用出的 T1 通道、T1 通道形成的串口及 T1 通过时隙捆绑形成的串口都有效。
- 关闭或启动 T1 通道，对于 T1 通道形成的串口及 T1 通过时隙捆绑形成的串口都有效。
- 如果只关闭或启动 T3 形成的串口、T1 通道形成的串口、以及 T1 通道时隙捆绑形成的串口，可在相应的串口视图下执行 **shutdown/undo shutdown** 命令。

目 录

1 帧中继配置 1-1
1.1 帧中继基本概念 1-1
1.1.1 帧中继协议简介 1-1
1.1.2 常用术语 1-1
1.1.3 虚电路介绍 1-1
1.1.4 帧中继协议参数 1-1
1.1.5 帧中继地址映射 1-2
1.2 帧中继配置任务简介 1-3
1.3 配置帧中继DTE侧 1-3
1.3.1 帧中继DTE侧基本配置 1-3
1.3.2 配置帧中继地址映射 1-4
1.3.3 配置帧中继本地虚电路 1-4
1.3.4 配置帧中继交换 1-5
1.3.5 配置帧中继子接口 1-5
1.4 配置帧中继DCE侧 1-6
1.4.1 帧中继DCE侧基本配置 1-6
1.4.2 配置帧中继地址映射 1-7
1.4.3 配置帧中继本地虚电路 1-7
1.4.4 配置帧中继交换 1-7
1.4.5 配置帧中继子接口 1-7
1.5 开启Trap功能 1-7
1.6 帧中继显示和维护 1-7
1.7 帧中继典型配置举例 1-8
1.7.1 通过帧中继网络互连局域网 1-8
1.7.2 通过专线互连局域网 1-9
1.8 帧中继常见配置错误举例 1-11
1.8.1 物理层处于down状态 1-11
1.8.2 物理层已经处于up状态，但链路层协议处于down状态 1-11
1.8.3 链路层up，但是ping不通对方 1-11
2 多链路帧中继 2-1
2.1 多链路帧中继简介 2-1
2.2 配置多链路帧中继 2-1
2.3 多链路帧中继显示和维护 2-2
2.4 多链路帧中继典型配置举例 2-3
2.4.1 多链路帧中继直连配置举例 2-3
2.4.2 多链路帧中继交换配置举例 2-4

1 帧中继配置

1.1 帧中继基本概念

1.1.1 帧中继协议简介

帧中继协议是一种简化的 X.25 广域网协议，是一种统计复用的协议，它在单一物理传输线路上能够提供多条虚电路。每条虚电路用 DLCI（Data Link Connection Identifier，数据链路连接标识）来标识，DLCI 只在本地接口和与之直接相连的对端接口有效，不具有全局有效性，即在帧中继网络中，不同的物理接口上相同的 DLCI 并不表示是同一个虚电路。

帧中继网络既可以是公用网络、某一企业的私有网络，也可以是两台设备之间直接连接构成的网络。

1.1.2 常用术语

帧中继网络提供了用户设备（如路由器和主机等）之间进行数据通信的能力，常见术语介绍如下：

- DTE：用户设备被称作 DTE（Data Terminal Equipment，数据终端设备）；
- DCE：为用户设备提供接入的设备，属于网络设备，被称为 DCE（Data Circuit-terminating Equipment，数据电路终接设备）；
- UNI：DTE 和 DCE 之间的接口被称为 UNI（User Network Interface，用户网络接口）；
- NNI：网络与网络之间的接口被称为 NNI（Network-to-Network Interface，网间网接口）。

1.1.3 虚电路介绍

根据虚电路建立方式的不同，虚电路分为两种类型：PVC（Permanent Virtual Circuit，永久虚电路）和 SVC（Switched Virtual Circuit，交换虚电路）。手工设置产生的虚电路称为永久虚电路。通过协议协商产生的虚电路称为交换虚电路，这种虚电路由帧中继协议自动创建和删除。目前在帧中继中使用最多的方式是永久虚电路方式。

在永久虚电路方式下，需要检测虚电路是否可用。LMI（Local Management Interface，本地管理接口）协议就是用来检测虚电路是否可用的。LMI 协议用于维护帧中继协议的 PVC 表，包括：通知 PVC 的增加、探测 PVC 的删除、监控 PVC 状态的变更、验证链路的完整性。系统支持三种本地管理接口协议：ITU-T 的 Q.933 附录 A、ANSI 的 T1.617 附录 D 以及非标准兼容协议。

LMI 协议的基本工作方式是：DTE 设备每隔一定的时间间隔发送一个状态请求（Status Enquiry）报文去查询虚电路的状态，DCE 设备收到状态请求报文后，立即用状态（Status）报文通知 DTE 当前接口上所有虚电路的状态。

对于 DTE 侧设备，永久虚电路的状态完全由 DCE 侧设备决定；对于 DCE 侧设备，永久虚电路的状态由网络来决定。在两台网络设备直接连接的情况下，DCE 侧设备的虚电路状态是由设备管理员来设置的。

1.1.4 帧中继协议参数

帧中继协议的参数以及含义如 表 1-1所示。

表1-1 帧中继协议参数含义

设备角色	参数含义	取值范围	缺省值
DTE	请求 PVC 状态的计数器（N391）	1～255	6
	错误门限（N392）	1～10	3
	事件计数器（N393）	1～10	4
	用户侧轮询定时器（T391），当为 0 时，表示禁止 LMI 协议	0～32767（单位：秒）	10（单位：秒）
DCE	错误门限（N392）	1～10	3
	事件计数器（N393）	1～10	4
	网络侧轮询定时器（T392）	5～30（单位：秒）	15（单位：秒）

这些参数由 Q.933 的附录 A 规定，各参数的含义如下。

1. 与 DTE 相关的参数含义

- N391：用来定义链路完整性验证报文和链路状态查询报文的发送比例，即（链路完整性验证报文数：链路状态查询报文数）= （N391-1：1）
- N392：表示在被观察的事件总数中发生错误的门限。
- N393：表示被观察的事件总数。
- T391：这是一个时间变量，它定义了 DTE 设备发送状态请求报文的时间间隔。

DTE 设备每隔一定的时间间隔（由 T391 决定）要发送一个状态请求报文（状态请求报文有两种类型：链路完整性验证报文和链路状态查询报文）去查询链路状态，DCE 设备收到该报文后应立即发送状态响应报文。如果 DTE 设备在规定的时间内没有收到响应，就记录该错误。如果错误次数超过门限，DTE 设备就认为物理通路不可用，所有的虚电路都不可用（上面 N392、N393 两个参数一起定义了“错误门限”。即：如果 DTE 设备发送 N393 个状态请求报文中，发生错误数达到 N392，DTE 设备就认为错误次数达到门限）。

2. 与 DCE 相关的参数含义

- N392、N393 两个参数的含义与 DTE 中的含义相似，区别在于：DCE 设备要求 DTE 设备发送状态请求报文的固定时间间隔由 T392 决定，不同于 DTE 由 T391 决定，若 DCE 在 T392 时间间隔内没有收到 DTE 的状态请求报文，则记录错误数。
- T392：这是一个时间变量，它定义了 DCE 设备等待一个状态请求报文的最长时间。

1.1.5 帧中继地址映射

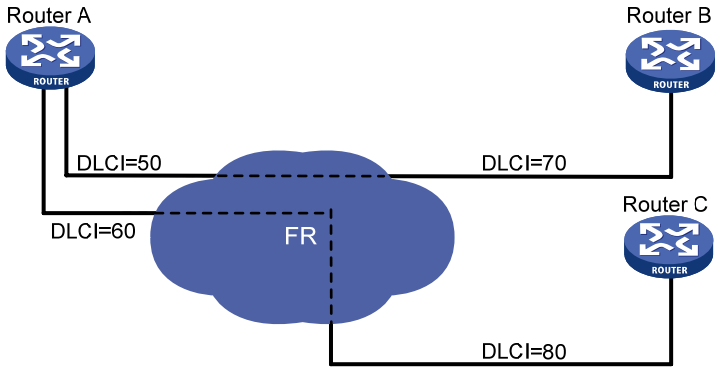
帧中继地址映射是把对端设备的协议地址与对端设备的帧中继地址（本地的 DLCI）关联起来，使高层协议能通过对端设备的协议地址寻址到对端设备。

帧中继主要用来承载 IP 协议，在发送 IP 报文时，根据路由表只能知道报文的下一跳地址，发送前必须由该地址确定它对应的 DLCI。这个过程可以通过查找帧中继地址映射表来完成，因为地址映射表中存放的是下一跳 IP 地址与其对应的 DLCI 的映射关系。

地址映射表可以由手工配置，也可以由 InARP（Inverse Address Resolution Protocol，逆向地址解析协议）动态维护。

如 图 1-1所示，通过帧中继网络可以实现局域网互联。

图1-1 通过帧中继网络实现局域网互联



1.2 帧中继配置任务简介

表1-2 帧中继配置任务简介

配置任务		说明	详细配置
配置帧中继 DTE 侧	帧中继 DTE 侧基本配置	必选	1.3.1
	配置帧中继地址映射	必选	1.3.2
	配置帧中继本地虚电路	必选	1.3.3
	配置帧中继交换	可选	1.3.4
	配置帧中继子接口	可选	1.3.5
配置帧中继 DCE 侧	帧中继 DCE 侧基本配置	必选	1.4.1
	配置帧中继地址映射	必选	1.4.2
	配置帧中继本地虚电路	必选	1.4.3
	配置帧中继交换	可选	1.4.4
	配置帧中继子接口	可选	1.4.5
开启 Trap 功能		可选	1.5

1.3 配置帧中继 DTE 侧

1.3.1 帧中继 DTE 侧基本配置

表1-3 帧中继 DTE 侧基本配置

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口封装的链路层协议为帧中继	**link-protocol fr [ietf | nonstandard]**	必选 缺省情况下，接口的链路层协议封装为 PPP，当封装帧中继协议时，缺省的封装格式为 IETF

操作	命令	说明
配置帧中继接口类型为 DTE	**fr interface-type dte**	可选 缺省情况下，帧中继接口类型为 DTE
配置帧中继 LMI 协议类型	**fr lmi type** { **ansi** | **nonstandard** | **q933a** }	可选 缺省情况下，接口的 LMI 协议类型为 **q933a**
配置 DTE 侧 N391 参数	**fr lmi n391dte** *n391-value*	可选 缺省情况下，该参数的值为 6
配置 DTE 侧 N392 参数	**fr lmi n392dte** *n392-value*	可选 缺省情况下，该参数值为 3
配置 DTE 侧 N393 参数	**fr lmi n393dte** *n393-value*	可选 缺省情况下，该参数值为 4
配置 DTE 侧 T391 参数	**timer hold** *seconds*	可选 缺省情况下，该参数为 10 秒

1.3.2 配置帧中继地址映射

帧中继地址映射可以通过下面两种方式建立：

- 静态配置：手工建立对端 IP 地址与本地 DLCI 的映射关系，一般用于对端主机较少或有缺省路由的情况。
- 动态建立：运行 InARP 后，可以动态地建立对端 IP 地址与本地 DLCI 的映射关系，适用于对端设备也支持 InARP 且网络较复杂的情况。

表1-4 配置帧中继地址映射

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	必选
增加一条静态地址映射	**fr map ip** { *ip-address* [*ip-mask* | *dlci-number*] | **default** [*dlci-number*] } [**broadcast** | [**ietf** | **nonstandard**]] *	可选 缺省情况下，系统没有静态地址映射
使能帧中继 InARP 以建立动态地址映射	**fr inarp** [**ip** [*dlci-number*]]	可选 缺省情况下，系统使能 InARP

1.3.3 配置帧中继本地虚电路

当帧中继接口类型是 DCE 或 NNI 时，必须为接口（不论是主接口还是子接口）手工配置虚电路。当帧中继接口类型是 DTE 时，如果接口是主接口，则系统会根据对端设备自动确定虚电路，也可以手工配置虚电路；如果是子接口，则必须手动为接口配置虚电路。

虚电路号在一个物理接口上是唯一的。

表1-5 配置帧中继本地虚电路

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口上配置虚电路	**fr dlci** *dlci-number*	必选 缺省情况下，接口上没有配置虚电路

1.3.4 配置帧中继交换

使能帧中继交换功能后设备可以作为帧中继交换机。当设备用做帧中继交换机时，帧中继接口类型应该为 NNI 或 DCE。而且，必须在设备用于帧中继交换的两个或两个以上接口上都进行了相应配置，帧中继交换才会起作用。

使能帧中继交换以后，需要配置帧中继交换的虚电路，该配置在系统视图下进行。

表1-6 配置帧中继交换

操作	命令	说明
进入系统视图	**system-view**	-
使能帧中继交换功能	**fr switching**	必选
进入接口视图	**interface** *interface-type interface-number*	-
设置担负帧中继交换功能的帧中继接口类型为 NNI 或 DCE	**fr interface-type** { **dce** | **nni** }	必选 缺省情况下，帧中继接口类型为 DTE 如果配置帧中继接口类型为 DTE，则帧中继交换功能不起作用
配置用于帧中继交换的虚电路	**quit**	退回系统视图
	fr switch *name* **interface** *interface-type interface-number* **dlci** *dlci1* **interface** *interface-type interface-number* **dlci** *dlci2*	必选
	fr switch *name*	可选 进入帧中继交换 PVC 的视图
	undo shutdown	可选 使能当前的交换 PVC

1.3.5 配置帧中继子接口

1. 帧中继子接口简介

帧中继有两种类型的接口：主接口和子接口。其中子接口是一个逻辑结构，可以配置协议地址和虚电路等，一个物理接口可以有多个子接口。虽然子接口是逻辑结构，并不实际存在，但对于网络层而言，子接口和主接口是没有区别的，都可以配置虚电路与远端设备相连。

帧中继的子接口又可以分为两种类型：点到点（point-to-point）子接口和点到多点（point-to-multipoint）子接口。点到点子接口用于连接单个远端目标，点到多点子接口用于连接多个远端目标。点到多点子接口在一个子接口上配置多条虚电路，每条虚电路都和它相连的远端网络地址建立一个地址映射，这样不同的虚电路就可以到达不同的远端而不会混淆。

地址映射的建立可以用手工配置的方法，也可以利用逆向地址解析协议来动态建立。点到点子接口和点到多点子接口配置虚电路以及地址映射的方法是不同的：

- 点到点（point-to-point）子接口

对点到点子接口而言，因为只有唯一的一个对端地址，所以在给子接口配置一条 PVC 时实际已经确定了对端地址，不必配置动态或静态地址映射。

- 点到多点（point-to-multipoint）子接口

对点到多点子接口，对端地址与本地 DLCI 映射可以通过配置静态地址映射或者通过逆向地址解析协议来确定（InARP 在主接口上配置即可）。如果要建立静态地址映射，则应该对每一条虚电路建立静态地址映射关系。

2. 配置帧中继子接口

表1-7 配置帧中继子接口

操作	命令	说明
进入系统视图	**system-view**	-
创建子接口并进入子接口视图	**interface** *interface-type interface-number.subnumber* [**p2mp** | **p2p**]	必选 缺省情况下，帧中继子接口类型为 **p2mp**
配置帧中继子接口的虚电路	请参见“1.3.3 配置帧中继本地虚电路”	必选
建立地址映射	请参见“1.3.2 配置帧中继地址映射”	可选 对于点到多点子接口需要配置该项

1.4 配置帧中继 DCE 侧

1.4.1 帧中继 DCE 侧基本配置

表1-8 帧中继 DCE 侧基本配置

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口封装的链路层协议为帧中继	**link-protocol fr** [**ietf** | **nonstandard**]	必选 缺省情况下，接口封装的链路层协议为 PPP
配置帧中继接口类型为 DCE 或者 NNI（Network-to-Network Interface）	**fr interface-type** { **dce** | **nni** }	必选 缺省情况下，帧中继接口类型为 DTE
配置帧中继 LMI 协议类型	**fr lmi type** { **ansi** | **nonstandard** | **q933a** }	可选 缺省情况下，接口的 LMI 协议类型为 **q933a**
配置 DCE 侧 N392 参数	**fr lmi n392dce** *n392-value*	可选 缺省情况下，该参数值为 3

操作	命令	说明
配置 DCE 侧 N393 参数	**fr lmi n393dce** *n393-value*	可选 缺省情况下，该参数值为 4
配置 DCE 侧 T392 参数	**fr lmi t392dce** *t392-value*	可选 缺省情况下，该参数值为 15 秒

1.4.2 配置帧中继地址映射

请参见“1.3.2 配置帧中继地址映射”。

1.4.3 配置帧中继本地虚电路

请参见“1.3.3 配置帧中继本地虚电路”。

1.4.4 配置帧中继交换

请参见“1.3.4 配置帧中继交换”。

1.4.5 配置帧中继子接口

请参见“1.3.5 配置帧中继子接口”。

1.5 开启 Trap 功能

开启帧中继的 Trap 功能后，会生成级别为 notifications 的 Trap 报文，用于报告帧中继的重要事件。生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-9 开启 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
开启帧中继的 Trap 功能	**snmp-agent trap enable fr**	可选 缺省情况下，帧中继的 Trap 功能处于开启状态



说明

snmp-agent trap enable fr 命令的详细介绍请参见“系统分册/SNMP 命令”中的 **snmp-agent trap enable** 命令。

1.6 帧中继显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后帧中继的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除所有自动建立的帧中继地址映射信息。

表1-10 帧中继显示和维护

操作	命令
显示各接口的帧中继协议状态	**display fr interface** [*interface-type* { *interface-number* | *interface-number.subnumber* }]
显示协议地址与帧中继地址映射表的信息	**display fr map-info** [**interface** *interface-type* { *interface-number* | *interface-number.subnumber* }]
显示帧中继 LMI 类型报文的收发统计	**display fr lmi-info** [**interface** *interface-type interface-number*]
显示帧中继数据收发统计信息	**display fr statistics** [**interface** *interface-type interface-number*]
显示帧中继永久虚电路表	**display fr pvc-info** [**interface** *interface-type* { *interface-number* | *interface-number.subnumber* }] [*dlci-number*]
显示帧中继逆向地址解析协议报文统计信息	**display fr inarp-info** [**interface** *interface-type interface-number*]
查看配置的帧中继交换的信息	**display fr dlci-switch** [**interface** *interface-type interface-number*]
清除所有自动建立的帧中继地址映射	**reset fr inarp**

1.7 帧中继典型配置举例

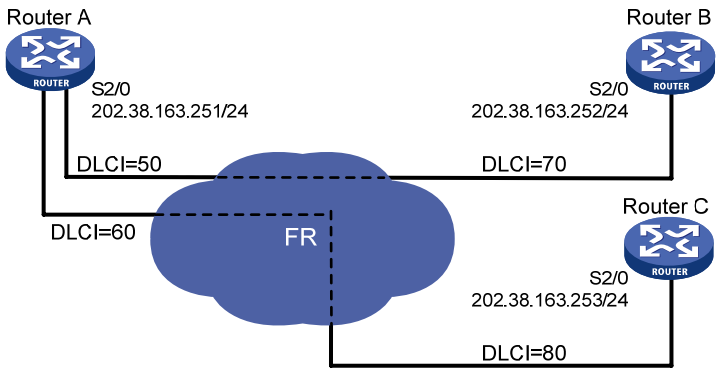
1.7.1 通过帧中继网络互连局域网

1. 组网需求

通过公用帧中继网络互连局域网，在这种方式下，路由器只能作为用户设备工作在帧中继的 DTE 方式。

2. 组网图

图1-2 通过帧中继网络互连局域网



3. 配置步骤

(1) 配置 Router A

配置接口 IP 地址。

```
<RouterA> system-view
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 202.38.163.251 255.255.255.0
```

配置接口封装为帧中继。

```
[RouterA-Serial2/0] link-protocol fr
[RouterA-Serial2/0] fr interface-type dte
```

如果对端路由器支持逆向地址解析功能，则配置动态地址映射。

```
[RouterA-Serial2/0] fr inarp
```

否则配置静态地址映射。

```
[RouterA-Serial2/0] fr map ip 202.38.163.252 50
[RouterA-Serial2/0] fr map ip 202.38.163.253 60
```

(2) 配置 Router B

配置接口 IP 地址。

```
<RouterB> system-view
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address 202.38.163.252 255.255.255.0
```

配置接口封装为帧中继。

```
[RouterB-Serial2/0] link-protocol fr
[RouterB-Serial2/0] fr interface-type dte
```

如果对端路由器支持逆向地址解析功能，则配置动态地址映射。

```
[RouterB-Serial2/0] fr inarp
```

否则配置静态地址映射。

```
[RouterB-Serial2/0] fr map ip 202.38.163.251 70
```

(3) 配置 Router C

配置接口 IP 地址。

```
<RouterC> system-view
[RouterC] interface serial 2/0
[RouterC-Serial2/0] ip address 202.38.163.253 255.255.255.0
```

配置接口封装为帧中继。

```
[RouterC-Serial2/0] link-protocol fr
[RouterC-Serial2/0] fr interface-type dte
```

如果对端路由器支持逆向地址解析功能，则配置动态地址映射。

```
[RouterC-Serial2/0] fr inarp
```

否则配置静态地址映射。

```
[RouterC-Serial2/0] fr map ip 202.38.163.251 80
```

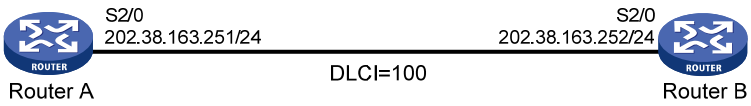
1.7.2 通过专线互连局域网

1. 组网需求

两台路由器通过串口直连，Router A 工作在帧中继的 DCE 方式，Router B 工作在帧中继的 DTE 方式。

2. 组网图

图1-3 通过专线互连局域网



3. 配置步骤

方法一：采用主接口方式

(1) 配置 Router A

配置接口 IP 地址。

```
<RouterA> system-view
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 202.38.163.251 255.255.255.0
```

配置接口的链路层协议为帧中继，工作在 DCE 方式。

```
[RouterA-Serial2/0] link-protocol fr
[RouterA-Serial2/0] fr interface-type dce
```

配置本地虚电路。

```
[RouterA-Serial2/0] fr dlci 100
```

(2) 配置 Router B

配置接口 IP 地址。

```
<RouterB> system-view
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address 202.38.163.252 255.255.255.0
```

配置接口的链路层协议为帧中继。

```
[RouterB-Serial2/0] link-protocol fr
[RouterB-Serial2/0] fr interface-type dte
```

方法二：采用子接口方式

(1) 配置 Router A

配置接口的链路层协议为帧中继，接口类型为 DCE。

```
<RouterA> system-view
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol fr
[RouterA-Serial2/0] fr interface-type dce
[RouterA-Serial2/0] quit
```

配置子接口 IP 地址及本地虚电路。

```
[RouterA] interface serial 2/0.1 p2p
[RouterA-Serial2/0.1] ip address 202.38.163.251 255.255.255.0
[RouterA-Serial2/0.1] fr dlci 100
```

(2) 配置 Router B

配置接口的链路层协议为帧中继，接口类型为缺省的 DTE。

```
<RouterB> system-view
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol fr
[RouterB-Serial2/0] quit
```

配置子接口 IP 地址及本地虚电路。

```
[RouterB] interface serial 2/0.1 p2p
[RouterB-Serial2/0.1] ip address 202.38.163.252 255.255.255.0
[RouterB-Serial2/0.1] fr dlci 100
```

1.8 帧中继常见配置错误举例

1.8.1 物理层处于 down 状态

1. 故障现象

物理层处于 down 状态。

2. 故障排除

- 检查物理线路是否正常。
- 检查对端设备是否正常运行。

1.8.2 物理层已经处于 up 状态，但链路层协议处于 down 状态

1. 故障现象

物理层已经处于 up 状态，但链路层协议处于 down 状态。

2. 故障排除

- 确认本地设备和对端设备是否都封装了帧中继协议。
- 如果两台设备直连，确认本地设备和对端设备是否配置成一端是帧中继 DTE 接口类型，一端是帧中继 DCE 接口类型。
- 确认两端配置的 LMI 协议类型是否相同。
- 如果以上检查都已经通过，可以打开帧中继 LMI 消息的监视开关，看状态请求报文与状态报文是否一一对应。如果不一一对应，说明物理层数据收发不正确，请检查物理层的问题。打开帧中继 LMI 消息的监视开关的命令请参见 **debugging fr lmi** 命令。

1.8.3 链路层 up，但是 ping 不通对方

1. 故障现象

链路层协议处于 UP 状态，但不能 **ping** 通对方。

2. 故障排除

- 确认两端设备是否都为对端配置（或产生）了正确的地址映射。
- 如果两端的 IP 地址不在同一个子网段，确认路由表是否有到达对端的路由。

2 多链路帧中继

2.1 多链路帧中继简介

多链路帧中继（Multilink Frame Relay，MFR）是为帧中继用户提供的一种性价比比较高的带宽解决方案，它基于帧中继论坛的 FRF.16 协议，实现在 DTE/DCE 接口下的多链路帧中继功能。

多链路帧中继特性提供一种逻辑接口：MFR 接口，由多个帧中继物理链路捆绑而成，每个 MFR 接口下还可以配置子接口，从而可以在帧中继网络上提供高速率、大带宽的链路。

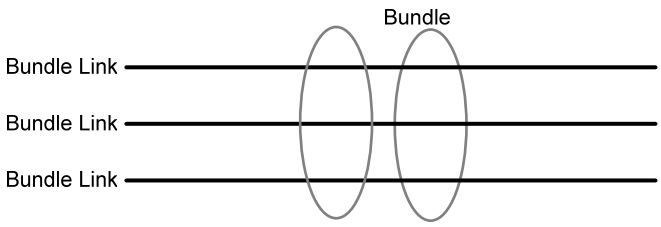
配置 MFR 接口时，为使捆绑后的接口带宽最大，建议对同一个 MFR 接口捆绑速率一致的物理接口，以减少管理开销。

1. Bundle 和 Bundle link

捆绑（bundle）和捆绑链路（bundle link）是多链路帧中继的两个基本概念。

一个 MFR 接口对应一个捆绑，一个捆绑中可以包含多个捆绑链路，一个捆绑链路对应着一个物理接口。捆绑对它的捆绑链路进行管理。二者的关系如下图所示：

图2-1 Bundle 和 Bundle link 示意图



对于实际的物理层可见的是捆绑链路；对于实际的数据链路层可见的是捆绑。

2. MFR 接口和物理接口

MFR 接口是逻辑接口，多个物理接口可以捆绑成一个 MFR 接口。一个 MFR 接口对应一个捆绑，一个物理接口对应一个捆绑链路。对捆绑和捆绑链路的配置实际就是对 MFR 接口和物理接口的配置。

MFR 接口的功能和配置与普通意义上的 FR 接口相同，也支持 DTE、DCE 接口类型，并支持 QoS 队列机制。当物理接口捆绑进 MFR 接口后，它原来配置的网络层和帧中继链路层参数将不再起作用，而是使用此 MFR 接口的参数。

2.2 配置多链路帧中继

表2-1 配置多链路帧中继

操作	命令	说明
进入系统视图	**system-view**	-
创建 MFR 接口并进入该 MFR 接口视图	**interface mfr** { *interface-number* | *interface-number.subnumber* }	必选 缺省情况下，没有创建 MFR 接口或子接口

操作	命令	说明
设置捆绑标识符	**mfr bundle-name** [*name*]	可选 缺省情况下，捆绑标识符是“mfr＋帧中继捆绑编号”。该标识符只在本地有效 需要注意的是，设置标识符时不允许出现“MFR＋数字”形式
使能 MFR 分片功能	**mfr fragment**	可选 缺省情况下，禁止多链路帧中继捆绑的分片功能
设置 MFR 滑动窗口的尺寸	**mfr window-size** *number*	可选 缺省情况下，滑动窗口尺寸等于 MFR 捆绑的物理接口数
设置捆绑链路允许的最大分片	**mfr fragment-size** *bytes*	可选 缺省情况下，最大分片是 300 字节
配置 MFR 接口的其它参数	请参见“1.2 帧中继配置任务简介”	可选
退回系统视图	**quit**	-
进入指定接口视图	**interface** *interface-type interface-number*	-
将当前接口捆绑到指定的 MFR 接口	**link-protocol fr mfr** *interface-number*	必选 缺省情况下，接口不与任何 MFR 接口捆绑
设置捆绑链路标识符名称	**mfr link-name** [*name*]	可选 缺省情况下，捆绑链路标识符是当前接口的名称
设置 MFR 捆绑链路的 hello 消息发送周期	**mfr timer hello** *seconds*	可选 缺省情况下，捆绑链路的 hello 消息发送周期为 10 秒
设置 MFR 捆绑链路重发 hello 消息前等待 hello 应答消息的时间	**mfr timer ack** *seconds*	可选 缺省情况下，重发 hello 消息前等待 hello 应答消息的时间为 4 秒
设置 MFR 捆绑链路最多可重发 hello 消息的次数	**mfr retry** *number*	可选 缺省情况下，最多可重发 hello 消息 2 次
配置子接口速率统计功能	**sub-interface rate-statistic**	可选 缺省情况下，接口的子接口速率统计功能处于关闭状态

2.3 多链路帧中继显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后多链路帧中继的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 MFR 接口的统计信息。

表2-2 多链路帧中继显示和维护

配置	命令
查看 MFR 接口的配置和状态信息	**display interface mfr** [*interface-number*]
查看多链路帧中继捆绑和捆绑链路的配置和统计信息	**display mfr** [**interface** *interface-type interface-number* | **verbose**]
清除 MFR 接口的统计信息	**reset counters interface** [**mfr** [*interface-number* | *interface-number.subnumber*]]

2.4 多链路帧中继典型配置举例

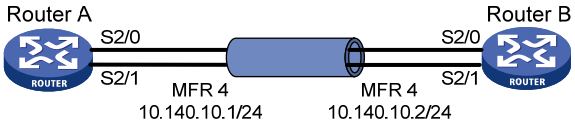
2.4.1 多链路帧中继直连配置举例

1. 组网需求

路由器 Router A 和 Router B 通过串口 Serial2/0 和 Serial2/1 直连，使用多链路帧中继协议将两个串口捆绑以提供更大的带宽。

2. 组网图

图2-2 多链路帧中继直连组网图



3. 配置步骤

(1) 配置路由器 Router A

创建并配置 MFR 接口 4。

```
<RouterA> system-view
[RouterA] interface mfr 4
[RouterA-MFR4] ip address 10.140.10.1 255.255.255.0
[RouterA-MFR4] fr interface-type dte
[RouterA-MFR4] fr map ip 10.140.10.2 100
[RouterA-MFR4] quit
```

将接口 Serial2/0 和 Serial2/1 捆绑至 MFR4。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol fr mfr 4
[RouterA-Serial2/0] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] link-protocol fr mfr 4
```

(2) 配置路由器 Router B

创建并配置 MFR 接口 4。

```
<RouterB> system-view
[RouterB] interface mfr 4
[RouterB-MFR4] ip address 10.140.10.2 255.255.255.0
```

```
[RouterB-MFR4] fr interface-type dce
[RouterB-MFR4] fr dlci 100
[RouterB-fr-dlci-MFR4-100] quit
[RouterB-MFR4] fr map ip 10.140.10.1 100
[RouterB-MFR4] quit
```

将接口 Serial2/0 和 Serial2/1 捆绑至 MFR4。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol fr mfr 4
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] link-protocol fr mfr 4
```

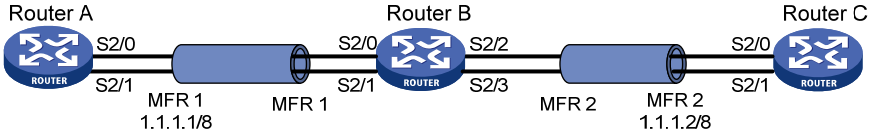
2.4.2 多链路帧中继交换配置举例

1. 组网需求

路由器 Router A 和 Router C 通过 MFR 连接到 Router B，Router B 使能多链路帧中继交换。

2. 组网图

图2-3 多链路帧中继交换组网图



3. 配置步骤

(1) 配置 Router A

配置 MFR1 接口。

```
<RouterA> system-view
[RouterA] interface mfr 1
[RouterA-MFR1] ip address 1.1.1.1 255.0.0.0
[RouterA-MFR1] quit
```

将接口 Serial2/0 和 Serial2/1 添加到 MFR1 接口。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol fr mfr 1
[RouterA-Serial2/0] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] link-protocol fr mfr 1
[RouterA-Serial2/1] quit
```

(2) 配置 Router B

使能帧中继交换功能。

```
<RouterB> system-view
[RouterB] fr switching
```

配置 MFR1 接口。

```
[RouterB] interface mfr 1
```

```
[RouterB-MFR1] fr interface-type dce
[RouterB-MFR1] fr dlci 100
[RouterB-fr-dlci-MFR1-100] quit
[RouterB-MFR1] quit
```

配置 MFR2 接口。

```
[RouterB] interface mfr 2
[RouterB-MFR2] fr interface-type dce
[RouterB-MFR2] fr dlci 200
[RouterB-fr-dlci-MFR2-200] quit
[RouterB-MFR2] quit
```

将接口 Serial2/0 和 Serial2/1 添加到 MFR1 接口。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol fr mfr 1
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] link-protocol fr mfr 1
[RouterB-Serial2/1] quit
```

将接口 Serial2/2 和 Serial2/3 添加到 MFR2 接口。

```
[RouterB] interface serial 2/2
[RouterB-Serial2/2] link-protocol fr mfr 2
[RouterB-Serial2/2] quit
[RouterB] interface serial 2/3
[RouterB-Serial2/3] link-protocol fr mfr 2
[RouterB-Serial2/3] quit
```

配置帧中继交换静态路由。

```
[RouterB] fr switch pvc1 interface mfr 1 dlci 100 interface mfr 2 dlci 200
```

(3) 配置 Router C

配置 MFR2 接口。

```
<RouterC> system-view
[RouterC] interface mfr 2
[RouterC-MFR2] ip address 1.1.1.2 255.0.0.0
[RouterC-MFR2] quit
```

将接口 Serial2/0 和 Serial2/1 添加到 MFR2 接口。

```
[RouterC] interface serial 2/0
[RouterC-Serial2/0] link-protocol fr mfr 2
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] link-protocol fr mfr 2
```

目 录

1 HDLC配置 1-1
1.1 HDLC协议简介 1-1
1.1.1 HDLC简介 1-1
1.1.2 HDLC的帧类型和帧格式 1-1
1.2 配置接口封装HDLC协议 1-1
1.3 配置接口IP地址 1-2
1.4 配置轮询时间 1-2
1.5 HDLC典型配置举例 1-3
1.5.1 配置HDLC基本功能举例 1-3
1.5.2 配置IP地址借用的HDLC举例 1-4

1 HDLC 配置

1.1 HDLC 协议简介

1.1.1 HDLC 简介

HDLC（High-level Data Link Control，高级数据链路控制）是一种面向比特的链路层协议，其最大特点是对任何一种比特流，均可以实现透明的传输。

- HDLC 协议只支持点到点链路，不支持点到多点。
- HDLC 不支持 IP 地址协商，不支持认证。协议内部通过 Keepalive 报文来检测链路状态。
- HDLC 协议只能封装在同步链路上，目前应用的接口为：Serial 接口和 POS 接口。

1.1.2 HDLC 的帧类型和帧格式

HDLC 有信息帧（I 帧）、监控帧（S 帧）和无编号帧（U 帧）3 种不同类型的帧。

- 信息帧用于传送有效信息或数据，通常简称为 I 帧。
- 监控帧用于差错控制和流量控制，通常称为 S 帧。
- 无编号帧用于提供对链路的建立、拆除以及多种控制功能，简称 U 帧。

HDLC 帧由标志、地址、控制、信息和帧校验序列等字段组成。

- 标志字段为 0111110，标志一个 HDLC 帧的开始和结束，所有的帧必须以 F 开头，并以 F 结束；在邻近两帧之间的 F，即作为前面帧的结束，又作为后续帧的开头；
- 地址字段是 8 比特，用于标识接收或发送 HDLC 帧的地址；
- 控制字段是 8 比特，用来实现 HDLC 协议的各种控制信息，并标识是否是数据；
- 信息字段可以是任意的二进制比特串，长度未作限定，其上限由 FCS 字段或通讯节点的缓冲容量来决定，目前国际上用得较多的是 1000-2000 比特，而下限可以是 0，即无信息字段。但是监控帧中不可有信息字段。
- 帧检验序列字段可以使用 16 位 CRC，对两个标志字段之间的整个帧的内容进行校验。

1.2 配置接口封装 HDLC 协议

表1-1 配置接口封装 HDLC 协议

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口封装 HDLC 协议	**link-protocol hdlc**	必选 缺省情况下，接口封装 PPP 协议

说明

- 可以通过 **display interface** 命令查看配置结果。
- 可以通过 **reset counters interface** 命令清除 HDLC 接口的统计信息，使接口重新开始统计流量。

1.3 配置接口 IP 地址

表1-2 配置接口 IP 地址

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图		**interface** *interface-type interface-number*	-
配置接口 IP 地址	配置接口的 IP 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	二者必选其一 缺省情况下，没有为接口配置 IP 地址也不借用其它接口的 IP 地址
	配置本接口借用指定接口的 IP 地址	**ip address unnumbered interface** *interface-type interface-number*	

如果对封装 HDLC 的接口配置 IP 地址借用，借用地址的一端必须能够学到对端的网络路由，否则将无法达到对端。

在配置 IP 地址借用时，可以使用路由协议或静态路由来学习到对端的路由，并注意以下原则：

- 如果使用路由协议，由于路由查找采用最长匹配原则，应确保学到路由的掩码长度大于被借用方 IP 地址的掩码长度；
- 如果使用静态路由，且被借用方的 IP 地址使用 32 位掩码，静态路由的掩码长度应小于被借用方 IP 地址的掩码长度；
- 如果使用静态路由，且被借用方的 IP 地址掩码小于 32 位，静态路由的掩码长度应大于被借用方 IP 地址的掩码长度。

说明

关于配置接口 IP 地址的详细介绍请参见“IP 业务分册”中的“IP 地址配置”。

1.4 配置轮询时间

HDLC 协议使用定时器来确认链路状态是否正常。在配置轮询时间的间隔时，应保证链路两端的设置相同。如果将两端的轮询时间间隔都设为 0，则禁止链路状态检测功能。

表1-3 配置轮询时间

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置状态轮询定时器的轮询时间间隔	**timer hold** *seconds*	可选 缺省情况下，接口的协议轮询时间间隔为 10 秒

说明

轮询时间可使用缺省配置，也可根据网络实际情况进行调整。如果网络的延迟比较大，或拥塞程度较高，可以适当加大轮询时间的间隔，以减少网络震荡的发生。

1.5 HDLC 典型配置举例

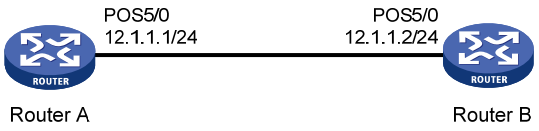
1.5.1 配置 HDLC 基本功能举例

1. 组网需求

路由器 Router A 和 Router B 通过 POS 接口相连，要求运行 HDLC。

2. 组网图

图1-1 配置 HDLC 基本功能组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface pos 5/0
[RouterA-Pos5/0] link-protocol hdlc
[RouterA-Pos5/0] ip address 12.1.1.1 24
[RouterA-Pos5/0] quit
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface pos 5/0
[RouterB-Pos5/0] link-protocol hdlc
[RouterB-Pos5/0] ip address 12.1.1.2 24
```

配置完成后 Router A 和 Router B 可以互相 ping 通。以 Router A 的显示为例。

```
[RouterA] ping 12.1.1.2
  PING 12.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 12.1.1.2: bytes=56 Sequence=1 ttl=255 time=126 ms
    Reply from 12.1.1.2: bytes=56 Sequence=2 ttl=255 time=1 ms
    Reply from 12.1.1.2: bytes=56 Sequence=3 ttl=255 time=1 ms
    Reply from 12.1.1.2: bytes=56 Sequence=4 ttl=255 time=1 ms
    Reply from 12.1.1.2: bytes=56 Sequence=5 ttl=255 time=10 ms
```

```
--- 12.1.1.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/27/126 ms
```

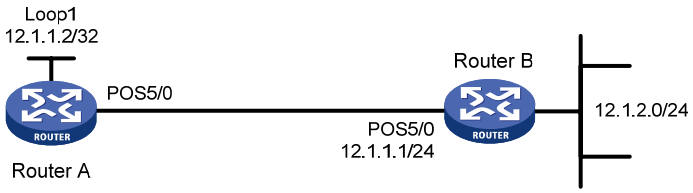
1.5.2 配置 IP 地址借用的 HDLC 举例

1. 组网需求

- 路由器 Router A 和 Router B 通过 POS 接口相连，要求运行 HDLC。
- Router A 的 POS0/0 接口借用本端 Loopback 接口的 IP 地址，Loopback 接口使用 32 位掩码。
- Router A 通过静态路由学习对端路由信息，可以到达 12.1.2.0/24 这个网段。

2. 组网图

图1-2 配置 IP 地址借用 HDLC



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface loopback 1
[RouterA-LoopBack1] ip address 12.1.1.2 32
[RouterA-LoopBack1] quit
[RouterA] interface pos 5/0
[RouterA-Pos5/0] link-protocol hdlc
[RouterA-Pos5/0] ip address unnumbered interface loopback 1
[RouterA-Pos5/0] quit
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface pos 5/0
[RouterB-Pos5/0] link-protocol hdlc
[RouterB-Pos5/0] ip address 12.1.1.1 24
```

(3) 在 Router A 上配置静态路由

```
[RouterA] ip route-static 12.1.1.0 24 pos 5/0
[RouterA] ip route-static 12.1.2.0 24 12.1.1.1
```

配置完成后，Router A 可以 ping 通 12.1.2.0/24 网段。

```
[RouterA] ping 12.1.2.1
  PING 12.1.2.1: 56  data bytes, press CTRL_C to break
    Reply from 12.1.2.1: bytes=56 Sequence=1 ttl=255 time=35 ms
    Reply from 12.1.2.1: bytes=56 Sequence=2 ttl=255 time=1 ms
    Reply from 12.1.2.1: bytes=56 Sequence=3 ttl=255 time=10 ms
```

```
  Reply from 12.1.2.1: bytes=56 Sequence=4 ttl=255 time=1 ms
  Reply from 12.1.2.1: bytes=56 Sequence=5 ttl=255 time=1 ms

--- 12.1.2.1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/9/35 ms
```

在 Router A 执行 **display ip routing-table** 命令可以看到路由表信息正确。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

12.1.1.0/24         Static 60   0            12.1.1.2        POS5/0
12.1.1.2/32         Direct 0    0            127.0.0.1       InLoop0
12.1.2.0/24         Static 60   0            12.1.1.1        POS5/0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

目 录

1 PPP、MP配置 1-1
1.1 PPP、MP协议简介 1-1
1.1.1 PPP简介 1-1
1.1.2 MP简介 1-3
1.2 配置PPP 1-4
1.2.1 配置PPP 1-4
1.2.2 配置PAP验证 1-5
1.2.3 配置CHAP验证 1-5
1.2.4 PPP协商参数 1-8
1.2.5 PPP计费统计功能 1-11
1.3 配置MP 1-11
1.3.1 通过虚拟模板接口配置MP 1-11
1.3.2 通过MP-group方式配置MP 1-13
1.3.3 配置MP短序协商方式 1-13
1.3.4 配置MP Endpoint选项 1-14
1.4 配置PPP链路效率机制 1-14
1.4.1 配置PPP链路效率机制 1-15
1.5 PPP、MP及PPP链路效率机制的显示和维护 1-15
1.6 PPP、MP典型配置举例 1-16
1.6.1 PAP验证举例 1-16
1.6.2 CHAP验证举例 1-17
1.6.3 MP配置举例 1-18
1.6.4 三种MP绑定方式的配置举例 1-20
1.7 PPP常见错误配置举例 1-28
1.7.1 链路始终不能转为up状态 1-28
1.7.2 物理链路不能转为up状态 1-28
1.7.3 IPv6CP协商down后不能重协商 1-29
2 PPPoE配置 2-1
2.1 PPPoE简介 2-1
2.1.1 PPPoE 2-1
2.1.2 PPPoE Server 2-1
2.1.3 PPPoE Client 2-1
2.2 配置PPPoE Server 2-1
2.3 PPPoE显示和维护 2-2
2.4 PPPoE典型配置举例 2-3
2.4.1 PPPoE Server典型配置举例 2-3

1 PPP、MP 配置

1.1 PPP、MP 协议简介

1.1.1 PPP 简介

PPP（Point to Point Protocol）协议是在点到点链路上承载网络层数据包的一种链路层协议，由于它能够提供用户验证、易于扩充，并且支持同/异步通信，因而获得广泛应用。

PPP 定义了一整套的协议，包括链路控制协议（LCP）、网络层控制协议（NCP）和验证协议（PAP 和 CHAP）。

- 链路控制协议（Link Control Protocol，LCP）：主要用来建立、拆除和监控数据链路。
- 网络控制协议（Network Control Protocol，NCP）：主要用来协商在该数据链路上所传输的数据包的格式与类型。
- 用于网络安全方面的验证协议族 PAP 和 CHAP。

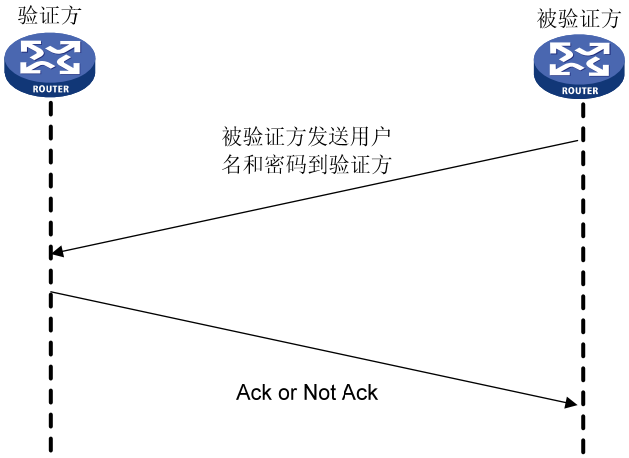
1. PAP 验证

PAP（Password Authentication Protocol）验证为两次握手验证，密码为明文，PAP 验证的过程如下：

(1) 被验证方发送用户名和密码到验证方；

(2) 验证方根据本端用户表查看是否有此用户以及密码是否正确，然后返回不同的响应（Acknowledge or Not Acknowledge）。

图1-1 PAP 验证示意图



PAP 不是一种安全的验证协议。当验证时，口令以明文方式在链路上发送，并且由于完成 PPP 链路建立后，被验证方会不停地在链路上反复发送用户名和口令，直到身份验证过程结束，所以不能防止攻击。

2. CHAP 验证

CHAP（Challenge-Handshake Authentication Protocol）验证为三次握手验证，密码为密文（密钥）。

CHAP 单向验证是指一端作为验证方，另一端作为被验证方。双向验证是单向验证的简单叠加，即两端都是既作为验证方又作为被验证方。在实际应用中一般只采用单向验证。

CHAP 单向验证过程分为两种情况：验证方配置了用户名和验证方没有配置用户名。推荐使用验证方配置用户名的方式，这样可以对验证方的身份进行确认。

- 验证方配置了用户名的 CHAP 验证过程如下：

(1) 验证方主动发起验证请求，验证方向被验证方发送一些随机产生的报文（Challenge），并同时将本端的用户名附带上一起发送给被验证方；

(2) 被验证方接到验证方的验证请求后，根据此报文中验证方的用户名在本端的用户表查找该用户对应的密码，如果在用户表找到了与验证方用户名相同的用户，便利用报文 ID、此用户的密钥（密码）和 MD5 算法对该随机报文进行加密，将生成的密文和被验证方自己的用户名发回验证方（Response）；

(3) 验证方用自己保存的被验证方密码和 MD5 算法对原随机报文加密，比较二者的密文，根据比较结果返回不同的响应（Acknowledge or Not Acknowledge）。

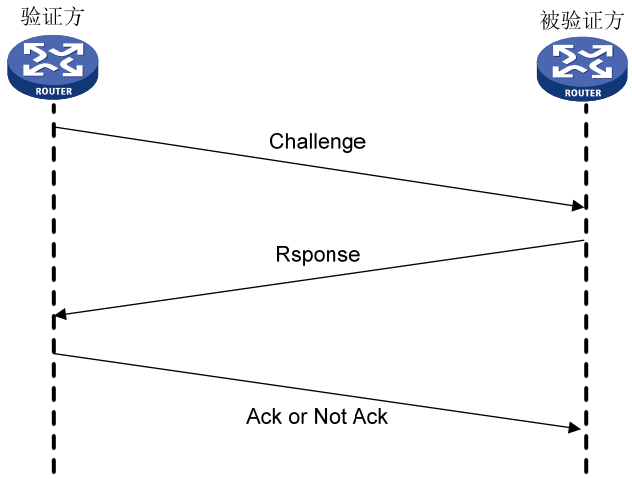
- 验证方没有配置用户名的 CHAP 验证过程如下：

(1) 验证方主动发起验证请求，验证方向被验证方发送一些随机产生的报文（Challenge）；

(2) 被验证方接到验证方的验证请求后，利用报文 ID、缺省的 CHAP 密码和 MD5 算法对该随机报文进行加密，将生成的密文和自己的用户名发回验证方（Response）；

(3) 验证方用自己保存的被验证方密码和 MD5 算法对原随机报文加密，比较二者的密文，根据比较结果返回不同的响应（Acknowledge or Not Acknowledge）。

图1-2 CHAP 验证示意图



3. PPP 运行过程：

PPP 运行过程（参见下图）如下：

(1) 在开始建立 PPP 链路时，先进入到 Establish 阶段。

(2) 在 Establish 阶段 PPP 链路进行 LCP 协商，协商内容包括工作方式（是 SP 还是 MP）、验证方式和最大传输单元等。LCP 协商成功后进入 Opened 状态，表示底层链路已经建立。

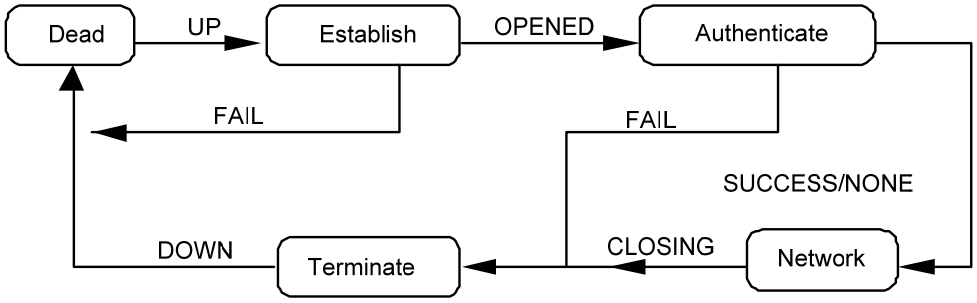
(3) 如果配置了验证（远端验证本地或者本地验证远端）则进入 Authenticate 阶段，开始 CHAP 或 PAP 验证。

(4) 如果验证失败进入 Terminate 阶段，拆除链路，LCP 状态转为 Down；如果验证成功就进入 Network 协商阶段（NCP），此时 LCP 状态仍为 Opened，而 IPCP 状态从 Initial 转到 Request。

(5) NCP 协商支持 IPCP 协商，IPCP 协商主要包括双方的 IP 地址。通过 NCP 协商来选择和配置一个网络层协议。只有相应的网络层协议协商成功后，该网络层协议才可以通过这条 PPP 链路发送报文。

(6) PPP 链路将一直保持通信，直至有明确的 LCP 或 NCP 帧关闭这条链路，或发生了某些外部事件（例如用户的干预）。

图1-3 PPP 运行流程图



有关 PPP 的详细说明，请参考 RFC1661。

1.1.2 MP 简介

为了增加带宽，可以将多个 PPP 链路捆绑使用，称为 MultiLink PPP，简称 MP。MP 会将报文分片（小于最小分片包长时不分片）后，从 MP 链路下的多个 PPP 通道发送到 PPP 对端，对端将这些分片组装起来递给网络层。

1. 实现方式

MP 的配置主要有两种方式，一种是通过虚拟模板接口（Virtual-Template，VT），VT 是用于配置一个虚拟访问接口（Virtual Access，VA）的模板，将多个 PPP 链路捆绑成 MP 之后，需要创建一个 VA 与对端交换数据。此时，系统将选择一个 VT，以便动态地创建一个 VA；一种是利用 MP-group 接口。这两种配置方式的区别主要是：

- 虚拟模板接口方式可以与验证相结合，可以根据对端的用户名找到指定的虚拟模板接口，从而利用模板上的配置，创建相应的捆绑（Bundle，系统中用 VT 通道来表示），以对应一条 MP 链路。
- 由一个虚拟模板接口还可以派生出若干个捆绑，每个捆绑对应一条 MP 链路。那么这样一来，从网络层看来，这若干条 MP 链路会形成一个点对多点的网络拓扑。从这个意义上讲，虚拟模板接口比 MP-group 接口更加灵活。
- 为区分虚拟模板接口派生出的多个捆绑，需要指定捆绑方式，系统在虚拟模板接口视图下提供了命令 **ppp mp binding-mode** 来指定绑定方式，绑定方式有 **authentication**、**both**、**descriptor** 三种，缺省是 **both**。**authentication** 是根据验证用户名捆绑，**descriptor** 是根据终端描述符捆绑（LCP 协商时，会协商出这个选项值），**both** 是要同时参考这两个值捆绑。
- MP-group 接口与虚拟模板接口相比则单纯许多，它是 MP 的专用接口，不能支持其他应用，也不能利用对端的用户名来指定捆绑，同时也不能派生多个捆绑。但正因为它的简单，导致了它的快速高效、配置简单、容易理解。

2. 协商过程

MP 的协商包括 LCP 协商和 NCP 协商两个过程：

- LCP 协商：两端首先进行 LCP 协商，除了协商一般的 LCP 参数外，还要验证对端接口是否也工作在 MP 方式下。如果两端工作方式不同，LCP 协商不成功。
- NCP 协商：在 LCP 协商成功后，根据 MP-group 接口或指定虚拟接口模板的各项 NCP 参数（如 IP 地址等）进行 NCP 协商，物理接口配置的 NCP 参数不起作用。

NCP 协商通过后，即可建立 MP 链路。

3. 作用

MP 的作用主要有：

- 增加带宽，结合 DCC（Dial Control Center，拨号控制中心）可以做到动态增加或减小带宽
- 负载分担
- 备份
- 利用分片降低时延

MP 能在任何支持 PPP 封装的接口下工作，例如：串口，也包括 PPPoX（PPPoE、PPPoA、PPPoFR 等）这类虚拟接口，建议用户尽可能将同一类的接口捆绑使用，不要将不同类的接口捆绑使用。

1.2 配置 PPP

1.2.1 配置 PPP

表1-1 配置 PPP

操作		命令	说明
进入系统视图		**system-view**	-
进入指定接口的视图		**interface** *interface-type interface-number*	-
配置接口封装的链路层协议为 PPP		**link-protocol ppp**	可选 缺省情况下，接口封装的链路层协议为 PPP
配置轮询时间间隔		**timer hold** *seconds*	可选 缺省情况下，轮询时间间隔为 10 秒
配置 PPP 验证方式	配置 PAP 验证	请参见 1.2.2 配置PAP验证	可选 二者只能选其一 缺省情况下，PPP 不进行验证
	配置 CHAP 验证	请参见 1.2.3 配置CHAP验证	
配置 PPP 协商参数		请参见 1.2.4 PPP协商参数	可选
配置 PPP 计费统计功能		请参见 1.2.5 PPP计费统计功能	可选

说明

本章只介绍本地认证方案，远端 AAA 认证方案请参见“安全分册”中的“AAA 配置”。

1.2.2 配置 PAP 验证

1. 配置验证方

表1-2 配置验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置本地验证对端的方式为 PAP	**ppp authentication-mode pap** [[**call-in**] **domain** *isp-name*]	必选 缺省情况下，PPP 协议不进行验证
退回系统视图	**quit**	-
创建本地用户，并进入本地用户视图	**local-user** *username*	必选
设置本地用户的密码	**password** { **cipher** | **simple** } *password*	必选
设置本地用户的服务类型以及其他属性	**service-type ppp**	必选
退回系统视图	**quit**	-
创建一个 ISP 域，或者进入已创建 ISP 域的视图	**domain** *isp-name*	可选
配置域用户使用本地认证方案	**authentication ppp local**	可选



说明

关于创建本地用户以及设置本地用户属性、创建域以及设置域的属性的详细说明，请参见“安全分册”中的“AAA 配置”。

2. 配置被验证方

表1-3 配置被验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置本地被对端以 PAP 方式验证时本地发送的 PAP 用户名和密码	**ppp pap local-user** *username* **password** { **cipher** | **simple** } *password*	必选 缺省情况下，被对端以 PAP 方式验证时，本地设备发送的用户名和密码均为空

1.2.3 配置 CHAP 验证

CHAP 验证分为两种：验证方配置了用户名和验证方没有配置用户名。

1. 验证方配置了用户名

(1) 配置验证方

表1-4 配置验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置本地验证对端的方式为 CHAP	**ppp authentication-mode chap** [[**call-in**] **domain** *isp-name*]	必选 缺省情况下，PPP 协议不进行验证
配置采用 CHAP 认证时验证方的用户名	**ppp chap user** *username*	可选 在被验证方上为验证方配置的本地用户的用户名必须跟此处配置的一致
退回系统视图	**quit**	-
为被验证方创建本地用户，并进入本地用户视图	**local-user** *username*	必选
设置本地用户的密码	**password** { **cipher** | **simple** } *password*	必选
设置本地用户的服务类型以及其他属性	**service-type ppp**	必选
退回系统视图	**quit**	-
创建一个 ISP 域，或者进入已创建 ISP 域的视图	**domain** *isp-name*	可选
配置域用户使用本地认证方案	**authentication ppp local**	可选



说明

关于创建本地用户以及设置本地用户属性、创建域以及设置域的属性的详细说明，请参见“安全分册”中的“AAA 配置”。

(2) 配置被验证方

表1-5 配置被验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置采用 CHAP 认证时被验证方的用户名	**ppp chap user** *username*	必选 在验证方上为被验证方配置的本地用户的用户名必须跟此处配置的一致
设置缺省的 CHAP 验证密码	**ppp chap password** { **cipher** | **simple** } *password*	可选 当选择验证方配置了用户名这种情况，不用配置；当选择验证方没有配置用户名这种情况，必须配置

操作		命令	说明
为验证方设置本地用户以及相应的密码	退回系统视图	**quit**	-
	创建本地用户，并进入本地用户视图	**local-user** *username*	可选
	设置本地用户的密码	**password** { **cipher** | **simple** } *password*	可选

2. 验证方没有配置用户名

(1) 配置验证方

表1-6 配置验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置本地验证对端的方式为CHAP	**ppp authentication-mode chap** [[**call-in**] **domain** *isp-name*]	必选 缺省情况下，PPP 协议不进行验证
退回系统视图	**quit**	-
为被验证方创建本地用户，并进入本地用户视图	**local-user** *username*	必选
设置本地用户的密码	**password** { **cipher** | **simple** } *password*	必选
设置本地用户的服务类型以及其他属性	**service-type ppp** [**callback-nocheck** | **callback-number** *callback-number* | **call-number** *call-number* [*:subcall-number*]]	必选
退回系统视图	**quit**	-
创建一个 ISP 域，或者进入已创建 ISP 域的视图	**domain** *isp-name*	可选
配置域用户使用本地认证方案	**authentication ppp local**	可选

说明

关于创建本地用户以及设置本地用户属性、创建域以及设置域的属性的详细说明，请参见“安全分册”中的“AAA 配置”。

(2) 配置被验证方

表1-7 配置被验证方

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置采用 CHAP 认证时被验证方的用户名	**ppp chap user** *username*	必选 在验证方上为被验证方配置的本地用户的用户名必须跟此处配置的一致
设置缺省的 CHAP 验证密码	**ppp chap password** { **cipher** | **simple** } *password*	必选

1.2.4 PPP 协商参数

1. PPP 协商参数介绍

可以配置的 PPP 协商参数包括：协商超时时间间隔、协商 IP 地址以及协商 DNS 地址。

协商超时时间间隔：在 PPP 协商过程中，如果在这个时间间隔内没有收到对端的应答报文，则 PPP 将会重发前一次发送的报文。超时时间间隔可选范围为 1 秒到 10 秒。

协商 IP 地址的方式有两种：

- 配置设备作为 Client 端：若本端接口封装的链路层协议为 PPP 但还未配置 IP 地址，而对端已有 IP 地址时，可为本端接口配置 IP 地址可协商属性，使本端接口接受 PPP 协商产生的由对端分配的 IP 地址。该配置主要用于在通过 ISP 访问 Internet 时，得到由 ISP 分配的 IP 地址。
- 配置设备作为 Server 端：若是设备作为 Server 为对端设备分配 IP 地址，则应首先在域视图或系统视图下配置本地 IP 地址池，指明地址池的地址范围，然后在接口视图下指定该接口使用的地址池。

设备在进行 PPP 地址协商的过程中可以进行 DNS 地址协商，此时设备既可以配置为接收对端分配的 DNS 地址，也可以配置为向对方提供 DNS 地址。一般情况下，当 PC 与设备通过 PPP 协议相连时（通常为 PC 机拨号连接设备），设备应为对端设备指定 DNS 地址，这样 PC 就可以通过域名直接访问 Internet；当设备通过 PPP 协议连接运营商的接入服务器时，设备应配置为被动接收或主动请求对端指定 DNS 地址，这样设备就可以使用接入服务器分配的 DNS 来解析域名。

2. 配置 PPP 协商参数

表1-8 配置 PPP 协商参数

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置协商超时时间间隔	**ppp timer negotiate** *seconds*	可选 缺省情况下，协商超时时间间隔为 3 秒
配置 PPP 协商 IP 地址	请参见 3. 配置PPP协商IP地址	可选
配置 DNS 服务器地址协商	请参见 4. 配置DNS服务器地址协商	可选

3. 配置 PPP 协商 IP 地址

表1-9 配置 PPP 协商 IP 地址

操作		命令	说明
进入系统视图		**system-view**	-
进入指定接口的视图		**interface** *interface-type interface-number*	-
配置 PPP 协商 IP 地址（Client 端和 Server 端选择其一）	配置 Client 端	请参见 表 1-10 配置Client端	必选
	配置 Server 端	请参见 表 1-11 配置Server端（对于不需要进行认证的PPP用户）或者 表 1-12 配置Server端（对于需要进行认证的PPP用户）	必选

(a) 配置 Client 端

表1-10 配置 Client 端

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
设置接口 IP 地址可协商属性	**ip address ppp-negotiate**	必选

(b) 配置 Server 端

对于不需要进行认证的 PPP 用户，Server 端的配置方式如下：

表1-11 配置 Server 端（对于不需要进行认证的 PPP 用户）

操作		命令	说明
进入系统视图		**system-view**	-
通过使用全局地址池给对端分配地址，或者直接为对端指定 IP 地址（两者选择其一）	首先定义全局 IP 地址池，然后在接口上使用全局地址池给 PPP 用户分配 IP 地址	**ip pool** *pool-number low-ip-address* [*high-ip-address*]	必选 当配置了 **remote address pool** 但没有指定 *pool-number* 时，缺省使用 0 号全局地址池
		interface *interface-type interface-number*	
		remote address pool [*pool-number*]	
	接口直接为对端指定 IP 地址	**interface** *interface-type interface-number* **remote address** *ip-address*	必选



说明

使用全局地址池分配地址时，如果存在多个全局地址池，只能使用指定的地址池，不能从多个地址池中轮询获取地址。

对于需要进行认证的PPP用户，Server端的配置方式如 表 1-12所示，定义地址池所使用的域就是配置进行PPP认证时指定的域：

表1-12 配置 Server 端（对于需要进行认证的 PPP 用户）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定的域视图	**domain** *domain-name*	必选
定义域地址池	**ip pool** *pool-number low-ip-address* [*high-ip-address*]	必选
退回系统视图	**quit**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
使用域地址池给 PPP 用户分配 IP 地址	**remote address pool** [*pool-number*]	必选 若配置该命令时不指定 *pool-number*，则在 IP 地址协商时依次使用该域下的地址池给用户分配 IP 地址
配置 PPP IPCP 不允许对端使用自行配置的固定 IP 地址	**ppp ipcp remote-address forced**	可选 缺省情况下，PPP IPCP 的 IP 地址协商情况为本端不具有地址分配的强制性，即设备本端允许对端自行配置地址。当对端明确请求本端分配地址时，本端给对端分配地址；若对端已自行配置 IP 地址时，本端不再强行给对端分配地址

4. 配置 DNS 服务器地址协商

(1) 配置 Client 端

表1-13 配置 Client 端

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置设备主动请求对端指定 DNS 地址	**ppp ipcp dns request**	必选 缺省情况下，禁止设备主动向对端请求 DNS 服务器地址
配置设备被动地接收对端指定的 DNS 服务器地址	**ppp ipcp dns admit-any**	可选 缺省情况下，设备不会被动地接收对端设备指定的 DNS 服务器的 IP 地址

(2) 配置 Server 端

表1-14 配置 Server 端

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
使能设备为对端设备指定 DNS 服务器地址	**ppp ipcp dns** *primary-dns-address* [*secondary-dns-address*]	必选 缺省情况下，设备不为对端设备指定 DNS 服务器的 IP 地址

1.2.5 PPP 计费统计功能

1. PPP 计费统计功能介绍

PPP 协议可以为每条 PPP 链路提供基于流量的计费统计功能，具体统计内容包括出入两个方向上流经本链路的报文数和字节数。AAA 应用模块可以获取这些流量统计信息用于计费控制使用。

2. 配置 PPP 计费统计功能

表1-15 配置 PPP 计费统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
使能 PPP 计费统计功能	**ppp account-statistics** enable	必选 缺省情况下，不启动计费统计功能

1.3 配置 MP

说明

不支持跨板卡、跨线卡进行 MP 捆绑。

1.3.1 通过虚拟模板接口配置 MP

1. 通过虚拟模板接口配置 MP 介绍

采用虚拟模板接口配置 MP 时，又可以细分为两种情况：

- 将物理接口与虚拟模板接口直接关联：通过命令 **ppp mp virtual-template** 直接将链路绑定到指定的虚拟模板接口上，这时可以配置验证也可以不配置验证。如果不配置验证，系统将通过对端的终端描述符捆绑出 MP 链路；如果配置了验证，系统将通过用户名和/或对端的终端描述符捆绑出 MP 链路。
- 将用户名与虚拟模板接口关联：根据验证通过后的用户名查找相关联的虚拟模板接口，然后根据用户名和对端终端描述符捆绑出 MP 链路。这种方式需在要绑定的接口下配置 **ppp mp** 及双向验证（CHAP 或 PAP），否则链路协商不通。

说明

- **ppp mp** 和 **ppp mp virtual-template** 命令互斥，即同一个接口只能配置成通过验证的绑定或直接绑定中的一种。
- 对于需要绑在一起的接口，必须采用同样的配置方式。
- 实际使用中也可以配置单向验证，即一端直接将物理接口绑定到虚拟模板接口，另一端则通过用户名查找虚拟模板接口。
- 不推荐使用同一个虚拟模板接口配置多种业务（如 MP、L2TP、PPPoE 等）。

在虚拟模板接口下指定捆绑方式时，可以使用用户名、终端标识符或者两者同时使用。用户名是指PPP链路进行PAP或CHAP验证时所接收到的对端用户名；终端标识符是用来唯一标识一台设备的标志，是指进行LCP协商时所接收到的对端终端标识符。系统可以根据接口接收到的用户名或终端标识符来进行MP捆绑，以此来区分虚模板接口下的多个MP捆绑（对应多条MP链路）。

2. 通过虚拟模板接口配置 MP

表1-16 通过虚拟模板接口配置 MP

操作		命令	说明
进入系统视图		**system-view**	-
创建并进入虚拟模板接口		**interface virtual-template** *number*	必选
退回系统视图		**quit**	-
将物理接口或用户名与虚拟模板接口关联（两者选择其一）	将物理接口与虚拟模板接口关联	**interface** *interface-type interface-number*	-
		ppp mp virtual-template *number*	必选 配置接口所要绑定的虚拟模板接口
		在接口下配置PPP认证，请参见1.2.1 配置PPP	可选 PPP认证对MP连接的建立没有影响
	将用户名与虚拟模板接口关联	**ppp mp user** *username* **bind virtual-template** *number*	必选 建立虚拟模板接口与MP用户的对应关系
		interface *interface-type interface-number*	-
		ppp mp	必选 配置封装PPP的接口工作在MP方式
		在接口下配置双向PPP认证，请参见1.2.1 配置PPP	必选
配置MP的其他可选参数		请参见3. 配置MP的其他可选参数	可选

3. 配置 MP 的其他可选参数

表1-17 配置 MP 的其他可选参数

操作	命令	说明
进入系统视图	**system-view**	-
创建并进入MP虚拟模板接口视图	**interface virtual-template** *number*	必选
设置虚拟接口模板支持发送组播或广播报文的最大链路数	**broadcast-limit link** *number*	可选 缺省情况下，虚拟接口模板支持发送组播或广播报文的最大链路数为30
虚模板接口下捆绑方式的指定	**ppp mp binding-mode { authentication | both | descriptor }**	可选 缺省情况下，同时根据PPP的验证用户名和终端描述符进行MP捆绑（**both**）
配置MP最大捆绑链路数	**ppp mp max-bind** *max-bind-num*	可选 该命令可以在虚拟模板接口视图下进行配置 缺省情况下，最大捆绑链路数为16

操作	命令	说明
设置 MP 出报文进行分片的最小报文长度	**ppp mp min-fragment** *size*	可选 缺省情况下，对 MP 报文进行分片的最小报文长度为 128

说明

- 在配置 **ppp mp max-bind** 或 **ppp mp min-fragment** 后，该命令不能立即生效，必须对所有已捆绑的物理接口执行 **shutdown** 和 **undo shutdown** 之后改变才会生效。
- 当只选择 **descriptor** 的绑定模式时，MP 捆绑时无法区分不同的用户，如果不同用户需要绑定到不同的捆绑组下时，应该选用 **both** 的绑定模式。
- 当只选择 **authentication** 的绑定模式时，无法区分各个对端设备，因此 MP 捆绑有多个对端设备时，应该选用 **both** 的绑定模式。
- 对于 VT 接口，如果配置静态路由，请指定下一跳而不要指定出接口。如果必须指定出接口的话，请保证 VT 下绑定的物理接口有效，从而保证报文能够正常传输。
- 关于在 Dailer 接口下设置 MP 参数的详细信息，请参见“接入分册”中的“DCC 配置”。

1.3.2 通过 MP-group 方式配置 MP

表1-18 通过 MP-group 方式配置 MP

操作	命令	说明
进入系统视图	**system-view**	-
创建 MP-group	**interface mp-group** *mp-number*	必选
退回系统视图	**quit**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
将接口加入指定的 MP-group	**ppp mp mp-group** *mp-number*	必选
配置 MP 最大捆绑链路数	**ppp mp max-bind** *max-bind-num*	可选 缺省情况下，最大捆绑链路数为 16
设置 MP 出报文进行分片的最小报文长度	**ppp mp min-fragment** *size*	可选 缺省情况下，对 MP 报文进行分片的最小报文长度为 128

说明

在配置 **ppp mp max-bind** 或 **ppp mp min-fragment** 后，该命令不能立即生效，必须对所有已捆绑的物理接口执行 **shutdown** 和 **undo shutdown** 之后改变才会生效。

1.3.3 配置 MP 短序协商方式

MP 捆绑组在收发报文时默认使用长序方式。

- 如果本端接收使用短序，则需要在协商 LCP 的过程添加短序请求，请求对端发送短序，协商通过后，对端使用短序发送；

- 如果本端发送使用短序，则需要对端发出短序协商请求，协商通过后，本端使用短序发送。

表1-19 配置 MP 短序协商方式

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置 MP 短序协商方式	**ppp mp short-sequence**	可选



说明

- MP 捆绑组使用的长短序方式由第一条加入该捆绑组中的子通道决定，后续加入捆绑组的子通道配置不能更改 MP 捆绑组的长短序方式。
- 如果想使用 MP 短序协商，对于拨号 MP，建议在 Dialer 接口及 ISDN 的 D 信道下均配置该命令；对于普通 MP，建议在所有的 MP 子通道下配置该命令，配置该命令会导致 PPP 重协商。

1.3.4 配置 MP Endpoint 选项

在 MP 的 LCP 协商过程会协商 Endpoint 选项（终端描述符）值，进行捆绑处理。

缺省情况下，接口发送报文中的 Endpoint 选项内容为设备名称；当使用 **ppp mp mp-group** 命令将接口加入指定 MP-group 时，接口发送报文中的 Endpoint 选项内容为 MP-goup 的接口名称。由于 Endpoint 选项内容最长为 20 字节，如果缺省内容超过 20 个字节，则截取前 20 个字节作为 Endpoint 选项内容。

用户也可以通过下面的命令来配置接口发送报文中的 Endpoint 选项内容。

表1-20 配置 MP Endpoint 选项

操作	命令	说明
进入系统视图	**system-view**	-
进入指定接口的视图	**interface** *interface-type interface-number*	-
配置 MP Endpoint 选项	**ppp mp endpoint string** *char-string*	必选

1.4 配置 PPP 链路效率机制

PPP 链路上提供了一种提高传输效率的机制：链路分片与交叉，详细介绍如下：

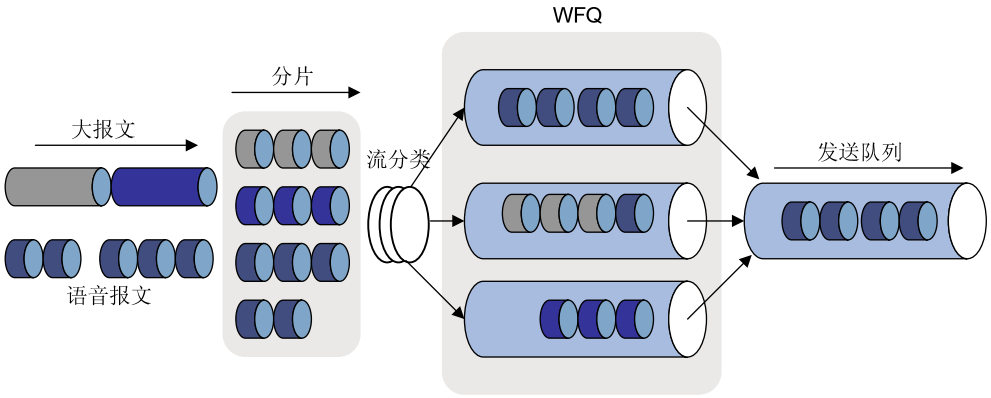
在低速串行链路上，实时交互式通信，如 Telnet 和 VoIP，往往会由于大型分组的发送而导致阻塞延迟。例如，正好在大报文被调度而等待发送时，语音报文到达，它需要等该大报文被传输完毕后才能被调度。对于诸如交互式语音等实时应用而言，大报文导致的这种阻塞延迟太长了，对端将听到断断续续的话音。交互式语音要求端到端的延迟不大于 100～150ms。

一个 1500bytes（即通常 MTU 的大小）的报文需要花费 215ms 穿过 56Kbps 的链路，这超过了人所能忍受的延迟限制。为了在相对低速的链路上限制实时报文的延迟时间，例如 56Kbps Frame Relay，需要一种方法将大报文进行分片，将小报文和大报文的分片一起加入到队列。

LFI（Link Fragmentation and Interleaving，链路分片与交叉）将大型数据帧分割成小型帧，与其他小片的报文一起发送，从而减少在速度较慢的链路上的延迟和抖动。被分割的数据帧在目的地被重组。

下图描述了 LFI 的处理过程。大报文和小的语音报文一起到达某个接口，将大报文分割成小的分片，如果在接口配置了 WFQ（Weighted Fair Queueing，加权公平队列），语音包与这些小的分片一起交叉放入 WFQ。

图1-4 LFI 的处理过程



1.4.1 配置 PPP 链路效率机制

表1-21 配置 PPP 链路效率机制

操作		命令	说明
进入系统视图		**system-view**	-
创建 MP-group		**interface mp-group** *mp-number*	必选
退回系统视图		**quit**	-
配置 PPP 上的链路分片和交叉	进入虚拟接口模板视图或 MP-group 接口视图	**interface virtual-template** *number* 或 **interface mp-group** *mp-number*	必选
	使能 LFI	**ppp mp lfi**	必选 缺省情况下，未使能 LFI
	配置 LFI 分片的最大时延	**ppp mp lfi delay-per-frag** *time*	必选 缺省情况下，分片时延为 10ms

1.5 PPP、MP 及 PPP 链路效率机制的显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 PPP 和 MP 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除相应的统计信息。

表1-22 PPP、MP 及 PPP 链路效率机制的显示和维护

操作	命令
显示已创建的 MP-group 接口的状态信息	**display interface mp-group** [*mp-number*]
显示虚拟访问接口的状态信息（集中式设备）	**display virtual-access** [*va-number* | **peer** *peer-address* | **user** *user-name* | **vt** *vt-number*] *
显示虚拟访问接口的状态信息（分布式设备）	**display virtual-access** [*va-number* | **peer** *peer-address* | **slot** *slot-number* | **user** *user-name* | **vt** *vt-number*] *
显示已创建的虚拟接口模板的状态信息	**display interface virtual-template** [*number*]
显示 MP 接口的接口信息和统计信息	**display ppp mp** [**interface** *interface-type interface-number*]
清除指定接口的统计信息	**reset counters interface** [*interface-type* [*interface-number*]]

1.6 PPP、MP 典型配置举例

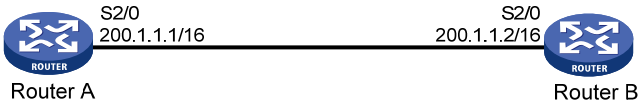
1.6.1 PAP 验证举例

1. 组网需求

如 图 1-5所示，Router A和Router B之间用接口Serial2/0 互连，要求Router A用PAP方式验证Router B。

2. 组网图

图1-5 PAP 验证、CHAP 验证举例组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] local-user user2
[RouterA-luser-user2] service-type ppp
[RouterA-luser-user2] password simple pass2
[RouterA-luser-user2] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol ppp
[RouterA-Serial2/0] ppp authentication-mode pap domain system
[RouterA-Serial2/0] ip address 200.1.1.1 16
[RouterA-Serial2/0] quit
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol ppp
```

```
[RouterB-Serial2/0] ppp pap local-user user2 password simple pass2
[RouterB-Serial2/0] ip address 200.1.1.2 16
```

1.6.2 CHAP 验证举例

1. 组网需求

在 图 1-5中，要求设备Router A用CHAP方式验证设备Router B。

2. 配置方法一（验证方配置用户名时以 CHAP 方式认证对端）

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] local-user user2
[RouterA-luser-user2] password simple hello
[RouterA-luser-user2] service-type ppp
[RouterA-luser-user2] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol ppp
[RouterA-Serial2/0] ppp chap user user1
[RouterA-Serial2/0] ppp authentication-mode chap domain system
[RouterA-Serial2/0] ip address 200.1.1.1 16
[RouterA-Serial2/0] quit
[RouterA]domain system
[RouterA-isp-system] authentication ppp local
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] local-user user1
[RouterB-luser-user1] service-type ppp
[RouterB-luser-user1] password simple hello
[RouterB-luser-user1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol ppp
[RouterB-Serial2/0] ppp chap user user2
[RouterB-Serial2/0] ip address 200.1.1.2 16
```

3. 配置方法二（验证方没有配置用户名时以 CHAP 方式认证对端）

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] local-user user2
[RouterA-luser-user2] password simple hello
[RouterA-luser-user2] service-type ppp
[RouterA-luser-user2] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ppp authentication-mode chap domain system
[RouterA-Serial2/0] ip address 200.1.1.1 16
[RouterA-Serial2/0] quit
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ppp chap user user2
[RouterB-Serial2/0] ppp chap password simple hello
[RouterB-Serial2/0] ip address 200.1.1.2 16
```

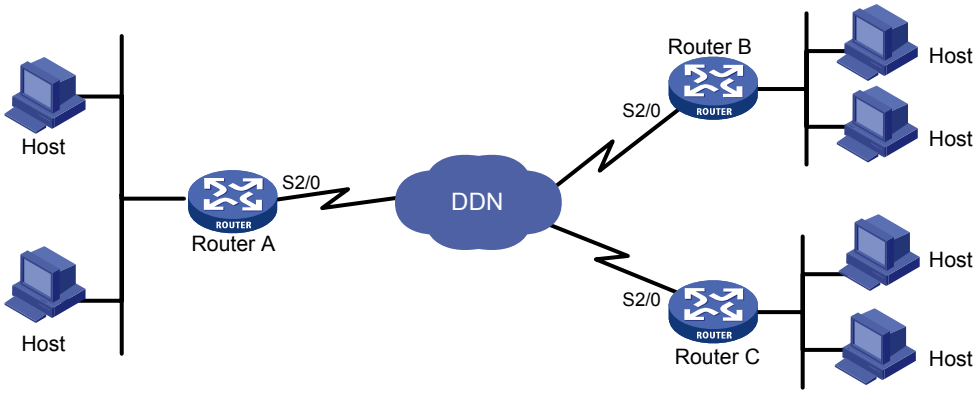
1.6.3 MP 配置举例

1. 组网需求

在 图 1-6中，Router A的E1 接口有两个通道连接到Router B的两个通道上，另外两个通道连接到Router C的两个通道上，采用验证绑定方式。假定Router A上的四个通道为：Serial2/0:1、Serial2/0:2、Serial2/0:3 和Serial2/0:4，Router B上的两个通道的接口名为：Serial2/0:1 和Serial2/0:2，Router C上的两个通道的接口名为：Serial2/0:1 和Serial2/0:2。

2. 组网图

图1-6 MP 配置举例组网图



3. 配置步骤

(1) 配置 Router A

在 Router A 上为 Router B 和 Router C 各增加一个用户。

```
<RouterA> system-view
[RouterA] local-user router-b
[RouterA-luser-router-b] password simple router-b
[RouterA-luser-router-b] service-type ppp
[RouterA-luser-router-b] quit
[RouterA] local-user router-c
[RouterA-luser-router-c] password simple router-c
[RouterA-luser-router-c] service-type ppp
[RouterA-luser-router-c] quit
```

为这两个用户指定虚拟接口模板。

```
[RouterA] ppp mp user router-b bind virtual-template 1
[RouterA] ppp mp user router-c bind virtual-template 2
```

配置虚拟接口模板。

```
[RouterA] interface virtual-template 1
[RouterA-Virtual-Template1] ip address 202.38.166.1 255.255.255.0
```

```
[RouterA-Virtual-Template1] quit
[RouterA] interface virtual-template 2
[RouterA-Virtual-Template2] ip address 202.38.168.1 255.255.255.0
[RouterA-Virtual-Template2] quit
```

将接口 Serial2/0:1、Serial2/0:2、Serial2/0:3 和 Serial2/0:4 加入 MP 通道，以 Serial2/0:1 为例，其它接口作同样配置。

```
[RouterA] interface serial 2/0:1
[RouterA-Serial2/0:1] link-protocol ppp
[RouterA-Serial2/0:1] ppp mp
[RouterA-Serial2/0:1] ppp authentication-mode pap domain system
[RouterA-Serial2/0:1] ppp pap local-user router-a password simple router-a
[RouterA-Serial2/0:1] quit
```

配置域用户使用本地认证方案。

```
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
```

(2) 配置 Router B

为 Router A 增加一个用户。

```
<RouterB> system-view
[RouterB] local-user router-a
[RouterB-luser-router-a] password simple router-a
[RouterB-luser-router-a] service-type ppp
[RouterB-luser-router-a] quit
```

为这个用户指定虚拟接口模板，将使用该模板的 NCP 信息进行 PPP 协商。

```
[RouterB] ppp mp user router-a bind virtual-template 1
```

配置虚拟接口模板的工作参数。

```
[RouterB] interface virtual-template 1
[RouterB-Virtual-Template1] ip address 202.38.166.2 255.255.255.0
[RouterB-Virtual-Template1] quit
```

将接口 Serial2/0:1 和 Serial2/0:2 加入 MP 通道，我们以 Serial2/0:1 为例，其它接口作同样配置。

```
[RouterB] interface serial 2/0:1
[RouterB-Serial2/0:1] ppp mp
[RouterB-Serial2/0:1] ppp authentication-mode pap domain system
[RouterB-Serial2/0:1] ppp pap local-user router-b password simple router-b
```

(3) 配置 Router C：

为 Router A 增加一个用户。

```
<RouterC> system-view
[RouterC] local-user router-a
[RouterC-luser-router-a] password simple router-a
[RouterC-luser-router-a] service-type ppp
[RouterC-luser-router-a] quit
```

为这个用户指定虚拟接口模板，将使用该模板的 NCP 信息进行 PPP 协商。

```
[RouterC] ppp mp user router-a bind virtual-template 1
```

配置虚拟接口模板的工作参数。

```
[RouterC] interface virtual-template 1
[RouterC-Virtual-Template1] ip address 202.38.168.2 255.255.255.0
[RouterC-Virtual-Template1] quit
```

将接口 Serial2/0:1 和 Serial2/0:2 加入 MP 通道，我们以 Serial2/0:1 为例，其它接口作同样配置。

```
[RouterC] interface serial 2/0:1
[RouterC-Serial2/0:1] ppp mp
[RouterC-Serial2/0:1] ppp authentication-mode pap domain system
[RouterC-Serial2/0:1] ppp pap local-user router-c password simple router-c
[RouterC-Serial2/0:1] quit
```

配置域用户使用本地认证方案。

```
[RouterC] domain system
[RouterC-isp-system] authentication ppp local
```

1.6.4 三种 MP 绑定方式的配置举例

1. 组网需求

在 图 1-7中，设备Router A和Router B的Serial2/1 和Serial2/0 分别对应连接。采用三种MP绑定方式，第一种是将链路直接绑定到VT上；第二种是按用户名查找VT，第三种是将链路绑定到Mp-group接口。

2. 组网图

图1-7 三种 MP 绑定方式的配置举例组网图



3. 第一种绑定方式（将物理接口直接绑定到 VT 接口上）

(1) 配置 Router A

配置对端设备 Router B 在 Router A 上的用户名和密码。

```
<RouterA> system-view
[RouterA] local-user rtb
[RouterA-luser-rtb] password simple rtb
[RouterA-luser-rtb] service-type ppp
[RouterA-luser-rtb] quit
```

创建虚拟接口模板，配置相应的 IP 地址。

```
[RouterA] interface virtual-template 1
[RouterA-Virtual-Template1] ip address 8.1.1.1 24
[RouterA-Virtual-Template1] ppp mp binding-mode authentication
```

配置串口 Serial2/1。

```
[RouterA-Virtual-Template1] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] link-protocol ppp
```

```
[RouterA-Serial2/1] ppp authentication-mode pap domain system
[RouterA-Serial2/1] ppp pap local-user rta password simple rta
[RouterA-Serial2/1] ppp mp virtual-template 1
[RouterA-Serial2/1] shutdown
[RouterA-Serial2/1] undo shutdown
[RouterA-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol ppp
[RouterA-Serial2/0] ppp authentication-mode pap domain system
[RouterA-Serial2/0] ppp pap local-user rta password simple rta
[RouterA-Serial2/0] ppp mp virtual-template 1
[RouterA-Serial2/0] shutdown
[RouterA-Serial2/0] undo shutdown
[RouterA-Serial2/0] quit
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
[RouterA-isp-system] quit
```

(2) 配置 Router B

配置对端设备 Router A 在 Router B 上的用户名和密码。

```
<RouterB> system-view
[RouterB] local-user rta
[RouterB-luser-rta] password simple rta
[RouterB-luser-rta] service-type ppp
[RouterB-luser-rta] quit
```

创建虚拟模板接口，配置相应的 IP 地址。

```
[RouterB] interface virtual-template 1
[RouterB-Virtual-Template1] ip address 8.1.1.2 24
[RouterB-Virtual-Template1] ppp mp binding-mode authentication
[RouterB-Virtual-Template1] quit
```

配置串口 Serial2/1。

```
[RouterB] interface serial 2/1
[RouterB-Serial2/1] link-protocol ppp
[RouterB-Serial2/1] ppp authentication-mode pap domain system
[RouterB-Serial2/1] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/1] ppp mp virtual-template 1
[RouterB-Serial2/1] shutdown
[RouterB-Serial2/1] undo shutdown
[RouterB-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol ppp
[RouterB-Serial2/0] ppp authentication-mode pap domain system
[RouterB-Serial2/0] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/0] ppp mp virtual-template 1
```

```
[RouterB-Serial2/0] shutdown
[RouterB-Serial2/0] undo shutdown
[RouterB-Serial2/0] quit
```

配置域用户使用本地认证方案。

```
[RouterB] domain system
[RouterB-isp-system] authentication ppp local
[RouterB-isp-system] quit
```

(3) 在 Router A 上查看绑定效果：

```
[RouterA] display ppp mp
Template is Virtual-Template1
Bundle rtb, 2 member, Master link is Virtual-Template1:0
0 lost fragments, 0 reordered, 0 unassigned, 0 interleaved,
sequence 0/0 rcvd/sent
The bundled member channels are:
      Serial2/1
      Serial2/0
```

(4) 查看 VA 状态：

```
[RouterA] display virtual-access
Virtual-Template1:0 current state: UP
Line protocol current state: UP
Description: Virtual-Template0:0 Interface
The Maximum Transmit Unit is 1500
Link layer protocol is PPP
LCP opened, MP opened, IPCP opened, OSICP opened
Physical is MP, baudrate: 64000 bps
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec 0 packets/sec
    Last 300 seconds output:  0 bytes/sec 0 packets/sec
    520 packets input, 44132 bytes, 0 drops
    527 packets output, 44566 bytes, 4 drops
```

Router B 上显示类似。

(5) 在 Router B 上 **ping** 对端 IP 8.1.1.1。

```
[RouterB] ping 8.1.1.1
  PING 8.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 8.1.1.1: bytes=56 Sequence=1 ttl=255 time=29 ms
    Reply from 8.1.1.1: bytes=56 Sequence=2 ttl=255 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=3 ttl=255 time=29 ms
    Reply from 8.1.1.1: bytes=56 Sequence=4 ttl=255 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=5 ttl=255 time=30 ms

  --- 8.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
```

```
round-trip min/avg/max = 29/30/31 ms
```

由于在物理接口下配置了 PPP 认证，故 **display ppp mp** 中的 Bundle 为对端用户名；若去掉认证的配置，Bundle 应为对端的终端描述符。

这里的虚拟访问接口 VA 的状态即对应 MP 虚通道的状态，可以用 **display virtual-access** 命令查看。

4. 第二种绑定方式（按用户名找 VT）

(1) 配置 Router A

配置对端设备 Router B 在 Router A 上的用户名和密码。

```
<RouterA> system-view
[RouterA] local-user rtb
[RouterA-luser-rtb] password simple rtb
[RouterA-luser-rtb] service-type ppp
[RouterA-luser-rtb] quit
```

指定用户 RTB 对应的 VT。

```
[RouterA] ppp mp user rtb bind virtual-template 1
```

创建 VT，配置相应的 IP 地址。

```
[RouterA] interface virtual-template 1
[RouterA-Virtual-Template1] ip address 8.1.1.1 24
[RouterA-Virtual-Template1] ppp mp binding authentication
[RouterA-Virtual-Template1] quit
```

配置串口 Serial2/1。

```
[RouterA] interface serial 2/1
[RouterA-Serial2/1] link-protocol ppp
[RouterA-Serial2/1] ppp authentication-mode pap domain system
[RouterA-Serial2/1] ppp pap local-user rta password simple rta
[RouterA-Serial2/1] ppp mp
[RouterA-Serial2/1] shutdown
[RouterA-Serial2/1] undo shutdown
[RouterA-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol ppp
[RouterA-Serial2/0] ppp authentication-mode pap domain system
[RouterA-Serial2/0] ppp pap local-user rta password simple rta
[RouterA-Serial2/0] ppp mp
[RouterA-Serial2/0] shutdown
[RouterA-Serial2/0] undo shutdown
[RouterA-Serial2/0] quit
```

配置域用户使用本地认证方案。

```
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
[RouterA-isp-system] quit
```

(2) 配置 Router B

配置对端设备 Router A 在 Router B 上的用户名和密码

```
<RouterB> system-view
[RouterB] local-user rta
[RouterB-luser-rta] password simple rta
[RouterB-luser-rta] service-type ppp
[RouterB-luser-rta] quit
```

指定用户 RTA 对应的 VT。

```
[RouterB] ppp mp user rta bind virtual-template 1
```

创建 VT，配置相应的 IP 地址。

```
[RouterB] interface virtual-template 1
[RouterB-Virtual-Template1] ip address 8.1.1.2 24
[RouterB-Virtual-Template1] ppp mp binding-mode authentication
[RouterB-Virtual-Template1] quit
```

配置串口 Serial2/1。

```
[RouterB] interface serial 2/1
[RouterB-Serial2/1] link-protocol ppp
[RouterB-Serial2/1] ppp authentication-mode pap domain system
[RouterB-Serial2/1] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/1] ppp mp
[RouterB-Serial2/1] shutdown
[RouterB-Serial2/1] undo shutdown
[RouterB-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol ppp
[RouterB-Serial2/0] ppp authentication-mode pap domain system
[RouterB-Serial2/0] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/0] ppp mp
[RouterB-Serial2/0] shutdown
[RouterB-Serial2/0] undo shutdown
[RouterB-Serial2/0] quit
```

配置域用户使用本地认证方案。

```
[RouterB] domain system
[RouterB-isp-system] authentication ppp local
[RouterB-isp-system] quit
```

(3) 在 Router A 上查看绑定效果：

```
[RouterA] display ppp mp
Template is Virtual-Template1
Bundle rtb, 2 member, Master link is Virtual-Template1:0
0 lost fragments, 0 reordered, 0 unassigned, 0 interleaved,
sequence 0/0 rcvd/sent
The bundled member channels are:
      Serial2/1
```

```
      Serial2/0
```

(4) 在 Router B 上查看绑定效果：

```
[RouterB] display ppp mp
Template is Virtual-Template1
Bundle rta, 2 member, Master link is Virtual-Template1:0
0 lost fragments, 0 reordered, 0 unassigned, 0 interleaved,
sequence 0/0 rcvd/sent
The bundled member channels are:
      Serial2/1
      Serial2/0
```

查看 VA 状态：

```
[RouterB] display virtual-access
Virtual-Template1:0 current state : UP
Line protocol current state : UP
Description : Virtual-Template1:0 Interface
The Maximum Transmit Unit is 1500
Link layer protocol is PPP
LCP opened, MP opened, IPCP opened, OSICP opened, MPLSCP opened
Physical is MP
Output queue : (Urgent queue : Size/Length/Discards)  0/50/0
Output queue : (Protocol queue : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec 0 packets/sec
    Last 300 seconds output:  0 bytes/sec 0 packets/sec
    0 packets input, 0 bytes, 0 drops
    0 packets output, 0 bytes, 0 drops
```

在 Router B 上 **ping** 对端 IP 地址 8.1.1.1。

```
[RouterB] ping 8.1.1.1
  PING 8.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 8.1.1.1: bytes=56 Sequence=1 ttl=255 time=29 ms
    Reply from 8.1.1.1: bytes=56 Sequence=2 ttl=255 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=3 ttl=255 time=30 ms
    Reply from 8.1.1.1: bytes=56 Sequence=4 ttl=255 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=5 ttl=255 time=30 ms

  --- 8.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
round-trip min/avg/max = 29/30/31 ms
```

错误配置：

如果想将 Serial2/1 和 Serial2/0 一起绑定到同一个 MP 中，不幸将一个串口配成了 **ppp mp**，而另一个配成 **ppp mp virtual-template** 1，这时系统会分别将两个串口绑定到不同的 MP 中，工作将不正常，不能达到预期的效果。

5. 第三种绑定方式（将链路绑定到 Mp-group 接口）

在系统中，还增加了 Mp-group 接口，可以将链路绑定到 Mp-group 接口。这种 MP 实现方法与绑定到 VT 的第一种方式类似。

(1) 配置 Router A

配置对端设备 Router B 在 Router A 上的用户名和密码。

```
<RouterA> system-view
[RouterA] local-user rtb
[RouterA-luser-rtb] password simple rtb
[RouterA-luser-rtb] service-type ppp
[RouterA-luser-rtb] quit
```

创建 Mp-group 接口，配置相应的 IP 地址。

```
[RouterA] interface mp-group 1
[RouterA-Mp-group1] ip address 111.1.1.1 24
```

配置串口 Serial2/1。

```
[RouterA-Mp-group1] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] link-protocol ppp
[RouterA-Serial2/1] ppp authentication-mode pap domain system
[RouterA-Serial2/1] ppp pap local-user rta password simple rta
[RouterA-Serial2/1] ppp mp mp-group 1
[RouterA-Serial2/1] shutdown
[RouterA-Serial2/1] undo shutdown
[RouterA-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol ppp
[RouterA-Serial2/0] ppp authentication-mode pap domain system
[RouterA-Serial2/0] ppp pap local-user rta password simple rta
[RouterA-Serial2/0] ppp mp mp-group 1
[RouterA-Serial2/0] shutdown
[RouterA-Serial2/0] undo shutdown
[RouterA-Serial2/0] quit
```

配置域用户使用本地认证方案。

```
[RouterA] domain system
[RouterA-isp-system] authentication ppp local
[RouterA-isp-system] quit
```

(2) 配置 Router B

配置对端设备 Router A 在 Router B 上的用户名和密码。

```
<RouterB> system-view
[RouterB] local-user rta
[RouterB-luser-rta] password simple rta
[RouterB-luser-rta] service-type ppp
[RouterB-luser-rta] quit
```

创建 Mp-group 接口，配置相应的 IP 地址。

```
[RouterB] interface mp-group 1
[RouterB-Mp-group1] ip address 111.1.1.2 24
[RouterB-Mp-group1] quit
```

配置串口 Serial2/1。

```
[RouterB] interface serial 2/1
[RouterB-Serial2/1] link-protocol ppp
[RouterB-Serial2/1] ppp authentication-mode pap domain system
[RouterB-Serial2/1] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/1] ppp mp mp-group 1
[RouterB-Serial2/1] shutdown
[RouterB-Serial2/1] undo shutdown
[RouterB-Serial2/1] quit
```

配置串口 Serial2/0。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol ppp
[RouterB-Serial2/0] ppp authentication-mode pap domain system
[RouterB-Serial2/0] ppp pap local-user rtb password simple rtb
[RouterB-Serial2/0] ppp mp mp-group 1
[RouterB-Serial2/0] shutdown
[RouterB-Serial2/0] undo shutdown
[RouterB-Serial2/0] quit
```

配置域用户使用本地认证方案。

```
[RouterB] domain system
[RouterB-isp-system] authentication ppp local
[RouterB-isp-system] quit
```

(3) 在 Router A 上查看绑定效果：

```
[RouterA] display ppp mp
Mp-group is Mp-group1
Bundle Multilink, 2 member, Master link is Mp-group1
0 lost fragments, 0 reordered, 0 unassigned, 0 interleaved,
sequence 0/0 rcvd/sent
The bundled member channels are:
      Serial2/1
      Serial2/0
```

查看 Mp-group1 状态：

```
[RouterA] display interface Mp-group 1
Mp-group1 current state : UP
Line protocol current state : UP
Description : Mp-group1 Interface
The Maximum Transmit Unit is 1500, Hold timer is 10(sec)
Internet Address is 111.1.1.1/24
Link layer protocol is PPP
LCP opened, MP opened, IPCP opened, MPLSCP opened
Physical is MP
```

```
Output queue : (Urgent queue : Size/Length/Discards)  0/50/0
Output queue : (Protocol queue : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec 0 packets/sec
    Last 300 seconds output:  0 bytes/sec 0 packets/sec
    0 packets input, 0 bytes, 0 drops
    0 packets output, 0 bytes, 0 drops
```

在 RouterA 上 **ping** 对端 IP：

```
[RouterA] ping 111.1.1.2
  PING 111.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 111.1.1.2: bytes=56 Sequence=1 ttl=255 time=29 ms
    Reply from 111.1.1.2: bytes=56 Sequence=2 ttl=255 time=31 ms
    Reply from 111.1.1.2: bytes=56 Sequence=3 ttl=255 time=29 ms
    Reply from 111.1.1.2: bytes=56 Sequence=4 ttl=255 time=30 ms
    Reply from 111.1.1.2: bytes=56 Sequence=5 ttl=255 time=30 ms
  --- 111.1.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 29/29/31 ms
```

需要注意的是，将链路绑定到Mp-group接口只有一种方式，就是在接口下直接指定相应的Mp-group接口。在这种 MP 绑定中，是将所有用户都绑定到一起，没有 VA 的概念。

1.7 PPP 常见错误配置举例

1.7.1 链路始终不能转为 up 状态

1. 故障现象

PPP 验证失败，链路始终不能转为 up 状态。

2. 故障分析

可能是由于 PPP 验证参数配置不正确，导致 PPP 验证失败。

3. 故障排除

打开 PPP 的调试开关，会看到 LCP 协商成功并转为 up 状态后进行 PAP 或 CHAP 协商，然后 LCP 转为 down 状态，则需要修改 PPP 验证参数的配置，保证验证方可以通过被验证方的验证。

1.7.2 物理链路不能转为 up 状态

1. 故障现象

物理链路始终是 down 状态。

2. 故障分析

物理链路始终是 down 状态，可能有以下原因：

- 接口没有被激活；
- 接口被管理员 down 掉；

- 物理接口已通，但是链路协商没有通过。

3. 故障排除

可以执行 **display interface** 命令来查看接口当前状态，判断故障原因，从而采取相应的方法使物理链路 up。

接口有五种状态：

`serial number is administratively down, line protocol is down`：表示该接口被管理员 down。

`serial number is down, line protocol is down`：表示该接口没有被激活或物理层没有转为 up 状态。

`Virtual-template number is down, line protocol is spoofing up`：表示该接口是拨号口，没有呼通。

`serial number is up, line protocol is up`：表示该接口链路协商即 LCP 协商已通过。

`serial number is up, line protocol is down`：表示该接口已激活，但链路协商仍没有通过。

1.7.3 IPv6CP 协商 down 后不能重协商

1. 故障现象

未使能 IPv6 的前提下在 PPP 链路上配置 IPv6 地址，IPv6CP 无法协商 up，使能 IPv6 功能后，依旧不能协商 up。

2. 故障分析

未使能 IPv6 的前提下，IPv6CP 无法协商成功，由于 IPv6CP 无重协商机制，所以后续再使能 IPv6 也无法使 IPv6CP 协商 up。

3. 故障排除

- 在 PPP 链路上配置 IPv6 地址时，建议首先使能系统的 IPv6 功能，然后再配置 IPv6 地址。
- 如果 IPv6CP 协商 down，可以通过命令 shutdown/undo shutdown 接口使其重新进行协商。

2 PPPoE 配置

2.1 PPPoE 简介

2.1.1 PPPoE

PPPoE 是 Point-to-Point Protocol over Ethernet 的简称，它可以通过一个远端接入设备为以太网上的主机提供因特网接入服务，并对接入的每个主机实现控制、计费功能。由于很好地结合了以太网的经济性及 PPP 良好的可扩展性与管理控制功能，PPPoE 在包括小区组网建设等一系列应用中被广泛采用。

PPPoE 协议采用 Client/Server 方式，它将 PPP 报文封装在以太网帧之内，在以太网上提供点对点的连接。

PPPoE 有两个阶段：Discovery 阶段和 PPP Session 阶段，具体如下：

- Discovery 阶段

当一个主机想开始 PPPoE 进程的时候，它必须先识别接入端的以太网 MAC 地址，建立 PPPoE 的 SESSION ID。这就是 Discovery 阶段的目的。

- PPP Session 阶段

当 PPPoE 进入 Session 阶段后 PPP 报文就可以作为 PPPoE 帧的净荷封装在以太网帧发到对端，SESSION ID 必须是 Discovery 阶段确定的 ID，MAC 地址必须是对端的 MAC 地址，PPP 报文从 Protocol ID 开始。在 Session 阶段，主机或服务器任何一方都可发 PADT（PPPoE Active Discovery Terminate）报文通知对方结束本 Session。

关于 PPPoE 的详细介绍，可以参考 RFC2516。

2.1.2 PPPoE Server

设备提供了 PPPoE Server 的功能，支持动态分配 IP 地址，提供本地认证、RADIUS/TACACS+等多种认证方式，配合访问包过滤防火墙及状态防火墙，可以对内部网络提供安全保障，适用于校园、智能小区等通过以太网接入 Internet 的组网应用。

这种组网方式需要在用户 Host 上安装 PPPoE 客户端拨号软件。

2.1.3 PPPoE Client

PPPoE 在 ADSL 宽带接入中被广泛使用。通常情况下，一台主机如果要通过 ADSL 接入 Internet，必须在主机上安装 PPPoE 客户端拨号软件。

2.2 配置 PPPoE Server

PPPoE Server 可以配置在物理以太网接口上。

表2-1 配置 PPPoE Server

操作	命令	说明
进入系统视图	**system-view**	-
创建虚拟接口模板并进入虚拟接口模板视图	**interface virtual-template** *number*	-
设置 PPP 的工作参数（包括验证方式、IP 地址获取方式），用户还可以设置为 PPP 对端分配的 IP 地址（或 IP 地址池）	请参见 1.2.1 配置PPP以及 1.2.4 PPP协商参数	可选
进入指定的以太网接口视图	**interface** *interface-type interface-number*	必选
在以太网接口上启用 PPPoE 协议	**pppoe-server bind virtual-template** *number*	必选 缺省情况下，禁止 PPPoE 协议
退回系统视图	**quit**	-
配置一个对端 MAC 地址上能创建的最大 PPPoE session 数目	**pppoe-server max-sessions remote-mac** *number*	可选 缺省情况下，*number* 数值为 100
配置一个本端 MAC 地址上能创建的最大 PPPoE session 数目	**pppoe-server max-sessions local-mac** *number*	可选 缺省情况下，*number* 数值为 100
配置本系统能创建的最大 PPPoE Session 数目（集中式设备）	**pppoe-server max-sessions total** *number*	可选 缺省情况下，*number* 数值为 18000
配置本系统能创建的最大 PPPoE Session 数目（分布式设备）	**pppoe-server max-sessions slot** *slot-number* **total** *number*	可选 不同 I/O 板的缺省值不同，请以设备的实际情况为准
PPPoE Server 对 PPP 用户进行认证、计费	相关内容请参考“安全分册”中的“AAA配置”	可选
退回系统视图	**quit**	-
关闭 PPPoE Server 产生的 PPP 日志信息的显示开关	**pppoe-server log-information off**	可选 缺省情况下，打开 PPPoE Server 产生的 PPP 日志信息的显示开关

说明

- 对于 Virtual-Template 接口，如果配置静态路由，请指定下一跳而不要指定出接口。如果必须指定出接口的话，请保证 Virtual-Template 下绑定的物理接口有效，从而保证报文能够正常传输。
- 配置 PPPoE Server 不进行认证时，将使用全局地址池分配地址，如果存在多个全局地址池，只能使用指定的地址池，不能从多个地址池中轮询获取地址。

2.3 PPPoE 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 PPPoE 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除相应的统计信息。

表2-2 PPPoE 显示和维护

配置	命令
显示 PPPoE Server 会话的状态和统计数据	**display pppoe-server session all**
终止 PPPoE Server 端的会话	**reset pppoe-server** { **all** | **interface** *interface-type interface-number* | **virtual-template** *number* }

2.4 PPPoE 典型配置举例

2.4.1 PPPoE Server 典型配置举例

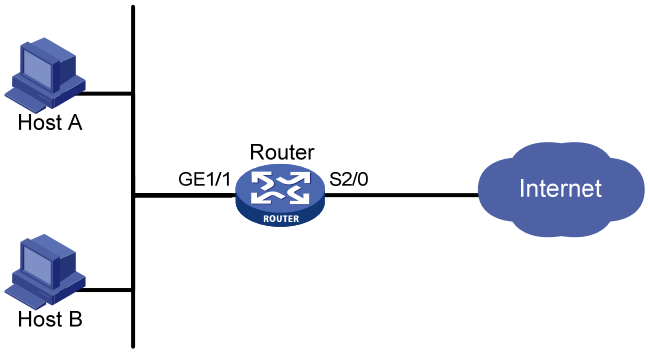
1. 组网需求

如 图 2-1所示，Host运行PPPoE Client，通过设备Router接入到Internet。设备作为PPPoE Server，配置本地认证，并通过地址池为用户分配IP地址。

2. 组网图

设备 Router 通过 GigabitEthernet1/1 接口连接以太网，Serial2/0 接口连接 Internet。

图2-1 PPPoE Server 典型配置举例组网图



3. 配置步骤

增加一个 PPPoE 用户。

```
<Sysname> system-view
[Sysname] local-user user1
[Sysname-luser-user1] password simple pass1
[Sysname-luser-user1] service-type ppp
[Sysname-luser-user1] quit
```

在设备 Router 上配置虚拟模板参数。

```
[Sysname] interface virtual-template 1
[Sysname-Virtual-Template1] ppp authentication-mode chap domain system
[Sysname-Virtual-Template1] ppp chap user user1
[Sysname-Virtual-Template1] remote address pool 1
[Sysname-Virtual-Template1] ip address 1.1.1.1 255.0.0.0
[Sysname-Virtual-Template1] quit
```

在设备 Router 上配置 PPPoE 参数。

```
[Sysname] interface gigabitethernet 1/1
```

```
[Sysname-GigabitEthernet1/1] pppoe-server bind virtual-template 1
[Sysname-GigabitEthernet1/1] qui
```

配置域用户使用本地认证方案。

```
[Sysname] domain system
[Sysname-isp-system] authentication ppp local
```

增加一个本地 IP 地址池（9 个 IP 地址）

```
[Sysname-isp-system] ip pool 1 1.1.1.2 1.1.1.10
```

这样，以太网上各主机安装 PPPoE 客户端软件后，配置好用户名和密码（此处为 user1 和 pass1）就能使用 PPPoE 协议，通过设备 Router 接入到 Internet。

若通过 **authentication ppp** 命令配置认证方案为 **radius-scheme** 或 **hwtacacs-scheme**，那么还需要配置 RADIUS/HWTACAS 参数，使系统可以进行认证、授权、计费，具体配置步骤请参见“安全分册”中的“AAA 配置”。

目 录

1 逻辑接口配置 1-1
1.1 逻辑接口简介 1-1
1.2 Loopback接口 1-1
1.2.1 Loopback接口简介 1-1
1.2.2 配置Loopback接口 1-1
1.3 Null接口 1-2
1.3.1 Null接口简介 1-2
1.3.2 配置Null接口 1-2
1.4 Tunnel接口 1-2
1.4.1 Tunnel接口简介 1-2
1.4.2 配置Tunnel接口 1-3
1.5 逻辑接口显示和维护 1-3

1 逻辑接口配置

说明

本部分只对逻辑接口中 LoopBack 接口、NULL 接口和 Tunnel 接口基本配置做简单介绍，其他逻辑接口及有关链路层、网络层协议参数以及一些特殊功能的配置，请参见“接入分册”和“IP 业务分册”的相关章节。

1.1 逻辑接口简介

逻辑接口指能够实现数据交换功能但物理上不存在的接口，又被称为虚接口。包括 LoopBack 接口、NULL 接口、子接口、Tunnel 接口、MP-group（Multilink Point to Point Protocol group，多链路点到点协议组）接口、MFR（Multilink Frame Relay，多链路帧中继）接口以及虚拟接口模板等。

1.2 Loopback 接口

1.2.1 Loopback 接口简介

Loopback 接口是一种纯软件性质的虚拟接口。Loopback 接口创建后除非手工关闭该接口，否则 Loopback 接口物理层状态和链路层协议永远处于 UP 状态。Loopback 接口可以配置 IP 地址，为了节约 IP 地址，系统会自动给 Loopback 接口的 IP 地址配置 32 位的子网掩码。Loopback 接口下也可以使能路由协议，可以收发路由协议报文。

Loopback 接口的应用非常广泛，其中最主要的是：将 Loopback 接口地址设置为该设备产生的所有 IP 数据包的源地址，因为 Loopback 接口地址稳定且是单播地址，所以通常将 Loopback 接口地址视为设备的标志，在认证或安全等服务器上设置允许或禁止携带 Loopback 接口地址的报文通过，就相当于允许或禁止某台设备产生的报文通过，这样可以简化报文过滤规则。但需要注意的是，将 Loopback 接口用于源地址绑定时，需确保 Loopback 接口到对端的路由可达，而且，任何送到 Loopback 接口的网络数据报文都会被认为是送往设备本身的，设备将不再转发这些数据包。

另外，因为 Loopback 接口状态稳定（永远处于 UP 状态），该接口还有特殊用途。在动态路由协议里，当没有配置 Router ID 时，将选取所有 Loopback 接口上数值最大的 IP 地址作为 Router ID。例如，为了使 BGP 会话不受物理接口故障的影响，可将发送 BGP 报文的源接口配置成 Loopback 接口。在使用 Loopback 接口作为 BGP 报文的源接口时，但必须注意 BGP 对等体的 Loopback 接口的地址是可达的。如果是 EBGP 连接，还要允许 EBGP 通过非直连建立邻居关系。

1.2.2 配置 Loopback 接口

表1-1 配置 Loopback 接口

操作	命令	说明
进入系统视图	**system-view**	-
创建 Loopback 接口并进入 Loopback 接口视图	**interface loopback** *interface-number*	-

操作	命令	说明
手工关闭 Loopback 接口	**shutdown**	可选 缺省情况下，Loopback 接口创建后永远处于开启状态

注意

- 在 Loopback 接口上只能配置 32 位的子网掩码。
- 在 Loopback 接口上可以配置 IP 地址、IP 路由等参数，具体配置请参见“IP 业务分册”。

1.3 Null 接口

1.3.1 Null 接口简介

Null 接口是一种纯软件性质的逻辑接口。它永远处于 up 状态，但不能转发数据包，也不能配置 IP 地址和链路层协议。如果在静态路由中指定到达某一网段的下一条为 Null 接口时，则任何送到该网段的网络数据报文都会被丢弃，因此设备通过 Null 接口提供了一种过滤报文的简单方法——将不需要的网络流量发送到 NULL 接口，从而免去配置 ACL（访问控制列表）的复杂工作。

例如：使用静态路由配置命令“ip route-static 92.101.0.0 255.255.0.0 null 0”将丢弃所有去往网段 92.101.0.0/16 的报文。

1.3.2 配置 Null 接口

表1-2 配置 Null 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入 Null 接口视图	**interface null 0**	必选 缺省情况下，设备上已经存在 Null0 接口，用户不能创建也不能删除

1.4 Tunnel 接口

1.4.1 Tunnel 接口简介

隧道技术是一种封装技术，它利用一种网络协议来传输另一种网络协议，即利用一种网络协议，将其他协议产生的数据报文封装在它自己的报文中，然后在网络中传输。

隧道技术可以：

- 作为过渡技术，实现 IPv4 和 IPv6 网络互通，如 IPv6 over IPv4 隧道技术。
- 创建 VPN（Virtual Private Network，虚拟私有网络），保证通信的安全性，如 GRE（Generic Routing Encapsulation，通用路由封装）。
- 实现流量工程，避免由于负载不均衡导致网络拥塞，如 MPLS TE（Multiprotocol Label Switching Traffic Engineering，多协议标记交换流量工程）。

隧道（Tunnel）是一个虚拟的点对点的连接，提供了一条使封装的数据报文能够传输的通路。隧道的两端需要对数据报文进行封装及解封装。

作为过渡技术的隧道、GRE 隧道、DVPN 隧道和 MPLS TE 隧道中，需要在隧道的两端创建虚拟的三层接口——Tunnel 接口，以便隧道两端的设备利用隧道发送报文、识别并处理来自隧道的报文。

1.4.2 配置 Tunnel 接口

表1-3 配置 Tunnel 接口

操作		命令	说明
进入系统视图		**system-view**	-
创建 Tunnel 接口视图，并进入 Tunnel 接口视图		**interface tunnel** *number*	必选
配置隧道模式		**tunnel-protocol** { **gre** [**ipv6**] | **ipv4-ipv4** | **ipv4-ipv6** | **ipv6-ipv4** [**6to4** | **auto-tunnel** | **isatap**] | **ipv6-ipv6** | **mpls te** }	可选 缺省情况下，采用 GRE 隧道模式
配置接口描述信息		**description** *text*	可选 缺省情况下，接口描述信息为“*该接口的接口名* Interface”
设置 MTU	设置隧道接口发送 IPv4 报文的 MTU 值	**mtu** *size*	可选 缺省情况下，接口处于开启状态
	设置隧道接口发送 IPv6 报文的 MTU 值	**ipv6 mtu** *size*	
关闭 Tunnel 接口		**shutdown**	可选 缺省情况下，接口处于开启状态
配置当前接口的转发流量业务处理板		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板



说明

上面仅列出了 Tunnel 接口的基本配置，关于 Tunnel 接口的详细介绍和应用，请参见“IP 业务分册”中的“隧道配置”、“MPLS 分册”中的“MPLS TE 配置”、以及“VPN 分册”中的“GRE 配置”。

1.5 逻辑接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示以太网子接口的配置情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除接口统计信息。

表1-4 逻辑接口显示和维护

操作	命令
显示 Loopback 接口的相关信息	**display interface loopback** [*interface-number*]
显示 Null 接口的状态信息	**display interface null** [**0**]

操作	命令
显示指定 Tunnel 接口的相关信息	**display interface tunnel** [*number*]
显示 Tunnel 接口的 IPv6 信息	**display ipv6 interface tunnel** [*number*] [**verbose**]
清除指定逻辑接口的统计信息	**reset counters interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]]

目 录

1 Modem管理配置 1-1
1.1 Modem管理简介 1-1
1.2 配置Modem管理 1-1
1.2.1 配置Modem的应答方式 1-1
1.2.2 通过AT命令配置Modem 1-2

1 Modem 管理配置

1.1 Modem 管理简介

Modem 是目前使用广泛的一种网络设备，实现对 Modem 的良好管理和控制是设备的一个重要功能，但是由于 Modem 的厂家众多而且类型各异，虽然都支持业界标准的 AT 命令集，但在具体的实现和命令的细节上存在着或多或少的差别。

设备提供以下的 Modem 管理功能：

(1) 与其他设备提供商的设备进行互通，即双方的异步串口都工作在 Flow 方式下并通过 Modem 进行互连；

(2) 提供了丰富的调试信息，便于 Modem 的监控和维护。

1.2 配置 Modem 管理

表1-1 配置 Modem 管理

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface aux 0**	-
允许 Modem 呼入和呼出、呼入或者呼出	**modem { both | call-in | call-out }**	必选 缺省情况下，禁止 Modem 呼入和呼出
配置呼入连接建立时，用户从摘机到拨号的有效间隔时间	**modem timer answer** *time*	可选 缺省情况下，有效间隔时间为 60 秒
配置 Modem 的应答方式	请参见 1.2.1 配置Modem的应答方式	可选 推荐用户不配置该任务
退回系统视图	**quit**	-
通过 AT 命令配置 Modem	请参见 1.2.2 通过AT命令配置Modem	可选
配置 Modem 回呼	**service modem-callback**	可选 缺省情况下，禁止 Modem 回呼

1.2.1 配置 Modem 的应答方式

建议根据设备外接 Modem 的当前应答状态配置 Modem 的应答方式。当 Modem 状态为自动应答（Modem 的 AA 灯亮）时，配置 **modem auto-answer**（以避免 Modem 自动应答后，设备又发出应答指令）；如果外接 Modem 为非自动应答方式，则可配置 **undo modem auto-answer**。

说明

当配置 Modem 的应答方式与 Modem 当前的应答方式不一致时，对于某些 Modem 可能会造成应答不正常。建议用户不要配置该命令。

表1-2 配置 Modem 的应答方式

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface aux 0**	-
配置 Modem 的应答方式为自动应答方式	**modem auto-answer**	必选 缺省情况下，Modem 为非自动应答方式

1.2.2 通过 AT 命令配置 Modem

表1-3 通过 AT 命令配置 Modem

操作	命令	说明
进入系统视图	**system-view**	-
进入相应的接口视图	**interface** *interface-type interface-number*	-
通过 AT 命令配置 Modem	**sendat** *at-string*	必选 该命令在异步串口（包括同异步串口工作在异步方式）下有效

目 录

1 ATM接口配置 1-1
1.1 ATM接口介绍 1-1
1.1.1 ATM 1-1
1.1.2 中高端路由器提供的ATM接口类型 1-1
1.1.3 ATM接口特性 1-1
1.2 ATM OC-3c/STM-1 接口 1-1
1.2.1 ATM OC-3c/STM-1 接口介绍 1-1
1.2.2 配置ATM OC-3c/STM-1 接口 1-2
1.3 ATM接口显示和维护 1-2
1.4 ATM接口故障的诊断与排除 1-3
1.4.1 ATM接口故障的诊断与排除 1-3

1 ATM 接口配置

1.1 ATM 接口介绍

1.1.1 ATM

ATM（Asynchronous Transfer Mode，异步传输模式）技术是一种主干网络技术，被设计用来传输语音、视频及数据信息。由于它的灵活性以及对多媒体业务的支持，被认为是实现宽带通信的核心技术。

ATM 物理层位于 ATM 协议参考模型的最底层，它涉及具体的传输介质，但其功能并不依赖于其所用的传输机制和速率，主要是在高层与传输介质之间传送有效的信元和相应的定时信号。对于接入的物理介质的速率，已由相应的国际标准组织进行标准化，例如有 ATM OC-3c/STM-1，ATM E3/T3，IMA-E1/T1 等。

1.1.2 中高端路由器提供的 ATM 接口类型

中高端路由器目前提供的 ATM 接口为基于 SONET/SDH 承载的 ATM OC-3c/STM-1 接口。

中高端路由器的 ATM 接口支持 IPoA、IPoEoA、PPPoA、PPPoEoA 这几种应用方式。相关的配置请参见“接入分册”中的“ATM 配置”部分。

1.1.3 ATM 接口特性

中高端路由器的 ATM 接口具有以下特性：

- 支持 nrt_VBR（Nonreal-time Variable Bit Rate，非实时可变比特率）；
- 支持 rt_VBR（Real-time Variable Bit Rate，实时可变比特率）；
- 支持 CBR（Constant Bit Rate，确定比特率）；
- 支持 UBR（Unspecified Bit Rate，不确定比特率）；
- 支持 PVC（Permanent Virtual Circuit，永久虚电路）；
- 支持基于 VC 的流量整形；
- 支持 UNI 接口（User-to-Network Interface，用户网络接口）；
- 支持 RFC1483：Multiprotocol Encapsulation over ATM Adaptation Layer 5；
- 支持 RFC2225：Classical IP and ARP over ATM；
- 支持 RFC 2390：Inverse Address Resolution Protocol；
- 支持 F5 End to End Loopback OAM；
- 支持 AAL5（ATM Adaptation Layer 5，ATM 适配层 5）。

1.2 ATM OC-3c/STM-1 接口

1.2.1 ATM OC-3c/STM-1 接口介绍

本节主要介绍 ATM OC-3c/STM-1 接口的物理配置，ATM 业务的详细配置（包括 PVC 的配置）请参考“接入分册”中的“ATM 配置”。

1.2.2 配置 ATM OC-3c/STM-1 接口

表1-1 配置 ATM OC-3c/STM-1 接口

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 ATM OC-3c/STM-1 接口视图	**interface atm** *interface-number*	必选
配置 ATM OC-3c/STM-1 接口的时钟模式	**clock** { **master** | **slave** }	可选 缺省情况下，ATM OC-3c/STM-1 接口的时钟模式为从时钟
配置 ATM OC-3c/STM-1 接口的帧格式	**frame-format** { **sdh** | **sonet** }	可选 缺省情况下，ATM OC-3c/STM-1 接口的帧格式为 SDH STM-1
配置 ATM OC-3c/STM-1 接口的加扰功能	**scramble**	可选 缺省情况下，使能加扰功能
配置 ATM OC-3c/STM-1 接口的环回方式	**loopback** { **local** | **remote** }	可选 缺省情况下，禁止环回功能
配置子接口速率统计功能	**sub-interface rate-statistic**	可选 缺省情况下，接口的子接口速率统计功能处于关闭状态

1.3 ATM 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 ATM 接口配置后的情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令清除相关统计信息。

表1-2 ATM 接口显示和维护

操作	命令
查看 ATM 接口的配置及状态	**display interface atm** [*interface-number*]
清除指定 ATM 接口下面的所有的 PVC 上的相关统计信息	**reset atm interface** [**atm** *interface-number*]
清除指定接口的统计信息	**reset counters interface** [**atm** [*interface-number*]]



说明

- 当路由器的某物理接口闲置，没有连接线缆时，请使用 **shutdown** 命令关闭该接口，以防止由于干扰导致接口异常。
- 有关 **reset counters interface** 命令的详细介绍，请参见“接入分册”中的“以太网命令”。

1.4 ATM 接口故障的诊断与排除

1.4.1 ATM 接口故障的诊断与排除

在对 ATM 接口进行故障诊断时，首先可使用 **ping** 命令或扩展的 **ping** 命令进行检测。

ping 命令用于检查基本的网络连通情况。扩展的 **ping** 命令还可以指定 IP 报文头中的一些选项。关于 **ping** 命令的具体使用，请参考“系统分册”中的“系统维护与调试配置”。

如果 **ping** 命令失败，检查是否是由于以下原因导致不能连通：

- 接口处于 down 的状态，导致路由表中没有路由。
- 如果是 ATM 155M 接口，可能是 PVC 的 AAL5 封装类型不正确。

目 录

1 ATM配置 1-1
1.1 ATM技术简介 1-1
1.1.1 ATM简介 1-1
1.1.2 ATM层次结构 1-1
1.2 IPoA、IPoEoA、PPPoA、PPPoEoA应用介绍 1-2
1.2.1 IPoA 1-2
1.2.2 IPoEoA 1-2
1.2.3 PPPoA 1-2
1.2.4 PPPoEoA 1-3
1.3 服务类型 1-3
1.3.1 恒定比特流业务CBR 1-3
1.3.2 实时的可变比特率业务VBR-RT 1-3
1.3.3 非实时的可变比特率业务VBR-NRT 1-3
1.3.4 未定义比特率业务UBR 1-3
1.4 InARP简介 1-3
1.5 ATM OAM 1-4
1.6 ATM配置任务简介 1-4
1.7 配置ATM接口 1-5
1.8 配置ATM子接口 1-5
1.8.1 配置ATM子接口 1-5
1.8.2 配置ATM P2P子接口协议状态与PVC相关 1-6
1.9 配置PVC及接口支持最大PVC数 1-6
1.9.1 配置PVC参数 1-6
1.9.2 配置PVC业务映射 1-7
1.9.3 配置ATM接口支持最大PVC数 1-8
1.10 配置ATM类 1-8
1.11 配置VP监管 1-10
1.12 配置ATM上承载的应用 1-10
1.12.1 配置IPoA 1-10
1.12.2 配置IPoEoA 1-11
1.12.3 配置PPPoA 1-12
1.12.4 配置PPPoEoA 1-12
1.13 ATM显示和维护 1-13
1.14 ATM典型配置举例 1-14
1.14.1 IPoA典型配置举例 1-14
1.14.2 IPoEoA典型配置举例 1-15
1.14.3 PPPoA典型配置举例 1-16
1.14.4 PPPoEoA Server典型配置举例 1-18
1.15 ATM故障的诊断与排除 1-20
1.15.1 采用IPoA时，链路状态为down 1-20
1.15.2 采用PPPoA时，链路不上报up 1-20

1.15.3 ping不通对方 1-20
1.15.4 ATM接口状态为down 1-21
1.15.5 ATM接口状态为up，但PVC状态为down 1-21
1.15.6 配置完PPPoA等应用之后，却无法ping通对端 1-21
1.15.7 出现大量报文丢弃和CRC校验错误，接口状态在up、down之间跳变 1-22

1 ATM 配置

1.1 ATM 技术简介

1.1.1 ATM 简介

ATM（Asynchronous Transfer Mode，异步传输模式）技术是以分组传输模式为基础并融合了电路传输模式高速化的优点发展而成的，可以满足各种通信业务的需求。ATM 已被 ITU-T 于 1992 年 6 月指定为 B-ISDN 的传输和交换模式。由于它的灵活性以及对多媒体业务的支持，被认为是实现宽带通信的核心技术。

根据 ITU-T 定义，ATM 是以信元为基本单位进行信息传输、复用和交换的。ATM 信元具有 53 字节的固定长度，其中前 5 个字节是信元头，其余 48 个字节是有效载荷。ATM 信元头的功能有限，主要用来标识虚连接，另外也完成了一些功能有限的流量控制，拥塞控制，差错控制等功能。

ATM 是面向连接的交换，其连接是逻辑连接，即虚电路。每条虚电路（Virtual Circuit，VC）用虚路径标识符（Virtual Path Identifier，VPI）和虚通道标识符（Virtual Channel Identifier，VCI）来标识。一个 VPI/VCI 值对只在 ATM 节点之间的一段链路上有局部意义，它在 ATM 节点上被翻译。当一个连接被释放时，与此相关的 VPI/VCI 值对也被释放，它被放回资源表，供其它连接使用。

目前 ATM 接口只支持永久虚电路（Permanent Virtual Circuit，PVC）。

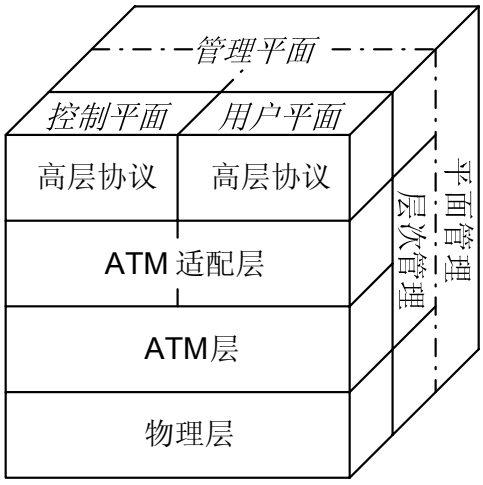
1.1.2 ATM 层次结构

ATM 基本协议框架分为 3 个平面，即用户平面、控制平面和管理平面。用户平面和控制平面又各分为 4 层，即物理层、ATM 层、ATM 适配层和高层，在各层中还有更精细的子层划分。

- 控制平面主要利用信令协议来完成连接的建立和拆除。
- 管理平面又分为层次管理和平面管理。其中层次管理负责各平面中各层的管理，具有与其它平面相对应的层次结构；平面管理负责系统的管理和各平面之间的通信。

各平面与各层的关系如 图 1-1：

图1-1 ATM 协议模型图



各层的具体功能如下：

- 物理层主要提供 ATM 信元的传输通道，将 ATM 层传来的信元加上其传输开销后形成连续的比特流；同时，在接收到物理媒介上传来的连续比特流后，取出有效信元传递给 ATM 层。
- ATM 层在物理层之上，利用物理层提供的服务，与对等层进行以信元为单位的通信。ATM 层与物理媒介的类型和物理层的具体实现无关，与具体传送的业务类型也无关。输入 ATM 层的是 48 字节的净荷，这 48 字节的净荷被称为分段和重组协议数据单元（SAR－PDU），而 ATM 层输出的则是 53 字节的信元，该信元将传送到物理层进行传输。ATM 层负责产生 5 个字节的信元头，信元头将加到净荷的前面。ATM 层的其他功能包括虚路径标识符/虚通道标识符（VPI/VCI）传输、信元多路复用/分用以及一般流量控制。
- ATM 适配层（ATM Adaption Layer，AAL）是高层协议与 ATM 层间的接口，它负责转接 ATM 层与高层协议之间的信息。目前，已经提出 4 种类型的 AAL：AAL1、AAL2、AAL3/4 和 AAL5，每一种类型分别支持 ATM 网中某些特征业务。大多数 ATM 设备制造商现在生产的产品普遍采用 AAL5 来支持数据通信业务。
- ATM 高层协议则主要具有 WAN 互连、语音互连、与现有 3 层协议互连、封装方式、局域网仿真、ATM 的多协议和经典 IP 等功能。

1.2 IPoA、IPoEoA、PPPoA、PPPoEoA 应用介绍

ATM 接口支持 IPoA、PPPoA、IPoEoA 以及 PPPoEoA 这四种应用方式。

1.2.1 IPoA

IPoA（IP over ATM）指的是在 ATM 上承载 IP 协议报文：ATM 为处在同一网络内的 IP 主机之间的通信提供数据链路层，同时将 IP 报文封装在 ATM 信元中。ATM 作为 IP 业务的承载网提供了优良的网络性能和完善、成熟的 QoS 保证。

1.2.2 IPoEoA

IPoEoA 是 IP over Ethernet over ATM 的简称。IPoEoA 有三层结构：最上层封装 IP 协议；中间为 IPoE，即以太网承载 IP 协议；最下一层为 ATM 承载 IPoE。通过设备高速连入远端的接入服务器（以便访问外部网络）时，由于距离较远，采用 ATM 的 PVC 承载，这就要求在服务器 ATM 端口承载以太网报文，这就是 IPoEoA。

对于 IPoEoA，设备实现的基本功能有：

- 在 IPoEoA 应用中使用 VE（Virtual Ethernet，虚拟以太网）接口，一个 VE（Virtual Ethernet，虚拟以太网）接口可以关联多个 PVC。
- 在同一个 VE 接口关联的 PVC 之间二层互通。

1.2.3 PPPoA

PPPoA（PPP over ATM）指的是在 ATM 上承载 PPP 协议报文：ATM 信元封装 PPP 报文，IP 或其它协议的报文则封装在 PPP 报文中。在这种视图下，可以将 ATM 简单地看成是 PPP 报文的承载层。PPPoA 的意义在于：PPPoA 的通讯过程由 PPP 协议管理，可以利用 PPP 的灵活性及其丰富的应用。为了在 ATM 上传送 PPP 报文，用户必须创建一个虚拟模板（Virtual Template，VT）接口。

1.2.4 PPPoEoA

PPPoEoA（PPPoE over ATM）指的是在 ATM 上承载 PPPoE（PPP over Ethernet）协议报文，其实质是用 ATM 信元封装以太网报文，这时候可以用一个 PVC 来模拟以太网的全部功能。为了实现在 ATM 上承载以太网报文，接口管理模块提供了一种新的虚拟以太网（Virtual Ethernet，VE）接口。这种接口具有以太网的特性，由用户通过配置命令动态创建。为这种接口配置的协议栈是：底层为 ATM 的 PVC，通过 PVC 收发报文；链路层为以太网协议；网络层及以上各层协议与普通以太网接口相同。

1.3 服务类型

ATM 支持四种服务类型：CBR、UBR、VBR-RT、VBR-NRT。这些服务类型的选择与网络的 QoS 需求有关。

1.3.1 恒定比特流业务 CBR

CBR（Constant Bit Rate）业务用于在连接的生命期中需要静态带宽的连接。这个带宽由 PCR（peak cell rate，峰值信元速率）值来确定。在 CBR 业务中，源端可以持续地以峰值信元速率发送信元。

CBR 业务一般用来支持对时延变化要求较高的实时业务（例如：语音、视频）。

1.3.2 实时的可变比特率业务 VBR-RT

VBR-RT（Variable Bit Rate- Real Time）业务也是一种实时的应用，对时延和抖动有严格的限制，VBR-RT 的主要应用有话音和视频业务。

VBR-RT 连接的指标主要靠峰值信元速率（PCR）、可持续信元速率（SCR）、最大突发长度（MBS）来描述。源端可以在平均信元速率为 SCR 的情况下，以 PCR 的速率发送最大长度为 MBS 的突发流量而不丢包。

1.3.3 非实时的可变比特率业务 VBR-NRT

VBR-NRT（Variable Bit Rate – Non Real Time）业务支持突发性的非实时的应用，该特性是通过 PCR、SCR 以及 MBS 来描述的。对那些满足流量合同的信元，VBR-NRT 业务可以保证很低的信元丢失率但是不保证时延。

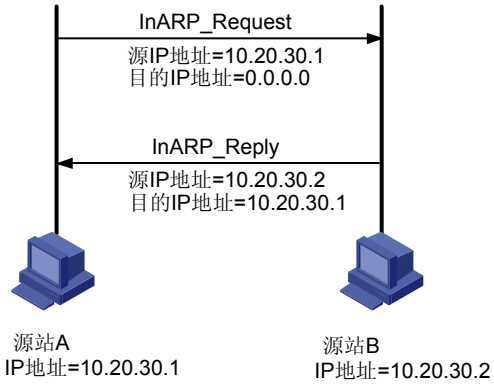
1.3.4 未定义比特率业务 UBR

UBR（Unspecified Bit Rate）用于对时延和带宽都要求不高的应用，也就是那些对时延和时延变化要求都不太严格的应用。UBR 业务不保证服务质量，连接的信元丢失率和信元传输时延均没有数值保证，如果发生拥塞，UBR 信元最先被丢弃。

1.4 InARP 简介

在ATM PVC环境中，可以使用反向地址解析协议（InARP）来解析与本PVC相连的对端接口的IP地址，这样不需要为PVC静态配置对端的IP地址。InARP交换过程如 图 1-2所示。

图1-2 InARP 工作过程示意图



1.5 ATM OAM

OAM 的名词存在两种不同解释，主要是针对不同的协议而言。

- OAM：Operation And Maintenance（ITU-T I.610 02/99）
- OAM：Operation Administration and Maintenance（LUCENT APC User Manual，03/99）

OAM 提供了一种不中断业务的故障检测、故障定位和性能检测功能。在用户信元流中间插入一些有着标准的信元结构的 OAM 信元，可以提供网络的一些特定信息。

OAM F5 LoopBack 检测：一端发送 OAM 信元给对端，如果对端收到后，则把这个 OAM 信元返回给发送方；发送方判断接收到自己的 OAM 信元，并在预先设置的时间内收到，则说明链路是通畅的；如果本端的计数器检查发送的 OAM 信元超时累计到设置的次数，则本端口认为链路有故障。

OAM F5 LoopBack 检测有手动（OAMPing）和自动（OAM Frequency）两种方式：前者手动指定发送一定量的 OAM 信元，常用于故障诊断；后者指定以一定的时间间隔发生 OAM 信元，用于链路自动检测。

OAM CC（Continuity Check）检测：一端发送 OAM 信元，对端进行检测。

1.6 ATM 配置任务简介

表1-1 ATM 配置任务简介

配置任务		说明	详细配置
配置 ATM 接口		必选	1.7
配置 ATM 子接口	配置 ATM 子接口	可选	1.8
	配置 ATM P2P 子接口协议状态与 PVC 相关		
配置 PVC 及接口支持最大 PVC 数	配置 PVC 参数	必选	1.9.1
	配置 PVC 业务映射	可选	1.9.2
	配置 ATM 接口支持最大 PVC 数	可选	1.9.3
配置 ATM 类		可选	1.10
配置 VP 监管		可选	1.11

配置任务		说明	详细配置
配置 ATM 上承载的应用（根据业务类型的不同进行选择配置）	配置 IPoA	四种方式必选其一	1.12.1
	配置 IPoEoA		1.12.2
	配置 PPPoA		1.12.3
	配置 PPPoEoA		1.12.4

1.7 配置 ATM 接口

根据实际组网环境和系统运行的要求，有时可能需要改变 ATM 接口的某些参数。需要注意的是，虽然这些参数同时作用于 ATM 主接口和子接口，但只能在 ATM 主接口视图下修改这些参数（**mtu** 命令除外，该命令也可以在子接口下执行）。

具体的 ATM 接口相关配置请参见“接入分册”中的“ATM 和 DSL 接口配置”。

1.8 配置 ATM 子接口

1.8.1 配置 ATM 子接口

表1-2 配置 ATM 子接口

操作	命令	说明
进入系统视图	**system-view**	-
创建并进入 ATM 子接口视图	**interface atm** *interface-number.subnumber* [**p2mp** | **p2p**]	必选 缺省情况下，子接口类型为 p2mp
设置 ATM 子接口的 MTU	**mtu** *mtu-number*	可选 缺省情况下，ATM 子接口最大传输单元（MTU）的大小为 1500 字节
关闭 ATM 物理接口	**shutdown**	可选 缺省情况下，ATM 物理接口为打开状态

注意

- 在创建 ATM 子接口时，**p2mp** 和 **p2p** 为可选项，命令形式为 **interface atm** *interface-number.subnumber* [**p2mp** | **p2p**];
- 在进入已创建的 ATM 子接口时，没有这两个选项，命令形式为 **interface atm** *interface-number.subnumber*。

1.8.2 配置 ATM P2P 子接口协议状态与 PVC 相关

表1-3 配置 ATM P2P 子接口协议状态与 PVC 相关

操作	命令	说明
进入系统视图	**system-view**	-
创建并进入 ATM 子接口视图	**interface atm** *interface-number.subnumber* **p2p**	必选 缺省情况下，子接口类型为 p2mp
配置 ATM P2P 子接口协议状态与 PVC 相关	**atm-link check**	必选 缺省情况下，ATM P2P 子接口的协议状态和物理接口的状态保持一致

1.9 配置 PVC 及接口支持最大 PVC 数

1.9.1 配置 PVC 参数

表1-4 配置 PVC 参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 ATM 接口视图或者 ATM 子接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC 并进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	必选 缺省情况下，未创建任何 PVC
配置 PVC 的 AAL5 封装协议类型	**encapsulation** *aal5-encap*	可选 缺省情况下，PVC 的 AAL5 封装协议类型为 aal5snap
启动 OAM F5 Loopback 信元的发送和重传检测	**oam frequency** *frequency* [**up** *up-count* **down** *down-count* **retry-frequency** *retry-frequency*]	可选 缺省情况下，不启动 OAM F5 Loopback 信元的发送，但如果收到 OAM F5 Loopback 信元，则要进行应答 缺省情况下，参数 *up-count* 为 3，*down-count* 为 5，*retry-frequency* 为 1 秒
启动 OAM CC（Continuity Check）功能	**oam cc end-to-end** { **both | sink | source** }	可选 缺省情况下，OAM 连续性检测功能处于关闭状态 • 在配置 OAM CC 功能时一端配置为 source，另一端配置为 sink • 启动 OAM CC 检测功能后，如果检测端 3 秒内收不到 CC 信元，PVC 状态变为 DOWN，当再次收到 CC 信元或者报文后，PVC 状态变为 UP • 配置了哪种启动方式，就只允许 undo 哪种启动方式，不允许 undo 其它启动方式，例如配置了 both 方式，就只允许执行 **undo oam cc end-to-end both**，不允许执行 **undo oam cc end-to-end sink** • 本命令不能在 PVC-Group 的从 PVC 下配置

操作		命令	说明
配置 AIS/RDI（Alarm Indication Signal/Remote Defect Indication）告警信元检测的相关参数		**oam ais-rdi up** *up-count* **down** *down-count*	可选 缺省情况下，系统启动 AIS/RDI 告警信元检测，其含义是当收到 *down-count* 个 AIS/RDI 告警信元后，PVC 状态转变为 down，当连续 *up-count* 秒没有收到 AIS/RDI 告警信元后，PVC 状态转变为 up
配置 PVC 的业务类型和相关速率参数	指定 PVC 的业务类型为确定速率（Constant Bit Rate，CBR）	**service cbr** *output-pcr* [**cdvt** *cdvt-value*]	可选 缺省情况下，PVC 的业务类型为 UBR 缺省情况下，信元时延变化容限 *cdvt-value* 的缺省值为 500μs 可以使用这四条命令来设置 PVC 的业务类型和相关速率参数。但需要注意的是，新指定的 PVC 业务类型将会覆盖已有的业务类型
	指定 PVC 的业务类型为非确定速率（Unspecified Bit Rate，UBR），并指定相关的速率参数	**service ubr** *output-pcr*	
	指定 PVC 的业务类型为非实时可变速率（Variable Bit Rate-Non Real Time，VBR-NRT），并指定相关的速率参数	**service vbr-nrt** *output-pcr output-scr output-mbs*	
	指定 PVC 的业务类型为实时可变速率（Variable Bit Rate - Real Time，VBR-RT），并指定相关的速率参数	**service vbr-rt** *output-pcr output-scr output-mbs*	

1.9.2 配置 PVC 业务映射

PVC 业务映射用于让同一个 PVC-Group 中的不同 PVC 承载不同优先级别的 IP 包。

表1-5 配置 PVC 业务映射

操作	命令	说明
进入系统视图	**system-view**	-
进入 ATM 子接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC，进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	-
退回 ATM 接口视图	**quit**	-
创建 PVC-Group 并进入 Group 视图	**pvc-group** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	必选 *pvc-name* 或者 *vpi/vci* 对应的 PVC 必须已经建立
在已经创建的 PVC-Group 中加入 PVC	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	可选
配置 PVC 承载的 IP 包的优先级	**ip-precedence** { *pvc-name* [*vpi/vci*] | *vpi/vci* } { *min* [*max*] | **default** }	可选

说明

- 主 PVC：创建 PVC-Group 时指定的 PVC（在 ATM 接口上创建）为基础 PVC，也称主 PVC。
- 从 PVC：在 PVC-Group 下创建的 PVC，称为从 PVC。

1.9.3 配置 ATM 接口支持最大 PVC 数

表1-6 配置 ATM 接口支持最大 PVC 数

操作	命令	说明
进入系统视图	**system-view**	-
进入 ATM 接口视图	**interface atm** *interface-number*	-
配置 ATM 接口 PVC 数目的最大值	**pvc max-number** *max-number*	可选 ATM 接口 PVC 数目的最大值跟接口类型有关，请以实际的接口为准。



说明

本命令虽然同时作用于 ATM 主接口和子接口，但只能在 ATM 主接口视图下使用此命令，ATM 子接口视图下无此命令。

1.10 配置 ATM 类

ATM 类主要的作用是方便对 ATM 的配置。可以在 ATM 类上设置 MAP、封装格式、OAM 环回、服务类型等，再将 ATM 类上的配置应用到到 PVC 中。具体使用时，先在系统视图下创建 ATM 类并设置所需要的参数，接着在 PVC 视图或 ATM 接口视图下调用该 ATM 类。ATM 类的参数设置步骤如下所示。

表1-7 配置 ATM 类

操作	命令	说明
进入系统视图	**system-view**	-
创建 ATM 类并进入 ATM 类视图	**atm class** *atm-class-name*	必选
配置 PVC 的 AAL5 封装协议类型	**encapsulation** *aal5-encap*	可选 缺省情况下，PVC 的 AAL5 封装协议类型为 aal5snap
启动 OAM F5 Loopback 信元的发送或者重传检测	**oam frequency** *frequency* [**up** *up-count* **down** *down-count* **retry-frequency** *retry-frequency*]	可选 缺省情况下，不启动 OAM F5 Loopback 信元的发送，但如果收到 OAM F5 Loopback 信元，则要进行应答，参数 *up-count* 为 3，*down-count* 为 5，*retry-frequency* 为 1 秒

操作		命令	说明
配置 PVC 的业务类型和相关速率参数	指定 PVC 的业务类型为确定速率	**service cbr** *output-pcr*	可选 缺省情况下，PVC 的业务类型为 UBR 可以使用这四条命令来设置 PVC 的业务类型和相关速率参数。但需要注意的是，新指定的 PVC 业务类型将会覆盖已有的业务类型
	指定 PVC 的业务类型为非确定速率，并指定相关的速率参数	**service ubr** *output-pcr*	
	指定 PVC 的业务类型为非实时可变速率，并指定相关的速率参数	**service vbr-nrt** *output-pcr output-scr output-mbs*	
	指定 PVC 的业务类型为实时可变速率，并指定相关的速率参数。	**service vbr-rt** *output-pcr output-scr output-mbs*	
配置应用类型（根据应用类型的不同选择配置）	配置 IPoA，在 PVC 上使能反向地址解析 InARP	**map ip inarp** [*minutes*] [*broadcast*]	可选 缺省情况下，不配置任何映射。如果配置，缺省不支持伪广播 在配置 InARP 时，必须保证使用的是 aal5snap 封装类型。当采用 aal5mux 和 aal5nlpid 封装时，也可以配置 InARP，但是在 PVC 下配置使用该 ATM 类时，系统会提示不支持
	创建 PVC 上的 PPPoA 映射	**map ppp virtual-template** *vt-number*	可选
	创建 PVC 上的 IPoEoA 映射或者 PPPoEoA 映射	**map bridge virtual-ethernet** *interface-number*	可选
退回系统视图		**quit**	-
进入 ATM 接口视图或 PVC 视图	进入 ATM 接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	必选
	进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	
将 ATM 类应用在接口上或者 PVC 上		**atm-class** *atm-class-name*	必选

需要注意的是：

- 如果配置的命令相同，但是参数不同，在 PVC 上直接配置的参数具有最高的优先级，其次是在 PVC 上应用的 ATM 类中配置的参数，最后是在 ATM 接口上应用的 ATM 类中配置的参数；
- 如果配置的命令不同，而且命令冲突，在 PVC 上直接配置的命令具有最高的优先级，其次是在 PVC 上应用的 ATM 类中配置的命令，最后是在 ATM 接口上应用的 ATM 类中配置的命令；
- 如果配置的命令不管相同与否，都不冲突，则 PVC 上直接配置的命令、PVC 上应用的 ATM 类中配置的命令和在 ATM 接口上应用的 ATM 类中配置的命令取并集生效；
- 如果 PVC 上/PVC 上应用的 ATM 类中/ATM 接口上应用的 ATM 类中配置的命令不同，而且命令冲突，先配置的命令生效，后配置的命令会出现配置冲突提示。
- ATM 类中的配置命令应用到 PVC 上时，无论成功与否，不进行提示。
- 在 PVC 上配置的命令不合法时（如配置参数超出指定范围），进行错误提示。

1.11 配置 VP 监管

VP 监管用来配置某个 VPI 的可承受速率。在应用 VP 监管时，PVC 的参数仍然有效，只有满足 PVC 的参数与 VP 监管的参数时，分组才会被发送或接收。在计算流量时，已经包括了 LLC/SNAP、MUX 和 NLPID 封装头部，但不包括 ATM 信元头。

表1-8 配置 VP 监管

操作	命令	说明
进入系统视图	**system-view**	-
进入 ATM 接口视图	**interface atm** *interface-number*	-
配置 VP 监管的参数	**pvp limit** *vpi output-scr*	必选

1.12 配置 ATM 上承载的应用

ATM PVC 支持同时承载多种协议，但某些类型的封装可能并不支持部分应用方式（即 IPoA、IPoEoA、PPPoA 和 PPPoEoA 中的一种或几种），当出现不能支持的情况时，系统会给出提示。

ATM PVC 封装方式和承载协议的关系及该封装是否支持多协议等信息见下表：

表1-9 各种端口角色具有的端口状态

承载协议/封装方式	aal5snap	aal5mux	aal5nlpid
IPoA	支持	支持	支持
IPoEoA	支持	支持	不支持
PPPoA	支持	支持	不支持
PPPoEoA	支持	支持	不支持



说明

- 受芯片限制，**service vbr-rt/service vbr-nrt** 命令参数有可能配置不成功，请适当调低 mbs 参数。
- aal5snap 可以支持同时封装两种以上协议，aal5nlpid 支持 IPoA 但是不支持 map ip inarp 配置。

1.12.1 配置 IPoA

表1-10 配置 IPoA

操作	命令	说明
进入系统视图	**system-view**	-
进入 ATM 接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC，进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	-

操作	命令	说明
为 PVC 配置 IPoA 映射，使 PVC 承载 IP 协议报文	**map ip** { *ip-address* [*ip-mask*] | **default** | **inarp** [*minutes*] } [**broadcast**]	必选 缺省情况下，不配置任何映射。如果配置了映射，缺省不支持伪广播（**broadcast**） 在配置 InARP 时，必须保证使用的是 aal5snap 封装类型，当采用 aal5mux 和 aal5nlpid 封装时，不能配置 InARP

说明

- **broadcast**：伪广播，可选参数。如果 PVC 上配置了一条具有此属性的映射，则该 PVC 所属接口上的广播报文都要在该 PVC 上发送一份。如果在 ATM PVC 上需要发送广播或者多播报文，如果使能广播或多播类协议，请务必配置此关键字。
- 当使用 P2P 类型 ATM 子接口时，该子接口下的 PVC 必须采用如下配置：**map ip default broadcast**。

1.12.2 配置 IPoEoA

下面的配置任务使得 PVC 能够承载以太网报文。

表1-11 配置 IPoEoA

操作	命令	说明
进入系统视图	**system-view**	-
创建一个虚拟以太网（VE）接口	**interface virtual-ethernet** *interface-number*	必选 IP 地址需要在 VE 接口下进行配置（ATM 接口下配置 IP 地址无效）
退回系统视图	**quit**	-
进入 ATM 接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC，进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	必选
配置 PVC 上的 IPoEoA 映射	**map bridge virtual-ethernet** *interface-number*	必选

说明

- 当同一设备的多个三层虚拟以太网接口通过不同的 PVC 连接到同一个 DHCP 服务器，并且 DHCP 服务器上采用静态绑定方式给三层虚拟以太网接口进行 IP 地址分配时，需要使用 **mac-address** 命令为不同的三层虚拟以太网接口配置不同的 MAC 地址。相关内容请参见“接入分册”中的“逻辑接口配置”。
- 在 IPoEoA 应用时，QoS 应该配置到 PVC 上。

1.12.3 配置 PPPoA

当设备采用 DSL 接口通过拨号方式互连时，路由器作 PPPoA Server 或 Client 使用均可，两侧配置的区别仅在于 PPPoA Server 端为 PPP Server，需要配置地址池，为对端分配 IP 地址；PPPoA Client 端为 PPP Client，需要配置地址协商，接受 Server 端分配的 IP 地址，相关信息请参见“接入分册”中的“PPP 配置”。

下面的配置任务使得 PVC 能够承载 PPP 协议，并为 PVC 配置一个 PPP 协议映射。

表1-12 配置 PPPoA

操作	命令	说明
进入系统视图	**system-view**	-
创建一个虚拟模板（VT）接口	**interface virtual-template** *vt-number*	必选 PPP 认证，IP 地址等均需要在 VT 接口下进行配置（ATM 接口下配置 IP 地址无效）
配置 PPP 的验证方式、IP 地址，如果 PPP Server，需要配置地址池，为对端分配 IP 地址；如果是 PPP Client，需要配置地址协商，接受 Server 端分配的 IP 地址	请参见“接入分册/PPP 配置”中的 PPP 的相关配置	必选
退回系统视图	**quit**	-
进入 ATM 接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC，进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	必选
为 PVC 配置 PPPoA 映射	**map ppp virtual-template** *vt-number*	必选

说明

对于 Virtual-Template 接口，如果配置静态路由，请指定下一跳而不要指定出接口。如果必须指定出接口的话，请保证 Virtual-Template 下绑定的物理接口有效，从而保证报文能够正常传输。

1.12.4 配置 PPPoEoA

PPPoE 协议采用 Client/Server 方式，它将 PPP 报文封装在以太网帧之内，在以太网上提供点对点的连接。下面的配置任务使得 PVC 能够承载 PPPoE 协议，并为 PVC 配置一个 PPPoE 协议地址映射。

表1-13 配置 PPPoEoA

操作	命令	说明
进入系统视图	**system-view**	-
创建一个虚拟模板（VT）接口	**interface virtual-template** *vt-number*	必选 PPP 认证、IP 地址等均需要 VT 接口下进行配置（ATM 接口下配置 IP 地址无效）

操作	命令	说明
配置 PPP 的验证方式、IP 地址，如果是 PPPoE Server，需要配置地址池，为对端分配 IP 地址；如果是 PPPoE Client，需要配置地址协商，接受 Server 端分配的 IP 地址	请参见“接入分册/PPP 配置”中的 PPP 的相关配置	必选
退回系统视图	**quit**	-
创建一个虚拟以太网（VE）接口	**interface virtual-ethernet** *interface-number*	必选
在 VE 接口下配置 PPPoE 的各项参数，根据角色的不同（PPPoE Server 或者 PPPoE Client）配置的内容也不同	请参见“接入分册/PPP 配置”中的 PPPoE 的相关配置	必选
退回系统视图	**quit**	-
进入 ATM 接口视图	**interface atm** { *interface-number* | *interface-number.subnumber* }	-
创建 PVC，进入 PVC 视图	**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }	必选
为 PVC 配置 PPPoEoA 映射	**map bridge virtual-ethernet** *interface-number*	必选 *interface-number* 是上面步骤中已经创建的虚拟以太网接口

说明

- SR6600 路由器仅支持 PPPoEoA 的 Server 端功能。
- 对于 Virtual-Template 接口，如果配置静态路由，请指定下一跳而不要指定出接口。如果必须指定出接口的话，请保证 Virtual-Template 下绑定的物理接口有效，从而保证报文能够正常传输。
- 当同一设备的多个三层虚拟以太网接口通过不同的 PVC 连接到同一个 DHCP 服务器，并且 DHCP 服务器上采用静态绑定方式给三层虚拟以太网接口进行 IP 地址分配时，需要使用 **mac-address** 命令为不同的三层虚拟以太网接口配置不同的 MAC 地址。相关内容请参见“接入分册”中的“逻辑接口配置”。

1.13 ATM 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 ATM 配置后的运行情况，通过查看显示信息验证配置的效果。

在 ATM 接口视图下执行 **oamping interface** 命令，在指定 ATM 接口的特定 PVC 上发送 oam 信元，根据在设定的时间内是否收到应答来判断链路的连接情况。

表1-14 ATM 显示和维护

操作	命令
显示 ATM 接口的相关信息	**display atm interface** [**atm** *interface-number*]
显示 PVC 的相关信息	**display atm pvc-info** [**interface** *interface-type interface-number* [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]]
显示 PVC 映射的相关信息	**display atm map-info** [**interface** *interface-type interface-number* [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]]

操作	命令
显示 PVC-Group 的信息	**display atm pvc-group** [**interface** *interface-type interface-number* [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]]
显示 ATM-Class 的相关信息	**display atm class** [*atm-class-name*]
指定 ATM 接口的特定 PVC 上发送 oam 信元，根据在设定的时间内是否收到应答来判断链路的连接情况	**oamping interface atm** *interface-number* **pvc** { *pvc-name* | *vpi /vci* } [*number timeout*]

1.14 ATM 典型配置举例

说明

以下举例中所涉及到的网络设备，DSLAM（数字用户线接入复接器）及其配置命令序列为 MA 5100 多业务接入设备及其配置环境下的命令序列，ADSL Router 根据实际组网中选用的设备进行配置。您可以参考相应的命令手册获取相关配置命令的全部细节。在实际组网中，不同的网络设备其组网能力和配置命令格式有可能与假定的网络设备不同，如有此种情况，恕不另行通知。

1.14.1 IPoA 典型配置举例

1. 组网需求

如 图 1-3所示，Router A、Router B和Router C接入到ATM网中互相通讯。要求：

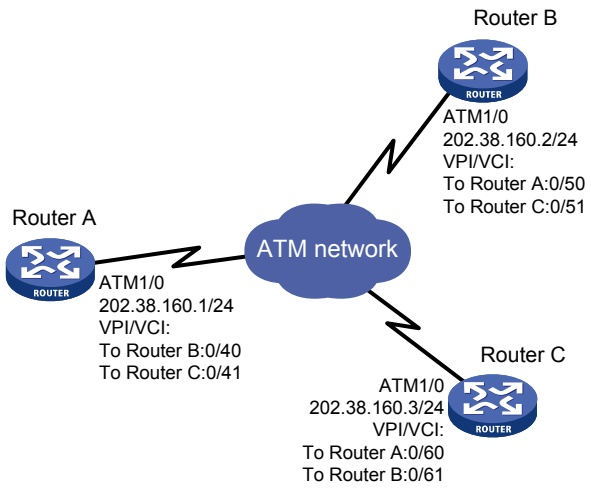
三台路由器 ATM 接口的 IP 地址分别是 202.38.160.1/24、202.38.160.2/24、202.38.160.3/24；

在 ATM 网中，Router A 的 VPI/VCI 是 0/40 和 0/41，分别连接 Router B 和 Router C；Router B 的 VPI/VCI 是 0/50 和 0/51，分别连接 Router A 和 RouterC；Router C 的 VPI/VCI 是 0/60 和 0/61，分别连接 Router A 和 Router B；

三台路由器的 ATM 接口上的所有 PVC 都采用 IPoA 应用方式。

2. 组网图

图1-3 IPoA 配置组网图



3. 配置步骤

(1) 配置 Router A

进入 ATM 接口，并为其配置 IP 地址。

```
<RouterA> system-view
[RouterA] interface atm 1/0
[RouterA-Atm1/0] ip address 202.38.160.1 255.255.255.0
```

创建 PVC，并指定承载 IP 协议。

```
[RouterA-Atm1/0] pvc to_b 0/40
[RouterA-atm-pvc-Atm1/0-0/40-to_b] map ip 202.38.160.2
[RouterA-atm-pvc-Atm1/0-0/40-to_b] quit
[RouterA-Atm1/0] pvc to_c 0/41
[RouterA-atm-pvc-Atm1/0-0/41-to_c] map ip 202.38.160.3
```

(2) 配置 Router B

进入 ATM 接口，并为其配置 IP 地址。

```
<RouterB> system-view
[RouterB] interface atm 1/0
[RouterB-Atm1/0] ip address 202.38.160.2 255.255.255.0
```

创建 PVC，并指定承载 IP 协议。

```
[RouterB-Atm1/0] pvc to_a 0/50
[RouterB-atm-pvc-Atm1/0-0/50-to_a] map ip 202.38.160.1
[RouterB-atm-pvc-Atm1/0-0/50-to_a] quit
[RouterB-Atm1/0] pvc to_c 0/51
[RouterB-atm-pvc-Atm1/0-0/51-to_c] map ip 202.38.160.3
```

(3) 配置 Router C

进入 ATM 接口，并为其配置 IP 地址。

```
<RouterC> system-view
[RouterC] interface atm 1/0
[RouterC-Atm1/0] ip address 202.38.160.3 255.255.255.0
```

创建 PVC，并指定承载 IP 协议。

```
[RouterC-Atm1/0] pvc to_a 0/60
[RouterC-atm-pvc-Atm1/0-0/60-to_a] map ip 202.38.160.1
[RouterC-atm-pvc-Atm1/0-0/60-to_a] quit
[RouterC-Atm1/0] pvc to_b 0/61
[RouterC-atm-pvc-Atm1/0-0/61-to_b] map ip 202.38.160.2
```

1.14.2 IPoEoA 典型配置举例

1. 组网需求

如 图 1-4所示，两个以太网中的多台主机各自通过一台ADSL Router接入ATM网，并通过DSLAM与Router C通讯。要求：

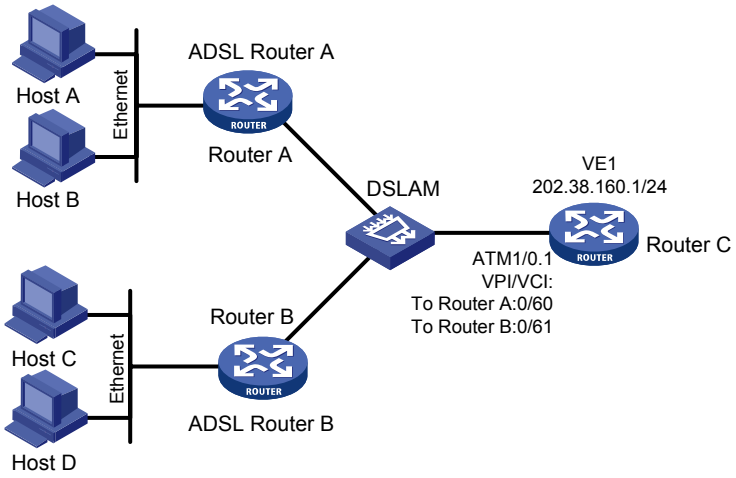
Router C 虚拟以太网接口的 IP 地址为 202.38.160.1；

Router C 连接至 DSLAM 的两条 PVC 的 VPI/VCI 为 0/60、0/61，分别指向 Router A 和 Router B；

Router C 广域网端口和 ADSL Router 的 DSL 接口均采用 IPoEoA 应用方式。

2. 组网图

图1-4 IPoEoA 配置组网图



3. 配置步骤

配置 Router C

创建虚拟以太网接口，并为其配置 IP 地址。

```
<RouterC> system-view
[RouterC] interface virtual-ethernet 1
[RouterC-Virtual-Ethernet1] ip address 202.38.160.1 255.255.255.0
[RouterC-Virtual-Ethernet1] quit
```

创建 PVC，配置 IPoEoA 的承载方式。

```
[RouterC] interface atm 1/0.1
[RouterC-Atm1/0.1] pvc to_adsl_a 0/60
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] map bridge virtual-ethernet 1
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] quit
[RouterC-Atm1/0.1] pvc to_adsl_b 0/61
[RouterC-atm-pvc-Atm1/0.1-0/61-to_adsl_b] map bridge virtual-ethernet 1
```

1.14.3 PPPoA 典型配置举例

1. 组网需求

如 图 1-5所示，两台主机各自通过ADSL Router A/B拨号接入ATM网，并通过DSLAM与Router C通讯。要求：

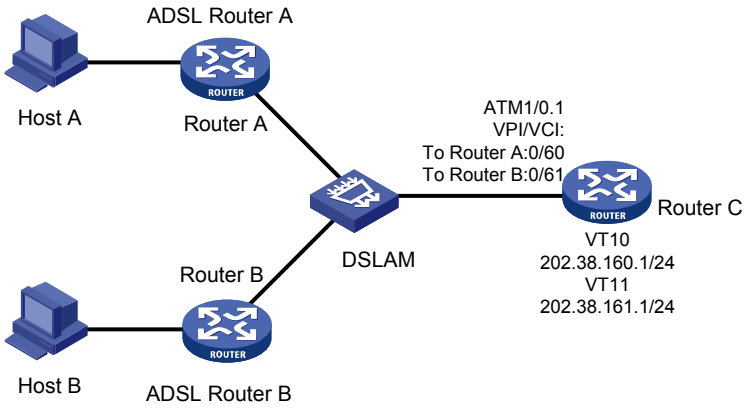
在 Router C 上为多用户创建虚模板，并在虚模板上配置 PPP 的映射；

Router C 连接至 DSLAM 的两条 PVC 的 VPI/VCI 为 0/60、0/61，分别指向 ADSL Router A 和 ADSL Router B；

Router C 广域网端口和 ADSL Router A/B 的 DSL 接口均采用 PPPoA 应用方式，并采用 PAP 对 ADSL Router A/B 进行验证，ADSL Router A/B 的 IP 地址由该路由器提供。

2. 组网图

图1-5 PPPoA 配置组网图



3. 配置步骤

(1) 配置 Router C（PPPoA Server）

为 PPP 验证建立用户，同时本地 IP 地址池。

```
<RouterC> system-view
[RouterC] local-user user1
[RouterC-luser-user1] service-type ppp
[RouterC-luser-user1] password simple pwd1
[RouterC-luser-user1] quit
[RouterC] local-user user2
[RouterC-luser-user2] service-type ppp
[RouterC-luser-user2] password simple pwd2
[RouterC-luser-user2] quit
[RouterC] domain system
[RouterC-isp-system] authentication ppp local
[RouterC-isp-system] ip pool 1 202.38.162.1 202.38.162.100
[RouterC-isp-system] quit
```

创建虚拟模板（VT）接口，配置 PAP 验证和 IP 地址，并为对端从 IP 地址池中分配 IP 地址。

```
[RouterC] interface virtual-template 10
[RouterC-Virtual-Template10] ip address 202.38.160.1 255.255.255.0
[RouterC-Virtual-Template10] ppp authentication-mode pap
[RouterC-Virtual-Template10] remote address pool 1
[RouterC-Virtual-Template10] quit
[RouterC] interface virtual-template 11
[RouterC-Virtual-Template11] ip address 202.38.161.1 255.255.255.0
[RouterC-Virtual-Template11] ppp authentication-mode pap
[RouterC-Virtual-Template11] remote address pool 1
[RouterC-Virtual-Template11] quit
```

创建 PVC，并指定承载 PPP 协议。

```
[RouterC] interface atm 1/0.1
[RouterC-Atm1/0.1] pvc to_adsl_a 0/60
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] map ppp virtual-template 10
```

```
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] quit
[RouterC-Atm1/0.1] pvc to_adsl_b 0/61
[RouterC-atm-pvc-Atm1/0.1-0/61-to_adsl_b] map ppp virtual-template 11
```

(2) 配置 ADSL Router A（PPPoA Client）

创建虚拟模板（VT）接口，配置 PAP 验证及 IP 地址协商。

```
<RouterA> system-view
[RouterA] interface virtual-Template 0
[RouterA-Virtual-Template0] ppp pap local-user user1 password simple pwd1
[RouterA-Virtual-Template0] ip address ppp-negotiate
[RouterA-Virtual-Template0] quit
```

创建 PVC，并指定承载 PPP 协议。

```
[RouterA] interface atm 1/0
[RouterA-Atm1/0] pvc pppoa 0/37
[RouterA-atm-pvc-Atm1/0-0/37-pppoa] map ppp virtual-template 0
```

ADSL Router B 的配置与 ADSL Router A 相似。

 注意

当客户端取消通过协商取得的 IP 地址或是配置了固定的 IP 地址，会造成两端无法互通。此时，需要首先将 ATM 接口 **shutdown**，然后删除服务器端 IP 地址池。

1.14.4 PPPoEoA Server 典型配置举例

1. 组网需求

如 图 1-6所示，两个以太网中的多主机各自通过一台ADSL Router拨号接入ATM网，并通过DSLAM与路由器通讯。要求：

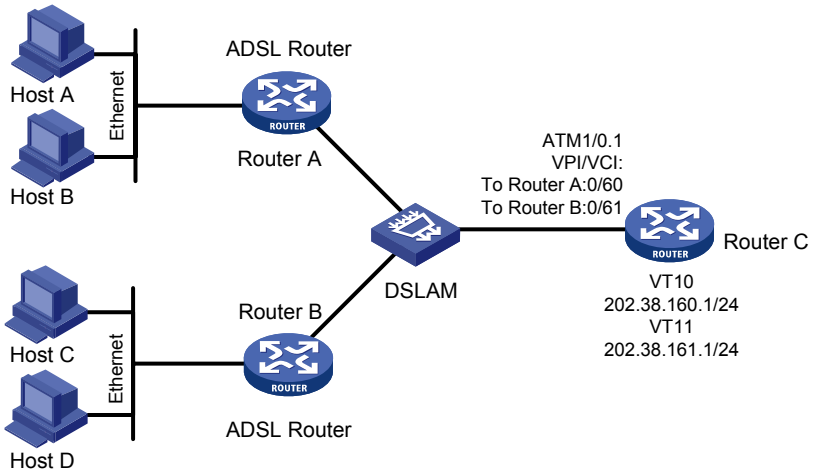
Router C 虚拟模板（VT）接口的 IP 地址分别为 202.38.160.1 和 202.38.161.1；

Router C 连接至 DSLAM 的两条 PVC 的 VPI/VCI 为 0/60、0/61，分别指向 ADSL Router A 和 ADSL Router B；

Router C 广域网端口和 ADSL Router 的 DSL 接口均采用 PPPoEoA 应用方式，两个以太网中的各主机通过事先安装好的 PPPoE 客户端程序与路由器进行 PAP 验证交互，并从路由器处获取 IP 地址。

2. 组网图

图1-6 PPPoEoA Server 配置组网图



3. 配置步骤

(1) 配置 Router C

为 PPP 验证建立用户，同时创建 system 域下的 IP 地址池。

```
<RouterC> system-view
[RouterC] local-user user1
[RouterC-luser-user1] service-type ppp
[RouterC-luser-user1] password simple pwd1
[RouterC-luser-user1] quit
[RouterC] local-user user2
[RouterC-luser-user2] service-type ppp
[RouterC-luser-user2] password simple pwd2
[RouterC-luser-user2] quit
[RouterC] domain system
[RouterC-isp-system] authentication ppp local
[RouterC-isp-system] ip pool 1 202.38.162.1 202.38.162.100
[RouterC-isp-system] quit
```

创建虚拟模板（VT）接口，封装 PPP 协议并配置 PAP 验证参数。

```
[RouterC] interface virtual-template 10
[RouterC-Virtual-Template10] ip address 202.38.160.1 255.255.255.0
[RouterC-Virtual-Template10] ppp authentication-mode pap
[RouterC-Virtual-Template10] quit
[RouterC] interface virtual-template 11
[RouterC-Virtual-Template11] ip address 202.38.161.1 255.255.255.0
[RouterC-Virtual-Template11] ppp authentication-mode pap
[RouterC-Virtual-Template11] quit
```

创建虚拟以太网（VE）接口，并用其封装 PPP 协议。

```
[RouterC] interface virtual-ethernet 1
[RouterC-Virtual-Ethernet1] pppoe-server bind virtual-template 10
[RouterC-Virtual-Ethernet1] quit
```

```
[RouterC] interface virtual-ethernet 2
[RouterC-Virtual-Ethernet2] pppoe-server bind virtual-template 11
[RouterC-Virtual-Ethernet2] quit
```

创建 PVC，并指定承载 PPPoE 协议。

```
[RouterC] interface atm 1/0.1
[RouterC-Atm1/0.1] pvc to_adsl_a 0/60
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] map bridge virtual-ethernet 1
[RouterC-atm-pvc-Atm1/0.1-0/60-to_adsl_a] quit
[RouterC-Atm1/0.1] pvc to_adsl_b 0/61
[RouterC-atm-pvc-Atm1/0.1-0/61-to_adsl_b] map bridge virtual-ethernet 2
```



说明

RADIUS 方案的具体配置方法请参见“安全分册”中的“AAA RADIUS HWTACACS 配置”。

1.15 ATM 故障的诊断与排除

1.15.1 采用 IPoA 时，链路状态为 down

1. 故障现象

采用 IPoA 时，链路状态为 down。

2. 故障排除

检查光纤是否正确连接。

检查本端 IP 地址是否配置。

检查是否 PVC 创建失败或板间通信失败。

1.15.2 采用 PPPoA 时，链路不上报 up

1. 故障现象

采用 PPPoA 时，链路不上报 up。

2. 故障排除

同 1.15.1 2.

1.15.3 ping 不通对方

1. 故障现象

接口物理层和线路协议都处于 up 状态，但是 **ping** 不通对方。

2. 故障排除

采用 IPOA 时，检查协议地址映射配置是否正确。如果是两台路由器接口背对背直连，本端上映射到对端 IP 地址的 PVC 的（VPI，VCI）必须和对端上映射到本端 IP 地址的 PVC 的（VPI，VCI）相同，且两端的 IP 地址必须在同一网段。

如果是两台路由器接口背对背直连，检查是否有一端的接口时钟设置成了 **master**，应至少有一端的时钟设置成 **master**（内部时钟）；如果路由器接入到 ATM 网络中，传输时钟应当设置为 **slave**（线路时钟）。

检查 ATM 端口，看两端的 ATM 端口是否同为多模光纤接口或单模光纤接口，或者两端使用的是多模光纤接口但使用了单模光纤进行连接。（注意：多数情况下，多模光纤口和单模光纤口直接对接是可以互通的，但有时会出现大量丢包和 CRC 错。）

如果两端是 PPPoA，检查两端的 IP 地址（应在同一网段）及验证的配置情况。

如果出现 **ping** 小包能通，**ping** 大包不能通的现象，请检查两端路由器接口的 **mtu** 配置是否一致。

1.15.4 ATM 接口状态为 down

1. 故障现象

ATM 接口状态为 down

2. 故障排除

请检查插接在 ATM 接口的光纤是否接错。应该有两根光纤，分别负责接收和发送，并且不能接反。如果接反，则 ATM 接口状态无法 up。

如果两台路由器之间采取直连方式对接（即所谓的“背靠背”连接），请检查是否两个 ATM 接口的时钟都为 **master**。路由器缺省采用 **slave** 时钟，但如果路由器之间采取直连方式对接，则应该有一方提供内部时钟即 **master**，命令为 **clock master**。

1.15.5 ATM 接口状态为 up，但 PVC 状态为 down

1. 故障现象

ATM 接口状态为 up，但 PVC 状态为 down

2. 故障排除

请检查是否由于启用了 OAM F5 Loopback 信元的发送和重传检测或 OAM CC 检测而导致这种现象。当两台 ATM 设备连接时，连接中的 PVC 在这两台设备上的 VPI/VCI 值对必须一致。如果直接连接的对端没有设置与本端相同（即 VPI/VCI 值对一致）的 PVC，则启用了 OAM F5 Loopback 信元的发送和重传检测或 OAM CC 检测后，本端 PVC 的状态无法转变成 up。

1.15.6 配置完 PPPoA 等应用之后，却无法 ping 通对端

1. 故障现象

PVC 状态为 up，但在配置完 PPPoA 等应用之后，却无法 **ping** 通对端

2. 故障排除

请查看对端是否支持所配置的应用方式。比如本侧采用 PPPoA 应用时，对端也应采用 PPPoA 应用。

如果对端支持所配置的应用方式，请检查两边的 AAL5 封装协议类型是否相同。比如一边使用 SNAP，而另一边却使用 MUX，则无法互通。可以打开 ATM 的报文调试开关，可以从中得到相应的提示信息。

1.15.7 出现大量报文丢弃和 CRC 校验错误，接口状态在 up、down 之间跳变

1. 故障现象

两台路由器采取直连方式对接，可以互相 **ping** 通，但有时会出现大量报文丢弃和 CRC 校验错误，或接口状态在 up、down 之间跳变

2. 故障排除

请检查两端的 ATM 接口，看其是否同为多模光纤接口或同为单模光纤接口。如果接口类型不相同，请予以更换。在多数情况下，多模光纤接口和单模光纤接口之间采取直连方式对接是可以互通的，但有时会出现上述现象。

目 录

1 E-CPOS接口配置 1-1
1.1 E-CPOS接口介绍 1-1
1.1.1 SONET 1-1
1.1.2 SDH 1-1
1.1.3 E-CPOS 1-4
1.2 配置E-CPOS接口 1-5
1.2.1 配置E-CPOS接口 1-5
1.2.2 配置E-CPOS接口工作模式 1-6
1.3 E-CPOS接口显示和维护 1-7
1.4 E-CPOS接口典型配置举例 1-7
1.4.1 E-CPOS典型举例 1-7
1.5 E-CPOS故障诊断与排除 1-8
1.5.1 E-CPOS物理up，通道化出来的POS通道物理up，但链路down 1-8

1 E-CPOS 接口配置

1.1 E-CPOS 接口介绍

1.1.1 SONET

SONET（Synchronous Optical Network，同步光网络）是 ANSI 定义的同步传输体制，是一种全球化的标准传输协议。它采用光传输，传输速率组成一个序列，包括 OC-1（51.84Mbps）、OC-3（155Mbps）、OC-12（622Mbps）和 OC-48（2.5Gbps）。由于是同步信号，因此 SONET 可以方便地实现多路信号的复用。

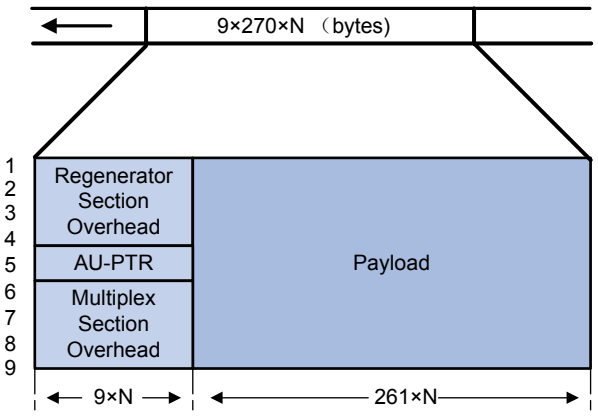
1.1.2 SDH

SDH（Synchronous Digital Hierarchy，同步数字系列）是 CCITT（现在的 ITU-T）定义的，采用同步复用方式和灵活的映射结构，可以从 SDH 信号中直接分插出低速的支路信号，而不需要使用大量的复接/分接设备，从而能够减少信号损耗和设备投资。

1. SDH 的帧结构

为方便地从高速信号中直接分/插低速支路信号，应尽可能使低速支路信号在一帧内均匀地、有规律地分布。ITU-T规定STM-N的帧采用以字节为单位的矩形块状结构，如 图 1-1所示：

图1-1 STM-N 帧结构



STM-N 是 9 行×270×N 列的块状帧结构，此处的 N 与 STM-N 的 N 相一致，取值范围：1，4，16，……，表示此信号由 N 个 STM-1 信号复用而成。

STM-N 的帧结构由 3 部分组成：段开销（SOH，Section Overhead），包括再生段开销（RSOH，Regenerator Section Overhead）和复用段开销（MSOH，Multiplex Section Overhead）；管理单元指针（AU-PTR）；信息净负荷（payload）。AU-PTR 是指示信息净负荷的第一个字节在 STM-N 帧内位置的指针，以便接收端能根据这个指针正确分离信息净负荷。

2. STM 帧的复用过程

STM-1 帧为 9 行 x 270 列的结构，其中前 10 列为开销，后 260 列为净负荷。STM-N 是由 N 个 STM-1 帧间插复用而成。

图1-2 STM-1 帧格式结构图



STM-1 section overhead

A1	A1	A1	A2	A2	A2	J0	Z0	Z0
B1			E1			F1		
D1			D2			D3		
H1	H1	H1	H2	H2	H2	H3	H3	H3
B2	B2	B2	K1			K2		
D4			D5			D6		
D7			D8			D9		
D10			D11			D12		
S1					M1	E2		

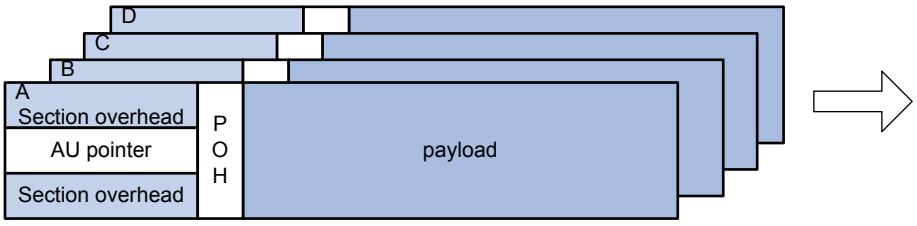
POH

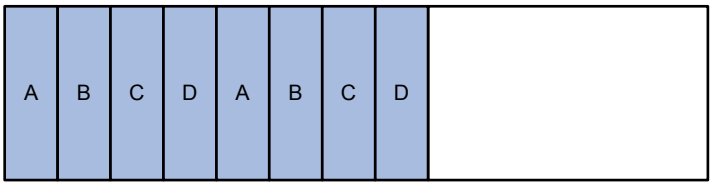
J1
B3
C2
G1
F2
H4
Z3
Z4
N1

- SOH：SDH 段层开销，负责对整个 STM-1 帧进行监控，本身不用于传输用户信息，可以分为 RSOH 和 MSOH。比如，B1，B2 的作用为进行帧的误码率校验。A1，A2 为帧同步字节。
- AU-PTR：管理单元指针，指示 Payload 净负荷在 STM-1 帧中的具体位置。
- POH：通道层开销，用于维护 Payload。比如，POH 中的 C2 字节用于指示净负荷类型；G1 字节用于指示 Payload 是否出现比特误码。
- Payload：Payload 部分有两种情况，如果不再通道化，则 Payload 部分放用户数据；如果继续通道化，则为复用后的低阶通道数据。

以下是 4 个 STM-1 复用成一个 STM-4，同样方法，STM-4 再复用成 STM-16。

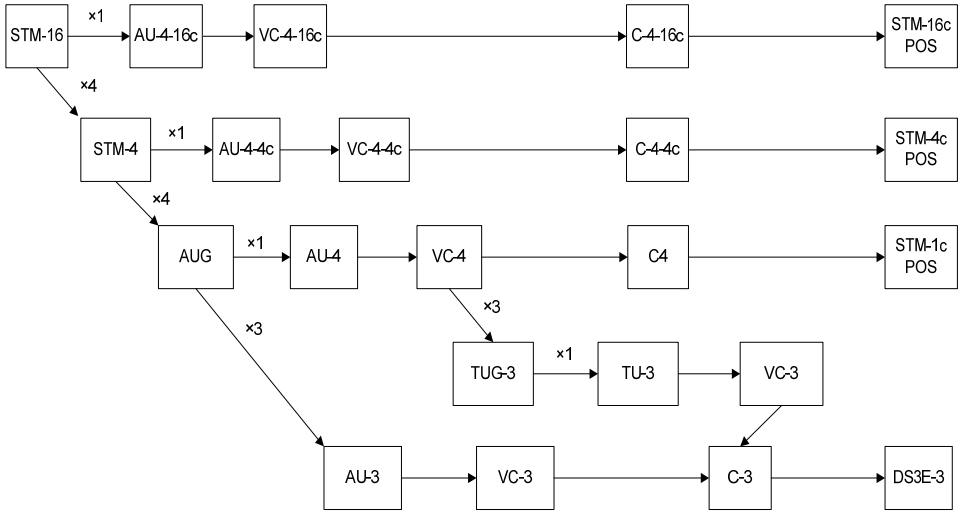
图1-3 STM-1 到 STM-4 的复用过程





对于复用报文。接收端接收报文后。会把报文解复用成 4 个STM-1 报文。对于再生字节字段中A1、A2、j0、Z0、B1、E1、F1、D1、D2、D3 字段为四帧复用。以第一个帧值为代表，其它帧上述字段为无效字节。其它字段分开复用在SDH帧结构下STM-16 的层次复用过程如 图 1-4所示：

图1-4 在 SDH 帧结构下 STM-16 的层次复用过程



3. 速率等级

不同速率的 SDH 帧的表示方法为 STM-N，N=1，4，16，64……，表示此信号通过 N 个 STM-1 信号字节间插复用而成。不同速率的 SONET 帧的表示方法为 STS-N，N=1,3,12,48……表示此信号通过 N 个 STS-1 信号字节间插复用而成。

表1-1 速率等级表

SONET	SDH	速率
STS-1	-	51.840Mbps
STS-3	STM-1	155.52Mbps
STS-12	STM-4	622.080Mbps
STS-48	STM-16	2,488.320Mbps

4. 开销字节

SDH 提供层层细化的监控管理功能。具体来讲，监控分为段监控和通道监控，段层次的监控分为再生段和复用段的监控，通道层次的监控分为高阶通道和低阶通道的监控。这些监控功能是通过不同的开销字节实现的。

说明

SDH 的开销字节非常丰富，本节只对在 E-CPOS 配置过程中用到的几种做简单介绍，如果希望更多了解 SDH 的开销字节，请查阅相关的专业书籍。

- 段开销

段开销（SOH）包括再生段开销（RSOH）和复用段开销（MSOH）。

再生段踪迹字节 J0（Regeneration Section Trace Message）包含在 RSOH 中，该字节被用来重复地发送段接入点标识符（Section Access Point Identifier），以便接收端能据此确认与指定的发送端处于持续连接状态。在同一个运营者的网络内该字节可为任意字符，而在不同两个运营者的网

络边界处要使设备收、发两端的 J0 字节匹配。通过 J0 字节可使运营者提前发现和解决故障，缩短网络恢复时间。

- 通道开销

STM-N 帧的 payload 部分包含对低速支路信号进行监控的开销字节 POH（Path Overhead，通道开销字节）。

段开销负责段层的监控功能，而通道开销负责的是通道层的监控功能。通道开销又分为高阶通道开销（Higher-Order Path Overhead）和低阶通道开销（Lower-Order Path Overhead）。

高阶通道开销对低阶级别的通道进行监测。

通道踪迹字节 J1（Higher-Order VC-N path trace byte）包含在高阶通道开销中，该字节的作用与 J0 字节类似，被用来重复发送高阶通道接入点标识符，使通道接收端能据此确认与指定的发送端处于持续连接（该通道处于持续连接）状态。要求收发两端 J1 字节匹配。

信号标记字节 C2（Path signal label byte）也包含在高阶通道开销中，C2 用来指示 VC 帧的复接结构和信息净负荷的性质，例如通道是否已装载、所载业务种类和它们的映射方式。要求收发两端 C2 字节匹配。

5. 相关术语

- 复用单元：SDH 的基本复用单元包括若干容器（C-n）、虚容器（VC-n）、支路单元（TU-n）、支路单元组（TUG-n）、管理单元（AU-n）和管理单元组（AUG-n），其中 n 为单元等级序号。
- Container（容器）：用来装载各种速率的业务信号的信息结构单元，G.709 定义了 C-11、C-12、C-2、C-3 和 C-4 五种标准容器的规范。
- VC（Virtual Container，虚容器）：用来支持 SDH 的通道层连接的信息结构单元，是 SDH 通道的信息终端。虚容器分为低阶虚容器和高阶虚容器两类，VC-4 和 AU-3 中的 VC-3 都属于高阶虚容器。
- TU（Tributary Unit，支路单元）和支路单元组（TUG，Tributary Unit Group）：TU 是提供低阶通道层和高阶通道层之间适配的信息结构。在高阶 VC 净负荷中，固定地占有规定位置的一个或多个 TU 的集合称为 TUG。
- AU（Administration Unit，管理单元）和管理单元组（AUG，Administration Unit Group）：AU 是提供高阶通道层和复用段层之间适配的信息结构。在 STM-N 的净负荷中，固定地占有规定位置的一个或多个 AU 的集合称为 AUG。
- OC（Optical Carrier，光载波）：为 SONET 光传输定义的一系列物理协议（OC-1，OC-2, 等）。OC 信号级别对应不同速率的 STS 帧。基本数率是 51.84Mbps（OC-1），OC-3 是 155.52Mbps 等。

1.1.3 E-CPOS

当把 SDH 信号看成由低速信号复用而成时，这些低速支路信号就称为通道。CPOS，即通道化的 POS 接口，它充分利用了 SDH 体制的特点，提供对带宽精细划分的能力，可减少组网中对设备低速物理端口的数量要求，增强设备的低速端口汇聚能力，并提高设备的专线接入能力。

E-CPOS（Enhanced Cpos）接口主要用于提高设备对低速接入的汇聚能力。通过对加强 CPOS 接口通道化出高速的 POS 子通道，从而达到减少下游接入设备的目的。

1. 通道化与非通道化

- 通道化是利用 STM-N 中的低速支路信号，通过一根光纤传送相互独立的多路数据，每一路独享带宽、有自己的起点终点和监控策略，称为一个通道。
- 非通道化是使用全部 STM-N 信号，在一根光纤上传送的所有数据属于同一路，具有相同的标识，相同的起点和终点，服从单一的监控策略。

当有多路低速信号需要传送时，通道化方式可以更有效地利用带宽。如果传送速率较高的单路信号，则宜采用非通道化方式。

2. E-CPOS 接口工作模式

E-CPOS 接口有两种工作模式：通道化模式和非通道化模式。

- 通道化模式（channelized mode）：当 E-CPOS 接口工作在此模式下时把一个高阶的 STM-N 帧看作四个低阶的 STM-N 帧时分复用而成。此时一个高阶的 STM-N 帧实际拆分成多个低阶 STM-N 帧处理。
- 非通道化模式（concatenated mode）：当 E-CPOS 接口工作在此模式下时，以一个 STM-N 帧做为处理单元，此时不再对报文进行拆分处理。

1.2 配置 E-CPOS 接口

当通过 SONET/SDH 光接口承载分组数据，并使用低速端口接入时，首先需要对 E-CPOS 接口进行配置。

1.2.1 配置 E-CPOS 接口

表1-2 配置 E-CPOS 接口

操作		命令	说明
进入系统视图		**system-view**	-
配置接口卡的工作模式	集中式设备	**card-mode slot** *slot-number* **e-cpos**	必选 缺省情况下，该接口卡工作在 POS 模式 模式切换后需要重启该接口卡，新配置的模式才会生效 有关接口卡的工作模式的介绍和配置，请参见“系统分册”中的“设备管理”
	分布式设备	**card-mode slot** *slot-number* **subslot** *subslot-number* **e-cpos**	
进入指定 E-CPOS 的接口视图		**controller e-cpos** *interface-number*	-
配置 E-CPOS 帧格式为 SDH 类型		**frame-format { sdh | sonet }**	可选 缺省情况下，CPOS 的帧格式为 **sdh**，即应用在 SDH 模式下
配置 E-CPOS 的时钟模式		**clock { master | slave }**	可选 缺省情况下，CPOS 的时钟模式为从时钟（**slave**）
配置 E-CPOS 的环回方式		**loopback { local | remote }**	可选 缺省情况下，禁止任何形式的环回

操作	命令	说明
配置段开销字节和高阶通道开销字节	**flag j0** { **sdh** *j0-string* | **sonet** *j0-value* }	可选 不同型号的设备支持的取值范围和缺省值不同，请以设备的实际情况为准
配置 E-CPOS 接口的 SD（信号劣化）和 SF（信号失败）的门限值	**threshold** { **sd** | **sf** } *value*	可选 缺省情况下，SD 门限值为 10e-6（即 value 取值为 6），SF 门限值为 10e-3（即 value 取值为 3）
关闭 E-CPOS 物理端口	**shutdown**	可选 缺省情况下，E-CPOS 物理接口处于打开状态

说明

- 当设备的某物理接口闲置，没有连接电缆时，请使用 **shutdown** 命令禁止该接口，以防止由于干扰导致接口异常。
- 使用 **shutdown** 命令关闭接口会导致接口停止工作，请慎用此命令。

1.2.2 配置 E-CPOS 接口工作模式

E-CPOS（Enhanced CPOS）接口和 CPOS 接口基本功能上是一样的，区别在端口的速率和通道化层次上不一样。目前支持的 E-CPOS 接口速率为 STM-16（2488Mbps），支持通道化 622Mbps/155Mbps POS。

E-CPOS 接口下可以创建 4 个 622Mbps 的通道，每个 622Mbps 的通道下又可以创建 4 个 155Mbps 的通道，使用 **using** 命令可以配置通道的工作模式，当将通道配置为非通道化模式时，将生成一个相应速率的 POS 接口。

表1-3 配置 2.5Gbps E-CPOS 接口工作模式（622Mbps POS 子通道）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 E-CPOS 的接口视图	**controller e-cpos** *interface-number*	-
在 2.5Gbps 接口下创建 622M 通道	**oc-12** *oc-12-number*	-
配置 622Mbps 通道工作模式为非通道化模式	**using oc-12c**	可选 缺省情况下，622Mbps 通道工作模式为通道化模式

表1-4 配置 2.5Gbps E-CPOS 接口工作模式（155Mbps POS 子通道）

操作	命令	说明
进入系统视图	**system-view**	-
进入指定 E-CPOS 的接口视图	**controller e-cpos** *interface-number*	-
在 2.5Gbps 接口下创建 622Mbps 通道	**oc-12** *oc-12-number*	-
在 622Mbps 通道下创建指定通道号的 155Mbps 通道	**oc-3** *oc-3-number*	-

操作	命令	说明
配置 155Mbps 通道工作模式为非通道化模式	**using oc-3c**	可选 缺省情况下，155Mbps 通道工作模式为通道化模式

1.3 E-CPOS 接口显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 E-CPOS 接口的运行情况，通过查看显示信息验证配置的效果。在用户视图下，用户可以执行 **reset** 命令清除 E-CPOS 接口的计数器信息。

表1-5 E-CPOS 接口显示和维护

操作	命令
显示 E-CPOS 的所有通道信息	**display controller e-cpos** *interface-number*
显示 E-CPOS 通道出 POS 通道信息	**display interface pos** *interface-number*
清除 E-CPOS 接口的 controller 计数器	**reset counters controller e-cpos** *interface-number*

1.4 E-CPOS 接口典型配置举例

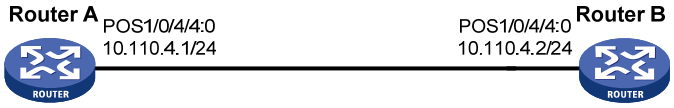
1.4.1 E-CPOS 典型举例

1. 组网需求

使用 2.5G E-CPOS 通过虚拟 155Mbps POS 子通道方式承载业务流量。

2. 组网图

图1-5 E-CPOS 配置组网图



3. 配置步骤

(1) 配置 Router A

配置 E-CPOS 接口的时钟模式。

```
<RouterA> system-view
[RouterA] controller e-cpos 1/0
[RouterA-E-Cpos1/0] clock master
```

创建 155Mbps POS 子通道。

```
[RouterA-E-Cpos1/0] oc-12 4
[RouterA-E-Cpos1/0-oc-12-4] oc-3 4
[RouterA-E-Cpos1/0-oc-12-4-oc-3-4] using oc-3c
```

配置通道化接口 POS1/0/4/4:0。

```
[RouterA-E-Cpos1/0-oc-12-4-oc-3-4] interface pos 1/0/4/4:0
[RouterA-pos1/0/4/4:0] ip address 10.110.4.1 255.255.255.0
```

(2) 配置 Router B

创建 155Mbps POS 子通道。

```
<RouterB> system-view
[RouterB] controller e-cpos 1/0
[RouterB-E-Cpos1/0] oc-12 4
[RouterB-E-Cpos1/0-oc-12-4] oc-3 4
[RouterB-E-Cpos1/0-oc-12-4-oc-3-4] using oc-3c
```

配置通道化接口 POS1/0/4/4:0。

```
[RouterB-E-Cpos1/0-oc-12-4-oc-3-4] interface pos 1/0/4/4:0
[RouterB-pos1/0/4/4:0] ip address 10.110.4.2 255.255.255.0
```

建立连接后，Router A 能够 ping 通 Router B。

1.5 E-CPOS 故障诊断与排除

1.5.1 E-CPOS 物理 up，通道化出来的 POS 通道物理 up，但链路 down

1. 故障现象

E-CPOS 物理 up，E-CPOS 通道化出来的 POS 通道物理 up，但链路 down。

2. 故障排除

- 两端互连的 E-CPOS 接口时钟、扰码等物理参数配置与对端不匹配；
- POS 通道链路层协议配置与对端不匹配；
- 本端或对端 POS 通道的 IP 地址没有配置；
- E-CPOS 通道化出来的 POS 通道和对端通道化出来的 POS 通道带宽不一样；
- E-CPOS 通道化出来的 POS 通道和对端通道化出来的 POS 通道对应的通道号不一样；
- 虚拟 POS 接口上 PPP 验证失败。可能是由于 PPP 验证参数配置不正确，导致 PPP 验证失败。

可以 **display interface pos** *interface-number* 命令确定 POS 通道复用路径及 PPP 链路协商信息

接口有四种状态：

- Pos1/0/1:0 current state: Administratively DOWN, Line protocol current state: DOWN：表示该接口被管理员 down。
- Pos1/0/1:0 current state: DOWN, Line protocol current state: DOWN：表示该接口没有被激活或物理层没有转为 up 状态。
- Pos1/0/1:0 current state: UP, Line protocol current state: UP：表示该接口链路协商即 LCP 协商已通过。
- Pos1/0/1:0 current state: UP, Line protocol current state: DOWN：表示该接口已激活，但链路协商仍没有通过。

目 录

1 VLAN终结配置 1-1
1.1 VLAN终结简介 1-1
1.1.1 VLAN终结概述 1-1
1.1.2 VLAN终结的应用场景 1-1
1.1.3 VLAN终结的分类 1-2
1.2 VLAN终结配置任务简介 1-2
1.3 配置报文的TPID 1-3
1.4 配置VLAN终结支持广播/组播 1-4
1.5 配置Dot1q终结 1-4
1.5.1 配置明确的Dot1q终结 1-5
1.5.2 明确的Dot1q终结配置举例 1-5
1.5.3 配置模糊的Dot1q终结 1-7
1.5.4 模糊的Dot1q终结配置举例 1-7
1.5.5 Dot1q终结支持PPPoE Server配置举例 1-9
1.6 配置QinQ终结 1-9
1.6.1 配置明确的QinQ终结 1-10
1.6.2 明确的QinQ终结配置举例 1-10
1.6.3 配置模糊的QinQ终结 1-11
1.6.4 模糊的QinQ终结配置举例 1-12
1.6.5 QinQ终结支持PPPoE Server配置举例 1-13
1.6.6 QinQ终结支持DHCP中继配置举例 1-13

1 VLAN 终结配置

1.1 VLAN 终结简介

1.1.1 VLAN 终结概述

VLAN 终结的实质就是：

- 对收到的 VLAN 报文，去除 VLAN tag 后进行三层转发或其他处理。转发出去的报文是否带有 tag 由出接口决定。
- 对将发送的报文，将相应的 VLAN 信息添加到报文中再发送。

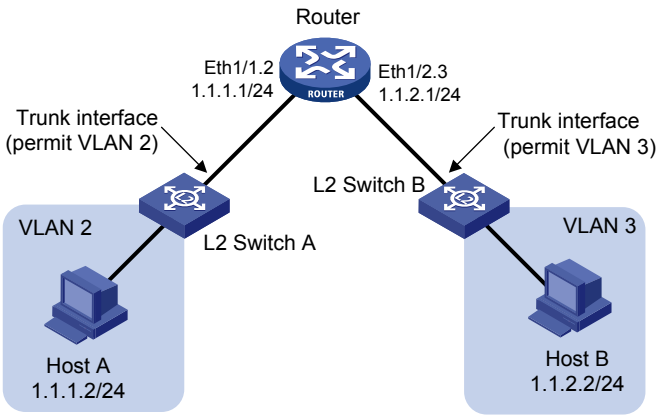
1.1.2 VLAN 终结的应用场景

VLAN 终结主要应用于以下情况：

- 不同 VLAN 间的互通。VLAN 因为能够隔离用户二层报文在网络中得到了广泛的应用。它将一个物理的 LAN 在逻辑上划分成多个广播域（对应多个 VLAN），同一 VLAN 内的主机可以直接互相通信，而不同 VLAN 间的主机二层不能互相通信。要实现不同 VLAN 之间报文的互通必须借用三层路由技术，目前有两种方法：一种是在三层交换机上通过 VLAN 接口来实现；另一种是在路由器上通过三层以太网接口来实现。但传统的三层以太网接口不支持 VLAN 报文，当它收到 VLAN 报文时，会将 VLAN 报文当成是非法报文而丢弃。为了实现 VLAN 间的互通，我们在三层以太网接口上开发了三层以太网子接口，通过在子接口上配置 VLAN 终结功能来实现 VLAN 间的互通。

如 图 1-1所示，Host A属于VLAN 2，Host B属于VLAN 3，将Host A的网关地址指定为 1.1.1.1/24，Host B的网关地址指定为 1.1.2.1/24，就可以通过Router的三层以太网子接口Ethernet1/1.2 和 Ethernet1/2.3 来实现Host A和Host B的三层报文互通了。

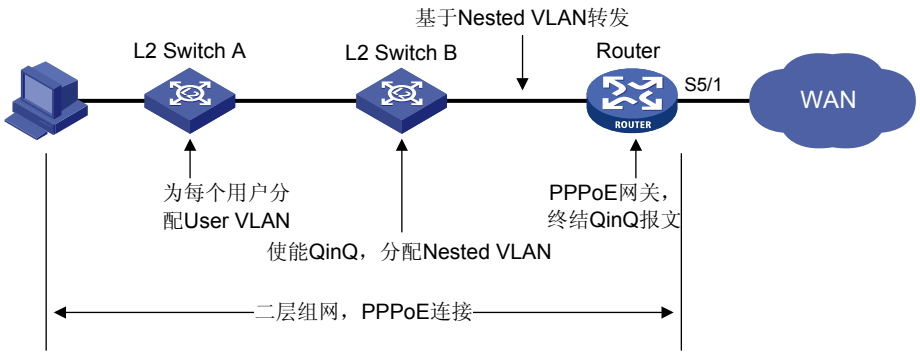
图1-1 VLAN 终结用于不同 VLAN 之间互通



- 用于实现局域网和广域网的互联。局域网内的报文大多数都带有 VLAN tag，但一些广域网协议并不识别 VLAN 报文，比如 ATM、FR 和 PPP 等，这种情况下，如果要局域网的 VLAN 报文转发到广域网，需要在本地记录报文的 VLAN 信息，去掉该信息后再转发。

如 图 1-2所示，用户的报文发送到L2 Switch A时，会被封装上一层VLAN Tag（User VLAN），报文发送到L2 Switch B时，会再被封装上一层VLAN Tag（Nested VLAN），在网络提供者的LAN内，基于Nested VLAN进行转发，如果报文要发往外部WAN，则需要在出口网关（Router）上对QinQ报文进行终结处理，再通过同异步串口Serial 5/1 发送到WAN。

图1-2 VLAN 终结用于 LAN 和 WAN 的互联



1.1.3 VLAN 终结的分类

根据 VLAN 报文携带 tag 的层数可以将 VLAN 报文分为 Dot1q 报文(带有一层 VLAN tag，即 802.1q 报文）和 QinQ 报文（带有两层 VLAN tag）。相应的 VLAN 终结也分为两种：

- Dot1q 终结：用来终结 Dot1q 报文。该方式在收到 Dot1q 报文时，会将报文里的 VLAN tag 去掉之后再继续处理。
- QinQ 终结：用来终结 QinQ 报文。该方式在收到 QinQ 报文时，会将报文里的内、外层 VLAN tag 都去掉之后再继续处理。

说明

主接口本身不能处理 VLAN 报文，在主接口创建子接口后，由子接口来处理。
子接口只能接收 VLAN 报文，发送的也都是带 VLAN tag 的报文。

1.2 VLAN 终结配置任务简介

表1-1 VLAN 终结配置任务简介

配置任务		说明	详细配置
配置报文的 TPID		可选	1.3
配置 Dot1q 终结	配置明确的 Dot1q 终结	请根据设备的支持情况选择一种方式	1.5.1
	配置模糊的 Dot1q 终结		1.5.3
配置 QinQ 终结	配置明确的 QinQ 终结		1.6.1
	配置模糊的 QinQ 终结		1.6.3

1.3 配置报文的 TPID

IEEE 802.1Q协议规定在以太网报文的目的MAC地址和源MAC地址之后封装了 4 个字节的VLAN Tag，用以标识VLAN的相关信息，如 图 1-3所示（这里的帧格式以Ethernet II型封装为例，以太网还支持 802.2 LLC、802.2 SNAP和 802.3 raw封装格式。对于这些封装格式的报文，也会添加VLAN Tag字段）。如果封装一个VLAN Tag则形成Dot1q报文，如果封装两个VLAN Tag则形成QinQ报文，如 图 1-4所示。

图1-3 VLAN Tag 的组成字段

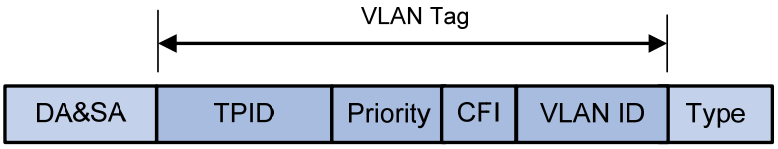
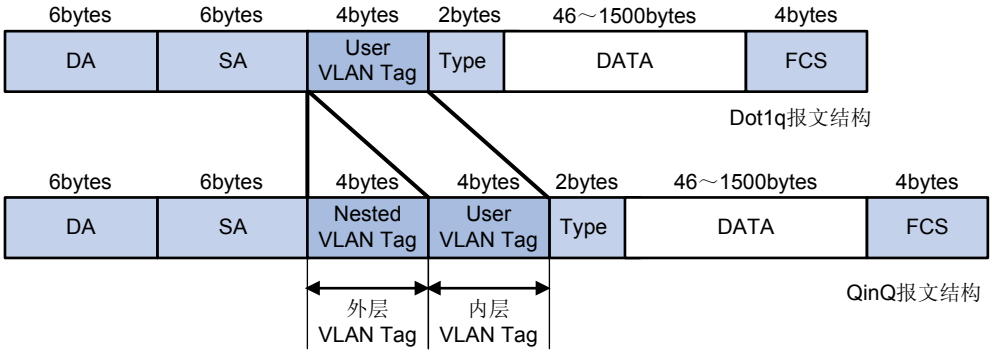


图1-4 Dot1q 报文和 QinQ 报文结构图



VLAN Tag 包含了四个字段，分别是 TPID（Tag Protocol Identifier，标签协议标识符）、Priority、CFI（Canonical Format Indicator，标准格式指示位）和 VLAN ID。其中 TPID 用来判断本太网报文是否带有 VLAN Tag（即是否为 VLAN 报文），长度为 16bit，一般取值为 0x8100，但在实际应用中，各个厂商可以自定义该字段的值。

为了能够识别 TPID 值不是 0x8100 的 VLAN 报文，与其它厂商的设备实现互通，设备提供了 TPID 值配置功能。配置 TPID 值之后，

- 当设备收到第一个 TPID 值为 0x8100 或用户指定值，第二个 TPID 值为 0x8100，系统会当成 QinQ 报文来处理；
- 当设备收到第一个 TPID 值为 0x8100 或用户指定值，第二个 TPID 值为其它值的报文时，系统会当成 Dot1q 报文来处理；
- 当设备收到 TPID 值为其它值（0x8100 和用户指定值以外的值）的报文时，系统会当成不带 VLAN Tag 的以太网报文来处理；
- 设备发送报文时，会将报文的 TPID 字段设置为用户指定的值，当用户没有指定时，就设置为 0x8100。

表1-2 配置 VLAN 报文的 TPID 值

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	必选 在以太网接口视图下配置，会对该接口的所有子接口生效
配置设备接收和发送报文 VLAN Tag 的 TPID 值	**dot1q ethernet-type** *hex-value*	可选 缺省情况下，设备接收和发送的报文 VLAN Tag 的 TPID 值均为 0X8100

说明

- 如果当前接口的状态为 up，配置 TPID 值后，接口会被禁用一次（down），再迅速恢复到 up 状态。
- 不管是否使用 **dot1q ethernet-type** 命令指定别的 TPID 值，当设备收到 TPID 值为 0X8100 的报文时，始终会把它当 VLAN 报文来处理，但发送的报文的 TPID 值会填入命令指定的值。
- 本端的 TPID 值必须和对端的 TPID 值配置一致，否则可能无法互通。

1.4 配置 VLAN 终结支持广播/组播

使能该功能后，模糊终结子接口支持对广播、组播报文的发送。缺省情况下，模糊终结子接口不支持对广播、组播报文的发送。

说明

本特性只对模糊终结子接口有效，对明确终结子接口无效。

表1-3 配置 VLAN 终结支持广播/组播

操作	命令	说明
进入系统视图	*system-view*	-
进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	-
使能模糊子接口的广播/组播发送功能	**vlan-termination broadcast enable**	可选 缺省情况下，不做广播/组播发送

1.5 配置 Dot1q 终结

根据每个子接口所能终结的 VLAN 报文中 VLAN ID 范围的不同，Dot1q 终结又分为：

- 明确的 Dot1q 终结：只能终结一个 VLAN 的报文。能接收的 VLAN 报文的 VLAN ID 必须是指定值，其它 VLAN 报文则不允许通过该子接口；发送报文时，会给报文添加一层 VLAN tag，VLAN ID 字段填上指定值。

- 模糊的 Dot1q 终结：可以终结多个 VLAN 的报文。能接收的 VLAN 报文的 VLAN ID 在指定的范围内，不属于该范围的 VLAN 报文则不允许通过该子接口；发送报文时，会给报文添加一层 VLAN tag，VLAN ID 字段填上指定值：对于 IPv4/MPLS 报文，通过查找 ARP 表项获取相应的 VLAN ID；对于 PPPoE 报文，通过查找 PPPoE 会话表项获取相应的 VLAN ID；对于 DHCP Relay 报文，通过查找 DHCP 会话表项获取相应的 VLAN ID。

1.5.1 配置明确的 Dot1q 终结

表1-4 配置明确的 Dot1q 终结

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	-
使能子接口的 Dot1q 终结，并指定当前子接口能够终结的 Dot1q 报文的 VLAN ID	**vlan-type dot1q vid** *vlan-id*	可选

1.5.2 明确的 Dot1q 终结配置举例

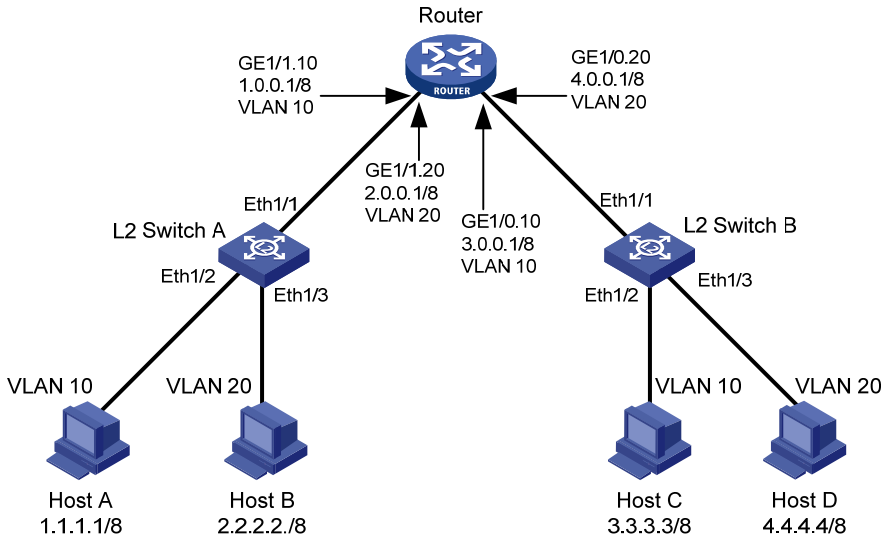
1. 组网需求

如下图所示，Host A、Host B 和 Switch A 相连，Host C、Host D 和 Switch B 相连。Host A 和 Host C 属于 VLAN 10，Host B 和 Host D 属于 VLAN 20，要求实现：

- 路由器子接口 GE1/1.10、GE1/1.20、GE1/0.10、GE1/0.20 的 IP 地址分别为 1.0.0.1/8、2.0.0.1/8、3.0.0.1/8 和 4.0.0.1/8；
- Host A 和 Host B 之间、Host C 和 Host D 之间能够互相通信，即同一交换机、不同 VLAN 之间能够互相通信；
- Host A 和 Host C 之间、Host B 和 Host D 之间能够互相通信，即不同交换机、同一 VLAN 之间能够互相通信；
- Host A 和 Host D 之间、Host B 和 Host C 之间能够互相通信，即不同交换机、不同 VLAN 之间能够互相通信。

2. 组网图

图1-5 配置以太网子接口组网图



3. 配置步骤

说明

必须配置 **vlan-type dot1q vid** 命令，因为以太网子接口只有在关联了 VLAN 后才能被激活，正常的收发报文。

- Host A、Host B、Host C、Host D 的配置

将 Host A 的 IP 地址指定为 1.1.1.1/8，网关地址指定为 1.0.0.1/8。

将 Host B 的 IP 地址指定为 2.2.2.2/8，网关地址指定为 2.0.0.1/8。

将 Host C 的 IP 地址指定为 3.3.3.3/8，网关地址指定为 3.0.0.1/8。

将 Host D 的 IP 地址指定为 4.4.4.4/8，网关地址指定为 4.0.0.1/8。

- L2 Switch A、L2 Switch B 的配置

下面以 L2 Switch A 的配置步骤为例，L2 Switch B 的配置与 L2 Switch A 的配置相同，不再赘述。

```
<L2_SwitchA> system-view
[L2_SwitchA] vlan 10
[L2_SwitchA-vlan10] port ethernet 1/2
[L2_SwitchA-vlan10] quit
[L2_SwitchA] vlan 20
[L2_SwitchA-vlan20] port ethernet 1/3
[L2_SwitchA-vlan20] quit
[L2_SwitchA] interface ethernet 1/1
[L2_SwitchA-Ethernet1/1] port link-type trunk
[L2_SwitchA-Ethernet1/1] port trunk permit vlan 10 20
 Please wait... Done.
```

- 对 Router 的配置

创建并进入以太网子接口 GE1/1.10、GE1/1.20、GE1/0.10、GE1/0.20，为其配置 IP 地址，GE1/1.10 和 GE1/0.10 用来终结 VLAN 10 的报文，GE1/1.20 和 GE1/0.20 用来终结 VLAN 20 的报文。

```
<Router> system-view
[Router] interface gigabitethernet 1/1.10
[Router-GigabitEthernet1/1.10] ip address 1.0.0.1 255.0.0.0
[Router-GigabitEthernet1/1.10] vlan-type dot1q vid 10
[Router-GigabitEthernet1/1.10] quit
[Router] interface gigabitethernet 1/1.20
[Router-GigabitEthernet1/1.20] ip address 2.0.0.1 255.0.0.0
[Router-GigabitEthernet1/1.20] vlan-type dot1q vid 20
[Router-GigabitEthernet1/1.20] quit
[Router] interface gigabitethernet 1/0.10
[Router-GigabitEthernet1/0.10] ip address 3.0.0.1 255.0.0.0
[Router-GigabitEthernet1/0.10] vlan-type dot1q vid 10
[Router-GigabitEthernet1/0.10] quit
[Router] interface gigabitethernet 1/0.20
[Router-GigabitEthernet1/0.20] ip address 4.0.0.1 255.0.0.0
[Router-GigabitEthernet1/0.20] vlan-type dot1q vid 20
[Router-GigabitEthernet1/0.20] quit
```

1.5.3 配置模糊的 Dot1q 终结

表1-5 配置模糊的 Dot1q 终结

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	-
使能子接口的 Dot1q 终结，并指定当前子接口能够终结的 Dot1q 报文的 VLAN 范围	**vlan-type dot1q vid** *vlan-id-list*	必选

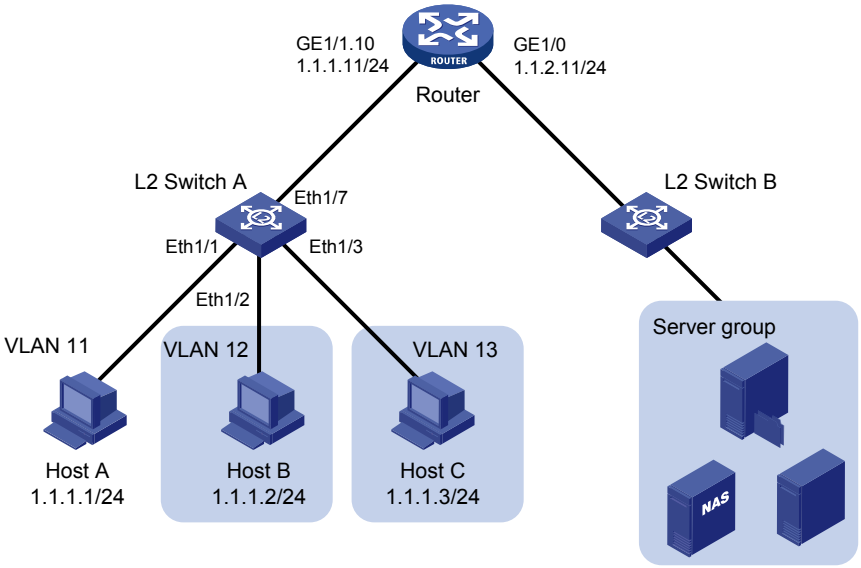
1.5.4 模糊的 Dot1q 终结配置举例

1. 组网需求

如下图所示，Host A、Host B、Host C 和 Switch A 相连，Server group 和 Switch B 相连。Host A、Host B、Host C 分别属于 VLAN 11、VLAN 12、VLAN 13；这些 Host 需要与 Server 群互相通信。

2. 组网图

图1-6 配置模糊的 Dot1q 终结组网图



3. 配置步骤

- Host A、Host B、Host C 的配置

将 Host A 的 IP 地址指定为 1.1.1.1/24， Host B 的 IP 地址为 1.1.1.2/24，Host C 的 IP 地址为 1.1.1.3/24，网关地址均指定为 1.1.1.11/24。

- L2 Switch A 的配置

```
<L2_SwitchA> system-view
[L2_SwitchA] vlan 11
[L2_SwitchA-vlan11] port ethernet 1/1
[L2_SwitchA-vlan11] quit
[L2_SwitchA] vlan 12
[L2_SwitchA-vlan12] port ethernet 1/2
[L2_SwitchA-vlan12] quit
[L2_SwitchA] vlan 13
[L2_SwitchA-vlan13] port ethernet 1/3
[L2_SwitchA-vlan13] quit
[L2_SwitchA] interface ethernet 1/7
[L2_SwitchA-Ethernet1/7] port link-type trunk
[L2_SwitchA-Ethernet1/7] port trunk permit vlan 11 to 13
 Please wait... Done.
```

- 对 Router 的配置

创建以太网子接口（如图所示为 GigabitEthernet1/1.10），为其配置 IP 地址，使能 Dot1q 终结功能，指定所能终结的 VLAN 范围。

```
<Router> system-view
[Router] interface gigabitethernet 1/1.10
[Router-GigabitEthernet1/1.10] ip address 1.1.1.11 255.255.255.0
[Router-GigabitEthernet1/1.10] vlan-type dot1q vid 11 to 13
[Router-GigabitEthernet1/1.10] quit
```

```
[Router] interface gigabitethernet 1/0
[Router-GigabitEthernet1/0] ip address 1.1.2.11 255.255.255.0
```

- L2 Switch B 的配置

L2 Switch B 采用出厂配置即可。

- Server group 的配置

将 Server group 里所有设备的 IP 地址配置在 1.1.2.0/24 网段，网关地址指定为 1.1.2.11/24 即可。

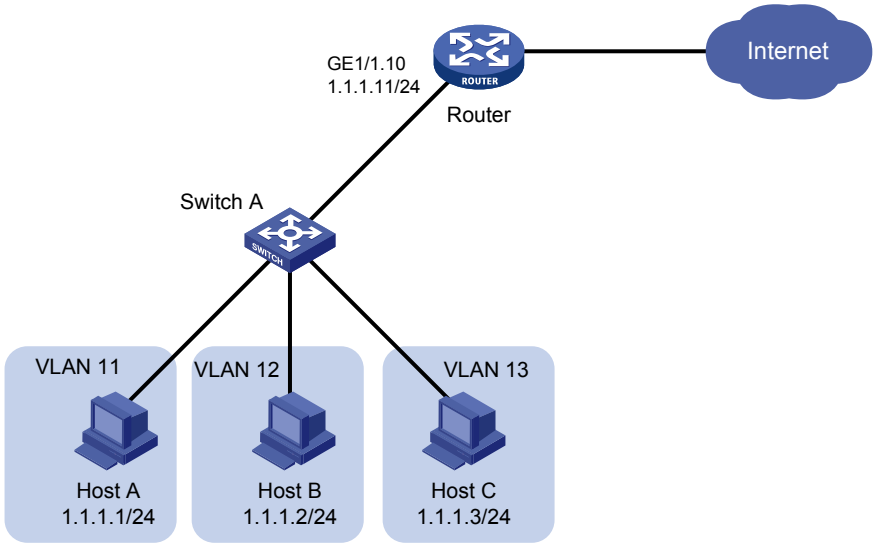
1.5.5 Dot1q 终结支持 PPPoE Server 配置举例

1. 组网需求

如下图所示，Host A、Host B、Host C 和 Switch A 相连，分别属于 VLAN 11、VLAN12、VLAN 13；这些 Host 需要使用拨号方式访问 Internet。

2. 组网图

图1-7 配置 Dot1q 终结支持 PPPoE Server 组网图



3. 配置步骤

关于VLAN的配置步骤请参见 1.5.4 ，另外Router作为网络中的PPPoE server，需要在Ethernet1/1.10 上进行PPPoE相关参数的配置，详情请参见“接入分册”中的“PPP配置”。

1.6 配置 QinQ 终结

根据每个子接口所能终结的 VLAN 报文的 VLAN ID 范围的不同，QinQ 终结又分为：

- 明确的 QinQ 终结：只能终结一种 QinQ 报文。能接收的 QinQ 报文的内外层 VLAN ID 必须是指定值，其它 QinQ 报文则不允许通过该子接口；发送报文时，会给报文添加两层 VLAN tag，内、外层 VLAN ID 字段分别填上指定值。
- 模糊的 QinQ 终结：可以终结多种 QinQ 报文。能接收的 QinQ 报文的外层 VLAN ID 必须是指定值，内层 VLAN ID 只需要在指定的范围内，不属于该范围的 VLAN 报文则不允许通过该子接口；发送报文时，会给报文添加两层 VLAN tag，内、外层 VLAN ID 字段分别填上指定值，内层VLAN ID通过以下方式决定：对于IPv4/MPLS报文，通过查找ARP表项获取相应的VLAN

ID；对于 PPPoE 报文，通过查找 PPPoE 会话表项获取相应的 VLAN ID；对于 DHCP Relay 报文，通过查找 DHCP 会话表项获取相应的 VLAN ID。

1.6.1 配置明确的 QinQ 终结

表1-6 配置明确的 QinQ 终结

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	-
使能子接口的 QinQ 终结功能，并指明子接口可以终结的 VLAN 报文的内外层 VLAN ID 值	**vlan-type dot1q vid** *vlan-id* **second-dot1q** *vlan-id*	必选

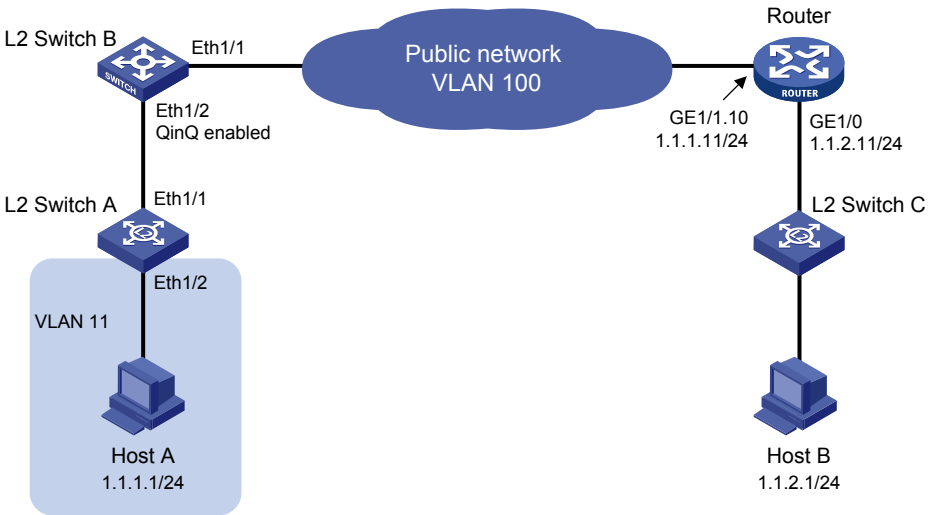
1.6.2 明确的 QinQ 终结配置举例

1. 组网需求

如 图 1-8所示，Host A和L2 Switch A相连，属于VLAN 11；Host B和L2 Switch C相连，L2 Switch C只支持一层VLAN标签；L2 Switch B上使能了QinQ功能，它会给收到的内层VLAN ID为 11 的Dot1q报文外面添加一层VLAN ID为 100 的VLAN Tag后再转发出去。现需要实现Host A与Host B之间能够互相通信。

2. 组网图

图1-8 配置明确的 QinQ 终结组网图



3. 配置步骤

- Host A、Host B 的配置

将 Host A 的 IP 地址指定为 1.1.1.1/24，网关地址指定为 1.1.1.11/24；Host B 的 IP 地址为 1.1.2.1/24，网关地址指定为 1.1.2.11/24。

- L2 Switch A 的配置

```
<L2_SwitchA> system-view
[L2_SwitchA] vlan 11
```

```
[L2_SwitchA-vlan11] port ethernet 1/2
[L2_SwitchA-vlan11] quit
[L2_SwitchA] interface ethernet 1/1
[L2_SwitchA-Ethernet1/1] port link-type trunk
[L2_SwitchA-Ethernet1/1] port trunk permit vlan 11
 Please wait... Done.
```

- L2 Switch B 的配置

```
<L2_SwitchB> system-view
[L2_SwitchB] interface ethernet 1/2
[L2_SwitchB-Ethernet1/2] port link-type trunk
[SwitchB-Ethernet1/2] port trunk permit vlan 11 100
 Please wait... Done.
[L2_SwitchB-Ethernet1/2] qinq vid 100
[L2_SwitchB-Ethernet1/2-vid-100] raw-vlan-id inbound 11
[L2_SwitchB-Ethernet1/2-vid-100] quit
[L2_SwitchB-Ethernet1/2] quit
[L2_SwitchB] interface ethernet 1/1
[L2_SwitchB-Ethernet1/1] port link-type trunk
[L2_SwitchB-Ethernet1/1] port trunk permit vlan 100
 Please wait... Done.
```

- 对 Router 的配置

创建并进入以太网子接口（如图所示为 GigabitEthernet1/1.10），为其配置 IP 地址，使能 QinQ 终结，指定所能终结的 VLAN 范围。

```
<Router> system-view
[Router] interface gigabitethernet 1/1.10
[Router-GigabitEthernet1/1.10] ip address 1.1.1.11 255.255.255.0
[Router-GigabitEthernet1/1.10] vlan-type dot1q vid 100 second-dot1q 11
[Router-GigabitEthernet1/1.10] quit
[Router] interface gigabitethernet 1/0
[Router-GigabitEthernet1/0] ip address 1.1.2.11 255.255.255.0
```

- L2 Switch C 的配置

L2 Switch C 用出厂配置即可。

1.6.3 配置模糊的 QinQ 终结

表1-7 配置模糊的 QinQ 终结

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网子接口视图	**interface** *interface-type interface-number.subnumber*	-
指定当前子接口能够终结的 QinQ 报文	**vlan-type dot1q vid** *vlan-id* **second-dot1q** { **any** | *vlan-id-list* }	必选 本命令的缺省情况与设备的型号有关，请以设备的实际情况为准

说明

- 在同一主接口的不同子接口下配置 QinQ 终结功能时，如果指定的外层 VLAN ID 相同内层 VLAN ID 必须不同；如果指定的外层 VLAN ID 不同，则内层 VLAN ID 可以相同也可以不同。不同主接口下的子接口可以终结的 QinQ 报文可以相同也可以不同。
- 多次执行 **vlan-type dot1q vid** 命令时，如果 vid 参数和当前生效的 vid 参数相同，则新的配置会覆盖旧的配置；如果 vid 参数和当前生效的 vid 参数不同，则必须先删除旧的配置新配置才能执行成功。

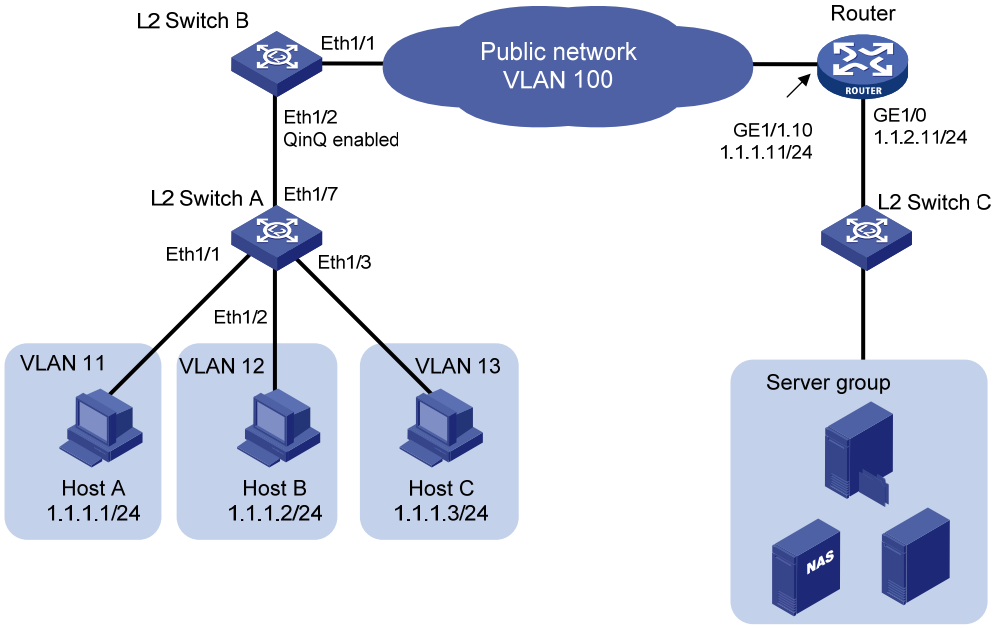
1.6.4 模糊的 QinQ 终结配置举例

1. 组网需求

如 图 1-9所示，Host A、Host B、Host C和Switch A相连，分别属于VLAN 11、VLAN12 和VLAN 13；Server群和Switch C相连；Switch B上使能了QinQ功能。Host A、Host B、Host C需要与Server群互相通信。

2. 组网图

图1-9 配置模糊的 QinQ 终结组网图



3. 配置步骤

- Host A、Host B、Host C 和 L2 Switch A 的配置

请参见 1.5.4 3. 中Host A、Host B、Host C和L2 Switch A的配置。

- L2 Switch B 的配置

请参见 1.6.2 3. 中L2 Switch B的配置。

- 对 Router 的配置

创建并进入以太网子接口（如图所示为 GigabitEthernet1/1.10），为其配置 IP 地址，当他收到内层 VLAN ID 为 11、12 或 13，外层 VLAN ID 为 100 的 QinQ 报文时，对该报文进行终结处理。

```
<Router> system-view
[Router] interface gigabitethernet 1/1.10
[Router-GigabitEthernet1/1.10] ip address 1.1.1.11 255.255.255.0
[Router-GigabitEthernet1/1.10] vlan-type dot1q vid 100 second-dot1q 11 to 13
[Router-GigabitEthernet1/1.10] quit
[Router] interface gigabitethernet 1/0
[Router-GigabitEthernet1/0] ip address 1.1.2.11 255.255.255.0
```

- L2 Switch C 和 Server group 的配置

请参见 1.5.4 3. 中L2 Switch C和Server group的配置。

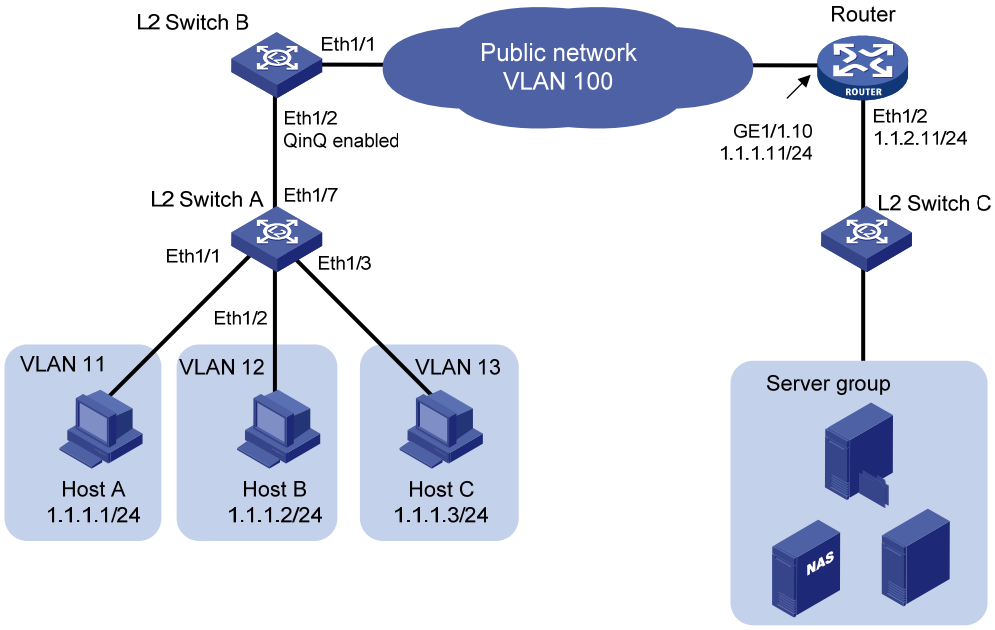
1.6.5 QinQ 终结支持 PPPoE Server 配置举例

1. 组网需求

如下图所示，Host A、Host B、Host C 和 Switch A 相连，分别属于 VLAN 11、VLAN12 和 VLAN 13；Switch B 上使能了 QinQ 功能。Host A、Host B、Host C 需要通过拨号方式访问 Internet。

2. 组网图

图1-10 配置 QinQ 终结支持 PPPOE Server 组网图



3. 配置步骤

关于VLAN和QinQ终结的配置步骤请参见 1.6.4 ，另外Router作为网络中的PPPoE server，需要在GigabitEthernet1/1.10 上进行PPPoE相关参数的配置，详情请参见“接入分册”中的“PPP配置”。

1.6.6 QinQ 终结支持 DHCP 中继配置举例

1. 组网需求

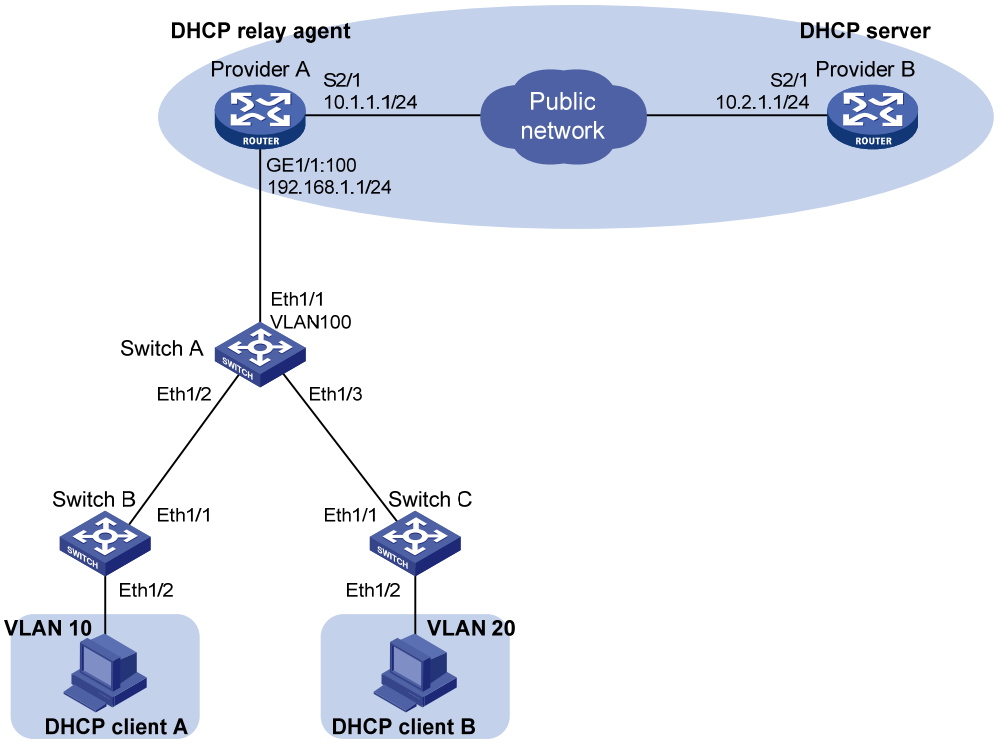
- Provider A、Provider B 作为运营商网络设备。
- DHCP client A 和 DHCP client B 作为用户网络设备。
- Provider A 作为 DHCP 中继，Provider B 作为 DHCP server。
- Provider A 和 Provider B 之间通过三层接口连接。

希望配置完成后达到下列要求：

- DHCP 中继接收到的客户端发送的报文包括两层 Tag，在 DHCP 中继上实现 QinQ 终结，即去掉报文中的两层 Tag 后，通过运营商网络将报文转发给 DHCP server。
- DHCP client A 和 DHCP client B 可以经过运营商网络从 DHCP server 申请到 IP 地址和网络配置参数。

2. 组网图

图1-11 QinQ 终结支持 DHCP 中继组网图



3. 配置步骤

- DHCP 中继 Provider A 的配置

使能 DHCP 服务。

```
<ProviderA> system-view
[ProviderA] dhcp enable
```

创建 DHCP 服务器组。

```
[ProviderA] dhcp relay server-group 1 ip 10.2.1.1
```

创建路由子接口。

```
[ProviderA] interface gigabitethernet 1/1.100
```

配置接口允许终结的内层 VLAN 范围为 10 和 20。

```
[ProviderA-GigabitEthernet1/1.100] vlan-type dot1q vid 100 second-dot1q 10 20
```

配置接口使能 DHCP 中继，并选择服务器组。

```
[ProviderA-GigabitEthernet1/1.100] dhcp select relay
[ProviderA-GigabitEthernet1/1.100] dhcp relay server-select 1
```

配置路由子接口的 IP 地址。

```
[ProviderA-GigabitEthernet1/1.100] ip address 192.168.1.1 24
[ProviderA-GigabitEthernet1/1.100] quit
```

配置与 DHCP 服务器连接接口的 IP 地址。

```
[ProviderA] interface serial 2/1
[ProviderA-Serial2/1] ip address 10.1.1.1 24
```

- DHCP 服务器 Provider B 的配置

配置 DHCP 服务器的 IP 地址。

```
<ProviderB> system-view
[ProviderB] interface serial 2/1
[ProviderB-Serial2/1] ip address 10.2.1.1 24
[ProviderB-Serial2/1] quit
```

使能 DHCP 服务。

```
[ProviderB] dhcp enable
```

配置 DHCP 服务器地址池。

```
[ProviderB] dhcp server ip-pool 1
[ProviderB-dhcp-pool-1] network 192.168.1.0 24
[ProviderB-dhcp-pool-1] quit
```

配置到 DHCP 中继路由子接口的静态路由。

```
[ProviderB] ip route static 192.168.1.1 24 10.1.1.1
```

- Switch A 的配置

配置 QinQ 上行口。

```
<SwitchA> system-view
[SwitchA] interface ethernet 1/1
[SwitchA-Ethernet1/1] qinq enable
[SwitchA-Ethernet1/1] port link-type trunk
```

配置 Trunk 口，允许 VLAN 100 的报文通过。

```
[SwitchA-Ethernet1/1] port trunk permit vlan 100
[SwitchA-Ethernet1/1] quit
```

配置 QinQ 下行口 Ethernet1/2。

```
[SwitchA] interface ethernet 1/2
[SwitchA-Ethernet1/2] qinq enable
[SwitchA-Ethernet1/2] quit
```

配置 QinQ 下行口 Ethernet1/3。

```
[SwitchA] interface Ethernet 1/3
[SwitchA-Ethernet1/3] qinq enable
[SwitchA-Ethernet1/3] quit
```

下行端口 Ethernet1/2 和 Ethernet1/3 加入 VLAN 100。

```
[SwitchA] vlan 100
[SwitchA-vlan100] port ethernet 1/2
[SwitchA-vlan100] port ethernet 1/3
```

- Switch B 的配置

端口 Ethernet1/2 加入 VLAN 20。

```
<SwitchB> system-view
[SwitchB] vlan 20
[SwitchB-vlan20] port ethernet 1/2
[SwitchB-vlan20] quit
```

配置端口 Ethernet1/1 为 Trunk 口，且允许 VLAN 20 的报文通过。

```
[SwitchB] interface ethernet 1/1
[SwitchB-Ethernet1/1] port link-type trunk
[SwitchB-Ethernet1/1] port trunk permit vlan 20
```

- Switch C 的配置

端口 Ethernet1/2 加入 VLAN 10。

```
<SwitchC> system-view
[SwitchC] vlan 10
[SwitchC-vlan10] port ethernet 1/2
[SwitchC-vlan10] quit
```

配置端口 Ethernet1/1 为 Trunk 口，且允许 VLAN 10 的报文通过。

```
[SwitchC] interface ethernet 1/1
[SwitchC-Ethernet1/1] port link-type trunk
[SwitchC-Ethernet1/1] port trunk permit vlan 10
```

目 录

1 端口镜像配置 1-1
1.1 端口镜像简介 1-1
1.1.1 端口镜像基本概念 1-1
1.1.2 端口镜像的分类 1-1
1.1.3 端口镜像的实现方式 1-1
1.1.4 端口镜像支持的其他功能 1-2
1.2 配置本地端口镜像 1-2
1.3 端口镜像显示和维护 1-3
1.4 镜像典型配置举例 1-3
1.4.1 本地镜像配置举例 1-3

1 端口镜像配置

说明

本章所涉及的特性仅在 SR6602 上支持。

1.1 端口镜像简介

端口镜像是将指定端口（源端口）的报文复制一份到目的端口，目的端口会与数据监测设备相连，管理员利用这些数据监测设备来分析复制到目的端口的报文，进行网络监控和故障排除。

1.1.1 端口镜像基本概念

为了更好地理解后面的内容，首先介绍一下端口镜像中涉及的基本概念。

1. 源端口

源端口是被监控的端口，用户需要对通过该端口的报文进行监控和分析。

2. 目的端口

目的端口也可称为监控端口，该端口将接收到的报文转发到数据监测设备，以便对报文进行监控和分析。

目的端口接收到的报文包括复制自源端口的报文、以及来自其它端口的正常转发报文。为了保证数据监测设备只对源端口的报文进行分析，建议目的端口仅用于端口镜像，不用做其它用途。

3. 镜像的方向

端口镜像的方向分为三种：

- 入方向：仅对源端口接收的报文进行镜像。
- 出方向：仅对源端口发送的报文进行镜像。
- 双向：对源端口接收和发送的报文都进行镜像。

1.1.2 端口镜像的分类

端口镜像分为两种镜像方式：

- 本地端口镜像：是指将设备的一个或多个源端口的报文复制到本设备的一个目的端口，用于报文的监控和分析。其中，源端口和目的端口必须在同一台设备上。
- 远程端口镜像：除了可以实现本地端口镜像的功能外，它还突破了源端口和目的端口必须在同一台设备上的限制，使源端口和目的端口间可以跨越多个网络设备。

目前 SR6602 支持本地端口镜像方式。

1.1.3 端口镜像的实现方式

端口镜像通过镜像组的方式实现。

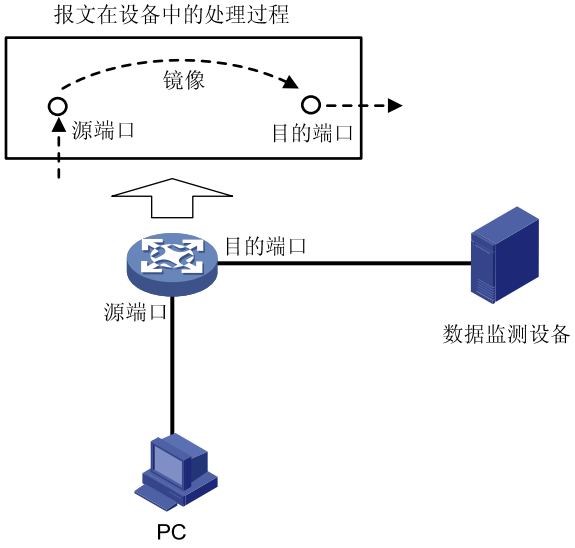
1. 本地端口镜像

本地端口镜像可以对所有报文进行镜像。

本地端口镜像通过本地镜像组的方式实现。源端口和目的端口在同一个本地镜像组中，设备将源端口的报文复制一份并转发到目的端口。

如 图 1-1所示，源端口的报文被镜像到目的端口，这样，接在目的端口上的数据监测设备就可以对源端口的报文进行监控和分析。

图1-1 本地端口镜像示意图



1.1.4 端口镜像支持的其他功能

镜像组支持一个目的端口监视多个源端口的功能。在某些配置情况下，目的端口会收到同一个报文的多个复制报文。例如，目的端口 P1 同时监控源端口 P2 和 P3 接收和发送的所有报文（P2 和 P3 在同一台设备上），如果一个报文从 P2 进入设备又从 P3 发送出去，那么这个报文将被复制两次送到目的端口 P1。

1.2 配置本地端口镜像

配置本地端口镜像时，用户首先要创建一个本地镜像组，然后为本地镜像组配置源端口和目的端口。

表1-1 配置本地端口镜像

操作	命令	说明
进入系统视图	**system-view**	-
创建本地镜像组	**mirroring-group** *groupid* **local**	必选

为镜像组配置源端口	在系统视图下配置源端口	**mirroring-group** *groupid* **mirroring-port** *mirroring-port-list* { **both** | **inbound** | **outbound** }	二者必选其一 用户可以在系统视图下同时配置多个源端口，也可以在具体的接口视图下配置源端口，两种视图下的配置效果相同
	在接口视图下配置源端口	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **mirroring-port** { **both** | **inbound** | **outbound** }	
		quit	
为镜像组配置目的端口	在系统视图下配置目的端口	**mirroring-group** *groupid* **monitor-port** *monitor-port-id*	二者必选其一 两种视图下的配置效果相同
	在接口视图下配置目的端口	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **monitor-port**	

说明

- 本地镜像组需要配置源端口、目的端口才能生效。
- 一个镜像组中可以配置多个源端口，但只能配置一个目的端口。
- 一个端口只能在一个镜像组中被配置。

1.3 端口镜像显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后镜像组的运行情况，通过查看显示信息验证配置的效果。

表1-2 端口镜像显示和维护

操作	命令
显示端口镜像组的配置信息	**display mirroring-group** { *groupid* | **all** | **local** }

1.4 镜像典型配置举例

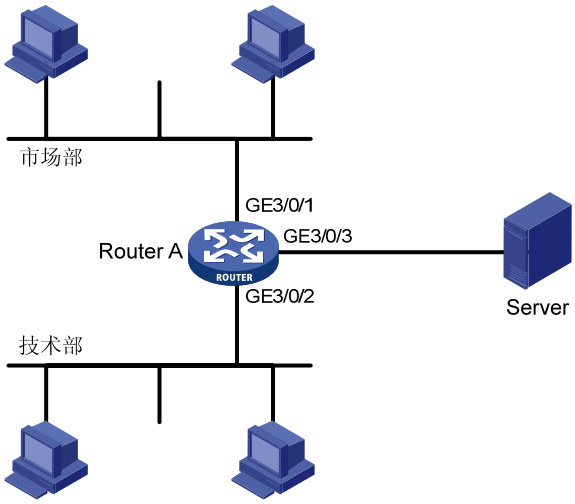
1.4.1 本地镜像配置举例

1. 组网需求

- Router A 通过端口 GigabitEthernet3/0/1 和 GigabitEthernet3/0/2 分别连接市场部和技术部，并通过端口 GigabitEthernet3/0/3 连接 Server。
- 通过配置源端口方式的本地镜像，使 Server 可以监控所有进、出市场部和技术部的报文。

2. 组网图

图1-2 本地镜像配置组网图



3. 配置步骤

(1) 配置本地镜像组

创建本地镜像组 1。

```
<RouterA> system-view
[RouterA] mirroring-group 1 local
```

配置本地镜像组 1 的源端口为 GigabitEthernet3/0/1 和 GigabitEthernet3/0/2，目的端口为 GigabitEthernet3/0/3。

```
[RouterA] mirroring-group 1 mirroring-port gigabitethernet 3/0/1 gigabitethernet 3/0/2 both
[RouterA] mirroring-group 1 monitor-port gigabitethernet 3/0/3
```

(2) 检验配置效果

显示所有镜像组的配置信息。

```
[RouterA] display mirroring-group all
mirroring-group 1:
    type: local
    status: active
    mirroring port:
        GigabitEthernet3/0/1  both
        GigabitEthernet3/0/2  both
    mirroring vlan:
    mirroring CPU:
    monitor port: GigabitEthernet3/0/3
```

配置完成后，用户可以通过 Server 监控所有进、出市场部和技术部的报文。

目 录

1 ARP配置 1-1
1.1 ARP简介 1-1
1.1.1 ARP作用 1-1
1.1.2 ARP报文结构 1-1
1.1.3 ARP地址解析过程 1-1
1.1.4 ARP表 1-2
1.2 配置ARP 1-3
1.2.1 手工添加静态ARP表项 1-3
1.2.2 配置接口学习动态ARP表项的最大数目 1-3
1.2.3 配置动态ARP表项的老化时间 1-3
1.2.4 使能ARP表项的检查功能 1-4
1.2.5 使能支持自然网段的ARP请求 1-4
1.2.6 ARP配置举例 1-4
1.3 配置免费ARP 1-5
1.3.1 免费ARP简介 1-5
1.3.2 配置免费ARP 1-6
1.4 配置ARP源抑制功能 1-6
1.4.1 ARP源抑制功能简介 1-6
1.4.2 配置ARP源抑制 1-6
1.4.3 ARP源抑制显示和维护 1-7
1.5 配置ARP防IP报文攻击功能 1-7
1.5.1 ARP防IP报文攻击功能介绍 1-7
1.5.2 启用ARP防IP报文攻击功能 1-7
1.6 ARP显示和维护 1-7
2 代理ARP配置 2-1
2.1 代理ARP简介 2-1
2.1.1 代理ARP 2-1
2.2 配置代理ARP功能 2-2
2.3 代理ARP显示和维护 2-2
2.4 代理ARP典型配置举例 2-2
2.4.1 代理ARP配置举例 2-2
3 ARP攻击防御配置 3-1
3.1 配置ARP主动确认功能 3-1
3.1.1 ARP主动确认功能简介 3-1
3.1.2 配置ARP主动确认功能 3-1
3.2 配置ARP报文源MAC一致性检查功能 3-1
3.2.1 简介 3-1
3.2.2 配置ARP报文源MAC一致性检查功能 3-2
3.3 配置ARP报文限速功能 3-2
3.3.1 简介 3-2

3.3.2 配置ARP报文限速功能 3-2
3.4 配置授权ARP功能 3-2
3.4.1 授权ARP功能简介 3-2
3.4.2 配置授权ARP功能 3-3
3.4.3 授权ARP功能在DHCP服务器上的典型配置举例 3-3
3.4.4 授权ARP功能在DHCP中继上的典型配置举例 3-4

1 ARP 配置

1.1 ARP 简介

1.1.1 ARP 作用

ARP（Address Resolution Protocol，地址解析协议）是将 IP 地址解析为以太网 MAC 地址（或称物理地址）的协议。

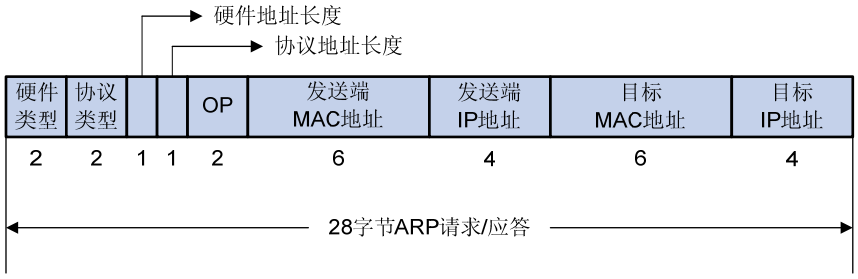
在局域网中，当主机或其它网络设备有数据要发送给另一个主机或设备时，它必须知道对方的网络层地址（即 IP 地址）。但是仅仅有 IP 地址是不够的，因为 IP 数据报文必须封装成帧才能通过物理网络发送，因此发送站还必须有接收站的物理地址，所以需要一个从 IP 地址到物理地址的映射。APR 就是实现这个功能的协议。

说明

本章中除特殊说明，以太网 MAC 地址均以 48bit 的以太网 MAC 地址为例。

1.1.2 ARP 报文结构

图1-1 ARP 报文结构



- 硬件类型：表示硬件地址的类型。它的值为 1 表示以太网地址；
- 协议类型：表示要映射的协议地址类型。它的值为 0x0800 即表示 IP 地址；
- 硬件地址长度和协议地址长度分别指出硬件地址和协议地址的长度，以字节为单位。对于以太网上 IP 地址的 ARP 请求或应答来说，它们的值分别为 6 和 4；
- 操作类型（OP）：1 表示 ARP 请求，2 表示 ARP 应答；
- 发送端 MAC 地址：发送方设备的硬件地址；
- 发送端 IP 地址：发送方设备的 IP 地址；
- 目标 MAC 地址：接收方设备的硬件地址。
- 目标 IP 地址：接收方设备的 IP 地址。

1.1.3 ARP 地址解析过程

假设主机A和B在同一个网段，主机A要向主机B发送信息。如 图 1-2所示，具体的地址解析过程如下：

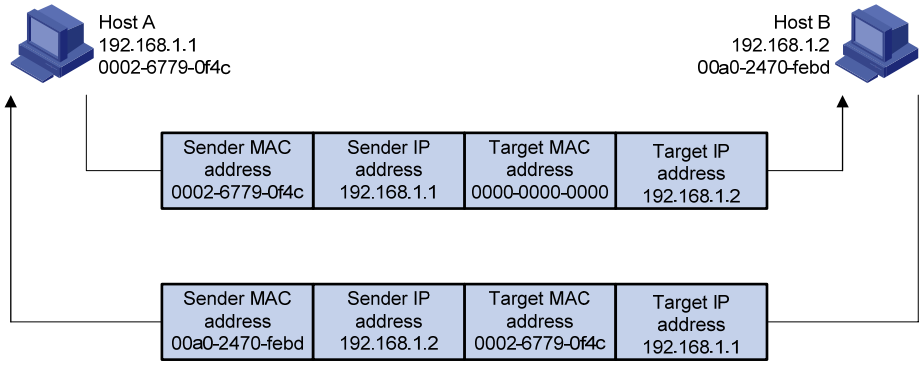
(1) 主机 A 首先查看自己的 ARP 表，确定其中是否包含有主机 B 对应的 ARP 表项。如果找到了对应的 MAC 地址，则主机 A 直接利用 ARP 表中的 MAC 地址，对 IP 数据包进行帧封装，并将数据包发送给主机 B。

(2) 如果主机 A 在 ARP 表中找不到对应的 MAC 地址，则将缓存该数据报文，然后以广播方式发送一个 ARP 请求报文。ARP 请求报文中的发送端 IP 地址和发送端 MAC 地址为主机 A 的 IP 地址和 MAC 地址，目标 IP 地址和目标 MAC 地址为主机 B 的 IP 地址和全 0 的 MAC 地址。由于 ARP 请求报文以广播方式发送，该网段上的所有主机都可以接收到该请求，但只有被请求的主机（即主机 B）会对该请求进行处理。

(3) 主机 B 比较自己的 IP 地址和 ARP 请求报文中的目标 IP 地址，当两者相同时进行如下处理：将 ARP 请求报文中的发送端（即主机 A）的 IP 地址和 MAC 地址存入自己的 ARP 表中。之后以单播方式发送 ARP 响应报文给主机 A，其中包含了自己的 MAC 地址。

(4) 主机 A 收到 ARP 响应报文后，将主机 B 的 MAC 地址加入到自己的 ARP 表中以用于后续报文的转发，同时将 IP 数据包进行封装后发送出去。

图1-2 ARP 地址解析过程



当主机 A 和主机 B 不在同一网段时，主机 A 就会先向网关发出 ARP 请求，ARP 请求报文中的目标 IP 地址为网关的 IP 地址。当主机 A 从收到的响应报文中获得网关的 MAC 地址后，将报文封装并发给网关。如果网关没有主机 B 的 ARP 表项，网关会广播 ARP 请求，目标 IP 地址为主机 B 的 IP 地址，当网关从收到的响应报文中获得主机 B 的 MAC 地址后，就可以将报文发给主机 B；如果网关已经有主机 B 的 ARP 表项，网关直接把报文发给主机 B。

1.1.4 ARP 表

设备通过 ARP 解析到目的 MAC 地址后，将会在自己的 ARP 表中增加 IP 地址到 MAC 地址的映射表项，以用于后续到同一目的地报文的转发。

ARP 表项分为动态 ARP 表项和静态 ARP 表项。

1. 动态 ARP 表项

动态 ARP 表项由 ARP 协议通过 ARP 报文自动生成和维护，可以被老化，可以被新的 ARP 报文更新，可以被静态 ARP 表项覆盖。当到达老化时间、接口 down 时会删除相应的动态 ARP 表项。

2. 静态 ARP 表项

静态 ARP 表项通过手工配置和维护，不会被老化，不会被动态 ARP 表项覆盖。配置静态 ARP 表项可以增加通信的安全性。静态 ARP 表项可以限制和指定 IP 地址的设备通信时只使用指定的 MAC

地址，此时攻击报文无法修改此表项的 IP 地址和 MAC 地址的映射关系，从而保护了本设备和指定设备间的正常通信。目前只能配置短静态 ARP 表项。

- 在配置短静态 ARP 表项时，只需要配置 IP 地址和 MAC 地址项。如果出接口是三层以太网接口，短静态 ARP 表项可以直接用于报文转发；如果出接口是 VLAN 虚接口，短静态 ARP 表项不能直接用于报文转发，当要发送 IP 数据包时，先发送 ARP 请求报文，如果收到的响应报文中的源 IP 地址和源 MAC 地址与所配置的 IP 地址和 MAC 地址相同，则将接收 ARP 响应报文的接口加入该静态 ARP 表项中，之后就可以用于 IP 数据包的转发。


说明

一般情况下，ARP 动态执行并自动寻求 IP 地址到以太网 MAC 地址的解析，无需管理员的介入。

1.2 配置 ARP

1.2.1 手工添加静态 ARP 表项

表1-1 手工添加静态 ARP 表项

操作	命令	说明
进入系统视图	**system-view**	-
手工添加短静态 ARP 表项	**arp static** *ip-address mac-address* [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，没有配置任何短静态 ARP 表项

1.2.2 配置接口学习动态 ARP 表项的最大数目

表1-2 配置接口学习动态 ARP 表项的最大数目

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口允许学习动态 ARP 表项的最大个数	**arp max-learning-num** *number*	可选 缺省情况下，接口允许学习动态 ARP 表项的最大个数是 1024

1.2.3 配置动态 ARP 表项的老化时间

为适应网络的变化，ARP 表需要不断更新。ARP 表中的动态 ARP 表项并非永远有效，每一条记录都有一个生存周期，到达生存周期仍得不到刷新的记录将被从 ARP 表中删除，这个生存周期被称作老化时间。如果在到达生存周期前纪录被刷新，则重新计算老化时间。用户可以根据网络实际情况调整老化时间。

表1-3 配置动态 ARP 的老化时间

操作	命令	说明
进入系统视图	**system-view**	-
配置动态 ARP 的老化时间	**arp timer aging** *aging-time*	可选 缺省情况下，动态 ARP 的老化时间为 20 分钟

1.2.4 使能 ARP 表项的检查功能

ARP 表项检查功能可以控制设备是否学习组播 MAC 地址。使能 ARP 表项的检查功能后，若设备接收到的 ARP 报文中的源 MAC 地址为组播 MAC，则不进行动态 ARP 表项的学习；且设备上不能配置 MAC 地址为组播 MAC 的静态 ARP 表项，否则会有错误提示。关闭 ARP 表项的检查功能后，可以对源 MAC 地址为组播 MAC 的 ARP 表项进行学习，且可以配置 MAC 地址为组播 MAC 的静态 ARP 表项。

表1-4 使能 ARP 表项的检查功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 ARP 表项的检查功能	**arp check enable**	可选 缺省情况下，使能 ARP 表项的检查功能，即不学习组播 MAC 地址

1.2.5 使能支持自然网段的 ARP 请求

学习 ARP 表项时，在发现 ARP 报文的源 IP 地址和入接口 IP 地址不在同一网段后，使用自然网段进行判断。

假设 GigabitEthernet0/1 接口的 IP 地址为 10.10.10.5/24，收到一个源 IP 地址为 10.11.11.1/8 的 ARP 报文，由于两个 IP 地址不在同一网段，GigabitEthernet0/1 接口无法处理这个报文。如果使能支持自然网段的 ARP 请求功能，则通过自然网段进行判断，由于 GigabitEthernet0/1 接口的 IP 地址为 A 类地址，因此默认掩码应该为 8 位，于是两个 IP 地址就在同一个网段，GigabitEthernet0/1 接口就可以学习源 IP 地址为 10.11.11.1 的 ARP 表项了。

表1-5 使能支持自然网段的 ARP 请求

操作	命令	说明
进入系统视图	**system-view**	-
使能支持自然网段的 ARP 请求	**naturemask-arp enable**	必选 缺省情况下，不使能支持自然网段的 ARP 请求

1.2.6 ARP 配置举例

1. 组网需求

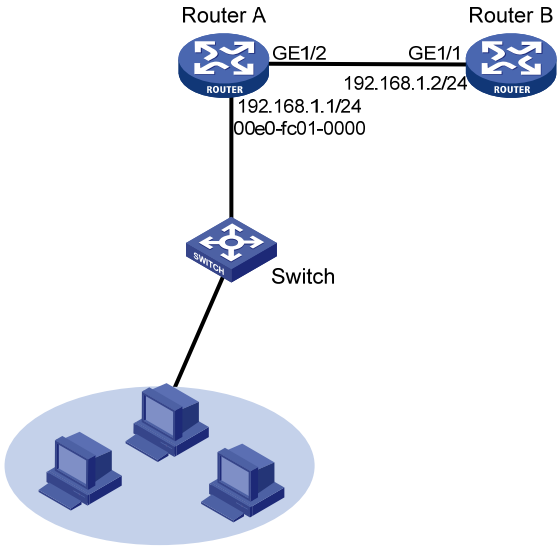
Router A 的 GE1/2 接口的 IP 地址为 192.168.1.1/24，MAC 地址为 00e0-fc01-0000。

Router B 的 GE1/1 接口的 IP 地址为 192.168.1.2/24。

为了增加 Router A 和 Router B 通信的安全性，可以在 Router B 上配置静态 ARP 表项。

2. 组网图

图1-3 配置静态 ARP 表项组网图



3. 配置步骤

在 Router A 上进行下列配置

为接口 GE1/2 配置 IP 地址

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/2
[RouterA- GigabitEthernet 1/2] ip address 192.168.1.1 24
[RouterA- GigabitEthernet 1/2]quit
```

在 Router B 上进行下列配置。

为接口 GE1/1 配置 IP 地址

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/1
[RouterB- GigabitEthernet 1/1] ip address 192.168.1.2 24
[RouterB- GigabitEthernet 1/1]quit
```

配置一条静态 ARP 表项，IP 地址为 192.168.1.1，对应的 MAC 地址为 00e0-fc01-0000

```
[RouterB] arp static 192.168.1.1 00e0-fc01-0000
```

查看静态 ARP 表项信息。

```
[RouterB] display arp static
                Type: S-Static    D-Dynamic    A-Authorized
IP Address       MAC Address     VLAN ID  Interface             Aging Type
192.168.1.1      00e0-fc01-0000   N/A       N/A                    N/A   S
```

1.3 配置免费 ARP

1.3.1 免费 ARP 简介

免费 ARP 报文是一种特殊的 ARP 报文，该报文中携带的发送者 IP 地址和目标 IP 地址都是本机 IP 地址，发送者 MAC 地址是本机 MAC 地址，目标 MAC 地址是广播地址。

设备通过对外发送免费 ARP 报文，实现以下功能：

- 确定其它设备的 IP 地址是否与本机 IP 地址冲突。
- 设备改变了硬件地址，通过发送免费 ARP 报文通知其他设备更新 ARP 表项。

设备通过学习免费 ARP 报文，实现以下功能：

对于收到的免费 ARP 报文，如果 ARP 表中没有与此报文对应的 ARP 表项，就将免费 ARP 报文中携带的信息添加到本地动态 ARP 映射表中。

1.3.2 配置免费 ARP

表1-6 配置免费 ARP

操作	命令	说明
进入系统视图	**system-view**	-
使能收到非同一网段 ARP 请求时发送免费 ARP 报文功能	**gratuitous-arp-sending enable**	必选 缺省情况下，设备收到非同一网段的 ARP 请求时不发送免费 ARP 报文
使能免费 ARP 报文学习功能	**gratuitous-arp-learning enable**	可选 缺省情况下，设备免费 ARP 报文的学习功能处于开启状态

1.4 配置 ARP 源抑制功能

1.4.1 ARP 源抑制功能简介

如果网络中有主机通过向设备发送大量目标 IP 地址不能解析的 IP 报文来攻击设备，则会造成下面的危害：

- 设备向目的网段发送大量 ARP 请求报文，加重目的网段的负载。
- 设备会不断解析目标 IP 地址，增加了 CPU 的负担。

为避免这种攻击所带来的危害，设备提供了 ARP 源抑制功能。开启该功能后，如果网络中某主机向设备某端口连续发送目标 IP 地址不能解析的 IP 报文（当每 5 秒内的 ARP 请求报文的流量超过设置的阈值），对于由此 IP 地址发出的 IP 报文，设备不允许其触发 ARP 请求，直至 5 秒后再处理，从而避免了恶意攻击所造成的危害。

1.4.2 配置 ARP 源抑制

表1-7 配置 ARP 源抑制

操作	命令	说明
进入系统视图	**system-view**	-
使能 ARP 源抑制功能	**arp source-suppression enable**	必选 缺省情况下，关闭 ARP 源抑制功能
配置 ARP 源抑制的阈值	**arp source-suppression limit** *limit-value*	可选 缺省情况下，ARP 源抑制的阈值为 10

1.4.3 ARP 源抑制显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 ARP 源抑制的运行情况，通过查看显示信息验证配置的效果。

表1-8 ARP 源抑制显示和维护

操作	命令
显示 ARP 源抑制的配置信息	**display arp source-suppression**

1.5 配置 ARP 防 IP 报文攻击功能

1.5.1 ARP 防 IP 报文攻击功能介绍

在进行 IP 报文转发过程中，设备需要依靠 ARP 解析下一跳 IP 地址的 MAC 地址。如果地址解析成功，报文可以直接通过硬件转发芯片直接转发出去，而不需要再由软件处理；如果地址解析不成功，需要由软件进行解析处理。这样，如果接收到大量下一跳 IP 地址循环变化并且该 IP 地址不可达的 IP 报文，由于下一跳 IP 地址解析不成功，软件会试图反复地对下一跳 IP 地址进行探测，导致 CPU 负荷过重，就造成了 IP 报文对设备的攻击。

用户可以通过配置 ARP 防 IP 报文攻击功能来预防这种可能存在的攻击情况。在防 IP 报文攻击功能启用后，一旦接收到下一跳地址不可达的 IP 报文（即 ARP 解析失败的 IP 报文），设备立即产生一个黑洞路由，使硬件转发芯片在一段时间内将去往该地址的报文直接丢弃。等待黑洞路由老化时间过后，如有报文触发则再次发起解析，如果解析成功则由硬件进行转发，否则仍然下发黑洞路由。这种方式能够有效的防止 IP 报文的攻击，减轻 CPU 的负担。

1.5.2 启用 ARP 防 IP 报文攻击功能

启用 ARP 防 IP 报文攻击功能对转发报文和本地产生的报文均有效。

表1-9 启用 ARP 防 IP 报文攻击功能

操作	命令	说明
进入系统视图	**system-view**	-
启用 ARP 防 IP 报文攻击功能	**arp resolving-route enable**	可选 缺省情况下，ARP 防 IP 报文攻击功能处于关闭状态

1.6 ARP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 ARP 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令清除 ARP 表中的 ARP 表项。

表1-10 ARP 显示和维护

操作		命令
显示 ARP 表项	集中式设备	**display arp** [[**all** | **dynamic** | **static**] | **interface** *interface-type interface-number*] [[**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*] | **count**]
	分布式设备	**display arp** [[**all** | **dynamic** | **static**] [**slot** *slot-number*] | **interface** *interface-type interface-number*] [[**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*] | **count**]
显示指定 IP 地址的 ARP 表项	集中式设备	**display arp** *ip-address* [**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
	分布式设备	**display arp** *ip-address* [**slot** *slot-number*] [**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示指定 VPN 实例的 ARP 表项		**display arp vpn-instance** *vpn-instance-name* [**|** { **begin** | **exclude** | **include** } *regular-expression* | **count**]
显示动态 ARP 表项的老化时间		**display arp timer aging**
清除 ARP 表项	集中式设备	**reset arp** { **all** | **dynamic** | **static** | **interface** *interface-type interface-number* }
	分布式设备	**reset arp** { **all** | **dynamic** | **static** | **slot** *slot-number* | **interface** *interface-type interface-number* }

2 代理 ARP 配置

2.1 代理 ARP 简介

如果 ARP 请求是从一个网络的主机发往同一网段却不在同一物理网络上的另一台主机，那么连接它们的具有代理 ARP 功能的设备就可以回答该请求，这个过程称作代理 ARP（Proxy ARP)。代理 ARP 功能屏蔽了分离的物理网络这一事实，使用户使用起来，好像在同一个物理网络上。

代理 ARP 分为普通代理 ARP 和本地代理 ARP，二者的应用场景有所区别：

- 普通代理 ARP 的应用环境为：想要互通的主机分别连接到设备的不同三层接口上，且这些主机不在同一个广播域中。
- 本地代理 ARP 的应用环境为：想要互通的主机连接到设备的同一个三层接口上，且这些主机不在同一个广播域中。

说明

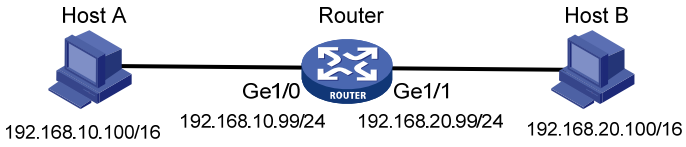
如无特殊说明，本章后续描述中的代理 ARP 均指普通代理 ARP。

2.1.1 代理 ARP

处于同一网段内的主机，当连接到设备的不同三层接口时，可以利用设备的代理 ARP 功能，通过三层转发实现互通。

代理 ARP 的典型应用环境如图 2-1 所示。设备 Router 通过两个三层接口 Ethernet1/1 和 Ethernet1/2 连接两个网络，两个三层接口的 IP 地址不在同一个网段，接口地址分别为 192.168.10.99/24、192.168.20.99/24。但是两个网络内的主机 Host A 和 Host B 的地址通过掩码的控制，既与相连设备的接口地址在同一网段，同时二者也处于同一个网段。

图2-1 代理 ARP 的应用环境



在这种组网情况下，当 Host A 需要与 Host B 通信时，由于目的 IP 地址与本机的 IP 地址为同一网段，因此 Host A 会直接发出请求 Host B 硬件地址的 ARP 请求。但是，此时的两台主机处于不同的广播域中，Host B 无法收到 Host A 的 ARP 请求报文，当然也就无法应答。

通过在 Router 上启用代理 ARP 功能，可以解决此问题。启用代理 ARP 后，Router 可以应答 Host A 的 ARP 请求。同时，Router 相当于 Host B 的代理，把从其他主机发送过来的报文转发给它。

代理 ARP 的优点是，它可以只被应用在一个设备上（此时该设备的作用相当于网关），不会影响到网络中其他设备的路由表。代理 ARP 功能可以在 IP 主机没有配置缺省网关或者 IP 主机没有任何路由能力的情况下使用。

2.2 配置代理 ARP 功能

代理 ARP 在以太网接口视图下进行配置。

表2-1 配置代理 ARP 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	必选
开启代理 ARP 功能	**proxy-arp enable**	必选 缺省情况下，关闭代理 ARP 功能

2.3 代理 ARP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后代理 ARP 的运行情况，查看显示信息验证配置的效果。

表2-2 代理 ARP 显示和维护

操作	命令
显示代理 ARP 的状态	**display proxy-arp** [**interface** *interface-type interface-number*]

2.4 代理 ARP 典型配置举例

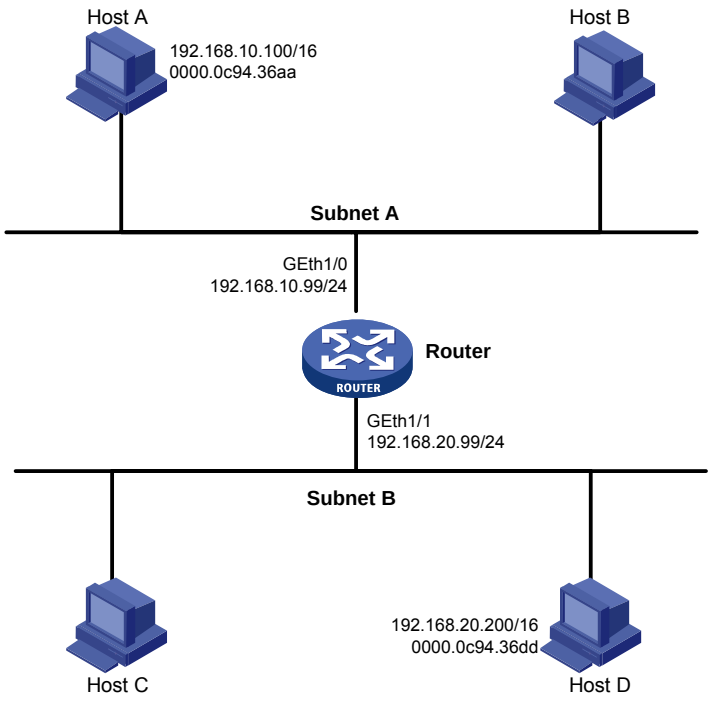
2.4.1 代理 ARP 配置举例

1. 组网需求

Host A 和 Host D 互相认为是同一子网，只是从设备上看属于不同的子网，通过在设备上启用代理 ARP 功能来实现 Host A 和 Host D 的通信。

2. 组网图

图2-2 配置代理 ARP 组网图



3. 配置步骤

配置代理 ARP，实现 Host A 和 Host D 之间的通信。

```
<Router> system-view
[Router] interface GigabitEthernet 1/0
[Router-GigabitEthernet1/0] ip address 192.168.10.99 255.255.255.0
[Router-GigabitEthernet1/0] proxy-arp enable
[Router-GigabitEthernet1/0] quit
[Router] interface GigabitEthernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.20.99 255.255.255.0
[Router-GigabitEthernet1/1] proxy-arp enable
[Router-GigabitEthernet1/1] quit
```

3 ARP 攻击防御配置

ARP 协议有简单、易用的优点，但是也因为其没有任何安全机制而容易被攻击发起者利用。目前 ARP 攻击和 ARP 病毒已经成为局域网安全的一大威胁，为了避免各种攻击带来的危害，设备提供了多种技术对攻击进行检测和解决。

下面将详细介绍一下这些技术的原理以及配置。

3.1 配置 ARP 主动确认功能

3.1.1 ARP 主动确认功能简介

ARP 的主动确认功能主要应用于网关设备上，防止攻击者仿冒用户欺骗网关设备。使能 ARP 主动确认功能之后，当收到的 ARP 报文中的源 MAC 地址和对应 ARP 表项中的不同时，设备首先判断 ARP 表项刷新时间是否超过 1 分钟，如果没有超过 1 分钟，则不更新 ARP 表项。否则向 ARP 表项对应的源发送一个单播 ARP 请求报文，如果在随后的 5 秒内收到 ARP 应答报文，则忽略之前收到的 ARP 攻击报文；如果没有收到 ARP 应答报文，则向之前收到的 ARP 报文对应的源发送一个单播 ARP 请求报文，如果在随后的 5 秒内收到了 ARP 应答报文，则根据之前收到的 ARP 报文更新 ARP 表项，否则 ARP 表项不会被修改。

3.1.2 配置 ARP 主动确认功能

表3-1 配置 ARP 主动确认功能

配置步骤	命令	说明
进入系统视图	**system-view**	-
使能 ARP 主动确认功能	**arp anti-attack active-ack enable**	必选 缺省情况下，关闭 ARP 主动确认功能

3.2 配置 ARP 报文源 MAC 一致性检查功能

3.2.1 简介

ARP 报文源 MAC 一致性检查功能主要应用于网关设备上，防御以太网数据帧首部中的源 MAC 地址和 ARP 报文中的源 MAC 地址不同的 ARP 攻击。

在 ARP Detection 中也对源 MAC 一致性进行了检查，但这两者功能不同。ARP Detection 中的源 MAC 一致性检查，是在接入设备上使能 ARP Detection，对上送的 ARP 报文进行源 MAC 一致性的检查。而这里的源 MAC 一致性检查，是网关设备在学习 ARP 之前，对要被学习的 ARP 报文进行检查。

3.2.2 配置 ARP 报文源 MAC 一致性检查功能

表3-2 配置 ARP 报文源 MAC 一致性检查功能

配置步骤	命令	说明
进入系统视图	**system-view**	-
使能ARP报文源MAC一致性检查功能	**arp anti-attack valid-check enable**	必选 缺省情况下，关闭 ARP 报文源 MAC 一致性检查功能

3.3 配置 ARP 报文限速功能

3.3.1 简介

为防止大量 ARP 报文对 CPU 进行冲击，需要对上送 CPU 的 ARP 报文进行限速。

3.3.2 配置 ARP 报文限速功能

表3-3 配置 ARP 报文限速功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 ARP 报文上送限速功能（集中式设备）	**arp rate-limit { disable | rate** *pps* **drop }**	必选 缺省情况下，开启 ARP 报文上送限速功能
开启 ARP 报文上送限速功能（分布式设备）	**arp rate-limit { disable | rate** *pps* **drop } [slot** *slot-number* **]**	必选 缺省情况下，开启 ARP 报文上送限速功能

3.4 配置授权 ARP 功能

说明

- 本特性目前仅支持三层以太网接口。
- 关于 DHCP 服务器和 DHCP 中继的介绍，请参见“IP 业务分册”中的“DHCP 配置”。

3.4.1 授权 ARP 功能简介

所谓授权 ARP（Authorized ARP），就是根据 DHCP 服务器生成的租约或者 DHCP 中继生成的安全表项同步生成 ARP 表项。

授权 ARP 功能可以检测用户的非正常下线，并且可以防止用户仿冒其他用户的 IP 地址或 MAC 地址对网络进行攻击，保证只有合法的用户才能使用网络资源，增加了网络的安全性。

静态 ARP 表项可以覆盖授权 ARP 表项，授权 ARP 表项可以覆盖动态 ARP 表项，但是授权 ARP 表项不能覆盖静态 ARP 表项，动态 ARP 表项不能覆盖授权 ARP 表项。

3.4.2 配置授权 ARP 功能

使能接口下的授权 ARP 功能后，会启动接口下授权 ARP 表项的老化探测功能，并禁止该接口学习动态 ARP 表项；关闭接口下的授权 ARP 功能后，会关闭该接口下授权 ARP 表项的老化探测功能，并允许该接口学习动态 ARP 表项。

表3-4 配置授权 ARP 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 DHCP 服务器（DHCP 中继）支持授权 ARP 功能	**dhcp update arp**	必选 缺省情况下，DHCP 服务器（DHCP 中继）不支持授权 ARP 功能
使能授权 ARP 功能	**arp authorized enable**	必选 缺省情况下，接口下没有使能授权 ARP 功能
配置授权 ARP 表项的老化时间	**arp authorized time-out** *seconds*	可选 缺省情况下，接口下授权 ARP 表项的老化时间为 1200 秒

说明

如果 DHCP 服务器（DHCP 中继）不支持授权 ARP 功能，当配置了 **arp authorized enable** 命令后，只是会禁止该接口学习动态 ARP 表项，不会同步生成授权 ARP 表项。

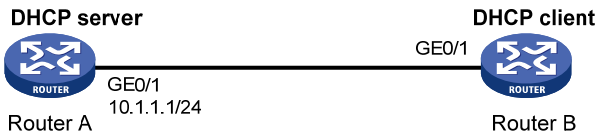
3.4.3 授权 ARP 功能在 DHCP 服务器上的典型配置举例

1. 组网需求

- Router A 是 DHCP 服务器，为同一网段中的客户端动态分配 IP 地址，地址池网段为 10.1.1.0/24。通过在接口 GigabitEthernet0/1 上启用授权 ARP 功能对客户端进行老化探测，并保证客户端的合法性。
- Router B 是 DHCP 客户端，通过 DHCP 协议从 DHCP 服务器获取 IP 地址。

2. 组网图

图3-1 授权 ARP 功能典型配置组网图



3. 配置步骤

(1) 配置 Router A

配置接口的 IP 地址。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 0/1
[RouterA-GigabitEthernet0/1] ip address 10.1.1.1 24
[RouterA-GigabitEthernet0/1] quit
```

使能 DHCP 服务。

```
[RouterA] dhcp enable
[RouterA] dhcp server ip-pool 1
[RouterA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.0
[RouterA-dhcp-pool-1] quit
```

进入三层以太网接口视图。

```
[RouterA] interface gigabitethernet 0/1
```

使能 DHCP 同步 ARP 表项功能。

```
[RouterA-GigabitEthernet0/1] dhcp update arp
```

使能接口授权 ARP 功能。

```
[RouterA-GigabitEthernet0/1] arp authorized enable
```

配置接口下授权 ARP 的老化时间。

```
[RouterA-GigabitEthernet0/1] arp authorized time-out 120
[RouterA-GigabitEthernet0/1] quit
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface gigabitethernet 0/1
[RouterB-GigabitEthernet0/1] ip address dhcp-alloc
[RouterB-GigabitEthernet0/1] quit
```

(3) Router B 获得 Router A 分配的 IP 后，在 Router A 查看授权 ARP 信息。

```
[RouterA] display arp all
                Type: S-Static    D-Dynamic    A-Authorized
IP Address       MAC Address     VLAN ID  Interface          Aging  Type
10.1.1.2         0012-3f86-e94c  N/A        GE0/1             1      A
```

从以上信息也可获知 Router A 为 Router B 动态分配的 IP 地址为 10.1.1.2。

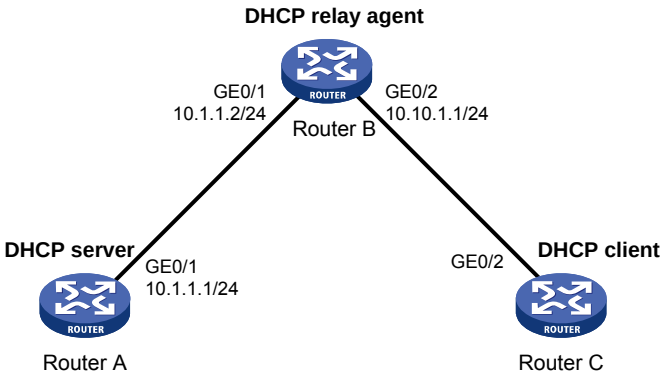
3.4.4 授权 ARP 功能在 DHCP 中继上的典型配置举例

1. 组网需求

- Router A 是 DHCP 服务器，为不同网段中的客户端动态分配 IP 地址，地址池网段为 10.10.1.0/24。
- Router B 是 DHCP 中继，通过在接口 GigabitEthernet0/2 上启用授权 ARP 功能对客户端进行老化探测，并保证客户端的合法性。
- Router C 是 DHCP 客户端，通过 DHCP 中继从 DHCP 服务器获取 IP 地址。

2. 组网图

图3-2 授权 ARP 功能典型配置组网图



3. 配置步骤

(1) 配置 Router A

配置接口的 IP 地址。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 0/1
[RouterA-GigabitEthernet0/1] ip address 10.1.1.1 24
[RouterA-GigabitEthernet0/1] quit
```

使能 DHCP 服务。

```
[RouterA] dhcp enable
[RouterA] dhcp server ip-pool 1
[RouterA-dhcp-pool-1] network 10.10.1.0 mask 255.255.255.0
[RouterA-dhcp-pool-1] gateway-list 10.10.1.1
[RouterA-dhcp-pool-1] quit
[RouterA] ip route-static 10.10.1.0 24 10.1.1.2
```

(2) 配置 Router B

使能 DHCP 服务。

```
<RouterB> system-view
[RouterB] dhcp enable
```

配置接口的 IP 地址。

```
[RouterB] interface gigabitethernet 0/1
[RouterB-GigabitEthernet0/1] ip address 10.1.1.2 24
[RouterB-GigabitEthernet0/1] quit
[RouterB] interface gigabitethernet 0/2
[RouterB-GigabitEthernet0/2] ip address 10.10.1.1 24
```

配置 GigabitEthernet0/2 接口工作在 DHCP 中继模式。

```
[RouterB-GigabitEthernet0/2] dhcp select relay
[RouterB-GigabitEthernet0/2] quit
```

配置 DHCP 服务器的地址。

```
[RouterB] dhcp relay server-group 1 ip 10.1.1.1
```

配置 GigabitEthernet0/2 接口对应 DHCP 服务器组 1。

```
[RouterB] interface gigabitethernet 0/2
[RouterB-GigabitEthernet0/2] dhcp relay server-select 1
```

使能 DHCP 同步 ARP 表项功能。

```
[RouterB-GigabitEthernet0/2] dhcp update arp
```

使能接口授权 ARP 功能。

```
[RouterB-GigabitEthernet0/2] arp authorized enable
```

配置接口下授权 ARP 的老化时间。

```
[RouterB-GigabitEthernet0/2] arp authorized time-out 120
[RouterB-GigabitEthernet0/2] quit
```

(3) 配置 Router C

```
<RouterC> system-view
[RouterC] interface gigabitethernet 0/2
[RouterC-GigabitEthernet0/2] ip address dhcp-alloc
[RouterC-GigabitEthernet0/2] quit
[RouterC] ip route-static 10.1.1.0 24 10.10.1.1
```

(4) Router B 获得 Router A 分配的 IP 后，在 Router B 查看授权 ARP 信息。

```
[RouterB] display arp all
                Type: S-Static    D-Dynamic    A-Authorized
IP Address       MAC Address     VLAN ID  Interface          Aging Type
10.10.1.2        0012-3f86-e94c  N/A      GE0/2              1     A
```

从以上信息也可获知 Router A 为 Router C 动态分配的 IP 地址为 10.10.1.2。

目 录

1 DHCP概述 1-1
1.1 DHCP简介 1-1
1.2 DHCP的IP地址分配 1-1
1.2.1 IP地址分配策略 1-1
1.2.2 IP地址动态获取过程 1-2
1.2.3 IP地址的租约更新 1-2
1.3 DHCP报文格式 1-3
1.4 DHCP选项 1-4
1.4.1 DHCP选项简介 1-4
1.4.2 DHCP常用选项介绍 1-4
1.4.3 自定义的选项格式 1-4
1.5 协议规范 1-7
2 DHCP服务器配置 2-1
2.1 DHCP服务器简介 2-1
2.1.1 DHCP服务器的应用环境 2-1
2.1.2 DHCP地址池 2-1
2.1.3 DHCP服务器分配IP地址的优先次序 2-2
2.2 DHCP服务器配置任务简介 2-2
2.3 配置DHCP服务器的地址池 2-3
2.3.1 DHCP服务器地址池配置任务简介 2-3
2.3.2 创建DHCP地址池 2-3
2.3.3 配置普通模式地址池的地址分配方式 2-3
2.3.4 配置DHCP客户端的域名后缀 2-5
2.3.5 配置DHCP客户端的DNS服务器地址 2-5
2.3.6 配置DHCP客户端的WINS服务器地址和NetBIOS节点类型 2-6
2.3.7 配置DHCP客户端的BIMS服务器信息 2-7
2.3.8 配置DHCP客户端的网关地址 2-7
2.3.9 配置DHCP客户端的TFTP服务器地址及启动文件名 2-7
2.3.10 配置DHCP自定义选项 2-8
2.4 使能DHCP服务 2-9
2.5 配置接口工作在DHCP服务器模式 2-9
2.6 配置DHCP服务的安全功能 2-10
2.6.1 配置准备 2-10
2.6.2 配置伪DHCP服务器检测功能 2-10
2.6.3 配置IP地址重复分配检测功能 2-11
2.7 配置Option 82 的处理方式 2-11
2.8 DHCP服务器显示和维护 2-12
2.9 DHCP服务器典型配置举例 2-13
2.9.1 静态绑定地址典型配置举例 2-13
2.9.2 动态分配地址典型配置举例 2-14

2.9.3 自定义选项典型配置举例 2-15
2.10 DHCP服务器常见配置错误举例 2-16
3 DHCP中继配置 3-1
3.1 DHCP中继简介 3-1
3.1.1 DHCP中继的应用环境 3-1
3.1.2 DHCP中继的基本原理 3-1
3.1.3 DHCP中继支持Option 82 功能 3-2
3.2 DHCP中继配置任务简介 3-3
3.3 配置DHCP中继 3-3
3.3.1 使能DHCP服务 3-3
3.3.2 配置接口工作在DHCP中继模式 3-3
3.3.3 配置DHCP服务器组 3-4
3.3.4 配置DHCP中继的安全功能 3-4
3.3.5 配置通过DHCP中继释放客户端的IP地址 3-6
3.3.6 配置DHCP中继支持Option 82 功能 3-7
3.4 DHCP中继显示和维护 3-8
3.5 DHCP中继典型配置举例 3-8
3.5.1 DHCP中继配置举例 3-8
3.5.2 DHCP中继支持Option 82 配置举例 3-9
3.6 DHCP中继常见配置错误举例 3-10
4 DHCP客户端配置 4-1
4.1 DHCP客户端简介 4-1
4.2 配置接口使用DHCP方式获取IP地址 4-1
4.3 DHCP客户端显示和维护 4-1
4.4 DHCP客户端典型配置举例 4-2

1 DHCP 概述

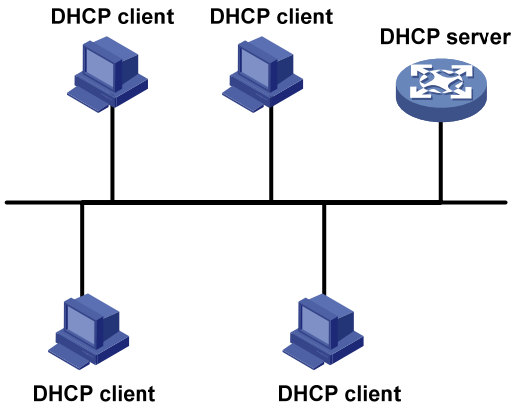
1.1 DHCP 简介

随着网络规模的不断扩大和网络复杂度的提高，计算机的数量经常超过可供分配的 IP 地址数量。同时随着便携机及无线网络的广泛使用，计算机的位置也经常变化，相应的 IP 地址也必须经常更新，从而导致网络配置越来越复杂。DHCP（Dynamic Host Configuration Protocol，动态主机配置协议）就是为解决这些问题而发展起来的。

DHCP 采用客户端/服务器通信模式，由客户端向服务器提出配置申请，服务器返回为客户端分配的 IP 地址等相应的配置信息，以实现 IP 地址等信息的动态配置。

在DHCP的典型应用中，一般包含一台DHCP服务器和多台客户端（如PC和便携机），如 图 1-1所示。

图1-1 DHCP 典型应用



说明

DHCP客户端和DHCP服务器处于不同物理网段时，客户端可以通过DHCP中继与服务器通信，获取IP地址及其他配置信息。DHCP中继的详细介绍，请参见“3.1 DHCP中继简介”。

1.2 DHCP 的 IP 地址分配

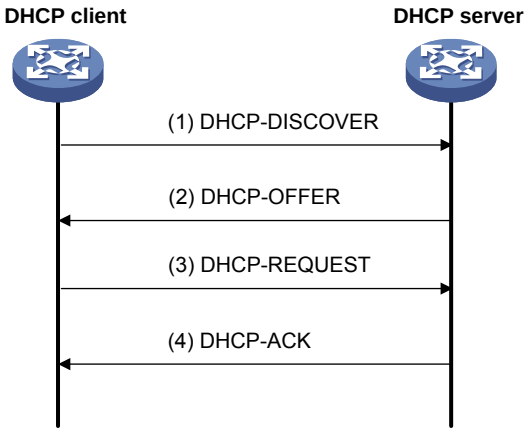
1.2.1 IP 地址分配策略

针对客户端的不同需求，DHCP 提供三种 IP 地址分配策略：

- 手工分配地址：由管理员为少数特定客户端（如 WWW 服务器等）静态绑定固定的 IP 地址。通过 DHCP 将配置的固定 IP 地址发给客户端。
- 自动分配地址：DHCP 为客户端分配租期为无限长的 IP 地址。
- 动态分配地址：DHCP 为客户端分配具有一定有效期限的 IP 地址，到达使用期限后，客户端需要重新申请地址。绝大多数客户端得到的都是这种动态分配的地址。

1.2.2 IP 地址动态获取过程

图1-2 IP 地址动态获取过程



如图 1-2所示，DHCP客户端从DHCP服务器动态获取IP地址，主要通过四个阶段进行：

(1) 发现阶段，即 DHCP 客户端寻找 DHCP 服务器的阶段。客户端以广播方式发送 DHCP-DISCOVER 报文。

(2) 提供阶段，即DHCP服务器提供IP地址的阶段。DHCP服务器接收到客户端的DHCP-DISCOVER报文后，根据IP地址分配的优先次序选出一个IP地址，与其他参数一起通过DHCP-OFFER报文发送给客户端。DHCP-OFFER报文的发送方式由DHCP-DISCOVER报文中的flag字段决定，具体请参见“1.3 DHCP报文格式”的介绍。

(3) 选择阶段，即 DHCP 客户端选择 IP 地址的阶段。如果有多台 DHCP 服务器向该客户端发来 DHCP-OFFER 报文，客户端只接受第一个收到的 DHCP-OFFER 报文，然后以广播方式发送 DHCP-REQUEST 报文，该报文中包含 DHCP 服务器在 DHCP-OFFER 报文中分配的 IP 地址。

(4) 确认阶段，即 DHCP 服务器确认 IP 地址的阶段。DHCP 服务器收到 DHCP 客户端发来的 DHCP-REQUEST 报文后，只有 DHCP 客户端选择的服务器会进行如下操作：如果确认将地址分配给该客户端，则返回 DHCP-ACK 报文；否则返回 DHCP-NAK 报文，表明地址不能分配给该客户端。

说明

- 客户端收到服务器返回的 DHCP-ACK 确认报文后，会以广播的方式发送免费 ARP 报文，探测是否有主机使用服务器分配的 IP 地址，如果在规定的时间内没有收到回应，客户端才使用此地址。否则，客户端会发送 DHCP-DECLINE 报文给 DHCP 服务器，并重新申请 IP 地址。
- 如果网络中存在多个 DHCP 服务器，除 DHCP 客户端选中的服务器外，其它 DHCP 服务器中本次未分配出的 IP 地址仍可分配给其他客户端。

1.2.3 IP 地址的租约更新

如果采用动态地址分配策略，则 DHCP 服务器分配给客户端的 IP 地址有一定的租借期限，当租借期满后服务器会收回该 IP 地址。如果 DHCP 客户端希望继续使用该地址，需要更新 IP 地址租约。

在 DHCP 客户端的 IP 地址租约期限达到一半时间时，DHCP 客户端会向为它分配 IP 地址的 DHCP 服务器单播发送 DHCP-REQUEST 报文，以进行 IP 租约的更新。如果客户端可以继续使用此 IP 地

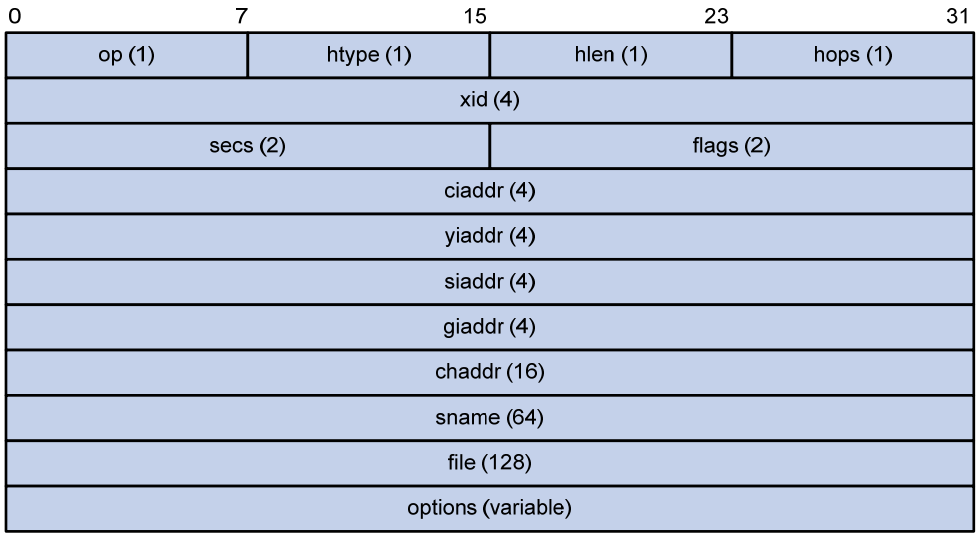
址，则 DHCP 服务器回应 DHCP-ACK 报文，通知 DHCP 客户端已经获得新 IP 租约；如果此 IP 地址不可以再分配给该客户端，则 DHCP 服务器回应 DHCP-NAK 报文，通知 DHCP 客户端不能获得新的租约。

如果在租约的一半时间进行的续约操作失败，DHCP 客户端会在租约期限达到 7/8 时，广播发送 DHCP-REQUEST 报文进行续约。DHCP 服务器的处理方式同上，不再赘述。

1.3 DHCP 报文格式

DHCP有 8 种类型的报文，每种报文的格式相同，只是某些字段的取值不同。DHCP报文格式基于BOOTP的报文格式，具体格式如 图 1-3所示（括号中的数字表示该字段所占的字节）。

图1-3 DHCP 报文格式



各字段的解释如下：

- op：报文的操作类型，分为请求报文和响应报文，1 为请求报文；2 为响应报文。具体的报文类型在 option 字段中标识。
- htype、hlen：DHCP 客户端的硬件地址类型及长度。
- hops：DHCP 报文经过的 DHCP 中继的数目。DHCP 请求报文每经过一个 DHCP 中继，该字段就会增加 1。
- xid：客户端发起一次请求时选择的随机数，用来标识一次地址请求过程。
- secs：DHCP 客户端开始 DHCP 请求后所经过的时间。目前没有使用，固定为 0。
- flags：第一个比特为广播响应标识位，用来标识 DHCP 服务器响应报文是采用单播还是广播方式发送，0 表示采用单播方式，1 表示采用广播方式。其余比特保留不用。
- ciaddr：DHCP 客户端的 IP 地址。
- yiaddr：DHCP 服务器分配给客户端的 IP 地址。
- siaddr：DHCP 客户端获取 IP 地址等信息的服务器 IP 地址。
- giaddr：DHCP 客户端发出请求报文后经过的第一个 DHCP 中继的 IP 地址。
- chaddr：DHCP 客户端的硬件地址。
- sname：DHCP 客户端获取 IP 地址等信息的服务器名称。
- file：DHCP 服务器为 DHCP 客户端指定的启动配置文件名称及路径信息。

- option：可选变长选项字段，包含报文的类型、有效租期、DNS 服务器的 IP 地址、WINS 服务器的 IP 地址等配置信息。

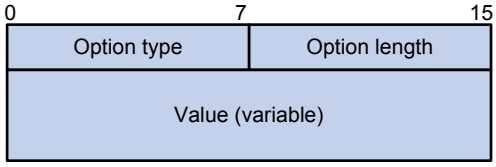
1.4 DHCP 选项

1.4.1 DHCP 选项简介

DHCP 利用 Option 字段传递控制信息和网络配置参数，实现地址的动态分配，为客户端提供更加丰富的网络配置信息。

DHCP选项的格式如 图 1-4所示。

图1-4 DHCP 选项格式



1.4.2 DHCP 常用选项介绍

常见的 DHCP 选项有：

- Option 3：路由器选项，用来指定为客户端分配的网关地址。
- Option 6：DNS 服务器选项，用来指定为客户端分配的 DNS 服务器地址。
- Option 51：IP 地址租约选项。
- Option 53：DHCP 消息类型选项，标识 DHCP 消息的类型。
- Option 55：请求参数列表选项。客户端利用该选项指明需要从服务器获取哪些网络配置参数。该选项内容为客户端请求的参数对应的选项值。
- Option 66：TFTP 服务器名选项，用来指定为客户端分配的 TFTP 服务器的域名。
- Option 67：启动文件名选项，用来指定为客户端分配的启动文件名。
- Option 150：TFTP 服务器地址选项，用来指定为客户端分配的 TFTP 服务器的地址。

更多 DHCP 选项的介绍，请参见 RFC 2132。

1.4.3 自定义的选项格式

有些选项的内容，RFC 2132 中没有统一规定，例如 Option 43。下面将介绍设备上定义的几种选项格式。

1. 厂商特定信息选项（Option 43）

Option 43 称为厂商特定信息选项。DHCP 服务器和 DHCP 客户端通过 Option 43 交换厂商特定的信息。当 DHCP 服务器接收到请求 Option 43 信息的 DHCP 请求报文（Option 55 中带有 43 参数）后，将在回复报文中携带 Option 43，为 DHCP 客户端分配厂商指定的信息。

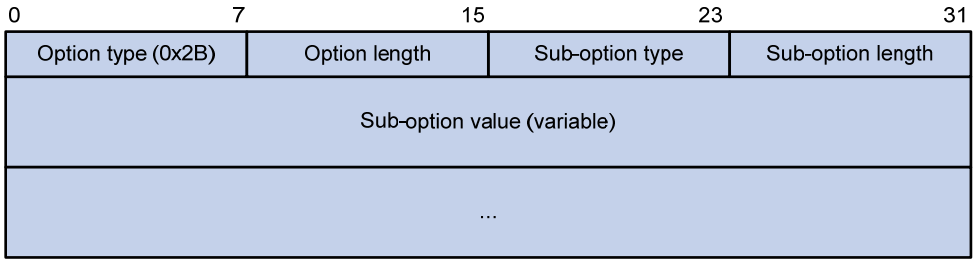
设备做为 DHCP 客户端时，可以通过 Option 43 获取：

- ACS（Auto-Configuration Server，自动配置服务器）的参数，包括 URL 地址、用户名和密码。

- 服务提供商标识，CPE（Customer Premises Equipment，用户侧设备）从 DHCP 服务器获取该信息后，将该信息通告给 ACS，以便 ACS 选择服务提供商特有的配置和参数等。
- PXE（Preboot eXecution Environment，预启动执行环境）引导服务器地址，以便客户端从 PXE 引导服务器获取启动文件或其他控制信息。

(1) Option 43 的格式

图1-5 Option 43 格式



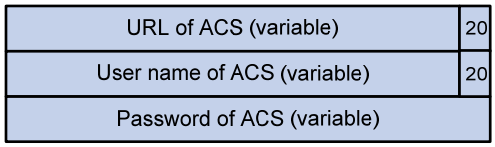
为了提供可扩展性，通过Option 43 为客户端分配更多的信息，Option 43 采用子选项的形式，通过不同的子选项为用户分配不同的网络配置参数，如 图 1-5所示。子选项中各字段的含义为：

- Sub-option type：子选项类型。目前，子选项类型值可以为 0x01 表示 ACS 参数子选项，0x02 表示服务提供商标识子选项，0x80 表示 PXE 引导服务器地址子选项。
- Sub-option length：子选项的长度，不包括子选项类型和子选项长度字段。
- Sub-option value 设备将 PXE 引导服务器地址作为 Option 43 的一个子选项，类型值为 0x80。：子选项的取值。

(2) Option 43 子选项取值字段的格式

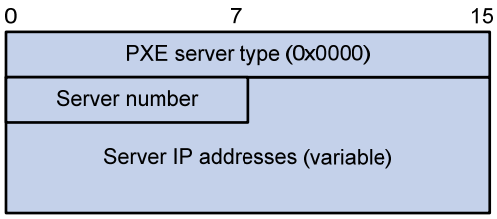
- ACS参数子选项的取值字段格式如 图 1-6所示。ACS的URL地址、用户名和密码长度可变，每个参数之间用空格（十六进制数为 0x20）隔开。

图1-6 ACS 参数子选项格式



- 服务提供商标识子选项的取值字段内容为服务提供商的标识。
- PXE引导服务器地址子选项的取值字段PXE服务器地址列表的格式如 图 1-7所示。其中，目前，PXE服务器类型目前取值只能为 0；Server number为子选项中包含的PXE服务器地址的数目；Server IP addresses为。PXE服务器的IP地址。

图1-7 PXE 引导服务器地址子选项格式



2. 中继代理信息选项（Option 82）

Option 82 称为中继代理信息选项，该选项记录了 DHCP 客户端的位置信息。DHCP 中继接收到 DHCP 客户端发送给 DHCP 服务器的请求报文后，在该报文中添加 Option 82，并转发给 DHCP 服务器。

管理员可以从 Option 82 中获得 DHCP 客户端的位置信息，以便定位 DHCP 客户端，实现对客户端的安全和计费等控制。支持 Option 82 的服务器还可以根据该选项的信息制定 IP 地址和其他参数的分配策略，提供更加灵活的地址分配方案。

Option 82 最多可以包含 255 个子选项。若定义了 Option 82，则至少要定义一个子选项。目前设备只支持两个子选项：sub-option 1（Circuit ID，电路 ID 子选项）和 sub-option 2（Remote ID，远程 ID 子选项）。

由于 Option 82 的内容没有统一规定，不同厂商通常根据需要进行填充。

设备上，可以通过两种方式配置 Option 82 的内容：

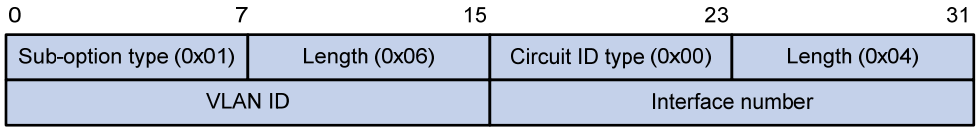
- 用户自定义方式：用户手工指定 Option 82 的内容；
- 非用户自定义方式：采用默认的 normal 模式或 verbose 模式填充 Option 82。

normal 和 verbose 填充模式中，子选项内容的填充格式可以是 ASCII 格式和 HEX 格式。子选项的内容如下：

(1) 采用 normal 模式填充

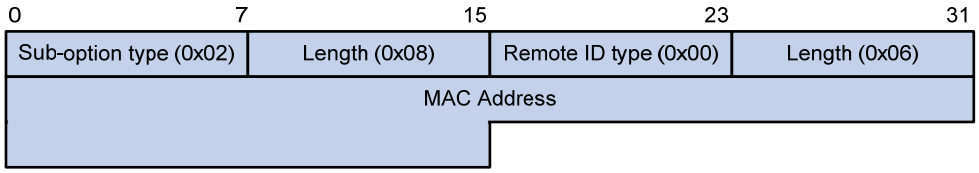
- sub-option 1 的内容是接收到DHCP客户端请求报文的接口属于的VLAN ID以及接口编号。如图 1-8所示，子选项类型值为 1，电路ID类型值为 0。

图1-8 normal 模式填充的 sub-option 1



- sub-option 2 的内容是接收到DHCP客户端请求报文的接口MAC地址（DHCP中继）。如 图 1-9所示，子选项类型值为 2，远程ID类型值为 0。

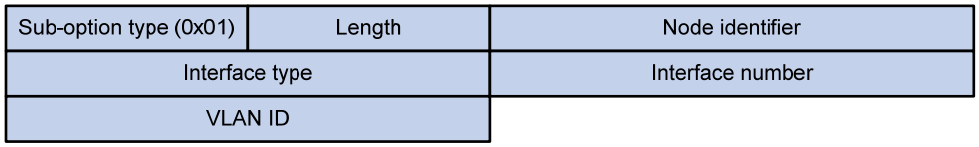
图1-9 normal 模式填充的 sub-option 2



(2) 采用 verbose 模式填充

- sub-option 1 的内容包括用户配置的接入节点标识（在报文中添加Option 82 的设备的标识）、接收到DHCP客户端请求报文的接口类型、接口号和VLAN ID，如 图 1-10所示。

图1-10 verbose 模式填充的 sub-option 1



说明

图 1-10中除VLAN ID固定为两字节外，其他sub-option 1 的填充内容均为可变长度。

- sub-option 2 的内容是接收到DHCP客户端请求报文的接口MAC地址（DHCP中继）或设备的桥MAC地址（DHCP Snooping设备）。verbose和normal填充模式的sub-option 2 内容相同，如 图 1-9所示。

1.5 协议规范

与 DHCP 相关的协议规范有：

- RFC 2131：Dynamic Host Configuration Protocol
- RFC 2132：DHCP Options and BOOTP Vendor Extensions
- RFC 1542：Clarifications and Extensions for the Bootstrap Protocol
- RFC 3046：DHCP Relay Agent Information Option

2 DHCP 服务器配置

说明

DHCP 服务器中对于接口的相关配置，目前只能在三层以太网接口（包括子接口）、串口、MP-group 接口和 Loopback 接口上进行。其中，DHCP 服务器的 subaddress 地址池配置不能在串口、MP-group 接口和 Loopback 接口上进行。

2.1 DHCP 服务器简介

2.1.1 DHCP 服务器的应用环境

在以下场合通常利用 DHCP 服务器来完成 IP 地址分配：

- 网络规模较大，手工配置需要很大的工作量，并难以对整个网络进行集中管理。
- 网络中主机数目大于该网络支持的 IP 地址数量，无法给每个主机分配一个固定的 IP 地址。例如，Internet 接入服务提供商限制同时接入网络的用户数目，大量用户必须动态获得自己的 IP 地址。
- 网络中只有少数主机需要固定的 IP 地址，大多数主机没有固定的 IP 地址需求。

2.1.2 DHCP 地址池

1. 普通模式地址池结构

DHCP 服务器从地址池中为客户端选择并分配 IP 地址及其他相关参数。

DHCP 服务器的普通模式地址池采用树状结构：树根是自然网段的地址池，分支是该网段的子网地址池，叶节点是手工绑定的客户端地址。同一级别地址池的顺序由配置的先后决定。这种树状结构实现了配置的继承性，即子网配置继承自然网段的配置，客户端的配置继承子网的配置。这样，对于一些通用参数（如 DNS 服务器地址），只需要在自然网段或者子网上配置即可。具体的继承情况如下：

(1) 在父子关系建立时，子地址池将会继承父地址池的已有配置。

(2) 在父子关系建立后，对父地址池进行的配置，子地址池是否会继承，则有下面两种情况：

- 如果子地址池没有该项配置，则继承父地址池的配置；
- 如果子地址池已有该项配置，则不会继承父地址池的配置。

说明

IP 地址的租用有效期限不具有继承关系。

2. 地址池的选取原则

DHCP 服务器为客户端分配 IP 地址时，地址池的选择原则如下：

(1) 如果存在将客户端MAC地址或客户端ID与IP地址静态绑定的地址池，则选择该地址池，并将静态绑定的IP地址分配给客户端。该地址池的配置方法请参见“2.3.3 1. 配置采用静态绑定方式进行地址分配”。

(2) 如果不存在静态绑定的地址池，则选择包含DHCP请求报文接收接口的IP地址（客户端与服务器在同一网段时）或DHCP请求报文中giaddr字段指定的IP地址（客户端与服务器不在同一网段，客户端通过DHCP中继获取IP地址时）、地址范围最小的普通模式地址池。如果该地址池中没有可供分配的IP地址，则服务器无法为客户端分配地址，服务器不会将父地址池中的IP地址分配给客户端。该地址池的配置方法请参见“2.3.3 2. 配置采用动态分配方式进行地址分配”。

例如，DHCP 服务器上配置了两个普通模式的地址池，动态分配的 IP 地址范围分别是 1.1.1.0/24 和 1.1.1.0/25，如果接收 DHCP 请求报文的接口 IP 地址为 1.1.1.1/25，服务器将从 1.1.1.0/25 地址池中选择 IP 地址分配给客户端，1.1.1.0/25 地址池中如果没有可供分配的 IP 地址，则服务器无法为客户端分配地址；如果接收 DHCP 请求报文的接口 IP 地址为 1.1.1.130/25，服务器将从 1.1.1.0/24 地址池中选择 IP 地址分配给客户端。

说明

配置地址池动态分配的 IP 地址范围时，请尽量保证该地址范围与 DHCP 服务器接口或 DHCP 中继接口地址的网段一致，以免分配错误的 IP 地址。

2.1.3 DHCP 服务器分配 IP 地址的优先次序

DHCP 服务器为客户端分配 IP 地址的优先次序如下：

(1) DHCP 服务器中与客户端 MAC 地址或客户端 ID 静态绑定的 IP 地址。

(2) DHCP 服务器记录的曾经分配给客户端的 IP 地址。

(3) 客户端发送的 DHCP-DISCOVER 报文中 Option 50 字段指定的 IP 地址。

(4) 选择合适的普通模式、动态分配地址池，从中顺序查找可供分配的 IP 地址，最先找到的 IP 地址。

(5) 如果未找到可用的 IP 地址，则依次查询租约过期、曾经发生过冲突的 IP 地址，如果找到则进行分配，否则将不予处理。

说明

Option 50 为客户端请求的 IP 地址选项(Requested IP Address)，客户端通过在 DHCP-DISCOVER 报文中添加该选项来指明客户端希望获取的 IP 地址。该选项的内容由客户端决定。

2.2 DHCP 服务器配置任务简介

表2-1 DHCP 服务器配置任务简介

操作	说明	详细配置
配置 DHCP 服务器的地址池	必选	2.3
使能 DHCP 服务	必选	2.4

操作	说明	详细配置
配置接口工作在 DHCP 服务器模式	必选	2.5
配置 DHCP 服务器的安全功能	可选	2.6
配置 Option 82 的处理方式	可选	2.7

2.3 配置 DHCP 服务器的地址池

2.3.1 DHCP 服务器地址池配置任务简介

表2-2 DHCP 服务器地址池配置任务简介

操作		说明	详细配置
创建 DHCP 地址池		必选	2.3.2
配置普通模式 DHCP 地址池的地址分配方式	配置采用静态绑定方式进行地址分配	对于普通模式地址池，二者必选其一，且对同一个地址池只能选一种方式	2.3.3
	配置采用动态分配方式进行地址分配		
配置 DHCP 客户端的域名后缀		可选	2.3.4
配置 DHCP 客户端的 DNS 服务器地址		可选	2.3.5
配置 DHCP 客户端的 WINS 服务器地址和 NetBIOS 节点类型		可选	2.3.6
配置 DHCP 客户端的 BIMS 服务器信息		可选	2.3.7
配置 DHCP 客户端的网关地址		可选	2.3.8
配置 DHCP 客户端的 TFTP 服务器地址及启动文件名		可选	2.3.9
配置 DHCP 自定义选项		可选	2.3.10

2.3.2 创建 DHCP 地址池

表2-3 创建 DHCP 地址池

操作	命令	说明
进入系统视图	**system-view**	-
创建 DHCP 地址池并进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	必选 缺省情况下，没有创建 DHCP 地址池

2.3.3 配置普通模式地址池的地址分配方式

注意

根据客户端的实际需要，可以将地址池配置为采用静态绑定或动态分配方式进行地址分配，但对一个 DHCP 地址池不能同时配置这两种方式。

动态地址分配需要指定用于分配的地址范围，而静态地址绑定则可以看作是只包含一个地址的特殊的 DHCP 地址池。

1. 配置采用静态绑定方式进行地址分配

某些客户端（如 WWW 服务器等）需要固定的 IP 地址，可以通过将客户端的 MAC 地址与 IP 地址绑定的方式实现。当具有此 MAC 地址的客户端申请 IP 地址时，DHCP 服务器将根据客户端的 MAC 地址查找到对应的 IP 地址，并分配给客户端。

某些客户端在向 DHCP 服务器发送 DHCP-DISCOVER 报文申请 IP 地址时，会构建客户端 ID 并添加到报文中一起发送。如果在 DHCP 服务器上将客户端 ID 与 IP 地址绑定，则当该客户端申请 IP 地址时，DHCP 服务器将根据客户端 ID 查找到对应的 IP 地址并分配给客户端。

目前一个 DHCP 地址池中只能配置一个静态绑定，可以是 IP 地址与 MAC 地址的绑定，也可以是 IP 地址与客户端 ID 的绑定。

表2-4 配置采用静态绑定方式进行地址分配

操作		命令	说明
进入系统视图		**system-view**	-
进入普通模式地址池视图		**dhcp server ip-pool** *pool-name*	-
配置静态绑定的 IP 地址		**static-bind ip-address** *ip-address* [*mask-length* | **mask** *mask*]	必选 缺省情况下，没有配置静态绑定的 IP 地址
配置静态绑定的 MAC 地址或客户端 ID	配置静态绑定的客户端 MAC 地址	**static-bind mac-address** *mac-address*	二者必选其一 缺省情况下，没有配置静态绑定的 MAC 地址和客户端 ID
	配置静态绑定的客户端 ID	**static-bind client-identifier** *client-identifier*	

说明

- **static-bind ip-address** 和 **static-bind mac-address** 或 **static-bind client-identifier** 命令必须配合使用。
- 在同一个 DHCP 地址池中，如果配置了 **static-bind mac-address** 命令后，再配置 **static-bind client-identifier** 命令，则后面的配置将会覆盖前面的配置，反之亦然。
- 在同一个 DHCP 地址池中，如果多次执行 **static-bind ip-address** 或 **static-bind mac-address** 或 **static-bind client-identifier** 命令，新的配置会覆盖已有配置。
- 静态绑定的 IP 地址不能是 DHCP 服务器的接口 IP 地址，否则会导致 IP 地址冲突，被绑定的客户端将无法正常获取到 IP 地址。
- 静态绑定的客户端 ID，要与在待绑定客户端通过 **display dhcp client verbose** 命令显示的客户端 ID 一致。否则，客户端无法成功获取 IP 地址。
- 如果作为 DHCP 客户端的设备，接口的 MAC 地址相同，则为了区分不同接口，采用静态绑定方式进行地址分配时，需要在服务器上配置静态绑定的客户端 ID，而不能配置静态绑定的客户端 MAC 地址，否则可能导致客户端无法成功获取 IP 地址。

2. 配置采用动态分配方式进行地址分配

对于采用动态地址分配方式的地址池，需要配置该地址池可分配的地址范围，地址范围的大小通过掩码来设定。目前，一个地址池中只能配置一个地址段。

DHCP 服务器在分配地址时，需要排除已经被占用的 IP 地址（如网关、FTP 服务器等）。否则，同一地址分配给两个客户端会造成 IP 地址冲突。

对于不同的地址池，DHCP 服务器可以指定不同的地址租用期限，但同一 DHCP 地址池中的地址具有相同的期限。地址租用有效期限不具有继承关系。

表2-5 配置采用动态分配方式进行地址分配

操作	命令	说明
进入系统视图	**system-view**	-
进入普通模式地址池视图	**dhcp server ip-pool** *pool-name*	-
配置动态分配的 IP 地址范围	**network** *network-address* [*mask-length* | **mask** *mask*]	必选 缺省情况下，没有配置动态分配的 IP 地址范围，即没有可供分配的 IP 地址
配置动态分配的 IP 地址的租用有效期限	**expired** { **day** *day* [**hour** *hour* [**minute** *minute*]] | **unlimited** }	可选 缺省情况下，IP 地址租用有效期限为 1 天
退回系统视图	**quit**	-
配置 DHCP 地址池中不参与自动分配的 IP 地址	**dhcp server forbidden-ip** *low-ip-address* [*high-ip-address*]	可选 缺省情况下，除 DHCP 服务器接口的 IP 地址外，DHCP 地址池中的所有 IP 地址都参与自动分配

说明

- 在同一个 DHCP 地址池中，如果多次执行 **network** 命令，新的配置会覆盖已有配置。
- 通过 **dhcp server forbidden-ip** 命令指定不参与自动分配的 IP 地址后，所有动态分配方式的地址池都不能分配这些 IP 地址。
- 多次执行 **dhcp server forbidden-ip** 命令，可以配置多个不参与自动分配的 IP 地址段。

2.3.4 配置 DHCP 客户端的域名后缀

在 DHCP 服务器上，可以为每个地址池指定客户端使用的域名后缀。在给客户端分配 IP 地址的同时，也将域名后缀发送给客户端。

在客户端进行域名解析时，用户只需要输入域名的部分字段，客户端会自动将输入的域名加上域名后缀进行解析。有关域名后缀的详细介绍，请参见“IP 业务分册”中的“域名解析配置”。

表2-6 配置 DHCP 客户端的域名后缀

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的域名后缀	**domain-name** *domain-name*	必选 缺省情况下，没有配置为 DHCP 客户端分配的域名后缀

2.3.5 配置 DHCP 客户端的 DNS 服务器地址

通过域名访问 Internet 上的主机时，需要将域名解析为 IP 地址，这是通过 DNS（Domain Name System，域名系统）实现的。为了使 DHCP 客户端能够通过域名访问 Internet 上的主机，DHCP

服务器应在为客户端分配 IP 地址的同时指定 DNS 服务器地址。目前，每个 DHCP 地址池最多可以配置 8 个 DNS 服务器地址。

表2-7 配置 DHCP 客户端的 DNS 服务器地址

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的 DNS 服务器地址	**dns-list** *ip-address*&<1-8>	必选 缺省情况下，没有配置为 DHCP 客户端分配的 DNS 服务器地址

2.3.6 配置 DHCP 客户端的 WINS 服务器地址和 NetBIOS 节点类型

对于使用 Microsoft Windows 操作系统的客户端，由 WINS（Windows Internet Naming Service，Windows Internet 名称服务）服务器为通过 NetBIOS 协议通信的主机提供主机名到 IP 地址的解析。所以，大部分 Windows 网络客户端需要进行 WINS 的设置。

为了使 DHCP 客户端实现主机名到 IP 地址的解析，DHCP 服务器应在为客户端分配 IP 地址的同时指定 WINS 服务器地址。目前，每个 DHCP 地址池最多可以配置 8 个 WINS 服务器地址。

DHCP 客户端在网络上使用 NetBIOS 协议通信时，需要在主机名和 IP 地址之间建立映射关系。根据获取映射关系方式的不同，NetBIOS 节点分为四种：

- b 类节点（b-node）：“b”代表广播（broadcast），即此类节点采用广播方式获取映射关系。源节点通过发送带有目的节点主机名的广播报文来获取目的节点的 IP 地址，目的节点收到广播报文后，就将自己的 IP 地址返回给源节点。
- p 类节点（p-node）：“p”代表端到端（peer-to-peer），即此类节点采用发送单播报文与 WINS 服务器通信的方式获取映射关系。源节点给 WINS 服务器发送单播报文，WINS 服务器收到单播报文后，返回源节点请求的目的节点名所对应的 IP 地址。
- m 类节点（m-node）：“m”代表混合（mixed），是具有部分广播特性的 p 类节点。即此类节点首先发送广播报文来获取映射关系，如果没有获取到，则再发送单播报文与 WINS 服务器通信来获取映射关系。
- h 类节点（h-node）：“h”代表混合（hybrid），是具备“端到端”通信机制的 b 类节点。即此类节点首先发送单播报文与 WINS 服务器通信来获取映射关系，如果没有获取到，再发送广播报文来获取映射关系。

表2-8 配置 DHCP 客户端的 WINS 服务器地址和 NetBIOS 节点类型

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的 WINS 服务器地址	**nbns-list** *ip-address*&<1-8>	必选 对于 b 类节点，为可选 缺省情况下，没有配置为 DHCP 客户端分配的 WINS 服务器地址

操作	命令	说明
配置为 DHCP 客户端分配的 NetBIOS 节点类型	**netbios-type { b-node | h-node | m-node | p-node }**	必选 缺省情况下，没有配置为 DHCP 客户端分配的 NetBIOS 节点类型


说明

如果配置 DHCP 客户端为 b 类 NetBIOS 节点，则不需要配置 WINS 服务器地址。

2.3.7 配置 DHCP 客户端的 BIMS 服务器信息

为了使 DHCP 客户端通过 BIMS（Branch Intelligent Management System，分支网点智能管理系统）服务器进行软件的备份和升级等操作，DHCP 服务器需要在为 DHCP 客户端分配 IP 地址的同时，将 BIMS 服务器的 IP 地址、端口号以及加密的共享密钥等信息也发给 DHCP 客户端。之后，DHCP 客户端就可以定期向 BIMS 服务器发送连接请求，从 BIMS 服务器上获取配置文件，进行软件的备份和升级等操作。

表2-9 配置 DHCP 客户端的 BIMS 服务器信息

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的 BIMS 服务器的 IP 地址、端口及共享密钥信息	**bims-server ip** *ip-address* [**port** *port-number*] **sharekey** *key*	必选 缺省情况下，没有配置为 DHCP 客户端分配的 BIMS 服务器信息

2.3.8 配置 DHCP 客户端的网关地址

DHCP 客户端访问本网段以外的服务器或主机时，数据必须通过网关进行转发。DHCP 服务器可以在为客户端分配 IP 地址的同时指定网关的地址。

在 DHCP 服务器上，可以为每个地址池分别指定客户端对应的网关地址。在给客户端分配 IP 地址的同时，也将网关地址发送给客户端。目前，每个 DHCP 地址池最多可以配置 8 个网关地址。

表2-10 配置 DHCP 客户端的网关地址

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的网关地址	**gateway-list** *ip-address*&<1-8>	必选 缺省情况下，没有配置为 DHCP 客户端分配的网关地址

2.3.9 配置 DHCP 客户端的 TFTP 服务器地址及启动文件名

设备在空配置启动时自动获取并执行配置文件的功能，被称为自动配置。具体过程如下：

(1) 设备在空配置启动时，系统会自动将处于 up 状态的接口（如缺省 VLAN 对应的虚接口或三层以太网接口）设置为 DHCP 客户端，并向 DHCP 服务器获取 IP 地址及后续获取配置文件所需要的信息（例如：TFTP 服务器的 IP 地址、TFTP 服务器名、启动文件名等）。

(2) 如果获取到相关信息，则 DHCP 客户端就可发起 TFTP 请求，从指定的 TFTP 服务器获取配置文件，之后设备就使用获取到的配置文件进行设备初始化工作。如果没有获取到相关信息，则设备在空配置的情况下正常启动。

自动配置功能在空配置启动的设备上不需要进行任何配置，但需要在 DHCP 服务器上配置一些必需的参数，包括 TFTP 服务器地址、TFTP 服务器名和启动文件名。

在 DHCP 服务器收到 DHCP 客户端发来的请求报文后，如果请求报文中的 Option 55 中有 Option 66、Option 67 或 Option 150 的参数，则会在为客户端分配 IP 地址的同时，将所配置的 TFTP 服务器的 IP 地址、TFTP 服务器名或启动文件名等信息也发给 DHCP 客户端。之后，DHCP 客户端就可以从 TFTP 服务器获取配置文件，完成设备的自动配置。

表2-11 配置 DHCP 客户端的 TFTP 服务器地址及启动文件名

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-
配置为 DHCP 客户端分配的 TFTP 服务器地址	**tftp-server ip-address** *ip-address*	二者至少选择其一 缺省情况下，没有配置为 DHCP 客户端分配的 TFTP 服务器地址和 TFTP 服务器名
配置为 DHCP 客户端分配的 TFTP 服务器名	**tftp-server domain-name** *domain-name*	
配置为 DHCP 客户端分配的启动文件名	**bootfile-name** *bootfile-name*	必选 缺省情况下，没有配置为 DHCP 客户端分配的启动文件名

2.3.10 配置 DHCP 自定义选项

通过配置 DHCP 自定义选项，用户可以：

- 定义新的 DHCP 选项。随着 DHCP 的不断发展，新的可选项会陆续出现，为了支持这些新的选项，可以通过手工定义的方式将新选项添加到 DHCP 服务器的属性列表中。
- 定义已有选项的内容。有些选项的内容，RFC 2132 中没有统一规定。厂商可以根据需要定义选项的内容，如 Option 43。通过配置 DHCP 自定义选项，可以为 DHCP 客户端提供厂商指定的信息。
- 扩展已有的 DHCP 选项。当前已提供的方式无法满足用户需求时（比如通过 **dns-list** 命令最多只能配置 8 个 DNS 服务器地址，如果用户需要配置的 DNS 服务器地址数目大于 8，则该命令无法满足需求），可以通过配置 DHCP 自定义选项的方式进行扩展。

表2-12 配置 DHCP 自定义选项

操作	命令	说明
进入系统视图	**system-view**	-
进入 DHCP 地址池视图	**dhcp server ip-pool** *pool-name*	-

操作	命令	说明
配置 DHCP 自定义选项	**option** *code* { **ascii** *ascii-string* | **hex** *hex-string*&<1-16> | **ip-address** *ip-address*&<1-8> }	必选 缺省情况下，没有配置 DHCP 自定义选项

表2-13 常用 Option 配置说明

选项代码	选项名称	对应的配置命令	option 命令参数选择
3	Router Option	**gateway-list**	**ip-address**
6	Domain Name Server Option	**dns-list**	**ip-address**
15	Domain Name	**domain-name**	**ascii**
44	NetBIOS over TCP/IP Name Server Option	**nbns-list**	**ip-address**
46	NetBIOS over TCP/IP Node Type Option	**netbios-type**	**hex**
66	TFTP server name	**tftp-server**	**ascii**
67	Bootfile name	**bootfile-name**	**ascii**
43	Vendor Specific Information	-	**hex**

注意

配置自定义选项可能会对 DHCP 的工作过程造成影响，请谨慎使用 DHCP 自定义选项。

2.4 使能 DHCP 服务

只有使能 DHCP 服务后，其它相关的 DHCP 配置才能生效。

表2-14 使能 DHCP 服务

操作	命令	说明
进入系统视图	**system-view**	-
使能 DHCP 服务	**dhcp enable**	必选 缺省情况下，DHCP 服务处于禁止状态

2.5 配置接口工作在 DHCP 服务器模式

配置接口工作在 DHCP 服务器模式后，当接口收到 DHCP 客户端发来的 DHCP 报文时，将从 DHCP 服务器的地址池中分配地址。

表2-15 配置接口工作在 DHCP 服务器模式

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置接口工作在 DHCP 服务器模式	**dhcp select server global-pool [subaddress]**	可选 缺省情况下，接口工作在 DHCP 服务器模式

说明

subaddress 关键字只在 DHCP 服务器与客户端在同一网段时有效。如果客户端与服务器之间通过 DHCP 中继相连，则不管是否输入 **subaddress** 关键字，当服务器为客户端分配 IP 地址时，都是从与 DHCP 中继接口（与客户端相连的接口）的主 IP 地址在同一网段的地址池中选择地址分配给客户端。

说明

当 DHCP 服务器与客户端在同一网段时，根据是否输入 **subaddress** 关键字，DHCP 服务器的处理不同：

- 如果输入 **subaddress** 关键字，当服务器为客户端分配 IP 地址时，将从与服务器接口（与客户端相连的接口）的从 IP 地址在同一网段的地址池中选择地址分配给客户端。如果接口有多个从 IP 地址，则从第一个从 IP 地址依次匹配；如果接口上没有配置从 IP 地址，则服务器无法为客户端分配 IP 地址。
- 如果不输入 **subaddress** 关键字，当服务器为客户端分配 IP 地址时，将从与服务器接口（与客户端相连的接口）的主 IP 地址在同一网段的地址池中选择地址分配给客户端。

2.6 配置 DHCP 服务的安全功能

在配置 DHCP 服务器后，为了提高 DHCP 服务的安全性，需要配置 DHCP 服务的安全功能。

2.6.1 配置准备

在配置 DHCP 服务的安全功能之前，需完成 DHCP 服务器的必配任务：

- 使能 DHCP 服务；
- 配置 DHCP 服务器的地址池。

2.6.2 配置伪 DHCP 服务器检测功能

如果网络中有私自架设的 DHCP 服务器，当客户端申请 IP 地址时，这台 DHCP 服务器就会与 DHCP 客户端进行交互，导致客户端获得错误的 IP 地址。这种私设的 DHCP 服务器称为伪 DHCP 服务器。

使能伪 DHCP 服务器检测功能后，DHCP 服务器会从接收到的 DHCP 报文中获取给客户端分配 IP 地址的服务器 IP 地址，并记录此 IP 地址及接收到报文的接口信息，以便管理员及时发现并处理伪 DHCP 服务器。

表2-16 配置伪 DHCP 服务器检测功能

操作	命令	说明
进入系统视图	**system-view**	-
使能伪 DHCP 服务器检测功能	**dhcp server detect**	必选 缺省情况下，禁止伪 DHCP 服务器检测功能



说明

- 使能伪 DHCP 服务器检测功能后，对所有 DHCP 服务器都会进行记录，包括合法的 DHCP 服务器。管理员需要从日志信息中查找伪 DHCP 服务器。
- 使能伪 DHCP 服务器检测功能后，对每个 DHCP 服务器只记录一次。

2.6.3 配置 IP 地址重复分配检测功能

为防止 IP 地址重复分配导致地址冲突，DHCP 服务器为客户端分配地址前，需要先对该地址进行探测。

地址探测是通过 ping 功能实现的，通过检测是否能在指定时间内得到 ping 响应来判断是否有地址冲突。DHCP 服务器发送目的地址为待分配地址的 ICMP 回显请求报文，如果在指定时间内收到回显响应报文，DHCP 服务器从地址池中选择新的 IP 地址，并重复上述操作；如果在指定时间内没有收到回显响应报文，则继续发送 ICMP 回显请求报文，直到发送的回显请求报文达到最大值，如果仍然没有收到回显响应报文，则将地址分配给客户端，从而确保客户端被分得的 IP 地址唯一。

DHCP 服务器通过 ping 操作来检测是否发生地址冲突，而 DHCP 客户端则通过发送免费 ARP 报文检测是否发生地址冲突。

表2-17 配置 IP 地址重复分配检测功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 DHCP 服务器发送回显请求报文的最大数目	**dhcp server ping packets** *number*	可选 缺省情况下，DHCP 服务器发送回显请求报文的最大数目为 1 0 表示不进行 ping 操作
配置 DHCP 服务器等待回显响应报文的超时时间	**dhcp server ping timeout** *milliseconds*	可选 缺省情况下，DHCP 服务器等待回显响应报文的超时时间为 500 毫秒 0 表示不进行 ping 操作

2.7 配置 Option 82 的处理方式

如果配置 DHCP 服务器处理 Option 82，则当 DHCP 服务器收到带有 Option 82 的报文后，会在响应报文中携带 Option 82，并为客户端分配 IP 地址等信息。

如果配置 DHCP 服务器忽略 Option 82，则当 DHCP 服务器收到带有 Option 82 的报文后，不会在响应报文中携带 Option 82，只为客户端分配 IP 地址等信息。

1. 配置准备

在配置 Option 82 的处理方式之前，需完成 DHCP 服务器的必配任务：

- 使能 DHCP 服务；
- 配置 DHCP 服务器的地址池。

2. 配置 Option 82 的处理方式

表2-18 配置 Option 82 的处理方式

操作	命令	说明
进入系统视图	**system-view**	-
配置 DHCP 服务器处理 Option 82	**dhcp server relay information enable**	可选 缺省情况下，DHCP 服务器处理 Option 82



说明

为使Option 82 功能正常使用，需要在DHCP服务器和DHCP中继上都进行相应配置。DHCP中继支持Option 82 功能的相关配置请参见“3.3.6 配置DHCP中继支持Option 82 功能”。

2.8 DHCP 服务器显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 DHCP 服务器的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令清除 DHCP 服务器的相关信息。

表2-19 DHCP 服务器显示和维护

操作	命令
显示 DHCP 的地址冲突统计信息	**display dhcp server conflict** { **all** | **ip** *ip-address* }
显示 DHCP 地址池中的租约超期信息	**display dhcp server expired** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }
显示 DHCP 地址池的可用地址信息	**display dhcp server free-ip**
显示DHCP地址池中不参与自动分配的IP地址	**display dhcp server forbidden-ip**
显示 DHCP 地址池中的地址绑定信息	**display dhcp server ip-in-use** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }
显示 DHCP 服务器的统计信息	**display dhcp server statistics**
显示 DHCP 地址池的树状结构信息	**display dhcp server tree** { **all** | **pool** [*pool-name*] }
清除 DHCP 地址冲突的统计信息	**reset dhcp server conflict** { **all** | **ip** *ip-address* }
清除 DHCP 动态地址绑定信息	**reset dhcp server ip-in-use** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }
清除 DHCP 服务器的统计信息	**reset dhcp server statistics**

说明

执行 **save** 命令时不会保存 DHCP 服务器的租约信息，故当系统重新启动或使用 **reset dhcp server ip-in-use** 命令清除租约后，配置文件中将没有任何租约信息。此时客户端如果发出续约请求将会被拒绝，客户端需要重新申请 IP 地址。

2.9 DHCP 服务器典型配置举例

常见的 DHCP 组网方式可分为两类：一种是 DHCP 服务器和客户端在同一个子网内，直接进行 DHCP 报文的交互；另一种是 DHCP 服务器和客户端处于不同的子网中，必须通过 DHCP 中继代理实现 IP 地址的分配。无论哪种情况下，DHCP 服务器的配置都是相同的。

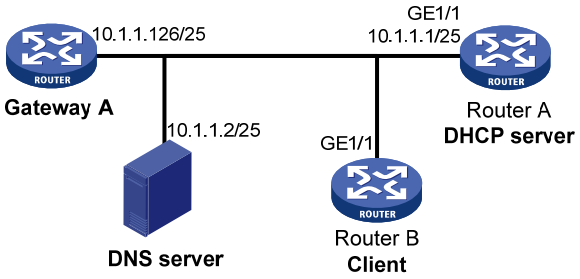
2.9.1 静态绑定地址典型配置举例

1. 组网需求

Router B 作为 DHCP 客户端，从 DHCP 服务器 Router A 获取静态绑定的 IP 地址、域名服务器、网关地址等信息。

2. 组网图

图2-1 静态绑定地址组网图



3. 配置步骤

(1) 配置接口的 IP 地址

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 10.1.1.1 25
[RouterA-GigabitEthernet1/1] quit
```

(2) 配置 DHCP 服务

使能 DHCP 服务。

```
[RouterA] dhcp enable
```

配置 DHCP 地址池 0，采用静态绑定方式分配 IP 地址。

```
[RouterA] dhcp server ip-pool 0
[RouterA-dhcp-pool-0] static-bind ip-address 10.1.1.5
[RouterA-dhcp-pool-0] static-bind mac-address 000f-e200-0002
[RouterA-dhcp-pool-0] dns-list 10.1.1.2
[RouterA-dhcp-pool-0] gateway-list 10.1.1.126
```

```
[RouterA-dhcp-pool-0] quit
```

2.9.2 动态分配地址典型配置举例

1. 组网需求

- 作为 DHCP 服务器的 Router A 为网段 10.1.1.0/24 中的客户端动态分配 IP 地址，该地址池网段分为两个子网网段：10.1.1.0/25 和 10.1.1.128/25；
- Router A 的两个以太网接口，Gigabit Ethernet1/1 和 Gigabit Ethernet1/2 的地址分别为 10.1.1.1/25 和 10.1.1.129/25；
- 10.1.1.0/25 网段内的地址租用期限为 10 天 12 小时，域名后缀为 aabbcc.com，DNS 服务器地址为 10.1.1.2/25，WINS 服务器地址为 10.1.1.4/25，网关的地址为 10.1.1.126/25；
- 10.1.1.128/25 网段内的地址租用期限为 5 天，域名后缀为 aabbcc.com，DNS 服务器地址为 10.1.1.2/25，无 WINS 服务器地址，网关的地址为 10.1.1.254/25。
- 10.1.1.0/25 网段与 10.1.1.128/25 网段的域名后缀、DNS 服务器地址相同，可以只配置 10.1.1.0/24 网段的域名后缀和 DNS 服务器地址，10.1.1.0/25 网段与 10.1.1.128/25 网段继承 10.1.1.0/24 网段的配置。

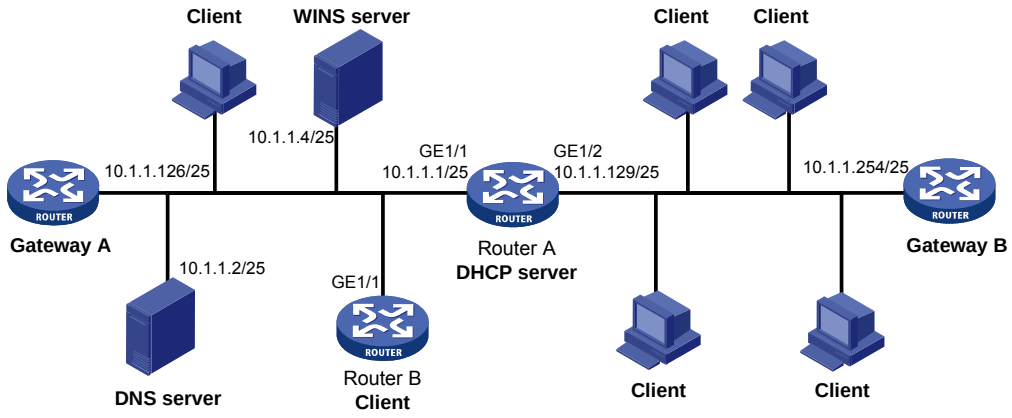
说明

在本例中，建议从 Gigabit Ethernet1/1 接口申请 IP 地址的客户端数目不要超过 122 个；从 Gigabit Ethernet1/2 接口申请 IP 地址的客户端不要超过 124 个。

---

2. 组网图

图2-2 DHCP 组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 DHCP 服务

使能 DHCP 服务。

```
<RouterA> system-view
[RouterA] dhcp enable
```

配置不参与自动分配的 IP 地址（DNS 服务器、WINS 服务器和网关地址）。

```
[RouterA] dhcp server forbidden-ip 10.1.1.2
[RouterA] dhcp server forbidden-ip 10.1.1.4
[RouterA] dhcp server forbidden-ip 10.1.1.126
[RouterA] dhcp server forbidden-ip 10.1.1.254
```

配置 DHCP 父地址池 0 的共有属性（地址池范围、客户端域名后缀、DNS 服务器地址）。

```
[RouterA] dhcp server ip-pool 0
[RouterA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[RouterA-dhcp-pool-0] domain-name aabbcc.com
[RouterA-dhcp-pool-0] dns-list 10.1.1.2
[RouterA-dhcp-pool-0] quit
```

配置 DHCP 子地址池 1 的属性（地址池范围、网关、地址租用期限、WINS 服务器地址）。

```
[RouterA] dhcp server ip-pool 1
[RouterA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.128
[RouterA-dhcp-pool-1] gateway-list 10.1.1.126
[RouterA-dhcp-pool-1] expired day 10 hour 12
[RouterA-dhcp-pool-1] nbns-list 10.1.1.4
[RouterA-dhcp-pool-1] quit
```

配置 DHCP 子地址池 2 的属性（地址池范围、地址租用期限、网关）。

```
[RouterA] dhcp server ip-pool 2
[RouterA-dhcp-pool-2] network 10.1.1.128 mask 255.255.255.128
[RouterA-dhcp-pool-2] expired day 5
[RouterA-dhcp-pool-2] gateway-list 10.1.1.254
```

2.9.3 自定义选项典型配置举例

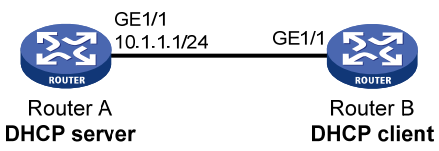
1. 组网需求

- DHCP 客户端 Router B 从 DHCP 服务器 Router A 获取 IP 地址和 PXE 引导服务器地址信息；
- Router B 获取的 IP 地址为 10.1.1.0/24 网段的地址；
- Router B 获取的 PXE 引导服务器地址为 1.2.3.4 和 2.2.2.2。

DHCP服务器需要通过自定义选项的方式配置Option 43 的内容，从而实现为客户端分配PXE引导服务器地址。Option 43 和PXE服务器地址列表的格式分别如 图 1-5和 图 1-7。DHCP服务器上配置的 Option 43 选项内容为 80 0B 00 00 02 01 02 03 04 02 02 02 02，其中 80 为子选项类型（Sub-option type），0B为子选项长度（Sub-option length），00 00 为PXE服务器类型（PXE server type），02 为服务器数目（Server number），01 02 03 04 02 02 02 02 为服务器的IP地址 1.2.3.4 和 2.2.2.2。

2. 组网图

图2-3 自定义选项典型配置举例



3. 配置步骤

(1) 配置接口 GigabitEthernet1/1 的 IP 地址（略）

(2) 配置 DHCP 服务

使能 DHCP 服务。

```
<RouterA> system-view
[RouterA] dhcp enable
```

配置 DHCP 地址池 0。

```
[RouterA] dhcp server ip-pool 0
[RouterA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[RouterA-dhcp-pool-0] option 43 hex 80 0B 00 00 02 01 02 03 04 02 02 02 02
```

2.10 DHCP 服务器常见配置错误举例

1. 故障现象

客户端从 DHCP 服务器动态获得的 IP 地址与其他主机 IP 地址冲突。

2. 故障分析

可能是网络上有主机私自配置了 IP 地址，导致冲突。

3. 处理过程

(1) 断开客户端的网线，从另外一台主机执行 ping 操作，设置较长超时时间，检查网络中是否已经存在该 IP 地址的主机。

(2) 如果能够收到 ping 操作的响应消息，则说明该 IP 地址已由用户静态配置。在 DHCP 服务器上执行 **dhcp server forbidden-ip** 命令，禁止该 IP 地址参与动态地址分配。

(3) 连接好客户端的网线，在客户端释放并重新获取 IP 地址。以 Windows XP 为例，在 Windows 环境下运行 **cmd** 进入 DOS 环境，使用 **ipconfig/release** 命令释放 IP 地址，之后使用 **ipconfig/renew** 重新获取 IP 地址。

3 DHCP 中继配置

说明

DHCP 中继中对于接口的相关配置，目前只能在三层以太网接口（包括子接口）、串口上进行。

3.1 DHCP 中继简介

3.1.1 DHCP 中继的应用环境

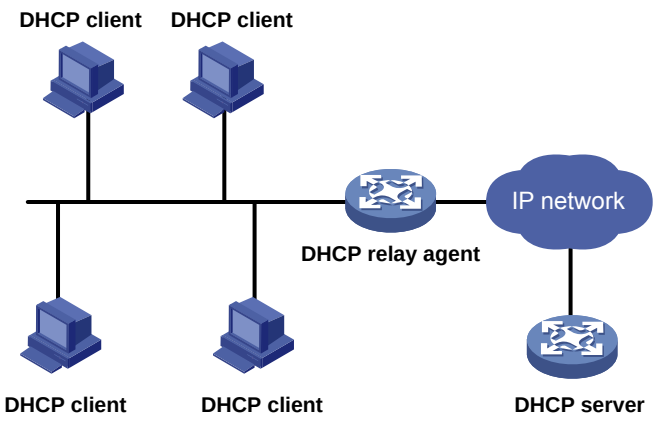
由于在 IP 地址动态获取过程中采用广播方式发送请求报文，因此 DHCP 只适用于 DHCP 客户端和服务器处于同一个子网内的情况。为进行动态主机配置，需要在所有网段上都设置一个 DHCP 服务器，这显然是很不经济的。

DHCP 中继功能的引入解决了这一难题：客户端可以通过 DHCP 中继与其他网段的 DHCP 服务器通信，最终获取到 IP 地址。这样，多个网络上的 DHCP 客户端可以使用同一个 DHCP 服务器，既节省了成本，又便于进行集中管理。

3.1.2 DHCP 中继的基本原理

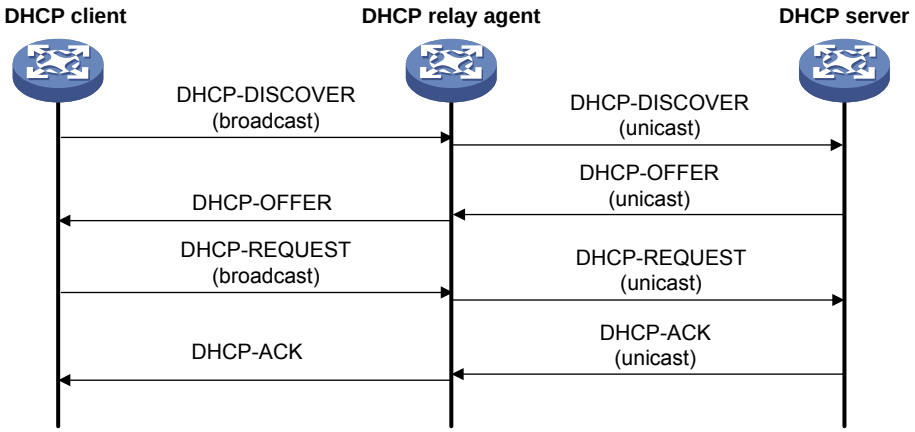
图 3-1是DHCP中继的典型应用示意图。

图3-1 DHCP 中继的典型组网应用



通过DHCP中继完成动态配置的过程中，DHCP客户端与DHCP服务器的处理方式与不通过DHCP中继时的处理方式基本相同。下面只说明DHCP中继的转发过程，报文的具体交互过程请参见“1.2.2 IP地址动态获取过程”。

图3-2 DHCP 中继的工作过程



如 图 3-2所示，DHCP中继的工作过程为：

(1) 具有 DHCP 中继功能的网络设备收到 DHCP 客户端以广播方式发送的 DHCP-DISCOVER 或 DHCP-REQUEST 报文后，将报文中的 giaddr 字段填充为 DHCP 中继的 IP 地址，并根据配置将报文单播转发给指定的 DHCP 服务器。

(2) DHCP 服务器根据 giaddr 字段为客户端分配 IP 地址等参数，并通过 DHCP 中继将配置信息转发给客户端，完成对客户端的动态配置。

3.1.3 DHCP 中继支持 Option 82 功能

Option 82 记录了DHCP客户端的位置信息。管理员可以利用该选项定位DHCP客户端，实现对客户端的安全和计费等控制。Option 82 的详细介绍请参见“1.4.3 2. 中继代理信息选项（Option 82）”。

如果DHCP中继支持Option 82 功能，则当DHCP中继接收到DHCP请求报文后，将根据报文中是否包含Option 82 以及用户配置的处理策略及填充模式对报文进行相应的处理，并将处理后的报文转发给DHCP服务器。具体的处理方式见 表 3-1。

如果 DHCP 中继收到的应答报文中带有 Option 82，则会将 Option 82 删除后再转发给 DHCP 客户端。

表3-1 DHCP 中继支持 Option 82 的处理方式

收到 DHCP 请求报文	处理策略	填充模式	DHCP 中继对报文的处理
收到的报文中带有 Option 82	Drop	任意	丢弃报文
	Keep	任意	保持报文中的 Option 82 不变并进行转发
	Replace	normal	采用 normal 模式填充 Option 82，替换报文中原有的 Option 82 并进行转发
		verbose	采用 verbose 模式填充 Option 82，替换报文中原有的 Option 82 并进行转发
		用户自定义	采用用户自定义的内容填充 Option 82，替换报文中原有的 Option 82 并进行转发
收到的报文中不带有 Option 82	-	normal	采用 normal 模式填充 Option 82 并进行转发
		verbose	采用 verbose 模式填充 Option 82 并进行转发
		用户自定义	采用用户自定义的内容填充 Option 82 并进行转发

3.2 DHCP 中继配置任务简介

表3-2 DHCP 中继配置任务简介

配置任务	说明	详细配置
使能 DHCP 服务	必选	3.3.1
配置接口工作在 DHCP 中继模式	必选	3.3.2
配置 DHCP 服务器组	必选	3.3.3
配置 DHCP 中继的安全功能	可选	3.3.4
配置通过 DHCP 中继释放客户端的 IP 地址	可选	3.3.5
配置 DHCP 中继支持 Option 82 功能	可选	3.3.6

3.3 配置 DHCP 中继

3.3.1 使能 DHCP 服务

只有使能 DHCP 服务后，其它相关的 DHCP 配置才能生效。

表3-3 使能 DHCP 服务

操作	命令	说明
进入系统视图	**system-view**	-
使能 DHCP 服务	**dhcp enable**	必选 缺省情况下，DHCP 服务处于禁止状态

3.3.2 配置接口工作在 DHCP 中继模式

配置接口工作在中继模式后，当接口收到 DHCP 客户端发来的 DHCP 报文时，会将报文转发给 DHCP 服务器，由服务器分配地址。

表3-4 配置接口工作在 DHCP 中继模式

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口工作在 DHCP 中继模式	**dhcp select relay**	必选 缺省情况下，使能 DHCP 服务后，接口工作在 DHCP 服务器模式

说明

- 以太网子接口支持 DHCP 中继时，以太网子接口接收到的客户端报文中必须包含 VLAN Tag，且 VLAN Tag 与子接口的 VLAN ID 一致，否则客户端报文将被丢弃。
- DHCP 客户端通过 DHCP 中继获取 IP 地址时，DHCP 服务器上需要配置与 DHCP 中继的 IP 地址所在网段（网络号和掩码）完全相同的地址池，否则会导致 DHCP 客户端无法获得正确的 IP 地址。

3.3.3 配置 DHCP 服务器组

为了提高可靠性，可以在一个网络中设置多个 DHCP 服务器。多个 DHCP 服务器构成一个 DHCP 服务器组。当接口与 DHCP 服务器组建立归属关系后，会将客户端发来的 DHCP 报文转发给服务器组中的所有服务器。

表3-5 配置 DHCP 服务器组

操作	命令	说明
进入系统视图	**system-view**	-
配置 DHCP 服务器组中 DHCP 服务器的 IP 地址	**dhcp relay server-group** *group-id* **ip** *ip-address*	必选 缺省情况下，没有配置 DHCP 服务器组中服务器的 IP 地址
进入接口视图	**interface** *interface-type interface-number*	-
配置接口与 DHCP 服务器组的归属关系	**dhcp relay server-select** *group-id*	必选 缺省情况下，接口没有与任何一个 DHCP 服务器组建立归属关系

说明

- 设备上最多可以配置 20 个 DHCP 服务器组。每个 DHCP 服务器组最多可以配置 8 个 DHCP 服务器地址。
- DHCP 服务器组中服务器的 IP 地址不能与 DHCP 中继的接口 IP 地址在同一网段。否则，可能导致客户端无法获得 IP 地址。
- 每个 DHCP 服务器组可以对应多个接口。但每个接口只能对应一个 DHCP 服务器组。在同一接口下多次执行 **dhcp relay server-select** 命令，新的配置会覆盖已有配置。但是，如果新指定的 DHCP 服务器组不存在时，新的归属关系配置不成功，接口还是维持与上一次配置的 DHCP 服务器组的归属关系。
- **dhcp relay server-select** 命令中所指定的组号，需事先通过 **dhcp relay server-group** 命令进行配置。

3.3.4 配置 DHCP 中继的安全功能

1. DHCP 中继的地址匹配检查功能

当客户端通过 DHCP 中继从 DHCP 服务器获取到 IP 地址时，DHCP 中继可以自动记录客户端 IP 地址与 MAC 地址的绑定关系，生成 DHCP 中继的动态用户地址表项。同时，为满足用户采用合法

固定 IP 地址访问外部网络的需求，DHCP 中继也支持静态用户地址表项配置，即在 DHCP 中继上手工配置 IP 地址与 MAC 地址的绑定关系。

为了防止非法主机静态配置一个 IP 地址并访问其他网络，设备支持 DHCP 中继的地址匹配检查功能。接口上使能该功能后，如果在 DHCP 中继的用户地址表中（包括 DHCP 中继动态记录的表项以及手工配置的用户地址表项）没有与主机 IP 地址和主机 MAC 地址匹配的表项，则该主机将不能通过 DHCP 中继访问外部网络。

表3-6 配置 DHCP 中继的地址匹配检查功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 DHCP 中继的静态用户地址表项	**dhcp relay security static** *ip-address mac-address* [**interface** *interface-type interface-number*]	可选 缺省情况下，没有配置 DHCP 中继的静态用户地址表项
进入接口视图	**interface** *interface-type interface-number*	-
配置 DHCP 中继的地址匹配检查功能	**dhcp relay address-check { disable | enable }**	必选 缺省情况，禁止 DHCP 中继的地址匹配检查功能

说明

- 目前，只能在三层以太网接口（包括子接口）上执行 **dhcp relay address-check** 命令。
- DHCP 中继的地址匹配检查功能与 DHCP 中继的其他配置无直接关系。即只要执行了 **dhcp relay address-check enable** 命令，地址匹配检查功能就可以生效，不需要配置 DHCP 中继的其他功能，如使能 DHCP、配置接口工作在 DHCP 中继模式等。
- 执行 **dhcp relay address-check enable** 命令后将只检查 IP 和 MAC 地址，不检查接口。
- 建议只在工作于 DHCP 中继模式的接口上配置地址匹配检查功能。否则，可能会导致合法的用户无法访问网络。
- 通过 **dhcp relay security static** 命令配置静态用户地址表项时，如果静态用户地址表项与接口绑定，配置的接口必须工作在 DHCP 中继模式，否则可能引起地址表项冲突。
- 同异步串口作为 DHCP 客户端申请 IP 地址时，DHCP 中继不会记录该客户端对应的用户地址表项。

2. DHCP 中继动态用户地址表项定时刷新功能

当 DHCP 客户端通过 DHCP 中继从 DHCP 服务器获取到 IP 地址时，DHCP 中继会记录 IP 地址与 MAC 地址的绑定关系。由于 DHCP 客户端释放该 IP 地址时，会给 DHCP 服务器发送单播 DHCP-RELEASE 报文，DHCP 中继不会处理该 DHCP 报文的内容，造成 DHCP 中继的用户地址项不能被实时刷新。为了解决这个问题，DHCP 中继支持动态用户地址表项的定时刷新功能。

每隔指定时间，DHCP 中继以客户端分配到的 IP 地址和 DHCP 中继接口的 MAC 地址向 DHCP 服务器发送 DHCP-REQUEST 报文：

- 如果 DHCP 中继接收到 DHCP 服务器响应的 DHCP-ACK 报文或在指定时间内没有接收到 DHCP 服务器的响应报文，则表明这个 IP 地址已经可以进行分配，DHCP 中继会将动态用户地址表中对应的表项老化掉；

- 如果 DHCP 中继接收到 DHCP 服务器响应的 DHCP-NAK 报文，则表示该 IP 地址的租约仍然存在，DHCP 中继不会老化该 IP 地址对应的表项。

表3-7 配置 DHCP 中继动态用户地址表项定时刷新周期

操作	命令	说明
进入系统视图	**system-view**	-
配置 DHCP 中继动态用户地址表项的定时刷新周期	**dhcp relay security tracker** { *interval* | **auto** }	可选 缺省情况下，定时刷新周期为 **auto**，即根据表项的数目自动计算刷新时间间隔

3. 配置伪 DHCP 服务器检测功能

如果网络中有私自架设的 DHCP 服务器，当客户端申请 IP 地址时，这台 DHCP 服务器就会与 DHCP 客户端进行交互，导致客户端获得错误的 IP 地址，这种私设的 DHCP 服务器称为伪 DHCP 服务器。

使能伪 DHCP 服务器检测功能后，DHCP 中继会从接收到的 DHCP 报文中获取给客户端分配 IP 地址的服务器 IP 地址，并记录此 IP 地址及接收到报文的接口信息，以便管理员及时发现并处理伪 DHCP 服务器。

表3-8 配置伪 DHCP 服务器检测功能

操作	命令	说明
进入系统视图	**system-view**	-
使能伪 DHCP 服务器检测功能	**dhcp relay server-detect**	必选 缺省情况下，禁止伪 DHCP 服务器检测功能

说明

- 使能伪 DHCP 服务器检测功能后，对所有 DHCP 服务器都会进行记录，包括合法的 DHCP 服务器，管理员需要从日志信息中查找伪 DHCP 服务器。
- 使能伪 DHCP 服务器检测功能后，对每个 DHCP 服务器只记录一次。记录的 DHCP 服务器信息被清除后，将重新记录。

3.3.5 配置通过 DHCP 中继释放客户端的 IP 地址

在某些情况下，可能需要通过 DHCP 中继手工释放客户端申请到的 IP 地址。配置通过 DHCP 中继释放客户端的 IP 地址后，DHCP 中继会主动向 DHCP 服务器发送 DHCP-RELEASE 报文，DHCP 服务器收到该报文后，将会释放指定 IP 地址的租约，同时 DHCP 中继上也会删除对应的用户地址表项。

表3-9 配置通过 DHCP 中继释放客户端的 IP 地址

操作	命令	说明
进入系统视图	**system-view**	-
向 DHCP 服务器请求释放客户端申请到的 IP 地址	**dhcp relay release ip** *client-ip*	必选

3.3.6 配置 DHCP 中继支持 Option 82 功能

1. 配置准备

在配置 DHCP 中继支持 Option82 功能之前，需完成 DHCP 中继的必配任务，即：

- 使能 DHCP 服务
- 配置接口工作在 DHCP 中继模式
- 配置 DHCP 服务器组

2. 配置 DHCP 中继支持 Option 82 功能

表3-10 配置 DHCP 中继支持 Option 82 功能

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图		**interface** *interface-type interface-number*	-
配置 DHCP 中继支持 Option 82 功能		**dhcp relay information enable**	必选 缺省情况下，DHCP 中继不支持 Option 82 功能
配置DHCP中继对包含Option 82的请求报文的处理策略		**dhcp relay information strategy** { **drop** | **keep** | **replace** }	可选 缺省情况下，处理策略为 **replace**
配置非用户自定义的 Option 82	配置 Option 82 的填充模式	**dhcp relay information format** { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] }	可选 缺省情况下，Opiton 82 的填充模式为 normal
	配置 Circuit ID 子选项的填充格式	**dhcp relay information circuit-id format-type** { **ascii** | **hex** }	可选 缺省情况下，Circuit ID 子选项的填充格式由 Option 82 的填充模式决定，每个字段的填充格式不同 配置的填充格式只对非用户自定义的填充内容有效
	配置 Remote ID 子选项的填充格式	**dhcp relay information remote-id format-type** { **ascii** | **hex** }	可选 缺省情况下，采用 HEX 格式填充 Remote ID 子选项 配置的填充格式只对非用户自定义的填充内容有效
配置用户自定义的 Option 82	配置 Circuit ID 子选项的内容	**dhcp relay information circuit-id string** *circuit-id*	可选 缺省情况下，Circuit ID 子选项的内容由 Option 82 的填充模式决定
	配置 Remote ID 子选项的内容	**dhcp relay information remote-id string** { *circuit-id* | **sysname** }	可选 缺省情况下，Remote ID 子选项的内容由 Option 82 的填充模式决定

说明

- 为使Option 82 功能正常使用，需要在DHCP服务器和DHCP中继上都进行相应配置。DHCP服务器的相关配置请参见“2.7 配置Option 82 的处理方式”。
- DHCP 中继对包含 Option 82 请求报文的处理策略为 **replace** 时，需要配置 Option 82 的填充格式；处理策略为 **keep** 或 **drop** 时，不需要配置 Option 82 的填充格式。
- 如果以节点的设备名称（**sysname**）作为节点标识填充 DHCP 报文的 Option 82，则设备名称中不能包含空格；否则，DHCP 中继将丢弃该报文。

3.4 DHCP 中继显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 DHCP 中继的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令清除 DHCP 中继的统计信息。

表3-11 DHCP 中继显示和维护

操作	命令
显示接口对应的 DHCP 服务器组的信息	**display dhcp relay** { **all** | **interface** *interface-type interface-number* }
显示 DHCP 中继上 Option 82 的配置信息	**display dhcp relay information** { **all** | **interface** *interface-type interface-number* }
显示 DHCP 中继的用户地址表项信息	**display dhcp relay security** [*ip-address* | **dynamic** | **static**]
显示 DHCP 中继用户地址表项的统计信息	**display dhcp relay security statistics**
显示 DHCP 中继动态用户地址表项的定时刷新周期	**display dhcp relay security tracker**
显示 DHCP 服务器组中服务器的 IP 地址。	**display dhcp relay server-group** { *group-id* | **all** }
显示 DHCP 中继的相关报文统计信息	**display dhcp relay statistics** [**server-group** { *group-id* | **all** }]
清除 DHCP 中继的统计信息	**reset dhcp relay statistics** [**server-group** *group-id*]

3.5 DHCP 中继典型配置举例

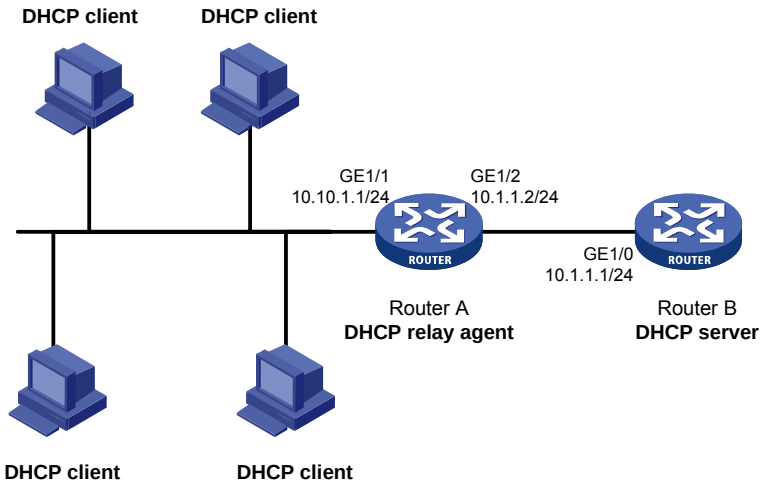
3.5.1 DHCP 中继配置举例

1. 组网需求

- 具有 DHCP 中继功能的 Router A 通过 Gigabit Ethernet1/1 接口连接到 DHCP 客户端所在的网络，Gigabit Ethernet1/1 接口的 IP 地址为 10.10.1.1/24，Gigabit Ethernet1/2 的 IP 地址为 10.1.1.2/24；
- DHCP 服务器的 IP 地址为 10.1.1.1/24；
- 通过 Router A 转发 DHCP 报文，DHCP 客户端可以从 DHCP 服务器上申请到 10.10.1.0/24 网段的 IP 地址及相关配置信息。

2. 组网图

图3-3 DHCP 中继组网示意图



3. 配置步骤

配置各接口的 IP 地址（略）。

使能 DHCP 服务。

```
<RouterA> system-view
[RouterA] dhcp enable
```

配置 DHCP 服务器的地址。

```
[RouterA] dhcp relay server-group 1 ip 10.1.1.1
```

配置 Gigabit Ethernet1/1 接口工作在 DHCP 中继模式。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] dhcp select relay
```

配置 Gigabit Ethernet1/1 接口对应 DHCP 服务器组 1。

```
[RouterA-GigabitEthernet1/1] dhcp relay server-select 1
```

说明

- 为了使DHCP客户端能从DHCP服务器获得IP地址，还需要在DHCP服务器上进行一些配置。DHCP服务器的配置方法，请参见“2.9 DHCP服务器典型配置举例”。
- 由于 DHCP 中继所在接口的 IP 地址与 DHCP 服务器的 IP 地址不在同一网段，因此需要在 DHCP 服务器上通过静态路由或动态路由协议保证两者之间路由可达。

3.5.2 DHCP 中继支持 Option 82 配置举例

1. 组网需求

- 在 DHCP 中继 Router A 上使能 Option 82 功能；
- 对包含 Option 82 的请求报文的处理策略为 **replace**；
- Ciruict ID 填充内容为 company001，Remote ID 填充内容为 device001；

- Router A 将添加 Option 82 的 DHCP 请求报文转发给 DHCP 服务器 Router B，使得 DHCP 客户端可以获取到 IP 地址。

2. 组网图

如 图 3-3所示。

3. 配置步骤

配置各接口的 IP 地址（略）。

使能 DHCP 服务。

```
<RouterA> system-view
[RouterA] dhcp enable
```

配置 DHCP 服务器的地址。

```
[RouterA] dhcp relay server-group 1 ip 10.1.1.1
```

配置 GigabitEthernet1/1 接口工作在 DHCP 中继模式。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] dhcp select relay
```

配置 GigabitEthernet1/1 接口对应 DHCP 服务器组 1。

```
[RouterA-GigabitEthernet1/1] dhcp relay server-select 1
```

配置 Option 82 的处理策略和填充内容。

```
[RouterA-GigabitEthernet1/1] dhcp relay information enable
[RouterA-GigabitEthernet1/1] dhcp relay information strategy replace
[RouterA-GigabitEthernet1/1] dhcp relay information circuit-id string company001
[RouterA-GigabitEthernet1/1] dhcp relay information remote-id string device001
```



说明

为使 Option 82 功能正常使用，DHCP 服务器也需要进行相应配置。

3.6 DHCP 中继常见配置错误举例

1. 故障现象

客户端不能通过 DHCP 中继获得配置信息。

2. 故障分析

DHCP 中继或 DHCP 服务器的配置可能有问题。可以打开调试开关显示调试信息，并通过执行 **display** 命令显示接口状态信息的方法来分析定位。

3. 处理过程

- 检查 DHCP 服务器和 DHCP 中继是否使能了 DHCP 服务。
- 检查 DHCP 服务器是否配置有 DHCP 客户端所在网段的地址池。
- 检查具有 DHCP 中继功能的网络设备和 DHCP 服务器是否配置有相互可达的路由。

- 检查具有 DHCP 中继功能的网络设备是否在连接 DHCP 客户端所在网段的接口上配置有正确的 DHCP 服务器组，且 DHCP 服务器组的 IP 地址配置正确。

4 DHCP 客户端配置

说明

- DHCP 客户端中对于接口的相关配置，目前只能在三层以太网接口（包括子接口）上进行。
- 建议不要在同一台设备上同时配置 DHCP 客户端和 DHCP Snooping 功能，否则可能无法生成 DHCP Snooping 表项，DHCP 客户端也可能申请不到 IP 地址。

4.1 DHCP 客户端简介

指定设备的接口作为 DHCP 客户端后，可以使用 DHCP 协议从 DHCP 服务器动态获得 IP 地址等参数，方便用户配置，也便于集中管理。

4.2 配置接口使用 DHCP 方式获取 IP 地址

表4-1 配置接口使用 DHCP 方式获取 IP 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口使用 DHCP 方式获取 IP 地址	**ip address dhcp-alloc** [**client-identifier mac** *interface-type interface-number*]	必选 缺省情况下，接口不使用 DHCP 方式获取 IP 地址

说明

- 接口可以采用多种方式获得 IP 地址，但不同方式是互斥的，新的配置方式会覆盖原有的配置方式。
- 当接口被配置为通过 DHCP 动态获取 IP 地址后，不能再给该接口配置从 IP 地址。
- 如果 DHCP 服务器为接口分配的 IP 地址与设备上其他接口的 IP 地址在同一网段，则该接口不再向 DHCP 服务器申请 IP 地址，除非手动删除冲突接口的 IP 地址，并重新使能接口（先后执行 **shutdown** 和 **undo shutdown** 命令）或重新配置接口使用 DHCP 方式获取 IP 地址（先后执行 **undo ip address dhcp-alloc** 和 **ip address dhcp-alloc** 命令）。

4.3 DHCP 客户端显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 DHCP 客户端的信息，通过查看显示信息验证配置的效果。

表4-2 DHCP 客户端显示和维护

操作	命令
显示 DHCP 客户端的相关信息	**display dhcp client** [**verbose**] [**interface** *interface-type interface-number*]

4.4 DHCP 客户端典型配置举例

1. 组网需求

Router B 的以太网接口 GigabitEthernet1/1 接入局域网，通过 DHCP 协议从 DHCP 服务器获取 IP 地址。

2. 组网图

如 图 2-2所示。

3. 配置步骤

下面只列出 图 2-2中，作为客户端的Router B的配置。

配置接口 Ethernet1/1 通过 DHCP 动态获取地址。

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ip address dhcp-alloc
```

说明

为了使DHCP客户端能从DHCP服务器获得IP地址，还需要在DHCP服务器上进行一些配置，具体内容请参见“2.9 DHCP服务器典型配置举例”。

目 录

1 域名解析配置 1-1
1.1 域名解析简介 1-1
1.1.1 静态域名解析 1-1
1.1.2 动态域名解析 1-1
1.1.3 DNS代理 1-3
1.2 配置DNS client 1-4
1.2.1 配置静态域名解析 1-4
1.2.2 配置动态域名解析 1-4
1.3 配置DNS proxy 1-4
1.4 域名解析显示和维护 1-5
1.5 域名解析典型配置举例 1-5
1.5.1 静态域名解析配置举例 1-5
1.5.2 动态域名解析配置举例 1-6
1.5.3 DNS proxy典型配置举例 1-9
1.6 域名解析常见配置错误举例 1-10

1 域名解析配置

说明

本文只涉及 IPv4 DNS 的配置，有关 IPv6 DNS 的详细介绍请参见“IP 业务分册”中的“IPv6 基础配置”。

1.1 域名解析简介

域名系统（DNS，Domain Name System）是一种用于 TCP/IP 应用程序的分布式数据库，提供域名与 IP 地址之间的转换。通过域名系统，用户进行某些应用时，可以直接使用便于记忆的、有意义的域名，而由网络中的域名解析服务器将域名解析为正确的 IP 地址。

域名解析分为静态域名解析和动态域名解析，二者可以配合使用。在解析域名时，首先采用静态域名解析（查找静态域名解析表），如果静态域名解析不成功，再采用动态域名解析。由于动态域名解析可能会花费一定的时间，且需要域名服务器的配合，因而可以将一些常用的域名放入静态域名解析表中，这样可以大大提高域名解析效率。

1.1.1 静态域名解析

静态域名解析就是手工建立域名和 IP 地址之间的对应关系。当用户使用域名进行某些应用（如 telnet 应用）时，系统查找静态域名解析表，从中获取指定域名对应的 IP 地址。

1.1.2 动态域名解析

1. 解析过程

动态域名解析是通过对域名服务器的查询完成的。解析过程如下：

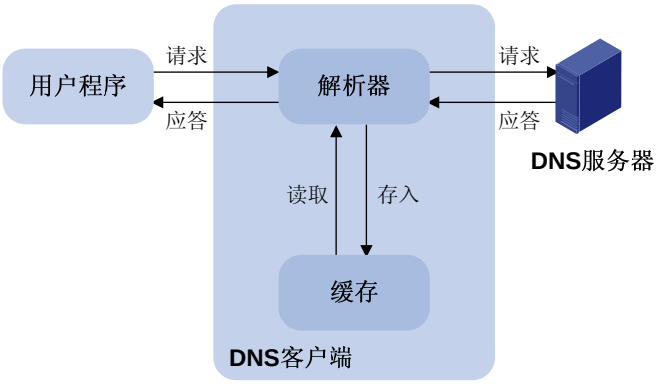
(1) 当用户使用域名进行某些应用时，用户程序首先向 DNS 客户端中的解析器发出请求。

(2) DNS 客户端收到请求后，首先查询本地的域名缓存。如果存在已解析成功的映射项，就将域名对应的 IP 地址返回给用户程序；如果没有发现所要查找的映射项，就向域名服务器（DNS Server）发送查询请求。

(3) 域名服务器首先从自己的数据库中查找域名对应的 IP 地址。如果判断该域名不属于本域范围之内，就将请求交给上一级的域名解析服务器处理，直到完成解析，并将解析的结果返回给 DNS 客户端。

(4) DNS 客户端收到域名服务器的响应报文后，将解析结果返回给应用程序。

图1-1 动态 DNS



用户程序、DNS客户端及域名服务器的关系如 图 1-1所示，其中解析器和缓存构成DNS客户端。用户程序、DNS客户端在同一台设备上，而DNS客户端和服务器一般分布在两台设备上。

动态域名解析支持缓存功能。每次动态解析成功的域名与 IP 地址的映射均存放在动态域名缓存区中，当下一次查询相同域名的时候，就可以直接从缓存区中读取，不用再向域名服务器进行请求。缓存区中的映射在一段时间后会被老化删除，以保证及时从域名服务器得到最新的内容。老化时间由域名服务器设置，DNS 客户端从协议报文中获得老化时间。

2. 域名后缀列表功能

动态域名解析支持域名后缀列表功能。用户可以预先设置一些域名后缀，在域名解析的时候，用户只需要输入域名的部分字段，系统会自动将输入的域名加上不同的后缀进行解析。举例说明，用户想查询域名 aabbcc.com，那么可以先在后缀列表中配置 com，然后输入 aabbcc 进行查询，系统会自动将输入的域名与后缀连接成 aabbcc.com 进行查询。

使用域名后缀的时候，根据用户输入域名方式的不同，查询方式分成以下几种情况：

- 如果用户输入的域名中没有“.”，比如 aabbcc，系统认为这是一个主机名，会首先加上域名后缀进行查询，如果所有加后缀的域名查询都失败，将使用最初输入的域名（如 aabbcc）进行查询。
- 如果用户输入的域名中间有“.”，比如 www.aabbcc，系统直接用它进行查询，如果查询失败，再依次加上各个域名后缀进行查询。
- 如果用户输入的域名最后有“.”，比如 aabbcc.com.，表示不需要进行域名后缀添加，系统直接用输入的域名进行查询，不论成功与否都直接返回。就是说，如果用户输入的字符中最后一个字符为“.”，就只根据用户输入的字符进行查找，而不会去匹配用户预先设置的域名后缀，因此最后这个“.”，也被称为查找终止符。带有查询终止符的域名，称为绝对域名或完全合格的域名 FQDN（Full Qualified Domain Name）。

目前，设备支持静态域名解析和动态域名解析的客户端功能。

说明

如果域名服务器上配置了域名的别名，设备可以通过别名来解析主机的 IP 地址。

1.1.3 DNS 代理

1. DNS 代理简介

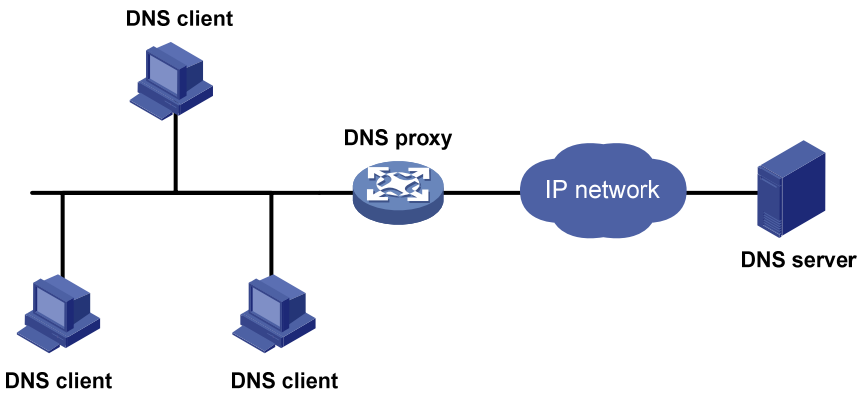
DNS 代理（DNS proxy）用来在 DNS client 和 DNS server 之间转发 DNS 请求和应答报文，代替 DNS client 进行域名解析。

局域网内的 DNS client 把 DNS proxy 当作 DNS server，将 DNS 请求报文发送给 DNS proxy。DNS proxy 将该请求报文转发到真正的 DNS server，并将 DNS server 的应答报文返回给 DNS client，从而实现域名解析。

使用 DNS proxy 功能后，当 DNS server 的地址发生变化时，只需改变 DNS proxy 上的配置，无需改变局域网内每个 DNS client 的配置，从而简化了网络管理。

DNS proxy的典型应用环境如 图 1-2所示。

图1-2 DNS 代理典型组网应用



2. DNS 代理的工作机制

DNS 代理的工作过程如下：

(1) DNS client 把 DNS proxy 当作 DNS server，将 DNS 请求报文发送给 DNS proxy，即请求报文的目的地址为 DNS proxy 的 IP 地址。

(2) DNS proxy 收到请求报文后，首先查找本地的静态域名解析表，如果静态域名解析表中存在请求的信息，则 DNS proxy 直接通过 DNS 应答报文，将域名解析结果返回给 DNS client。

(3) 如果静态域名解析表中没有请求的信息，则 DNS proxy 将报文转发给 DNS server，通过 DNS server 进行域名解析。

(4) DNS proxy 收到 DNS server 的应答报文后，将报文转发给 DNS client。DNS client 利用域名解析的结果进行相应的处理。

1.2 配置 DNS client

1.2.1 配置静态域名解析

表1-1 配置静态域名解析

操作	命令	说明
进入系统视图	**system-view**	-
配置静态域名解析表中主机名和对应 IP 地址	**ip host** *hostname ip-address*	必选 缺省情况下，静态域名解析表中没有主机名及其 IP 地址的对应关系



说明

- 每个主机名只能对应一个 IP 地址，当对同一主机名进行多次配置时，最后配置的 IP 地址有效。
- 最多可配置 50 条静态域名解析信息。

1.2.2 配置动态域名解析

表1-2 配置动态域名解析

操作	命令	说明
进入系统视图	**system-view**	-
开启动态域名解析功能	**dns resolve**	必选 缺省情况下，动态域名解析功能处于关闭状态
配置域名服务器的 IP 地址	**dns server** *ip-address*	必选 缺省情况下，没有配置域名服务器的 IP 地址
配置域名后缀	**dns domain** *domain-name*	可选 缺省情况下，没有配置域名后缀，即只根据用户输入的域名信息进行解析



说明

最多可配置 6 个域名服务器和 10 个域名后缀。

1.3 配置 DNS proxy

表1-3 配置 DNS proxy

操作	命令	说明
进入系统视图	**system-view**	-
开启 DNS proxy 功能	**dns proxy enable**	必选 缺省情况下，DNS proxy 功能处于关闭状态

1.4 域名解析显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示域名解析配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除动态域名缓存区信息。

表1-4 域名解析显示和维护

操作	命令
显示静态域名解析表	**display ip host**
显示域名服务器信息	**display dns server [dynamic]**
显示域名后缀列表信息	**display dns domain [dynamic]**
显示动态域名缓存区的信息	**display dns dynamic-host**
清空动态域名缓存区	**reset dns dynamic-host**

1.5 域名解析典型配置举例

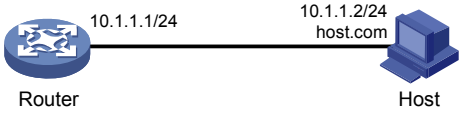
1.5.1 静态域名解析配置举例

1. 组网需求

Router 利用静态域名解析功能，实现通过主机名 host.com 访问 IP 地址为 10.1.1.2 的主机 Host。

2. 组网图

图1-3 静态域名解析配置组网图



3. 配置步骤

配置主机名 host.com 对应的 IP 地址为 10.1.1.2。

```
<Sysname> system-view
[Sysname] ip host host.com 10.1.1.2
```

执行 ping host.com 命令，Router 通过静态域名解析可以解析到 host.com 对应的 IP 地址为 10.1.1.2。

```
[Sysname] ping host.com
  PING host.com (10.1.1.2):
  56  data bytes, press CTRL_C to break
    Reply from 10.1.1.2: bytes=56 Sequence=1 ttl=128 time=1 ms
    Reply from 10.1.1.2: bytes=56 Sequence=2 ttl=128 time=4 ms
    Reply from 10.1.1.2: bytes=56 Sequence=3 ttl=128 time=3 ms
    Reply from 10.1.1.2: bytes=56 Sequence=4 ttl=128 time=2 ms
    Reply from 10.1.1.2: bytes=56 Sequence=5 ttl=128 time=3 ms
```

```
--- host.com ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/2/4 ms
```

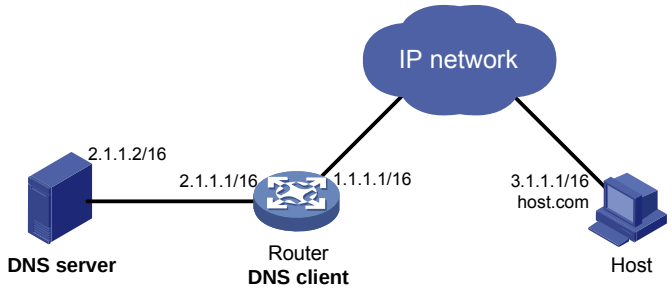
1.5.2 动态域名解析配置举例

1. 组网需求

- 域名服务器的 IP 地址是 2.1.1.2/16，配置域名后缀为 com，且 com 域中包含域名“host”和 IP 地址 3.1.1.1/16 的对应关系。
- Router 作为 DNS 客户端，使用动态域名解析和域名后缀列表功能，实现通过输入 host 来访问域名为 host.com、IP 地址为 3.1.1.1/16 的主机 Host。

2. 组网图

图1-4 动态域名解析组网图



3. 配置步骤

说明

- 在开始下面的配置之前，假设设备与主机之间的路由可达，设备和主机都已经配置完毕，接口IP地址如 图 1-4所示。
- 不同域名服务器的配置方法不同，下面仅以 Windows Server 2000 为例，说明域名服务器的配置方法。

(1) 配置域名服务器

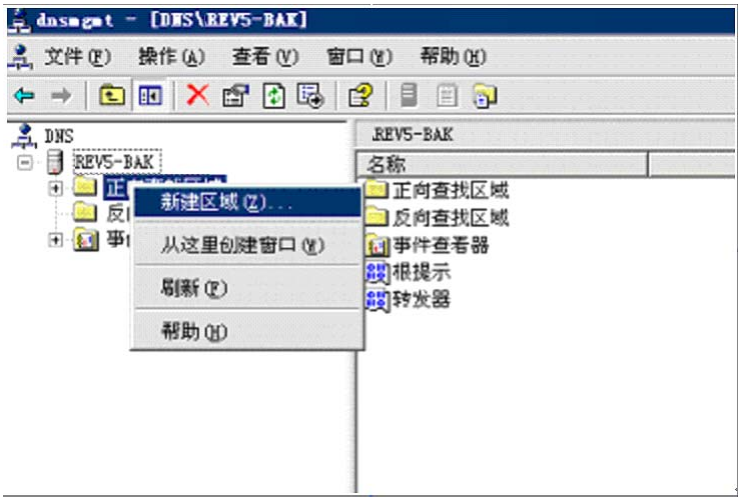
进入域名服务器配置界面。

在开始菜单中，选择[程序/管理工具/DNS]。

创建区域 com。

如 图 1-5所示，右键点击[正向查找区域]，选择[新建区域]，按照提示创建新的区域com。

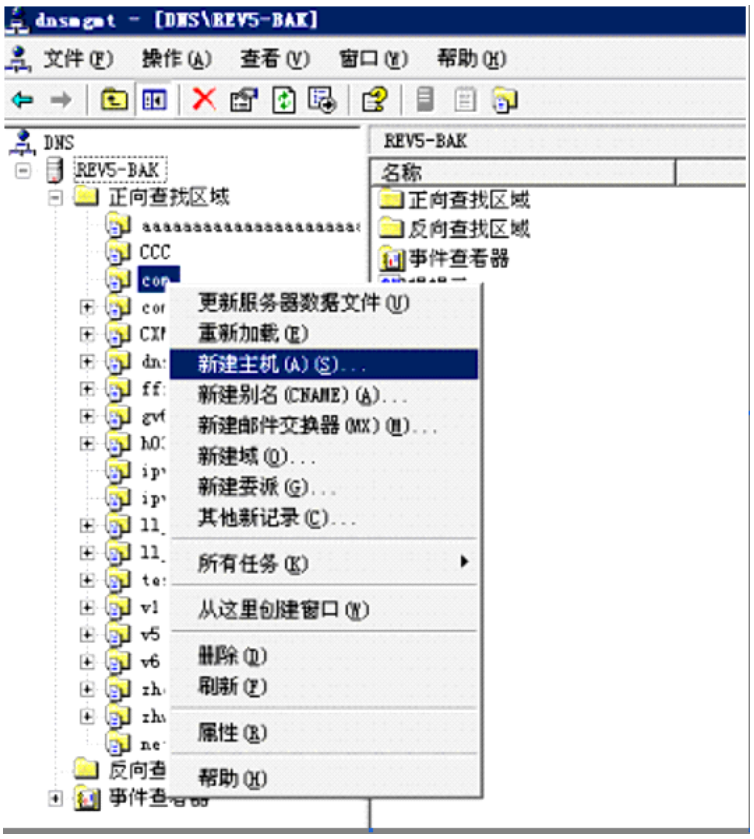
图1-5 创建区域



添加域名和 IP 地址的映射。

如 图 1-6所示，右键点击区域com。

图1-6 新建主机



选择[新建主机]，弹出如 图 1-7的对话框。按照 图 1-7输入域名host和IP地址 3.1.1.1。

图1-7 添加域名和 IP 地址的映射



(2) 配置 DNS 客户端 Router

开启动态域名解析功能。

```
<Sysname> system-view
[Sysname] dns resolve
```

配置域名服务器的 IP 地址为 2.1.1.2。

```
[Sysname] dns server 2.1.1.2
```

配置域名后缀 com。

```
[Sysname] dns domain com
```

(3) 验证配置结果

在设备上执行 ping host 命令，可以 ping 通主机，且对应的目的地址为 3.1.1.1。

```
[Sysname] ping host
 Trying DNS resolve, press CTRL_C to break
 Trying DNS server (2.1.1.2)
  PING host.com (3.1.1.1):
  56  data bytes, press CTRL_C to break
    Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=128 time=3 ms
    Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=128 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=128 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=128 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=128 time=1 ms

  --- host.com ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/3 ms
```

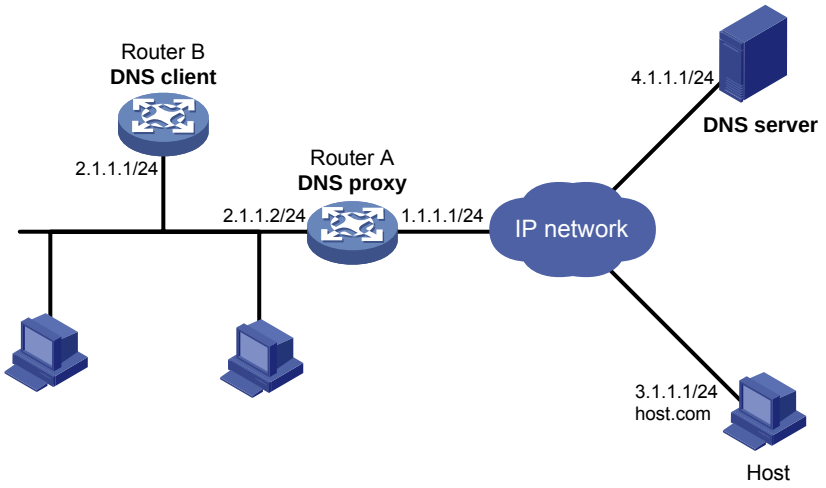
1.5.3 DNS proxy 典型配置举例

1. 组网需求

- Router B 作为 DNS 客户端，指定域名服务器为 Router A。
- Router A 作为 DNS 代理。实际域名服务器的 IP 地址为 4.1.1.1。
- Router B 通过 DNS 代理 Router A 实现域名解析。

2. 组网图

图1-8 DNS proxy 组网图



3. 配置步骤

说明

在开始下面的配置之前，假设设备与域名服务器、主机之间的路由可达，并已按照 图 1-8配置各接口的IP地址。

(1) 配置域名服务器

不同的域名服务器的配置方法不同。Windows Server 2000 作为域名服务器时，配置方法请参见“1.5.2 动态域名解析配置举例”。

(2) 配置 DNS 代理 Router A

配置域名服务器的 IP 地址为 4.1.1.1。

```
<RouterA> system-view
[RouterA] dns server 4.1.1.1
```

开启 DNS proxy 功能。

```
[RouterA] dns proxy enable
```

(3) 配置 DNS 客户端 Router B

开启动态域名解析功能。

```
<RouterB> system-view
[RouterB] dns resolve
```

配置域名服务器的 IP 地址为 2.1.1.2。

```
[RouterB] dns server 2.1.1.2
```

(4) 验证配置结果

在 Router B 上执行 ping host.com 命令，可以 ping 通主机，且对应的目的地址为 3.1.1.1。

```
[RouterB] ping host.com
Trying DNS resolve, press CTRL_C to break
 Trying DNS server (2.1.1.2)
  PING host.com (3.1.1.1):
  56  data bytes, press CTRL_C to break
    Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=126 time=3 ms
    Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=126 time=1 ms

  --- host.com ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/3 ms
```

1.6 域名解析常见配置错误举例

1. 现象描述

配置了动态域名解析，但不能根据域名解析到正确的 IP 地址。

2. 故障分析

DNS 客户端需要和域名服务器配合使用，才能根据域名解析到正确的 IP 地址。

3. 故障排除

- 执行命令 **display dns dynamic-host**，检查动态域名缓存区中的信息是否存在指定域名。
- 如果不存在要解析的域名，检查 DNS 客户端是否和域名服务器通信正常，域名服务器是否工作正常，动态域名解析功能是否已经开启。
- 如果存在要解析的域名，但地址不对，则检查 DNS 客户端所配置的域名服务器的 IP 地址是否正确。
- 检查域名服务器所设置的域名和地址映射表是否正确。

目 录

1 IP地址配置 1-1
1.1 IP地址简介 1-1
1.1.1 IP地址的分类和表示 1-1
1.1.2 特殊的IP地址 1-1
1.1.3 子网和掩码 1-2
1.1.4 IP地址借用 1-2
1.2 配置IP地址 1-2
1.2.1 配置接口的IP地址 1-3
1.2.2 IP地址配置举例 1-3
1.3 配置接口借用IP地址 1-5
1.3.1 配置准备 1-5
1.3.2 配置接口借用IP地址 1-5
1.3.3 接口借用IP地址配置举例 1-6
1.4 IP地址的显示和维护 1-7

1 IP 地址配置

1.1 IP 地址简介

1.1.1 IP 地址的分类和表示

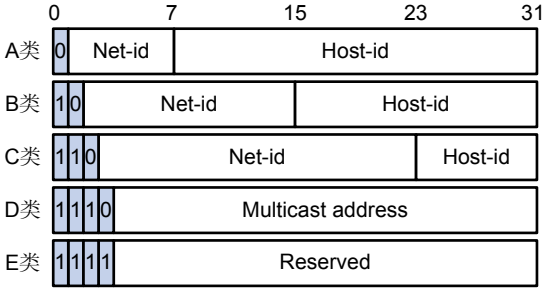
连接到 Internet 上的设备接口必须有一个全球唯一的 IP 地址。IP 地址长度为 32 比特，通常采用点分十进制方式表示，即每个 IP 地址被表示为以小数点隔开的 4 个十进制整数，每个整数对应一个字节，如 10.1.1.1。

IP 地址由两部分组成：

- 网络号码字段（Net-id）：用于区分不同的网络。网络号码字段的前几位称为类别字段（又称为类别比特），用来区分 IP 地址的类型。
- 主机号码字段（Host-id）：用于区分一个网络内的不同主机。

为了方便管理及组网，IP地址分成五类，如 图 1-1所示，其中蓝色部分为类别字段。

图1-1 五类 IP 地址



上述五类IP地址的地址范围如 表 1-1所示。目前大量使用的IP地址属于A、B、C三类。

表1-1 IP 地址分类及范围

地址类型	地址范围	说明
A	0.0.0.0～127.255.255.255	IP 地址 0.0.0.0 仅用于主机在系统启动时进行临时通信，并且永远不是有效目的地址 127.0.0.0 网段的地址都保留作环回测试，发送到这个地址的分组不会输出到链路上，它们被当作输入分组在内部进行处理
B	128.0.0.0～191.255.255.255	-
C	192.0.0.0～223.255.255.255	-
D	224.0.0.0～239.255.255.255	组播地址
E	240.0.0.0～255.255.255.255	255.255.255.255 用于广播地址，其他地址保留今后使用

1.1.2 特殊的 IP 地址

下列 IP 地址具有特殊的用途，不能作为主机的 IP 地址。

- Net-id 为全 0 的地址：表示本网络内的主机。例如，0.0.0.16 表示本网络内 Host-id 为 16 的主机。

- Host-id 为全 0 的地址：网络地址，用于标识一个网络。
- Host-id 为全 1 的地址：网络广播地址。例如，目的地址为 192.168.1.255 的报文，将转发给 192.168.1.0 网络内所有的主机。

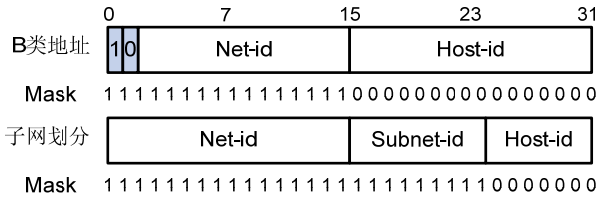
1.1.3 子网和掩码

随着 Internet 的快速发展，IP 地址已近枯竭。为了充分利用已有的 IP 地址，可以使用子网掩码将网络划分为更小的部分（即子网）。通过从主机号码字段部分划出一些比特位作为子网号码字段，能够将一个网络划分为多个子网。子网号码字段的长度由子网掩码确定。

子网掩码是一个长度为 32 比特的数字，由一串连续的“1”和一串连续的“0”组成。“1”对应于网络号码字段和子网号码字段，而“0”对应于主机号码字段。

图 1-2所示是一个B类地址划分子网的情况。

图1-2 IP 地址子网划分



多划分出一个子网号码字段会浪费一些 IP 地址。例如，一个 B 类地址可以容纳 65534（2^{16}-2，去掉主机号码字段全 1 的广播地址和主机号码字段全 0 的网段地址）个主机号码。但划分出 9 比特长的子网字段后，最多可有 512（2^9）个子网，每个子网有 7 比特的主机号码，即每个子网最多可有 126（2^7-2，去掉主机号码字段全 1 的广播地址和主机号码字段全 0 的网段地址）个主机号码。因此主机号码的总数是 512*126=64512 个，比不划分子网时要少 1022 个。

若不进行子网划分，则子网掩码为默认值，此时子网掩码中“1”的长度就是网络号码的长度，即 A、B、C 类 IP 地址对应的子网掩码默认值分别为 255.0.0.0、255.255.0.0 和 255.255.255.0。

1.1.4 IP 地址借用

所谓“IP 地址借用”，是指一个接口上没有配置 IP 地址，但为了使该接口能正常使用，就向同一设备上其它有 IP 地址的接口借用一个 IP 地址。IP 地址借用使用的场合如下：

- 在 IP 地址资源比较缺乏的环境下，为了节约 IP 地址资源，可以配置某个接口借用其他接口的 IP 地址。
- 如果某个接口只是偶尔使用，可以配置该接口借用其他接口的 IP 地址，而不必让其一直占用一个单独的 IP 地址。

1.2 配置 IP 地址

接口获取 IP 地址有以下几种方式：

- 通过手工指定 IP 地址
- 通过 DHCP 分配得到 IP 地址
- 通过 PPP 协商获得 IP 地址

这几种方式是互斥的，通过新的配置方式获取的 IP 地址会覆盖通过原有方式获取的 IP 地址。例如，

首先通过手工指定了 IP 地址，然后使用 DHCP 协议获得 IP 地址，那么手工指定的 IP 地址会被删除，接口的 IP 地址是通过 DHCP 协议获得的。

说明

本节只介绍通过手工指定 IP 地址的方式。其他两种方式请参见“IP 业务分册”中的“DHCP 配置”以及“接入分册”中的“PPP 配置”。

1.2.1 配置接口的 IP 地址

在一般情况下，一个接口配置一个 IP 地址。为了使设备的一个接口与多个子网相连，可以为一个接口配置多个 IP 地址，其中一个为主 IP 地址，其余为从 IP 地址。

表1-2 配置接口的 IP 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 IP 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	必选 缺省情况下，没有为接口配置 IP 地址

注意

- 一个接口只能有一个主 IP 地址。新配置的主 IP 地址将覆盖原有主 IP 地址。
- 当接口被配置为通过 DHCP、PPP 方式获取 IP 地址或借用其他接口的 IP 地址后，则不能再给该接口配置从 IP 地址。
- 同一接口的主、从 IP 地址可以在同一网段，但不同接口之间、主接口及其子接口之间、同一主接口下不同子接口之间的 IP 地址不可以在同一网段。

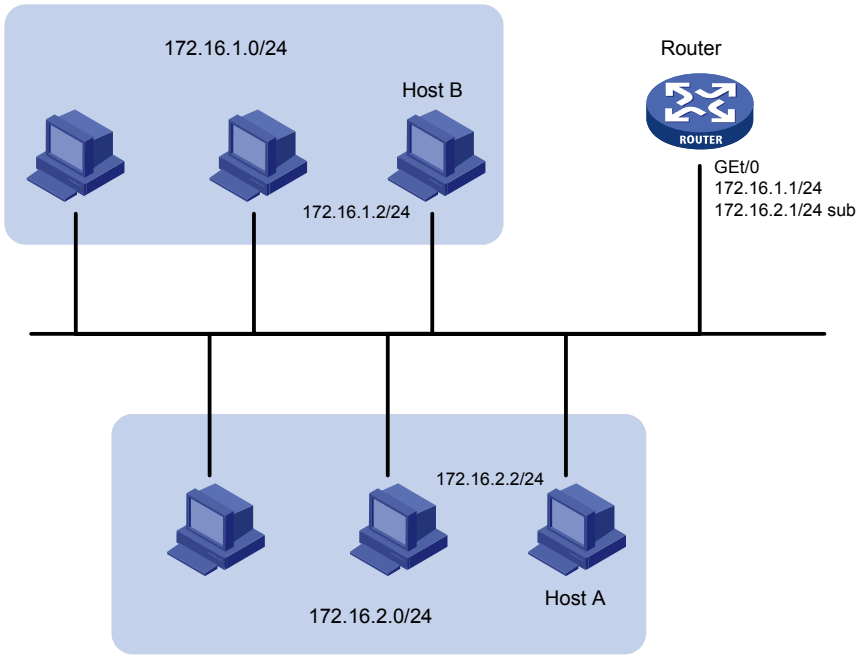
1.2.2 IP 地址配置举例

1. 组网需求

Router 的以太网接口 GigabitEthernet1/0 连接一个局域网，该局域网中的计算机分别属于 2 个网段：172.16.1.0/24 和 172.16.2.0/24。要求这两个网段的主机都可以通过 Router 与外部网络通信，且这两个网段中的主机能够互通。

2. 组网图

图1-3 IP 地址配置组网图



3. 配置步骤

针对上述的需求，如果在 Router 的接口上只配置一个 IP 地址，则只有一部分主机能够通过 Router 与外部网络通信。为了使局域网内的所有主机都能够通过 Router 访问外部网络，需要配置接口的从 IP 地址。为了使两个网段中的主机能够互通，两个网段中的主机都需要将 Router 设置为网关。

配置接口 GigabitEthernet1/0 的主 IP 地址和从 IP 地址。

```
<Router> system-view
[Router] interface GigabitEthernet 1/0
[Router-GigabitEthernet1/0] ip address 172.16.1.1 255.255.255.0
[Router-GigabitEthernet1/0] ip address 172.16.2.1 255.255.255.0 sub
```

在 172.16.1.0/24 网段中的主机上配置网关为 172.16.1.1；在 172.16.2.0/24 网段中的主机上配置网关为 172.16.2.1。

使用 **ping** 命令检测 Router 与网络 172.16.1.0/24 内主机的连通性。

```
<Router> ping 172.16.1.2
  PING 172.16.1.2: 56  data bytes, press CTRL_C to break
    Reply from 172.16.1.2: bytes=56 Sequence=1 ttl=255 time=25 ms
    Reply from 172.16.1.2: bytes=56 Sequence=2 ttl=255 time=27 ms
    Reply from 172.16.1.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 172.16.1.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 172.16.1.2: bytes=56 Sequence=5 ttl=255 time=26 ms

  --- 172.16.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 25/26/27 ms
```

显示信息表示 Router 与网络 172.16.1.0/24 内的主机可以互通。

使用 **ping** 命令检测 Router 与网络 172.16.2.0/24 内主机的连通性。

```
<Router> ping 172.16.2.2
  PING 172.16.2.2: 56  data bytes, press CTRL_C to break
    Reply from 172.16.2.2: bytes=56 Sequence=1 ttl=255 time=25 ms
    Reply from 172.16.2.2: bytes=56 Sequence=2 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=5 ttl=255 time=26 ms

  --- 172.16.2.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 25/25/26 ms
```

显示信息表示 Router 与网络 172.16.2.0/24 内的主机可以互通。

使用 **ping** 命令检测网络 172.16.1.0/24 和网络 172.16.2.0/24 内主机的连通性。在 Host A 上可以 ping 通 Host B。

1.3 配置接口借用 IP 地址

1.3.1 配置准备

被借用接口的主 IP 地址已经配置，配置方法可以为手工指定、通过 DHCP 动态获取或通过 PPP 协商分配。

1.3.2 配置接口借用 IP 地址

表1-3 配置接口借用 IP 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置本接口借用指定接口的 IP 地址	**ip address unnumbered interface** *interface-type interface-number*	必选 缺省情况下，不借用其它接口的 IP 地址

注意

- 通常，同步串口、POS 接口可以借用三层以太网接口或其他接口的 IP 地址。
- 三层以太网接口、Tunnel 接口和 Loopback 接口的 IP 地址可被其它接口借用，但本身不能借用其它接口的地址。
- 被借用接口的地址本身不能为借用地址。
- 一个接口的地址可以借给多个接口。
- 如果被借用接口有多个 IP 地址，则只有主 IP 地址能被借用。
- 借用接口的 IP 地址始终与被借用接口的 IP 地址保持一致，并且随着被借用接口 IP 地址的变化而变化。即如果被借用接口已经配置 IP 地址，则借用接口的 IP 地址与被借用接口相同；如果被借用接口没有配置 IP 地址，则借用接口的 IP 地址也处于未配置状态。

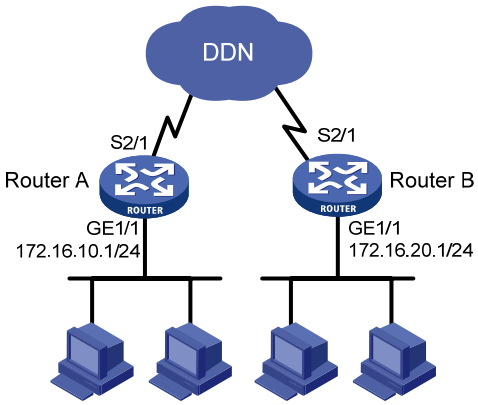
1.3.3 接口借用 IP 地址配置举例

1. 组网需求

某企业通过 DDN 组建内部网，节点路由器之间通过同步串口相连，并分别通过以太网接口连接本地的局域网。为了节省 IP 地址，规划串口借用以太网接口的 IP 地址。

2. 组网图

图1-4 借用 IP 地址示例的组网图



3. 配置步骤

(1) 配置 Router A

配置被借用以太网接口的主 IP 地址。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 172.16.10.1 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
```

配置串口借用以太网接口的 IP 地址。

```
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ip address unnumbered interface GigabitEthernet 1/1
[RouterA-Serial2/1] quit
```

配置到 Router B 所连的局域网的路由，指定出接口为串口 Serial2/1。

```
[RouterA] ip route-static 172.16.20.0 255.255.255.0 serial 2/1
```

(2) 配置 Router B

配置被借用以太网接口的主 IP 地址。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 172.16.20.1 255.255.255.0
[RouterB-GigabitEthernet1/1] quit
```

配置串口借用以太网接口的 IP 地址。

```
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ip address unnumbered interface GigabitEthernet 1/1
[RouterB-Serial2/1] quit
```

配置到 Router A 所连的局域网的路由，指定出接口为串口 Serial2/1。

```
[RouterB] ip route-static 172.16.10.0 255.255.255.0 serial 2/1
```

(3) 验证配置结果

在 Router A 上可以 Ping 通与 Router B 相连的局域网中主机。

```
[RouterA] ping 172.16.20.2
  PING 172.16.20.2: 56  data bytes, press CTRL_C to break
    Reply from 172.16.20.2: bytes=56 Sequence=1 ttl=255 time=25 ms
    Reply from 172.16.20.2: bytes=56 Sequence=2 ttl=255 time=25 ms
    Reply from 172.16.20.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 172.16.20.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 172.16.20.2: bytes=56 Sequence=5 ttl=255 time=26 ms

  --- 172.16.20.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 25/25/26 ms
```

1.4 IP 地址的显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IP 地址的运行情况，通过查看显示信息验证配置的效果。

表1-4 IP 地址的显示和维护

操作	命令
显示三层接口的相关信息	**display ip interface** [*interface-type interface-number*]
显示三层接口的 IP 基本配置信息	**display ip interface brief** [*interface-type interface-number*]

目 录

1 IP性能优化配置 1-1
1.1 IP性能优化简介 1-1
1.2 配置允许发送定向广播报文 1-1
1.2.1 配置允许转发定向广播报文 1-1
1.2.2 允许发送定向广播报文配置举例 1-2
1.3 配置TCP属性 1-3
1.3.1 配置接口的TCP最大报文段长度 1-3
1.3.2 配置SYN Cookie功能 1-3
1.3.3 配置防止Naptha攻击功能 1-4
1.3.4 配置TCP的可选参数 1-5
1.4 配置ICMP差错报文发送功能 1-5
1.5 配置IP转发负载分担模式 1-7
1.6 IP性能优化显示和维护 1-8

1 IP 性能优化配置

1.1 IP 性能优化简介

在一些特定的网络环境里，可以通过需要调整 IP 的参数，以使网络性能达到最佳。IP 性能的优化配置包括：

- 配置允许接收和发送定向广播报文
- 配置接口的 TCP 最大报文段长度
- 配置 TCP 的 SYN Cookie 和防止 Naptha 攻击功能
- 配置 TCP 定时器
- 配置 TCP 连接的接收和发送缓冲区的大小
- 配置 ICMP 差错报文发送功能

1.2 配置允许发送定向广播报文

定向广播报文是指发送给特定网络的广播报文。该报文的目的 IP 地址中网络号码字段为特定网络的网络号，主机号码字段为全 1。

如果允许设备接收并转发目的地址为接口所在网络的定向广播报文，黑客就可以利用这样的报文来攻击网络系统，给网络的安全带来了很大的隐患。但在某些应用环境下，设备又需要转发定向广播报文，例如使用 Wake on LAN（网络唤醒）功能，发送定向广播报文唤醒远程网络中的计算机。在上述情况下，用户可以通过命令配置设备允许转发定向广播报文。

1.2.1 配置允许转发定向广播报文

表1-1 配置允许转发定向广播报文

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置允许接口转发定向广播报文	**ip forward-broadcast** [**acl** *acl-number*]	必选 缺省情况下，禁止接口转发定向广播报文



说明

- 允许接口转发定向的广播报文时，如果配置了 ACL 规则，则在转发广播报文的同时还需要对报文进行过滤，不符合 ACL 规则的报文将被丢弃，只转发符合 ACL 规则的报文。
- 如果在同一接口下重复执行 **ip forward-broadcast acl** 命令，则后面配置的 ACL 会覆盖以前配置的 ACL；如果后配置的命令不带 **acl** *acl-number*，则以前配置中的 ACL 规则将被取消。

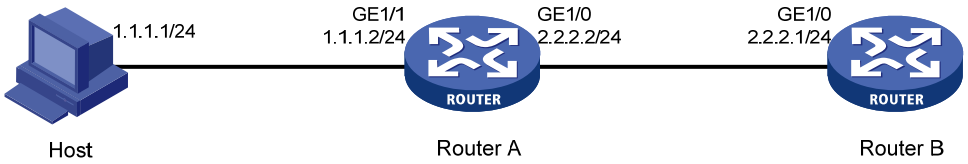
1.2.2 允许发送定向广播报文配置举例

1. 组网需求

如 图 1-1所示，Host的接口和Router A的接口GigabitEthernet1/1 处于同一个网段（1.1.1.0/24），Router A的接口GigabitEthernet1/0 和Router B的接口GigabitEthernet1/0 处于另外一个网段（2.2.2.0/24）。Host上配置默认网关为Router A的接口GigabitEthernet1/1 的地址（1.1.1.2/24），Router B上配置静态路由使得Host与Router B之间路由可达。

2. 组网图

图1-1 配置收发定向广播报文组网图



3. 配置步骤

(1) 配置 RouterA

配置允许 RouterA 的 GigabitEthernet1/0 接口转发定向的广播报文。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ip forward-broadcast
```

配置接口 GigabitEthernet1/1 和 GigabitEthernet1/0 的 IP 地址。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 1.1.1.2 24
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ip address 2.2.2.2 24
```

(2) 配置 RouterB

配置 RouterB 到 Host 的静态路由。

```
<RouterB> system-view
[RouterB] ip route-static 1.1.1.1 24 2.2.2.2
```

配置接口 GigabitEthernet1/0 的 IP 地址。

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] ip address 2.2.2.1 24
[RouterB-GigabitEthernet1/0] quit
```

(3) 配置 Host

配置 Host 与 Router A 相连的网卡 IP 地址为 1.1.1.1，掩码为 255.255.255.0，网关为 1.1.1.2。

配置完成以后，在 Host 上 ping RouterA 的接口 GigabitEthernet1/0 所在子网网段的广播地址（2.2.2.255）时，RouterB 的接口 GigabitEthernet1/0 可以收到该报文。取消掉 **ip forward-broadcast** 的配置，RouterB 的接口 GigabitEthernet1/0 不能接收到该报文。

1.3 配置 TCP 属性

1.3.1 配置接口的 TCP 最大报文段长度

TCP 最大报文段长度（Max Segment Size，MSS）表示 TCP 连接的对端发往本端的最大 TCP 报文段的长度，目前作为 TCP 连接建立时的一个选项来协商：当一个 TCP 连接建立时，连接的双方要将 MSS 作为 TCP 报文的一个选项通告给对端，对端会记录下这个 MSS 值，后续在发送 TCP 报文时，会限制 TCP 报文的大小不超过该 MSS 值。当对端发送的 TCP 报文的长度小于本端的 TCP 最大报文段长度时，TCP 报文不需要分段；否则，对端需要对 TCP 报文按照最大报文段长度进行分段处理后再发给本端。

用户可以通过下面的命令配置接口的 TCP 最大报文段长度，配置后该接口接收和发送的 TCP 报文的大小都不能超过该值。

表1-2 配置接口的 TCP 最大报文段长度

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 TCP 最大报文段长度	**tcp mss** *value*	可选 缺省情况下，TCP 最大报文段长度为 1460 字节

说明

- 该配置仅对新建的 TCP 连接生效，对于配置前已建立的 TCP 连接不生效。
- 该配置仅对 IP 报文生效，当接口上配置了 MPLS 功能后，不建议再配置本功能。

1.3.2 配置 SYN Cookie 功能

一般情况下，TCP 连接的建立需要经过三次握手，即：

(1) TCP 连接请求的发起者向目标服务器发送 SYN 报文；

(2) 目标服务器收到 SYN 报文后，建立处于 SYN_RECEIVED 状态的 TCP 半连接，并向发起者回复 SYN ACK 报文，等待发起者的回应；

(3) 发起者收到 SYN ACK 报文后，回应 ACK 报文，这样 TCP 连接就建立起来了。

利用 TCP 连接的建立过程，一些恶意的攻击者可以进行 SYN Flood 攻击。攻击者向服务器发送大量请求建立 TCP 连接的 SYN 报文，而不回应服务器的 SYN ACK 报文，导致服务器上建立了大量的 TCP 半连接。从而，达到耗费服务器资源，使服务器无法处理正常业务的目的。

SYN Cookie 功能用来防止 SYN Flood 攻击。当服务器收到 TCP 连接请求时，不建立 TCP 半连接，而直接向发起者回复 SYN ACK 报文。服务器接收到发起者回应的 ACK 报文后，才建立连接，并进入 ESTABLISHED 状态。通过这种方式，可以避免在服务器上建立大量的 TCP 半连接，防止服务器受到 SYN Flood 攻击。

表1-3 配置 SYN Cookie 功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 SYN Cookie 功能	**tcp syn-cookie enable**	必选 缺省情况下，SYN Cookie 功能处于关闭状态

说明

- 如果启用了 MD5 加密验证功能，则 SYN Cookie 功能不会生效。取消 MD5 加密验证功能后，之前配置的 SYN Cookie 功能会自动生效。
- 使能 SYN Cookie 功能后，建立 TCP 连接时只协商最大报文段长度选项，而不协商窗口缩放因子和时间戳选项。

1.3.3 配置防止 Naptha 攻击功能

Naptha 攻击类似于 SYN Flood 攻击，所不同的是 Naptha 攻击可利用 TCP 连接的 CLOSING、ESTABLISHED、FIN_WAIT_1、FIN_WAIT_2、LAST_ACK 和 SYN_RECEIVED 六种状态来达到攻击目的，而 SYN Flood 只是利用 SYN_RECEIVED 状态。

Naptha 攻击通过控制大量主机与服务器建立 TCP 连接，并使这些连接处于同一种状态（上述六种状态中的一种），而不请求任何数据，从而达到消耗服务器的内存资源，导致服务器无法处理正常业务的目的。

防止 Naptha 攻击功能通过加速 TCP 状态的老化，来降低服务器遭受 Naptha 攻击的风险。使能防止 Naptha 攻击功能后，设备周期性地检测处于上述六种状态的 TCP 连接数，如果检测到某个状态的 TCP 连接数目超过设定的最大连接数，则加速该状态下 TCP 连接的老化。

表1-4 配置防止 Naptha 攻击功能

操作	命令	说明
进入系统视图	**system-view**	-
使能防止 Naptha 攻击功能	**tcp anti-naptha enable**	必选 缺省情况下，防止 Naptha 攻击功能处于关闭状态
配置某一状态下的最大 TCP 连接数	**tcp state** { **closing** | **established** | **fin-wait-1** | **fin-wait-2** | **last-ack** | **syn-received** } **connection-number** *number*	可选 缺省情况下，六种 TCP 状态下的最大连接数均为 5 如果 TCP 状态下的最大连接数为 0，则表示不会加速该状态的老化
配置 TCP 连接状态的轮询检测时间间隔	**tcp timer check-state** *timer-value*	可选 缺省情况下，TCP 状态轮询检测的时间间隔为 30 秒

说明

- 使能防止 Naptha 攻击功能后，设备会周期性地检测，并记录 TCP 连接数。设备只记录其作为 TCP 服务器的 TCP 连接数。
- 使能防止 Naptha 攻击功能后，如果检测到某一 TCP 状态下作为 TCP 服务器的连接数大于设定的最大连接数，则认为设备受到 Naptha 攻击，就会设置加速老化标志，加速该状态下 TCP 连接的老化。当该状态下的 TCP 连接数低于最大连接数的 80％（最小值为 1），则取消该状态的加速老化标志。

1.3.4 配置 TCP 的可选参数

可以配置的 TCP 可选参数包括：

- synwait 定时器：当发送 SYN 报文时，TCP 启动 synwait 定时器，如果 synwait 超时前未收到回应报文，则 TCP 连接建立不成功。
- finwait 定时器：当 TCP 的连接状态为 FIN_WAIT_2 时，启动 finwait 定时器，如果在定时器超时前没有收到报文，则TCP连接终止；如果收到FIN报文，则TCP连接状态变为TIME_WAIT状态；如果收到非 FIN 报文，则从收到的最后一个非 FIN 报文开始重新计时，在超时后中止连接。
- TCP 连接的接收和发送缓冲区的大小。

表1-5 配置 TCP 属性

操作	命令	说明
进入系统视图	**system-view**	-
配置 TCP 的 synwait 定时器超时时间	**tcp timer syn-timeout** *time-value*	可选 缺省情况下，synwait 定时器超时时间为 75 秒
配置 TCP 的 finwait 定时器超时时间	**tcp timer fin-timeout** *time-value*	可选 缺省情况下，finwait 定时器超时时间为 675 秒
配置 TCP 连接的接收和发送缓冲区的大小	**tcp window** *window-size*	可选 缺省情况下，TCP 连接的接收和发送缓冲区大小为 8KB

注意

finwait 定时器的实际超时时间由如下公式决定：

finwait 定时器的实际超时时间 =（配置的 finwait 定时器超时时间 − 75）+ 配置的 synwait 定时器超时时间。

1.4 配置 ICMP 差错报文发送功能

发送差错报文是 ICMP（Internet Control Message Protocol，互联网控制报文协议）的主要功能之一。ICMP 报文通常被网络层或传输层协议用来在异常情况发生时通知相应设备，从而便于进行控制管理。

1. ICMP 差错报文发送功能的作用

重定向报文、超时报文、目的不可达报文是 ICMP 差错报文中的三种。下面分别介绍这三种差错报文发送的条件及作用。

(1) ICMP 重定向报文发送功能

主机启动时，它的路由表中可能只有一条到缺省网关的缺省路由。当满足一定的条件时，缺省网关会向源主机发送 ICMP 重定向报文，通知主机重新选择正确的下一跳进行后续报文的发送。

满足下列条件时，设备会发送 ICMP 重定向报文：

- 接收和转发数据报文的接口是同一接口；
- 被选择的路由本身没有被 ICMP 重定向报文创建或修改过；
- 被选择的路由不是设备的默认路由；
- 数据报文中没有源路由选项。

ICMP 重定向报文发送功能可以简化主机的管理，使具有很少选路信息的主机逐渐建立较完善的路由表，从而找到最佳路由。

(2) ICMP 超时报文发送功能

ICMP 超时报文发送功能是在设备收到 IP 数据报文后，如果发生超时差错，则将报文丢弃并给源端发送 ICMP 超时差错报文。

设备在满足下列条件时会发送 ICMP 超时报文：

- 设备收到 IP 数据报文后，如果报文的目的地不是本地且报文的 TTL 字段是 1，则发送“TTL 超时”ICMP 差错报文；
- 设备收到目的地址为本地的 IP 数据报文的第一个分片后，启动定时器，如果所有分片报文到达之前定时器超时，则会发送“重组超时”ICMP 差错报文。

(3) ICMP 目的不可达报文发送功能

ICMP 目的不可达报文发送功能是在设备收到 IP 数据报文后，如果发生目的不可达的差错，则将报文丢弃并给源端发送 ICMP 目的不可达差错报文。

设备在满足下列条件时会发送目的不可达报文：

- 设备在转发报文时，如果在路由表中没有找到对应的转发路由，且路由表中没有缺省路由，则给源端发送“网络不可达”ICMP 差错报文；
- 设备收到目的地址为本地的数据报文时，如果设备不支持数据报文采用的传输层协议，则给源端发送“协议不可达”ICMP 差错报文；
- 设备收到目的地址为本地、传输层协议为 UDP 的数据报文时，如果报文的端口号与正在使用的进程不匹配，则给源端发送“端口不可达”ICMP 差错报文；
- 源端如果采用“严格的源路由选择”发送报文，当中间设备发现源路由所指定的下一个设备不在其直接连接的网络上，则给源端发送“源站路由失败”的 ICMP 差错报文；
- 设备在转发报文时，如果转发接口的 MTU 小于报文的长度，但报文被设置了不可分片，则给源端发送“需要进行分片但设置了不分片比特”ICMP 差错报文。

2. ICMP 差错报文发送功能的弊端

ICMP 差错报文的发送虽然方便了网络的控制管理，但也存在一定的弊端：

- 由于发送大量的 ICMP 报文，增大了网络流量。
- 如果设备接收到大量需要发送 ICMP 差错报文的恶意攻击报文，设备会因为处理大量该类报文而导致性能降低。
- 由于重定向功能会在主机的路由表中增加主机路由，当增加的主机路由很多时，会降低主机性能。
- 由于 ICMP 目的不可达报文传递给用户进程的信息为不可达信息，如果有用户恶意攻击，可能会影响终端用户的正常使用。

为了避免上述现象发生，可以关闭设备的 ICMP 差错报文发送功能，从而减少网络流量、防止遭到恶意攻击。

表1-6 配置 ICMP 差错报文发送功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 ICMP 重定向报文发送功能	**ip redirects enable**	必选 缺省情况下，ICMP 重定向报文发送功能处于关闭状态
开启 ICMP 超时报文发送功能	**ip ttl-expires enable**	必选 缺省情况下，ICMP 超时报文发送功能处于关闭状态
开启 ICMP 目的不可达报文发送功能	**ip unreachables enable**	必选 缺省情况下，ICMP 目的不可达报文发送功能处于关闭状态



说明

关闭 ICMP 超时报文发送功能后，设备不会再发送“TTL 超时”ICMP 差错报文，但“重组超时”ICMP 差错报文仍会正常发送。

1.5 配置 IP 转发负载分担模式

目前多核设备的等价路由负载分担方式是基于流标来实现的。而在某些情况下可能出现相同五元组的条流对应多个流标的情况。这样在等价路由负载分担的组网中，会有可能使得同一条流被分担到不同的出接口，从而导致业务模块无法处理。在这种情况下，可以通过配置基于 IP 地址的等价路由负载分担方式，以保证相同的流量从同一接口发送。

表1-7 配置 IP 转发负载分担模式

操作	命令	说明
进入系统视图	**system-view**	-
开启基于源 IP 地址和目的 IP 地址的等价路由负载分担方式	**ip load-sharing mode per-flow**	可选 缺省情况下，基于源 IP 地址和目的 IP 地址的等价路由负载分担方式处于关闭状态

1.6 IP 性能优化显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置 IP 性能后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令清除 IP、TCP 和 UDP 的流量统计信息。

表1-8 IP 性能优化显示和维护

操作		命令
显示所有 TCP 连接的状态		**display tcp status**
显示 TCP 连接的流量统计信息		**display tcp statistics**
显示 UDP 流量统计信息		**display udp statistics**
显示 IP 报文统计信息	集中式设备	**display ip statistics**
	分布式设备	**display ip statistics** [**slot** *slot-number*]
显示 ICMP 流量统计信息	集中式设备	**display icmp statistics**
	分布式设备	**display icmp statistics** [**slot** *slot-number*]
显示套接口信息	集中式设备	**display ip socket** [**socktype** *sock-type*] [*task-id socket-id*]
	分布式设备	**display ip socket** [**socktype** *sock-type*] [*task-id socket-id*] [**slot** *slot-number*]
显示 FIB 信息		**display fib** [vpn-instance *vpn-instance-name*] [**|** { **begin** | **include** | **exclude** } *regular-expression* | **acl** *acl-number* | **ip-prefix** *ip-prefix-name*]
显示与指定目的 IP 地址匹配的 FIB 信息		**display fib** [vpn-instance *vpn-instance-name*] *ip-address* [*mask* | *mask-length*]
清除 IP 报文统计信息	集中式设备	**reset ip statistics**
	分布式设备	**reset ip statistics** [**slot** *slot-number*]
清除 TCP 连接的流量统计信息		**reset tcp statistics**
清除 UDP 流量统计信息		**reset udp statistics**

目 录

1 IP单播策略路由配置 1-1
1.1 IP单播策略路由简介 1-1
1.1.1 策略路由 1-1
1.1.2 策略路由与Track关联 1-1
1.2 配置IP单播策略路由 1-1
1.2.1 配置策略 1-1
1.2.2 配置系统策略路由 1-4
1.2.3 配置接口策略路由 1-4
1.3 IP单播策略路由显示和维护 1-4
1.4 IP单播策略路由典型配置举例 1-5
1.4.1 基于源地址的策略路由的配置举例 1-5
1.4.2 基于报文大小的策略路由的配置举例 1-6
1.4.3 反向入接口策略路由配置举例 1-7

1 IP 单播策略路由配置

1.1 IP 单播策略路由简介

1.1.1 策略路由

策略路由（policy-based-route）是一种依据用户制定的策略进行路由选择的机制。与单纯依照 IP 报文的目的地址查找路由表进行转发不同，策略路由基于到达报文的源地址、长度等信息灵活地进行路由选择。

策略路由可分为系统策略路由和接口策略路由：

- 系统策略路由对本地产生的报文进行策略路由，它只对本地产生的报文起作用，对转发的报文不起作用；
- 接口策略路由作用于到达该接口的报文，它只对转发的报文起作用，对本地产生的报文（比如本地的 ping 报文）不起作用。

一般来讲，策略路由的优先级要高于普通路由，即报文先按照策略路由进行转发。如果无法匹配所有的策略路由条件，再按照普通路由进行转发。但对于配置了缺省出接口（下一跳）的情况，则是先进行普通路由的转发，如果无法匹配，再进行策略路由转发。

1.1.2 策略路由与 Track 关联

IP 单播策略路由通过关联 Track 机制，增强了应用的灵活性和策略路由对网络环境的动态感知能力。Track 的状态分为 Positive、Negative 和 Invalid，策略路由配置仅在关联的 Track 项状态为 Positive 或 Invalid 时生效。

Track 相关配置可参考“系统分册/Track 命令”中的命令 track。

1.2 配置 IP 单播策略路由

1.2.1 配置策略

策略可以包含若干节点，每个策略节点由 *node-number* 来指定，*node-number* 的值越小优先级越高，优先级高的策略优先被执行。一个策略用来引入一条路由，对 IP 报文转发进行路由选择。策略的具体内容由 **if-match** 和 **apply** 子句来指定。

if-match 子句定义该节点的匹配规则，**apply** 子句定义通过该节点过滤后进行的动作。**if-match** 子句定义了那些需要使用策略的报文，当报文满足 **if-match** 子句时，则执行策略对应节点的 **apply** 子句，以完成报文的转发。

在一个策略中，每个节点的 **if-match** 子句之间的过滤关系是“与”的关系，即报文必须满足该节点的所有 **if-match** 子句才能执行该节点的 **apply** 子句。

目前 IP 单播策略路由提供了三种 **if-match** 子句，分别为 **if-match packet-length**、**if-match acl** 和 **if-match reverse-input-interface**，同一类型的 **if-match** 子句只能有一条；IP 单播策略路由同时提供六种 **apply** 子句：**apply access-vpn vpn-instance**，**apply ip-precedence**，**apply output-interface**，**apply ip-address next-hop**，**apply default output-interface**，**apply**

ip-address default next-hop，同一类型的 **apply** 子句最多只能配置一个。在满足所有 **if-match** 子句的情况下，**apply** 子句执行的优先级情况如下：

- 配置转发报文的 VPN 实例：apply access-vpn vpn-instance，只要配置了该子句，就不会执行其他 apply 子句。报文如果匹配了其中一个 VPN 实例下的转发表，报文将在该 VPN 实例中进行转发，如果对于所有配置的 VPN 实例，报文都未能匹配其下的转发表，该报文将被丢弃；
- 配置报文的优先级：**apply ip-precedence**，在公网转发中，即在未配置 **apply access-vpn vpn-instance** 的情况下，只要配置了该子句，该子句就一定会执行；
- 配置策略路由出接口和下一跳：**apply output-interface** 和 **apply ip-address next-hop**，其中 **apply ip-address next-hop** 命令的优先级高于 **apply output-interface**。当两条命令同时配置并且都有效时，系统只会执行 **apply ip-address next-hop** 命令；
- 配置策略路由缺省出接口和下一跳：**apply default output-interface** 和 **apply ip-address default next-hop**，同样，**apply ip-address default next-hop** 命令的优先级高于 **apply default output-interface**。当两条命令同时配置并且都有效时，系统只会执行 **apply ip-address default next-hop** 命令。执行缺省出接口和下一跳命令的前提是，在策略路由中报文没有配置出接口或者下一跳，或者配置的出接口和下一跳无效，并且报文目的 IP 地址在路由表中没有查到相应的路由，这时才会使用策略路由配置的缺省下一跳或者出接口。

在一个策略中，节点之间的过滤关系是“或”的关系，即只要通过一个节点的过滤，就意味着通过该策略路由的过滤。

在配置策略节点时，可能需要指定节点的匹配模式为 **permit** 或者 **deny**：

- **permit** 指定所定义的策略节点的匹配模式为允许模式。当报文满足该节点的所有 **if-match** 子句时被允许通过该节点的过滤并执行该节点的 **apply** 子句，如报文不满足该节点的 **if-match** 子句，报文将会使用该条策略的下一个节点进行匹配。
- **deny** 指定所定义的策略节点的匹配模式为拒绝模式，当报文满足该节点的所有 **if-match** 子句时，被拒绝通过该节点的过滤，并且不会进行下一个节点的匹配。

通过一个节点所定义的策略的报文将不再参与其他策略的过滤和处理。如果报文不能通过一个策略所有节点的过滤，则认为没有通过该策略。该报文按正常转发流程处理。

在一个策略中最多可以指定两个下一跳或者两个出接口，此时，报文的转发将在两个合法参数中负载分担，轮流在每一个下一跳或者出接口上发送一个报文。

策略路由可以在配置报文的发送接口、缺省发送接口、下一跳、缺省下一跳时与 Track 项关联，通过 Track 项的状态来动态地决定策略的可用性。当应用动作增加 Track 关联后，如果事件发生时 Track 项状态为 Positive 或 Invalid，则该配置项有效，可以指导转发；事件发生时，Track 项状态为 Negative，则该配置项无效，转发时忽略该配置项。

表1-1 配置策略

操作	命令	说明
进入系统视图	**system-view**	-
创建策略或一个策略节点，并进入该策略视图	**policy-based-route** *policy-name* [**deny** | **permit**] **node** *node-number*	必选
设置 IP 报文长度匹配条件	**if-match packet-length** *min-len max-len*	可选

操作	命令	说明
设置 ACL 匹配条件	**if-match acl** *acl-number*	可选
设置反向入接口匹配条件	**if-match reverse-input-interface** *interface-type interface-number*	可选
设置报文在指定 VPN 实例进行转发	**apply access-vpn vpn-instance** *vpn-instance-name*&<1-6>	可选
设置报文的优先级	**apply ip-precedence** { *type* | *value* }	可选
设置报文的发送接口	**apply output-interface** *interface-type interface-number* [**track** *track-entry-number*] [*interface-type interface-number* [**track** *track-entry-number*]]	可选 用户可以同时配置两个发送接口，这两个发送接口同时有效，可以起到负载分担的作用 对于非 P2P 接口（广播类型的接口和 NBMA 类型的接口），比如以太网接口，由于有多个可能的下一跳，可能会造成报文转发不成功的现象
设置报文的下一跳	**apply ip-address next-hop** *ip-address* [**direct**] [**track** *track-entry-number*] [*ip-address* [**direct**] [**track** *track-entry-number*]] [**standby**]	可选 用户可以同时配置两个下一跳，这两个下一跳同时有效，可以起到负载分担的作用
设置报文缺省发送接口	**apply default output-interface** *interface-type interface-number* [**track** *track-entry-number*] [*interface-type interface-number* [**track** *track-entry-number*]]	可选 用户可以同时配置两个缺省发送接口，这两个发送接口同时有效，可以起到负载分担的作用
设置报文缺省下一跳	**apply ip-address default next-hop** *ip-address* [**direct**] [**track** *track-entry-number*] [*ip-address* [**direct**] [**track** *track-entry-number*] [**standby**]]	可选 用户可以同时配置两个缺省下一跳，这两个下一跳同时有效，可以起到负载分担的作用

注意

- **apply output-interface** 和 **apply ip-address next-hop** 可以配置两个出接口或者下一跳，当用户希望修改其中一个时，只需要继续执行 **apply output-interface** *interface-type interface-number* 命令或者 **apply ip-address next-hop** *ip-address* 命令，系统就会将最早配置的出接口或者下一跳替换；如果先前配置的两个出接口或者下一跳都需要修改，则直接在 **apply output-interface** 命令或者 **apply ip-address next-hop** 命令后配置两个参数即可。
- 当使用 **if-match reverse-input-interface** *interface-type interface-number* 命令配置接口后，若接口所在接口卡被拔出或该接口被删除的情况下，则显示信息中策略节点上 `if-match reverse-input-interface` 后配置的接口会消失，该命令配置失效，此时，任何报文都不能被该策略路由节点匹配。
- **if-match reverse-input-interface** 命令需要和会话(session)配合使用。配置 **if-match reverse-input-interface** 命令之前，可通过执行 **firewall** 或 **nat server** 等命令使能会话功能(分布式设备需要在流量的入出接口板都使能会话),这样反向策略路由功能才能生效。关于 **firewall**、**nat server** 等命令的详细介绍请参见“安全分册”中“防火墙配置”、“NAT 配置”或其它相关手册。

1.2.2 配置系统策略路由

策略路由可以分为系统策略路由和接口策略路由。对于一般转发和安全等方面的使用需求，大多数情况下使用的是接口策略路由。

系统策略路由对本机产生的报文进行路由选择。用户可以分别配置接口策略路由和系统策略路由，在使能系统策略路由时，只能引用一个策略。

表1-2 配置系统策略路由

操作	命令	说明
进入系统视图	**system-view**	-
使能系统策略路由	**ip local policy-based-route** *policy-name*	必选 缺省情况下，没有配置系统策略路由

1.2.3 配置接口策略路由

接口策略路由（应用于报文到达的接口上），作用于到达该接口的报文。用户可以分别配置接口策略路由和系统策略路由，在使能接口策略路由时，一个接口只能引用一个策略。

表1-3 配置接口策略路由

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能接口策略路由	**ip policy-based-route** *policy-name*	必选 缺省情况下，没有配置接口策略路由

1.3 IP 单播策略路由显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置 IP 单播策略路由后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令重置策略路由的统计信息。

表1-4 IP 单播策略路由显示和维护

操作	命令
显示系统和接口设置的所有策略路由的信息	**display ip policy-based-route**
显示系统的策略路由的设置情况	**display ip policy-based-route setup** { **interface** *interface-type interface-number* | **local** | *policy-name* }
显示策略路由的统计信息	**display ip policy-based-route statistics** { **interface** *interface-type interface-number* | **local** }
显示已经配置的策略路由	**display policy-based-route** [*policy-name*]
重置策略路由的统计信息	**reset policy-based-route statistics** [*policy-name*]

1.4 IP 单播策略路由典型配置举例

1.4.1 基于源地址的策略路由的配置举例

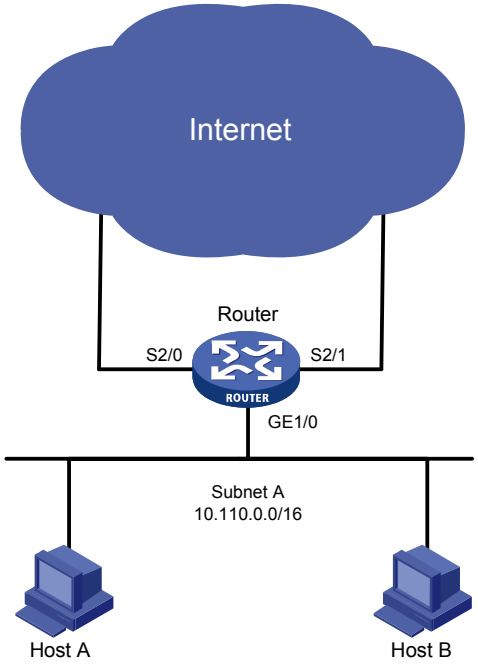
1. 组网需求

定义策略 aaa 的策略路由为控制所有从以太网接口 GigabitEthernet1/0 接收的 TCP 报文，使用串口 Serial2/0 发送，对其它报文，按照查找路由表的方式进行转发。

- 5 号节点，表示匹配 ACL 3101 的报文将被发往串口 Serial2/0；
- 10 号节点，表示匹配 ACL 3102 的报文不做策略路由处理；

2. 组网图

图1-1 基于源地址的策略路由的配置举例组网图



3. 配置步骤

如果设备支持防火墙功能，首先设置防火墙的过滤方式为 **deny**。

```
<Router> system-view
[Router] firewall default deny
```

定义访问控制列表。

```
[Router] acl number 3101
[Router-acl-adv-3101] rule permit tcp
[Router-acl-adv-3101] quit
[Router] acl number 3102
[Router-acl-adv-3102] rule permit ip
[Router-acl-adv-3102] quit
```

定义 5 号节点，使匹配 ACL 3101 的任何 TCP 报文被发往串口 Serial2/0。

```
[Router] policy-based-route aaa permit node 5
[Router-pbr-aaa-5] if-match acl 3101
```

```
[Router-pbr-aaa-5] apply output-interface serial 2/0
[Router-pbr-aaa-5] quit
```

定义 10 号节点，表示匹配 ACL 3102 的报文不做策略路由处理，按照查找路由表的方式进行转发。

```
[Router] policy-based-route aaa deny node 10
[Router-pbr-aaa-10] if-match acl 3102
[Router-pbr-aaa-10] quit
```

在以太网接口上应用策略 aaa。

```
[Router] interface GigabitEthernet 1/0
[Router-GigabitEthernet1/0] ip policy-based-route aaa
```

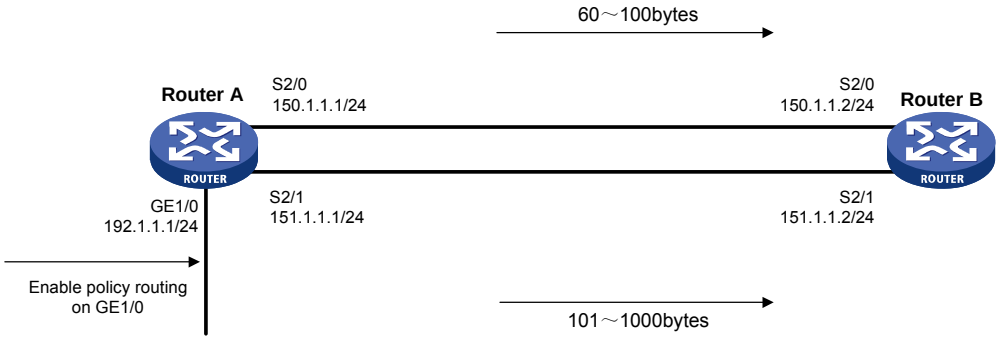
1.4.2 基于报文大小的策略路由的配置举例

1. 组网需求

在 Router A 的 GigabitEthernet1/0 接口上应用策略 lab1。这个策略将大小为 64～100 字节的报文设置 150.1.1.2/24 作为下一跳 IP 地址；而将大小为 101～1000 字节的报文设置 151.1.1.2/24 作为下一跳 IP 地址。所有其它长度的报文都按照查找路由表的方式转发。

2. 组网图

图1-2 基于报文大小的策略路由的配置举例组网图



3. 配置步骤

(1) 配置 Router A

配置动态路由协议 RIP。

```
<RouterA> system-view
[RouterA] rip
[RouterA-rip-1] network 192.1.1.0
[RouterA-rip-1] network 150.1.0.0
[RouterA-rip-1] network 151.1.0.0
[RouterA-rip-1] quit
```

在以太网口 GigabitEthernet1/0 上应用定义的策略 lab1，处理此接口接收的报文。

```
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ip address 192.1.1.1 255.255.255.0
[RouterA-GigabitEthernet1/0] ip policy-based-route lab1
[RouterA-GigabitEthernet1/0] quit
```

将大小为 64～100 字节的报文转发到下一跳 150.1.1.2；而将大小为 101～1000 字节的报文转发到下一跳 151.1.1.2。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 150.1.1.1 255.255.255.0
[RouterA-Serial2/0] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ip address 151.1.1.1 255.255.255.0
[RouterA-Serial2/1] quit
[RouterA] policy-based-route lab1 permit node 10
[RouterA-pbr-lab1-10] if-match packet-length 64 100
[RouterA-pbr-lab1-10] apply ip-address next-hop 150.1.1.2
[RouterA-pbr-lab1-10] quit
[RouterA] policy-based-route lab1 permit node 20
[RouterA-pbr-lab1-20] if-match packet-length 101 1000
[RouterA-pbr-lab1-20] apply ip-address next-hop 151.1.1.2
```

(2) 配置 Router B

配置动态路由协议 RIP。

```
<RouterB> system-view
[RouterB] rip
[RouterB-rip-1] network 150.1.0.0
[RouterB-rip-1] network 151.1.0.0
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address 150.1.1.2 255.255.255.0
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ip address 151.1.1.2 255.255.255.0
[RouterB-Serial2/1] quit
```

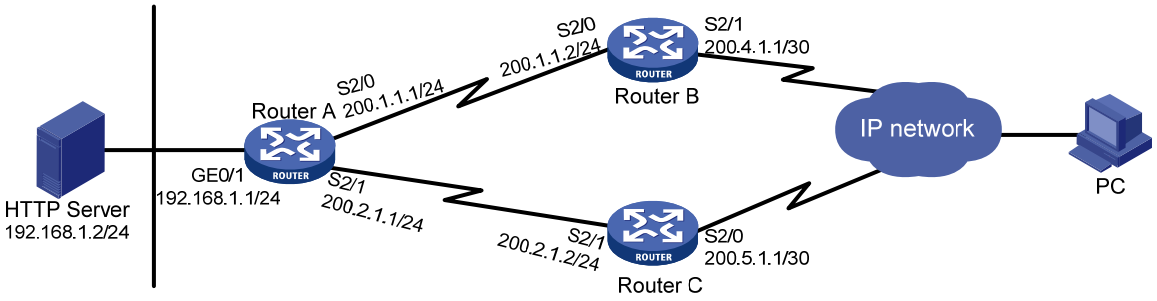
1.4.3 反向入接口策略路由配置举例

1. 组网需求

网关设备 Router A 通过两个接口（Serail2/0 和 Serial2/1）和公网连接。用户 PC 从公网访问内网的 HTTP Server 服务，不妨设 PC 请求报文从 Router A 接口 Serial2/0 进入，通过网关设备转发，从 Router A 的私网接口 GigabitEthernet 0/1 进入内网访问 HTTP Server 服务器。要求从私网返回的响应报文从 Router A 的接口 GigabitEthernet 0/1 进入，经 Router A 转发时，能够从原来请求报文的入接口 Serial2/0 进入公网，返回用户 PC。

2. 组网图

图1-3 反向入接口策略路由配置举例组网图



3. 配置步骤

配置 Router A 各接口 IP 地址，并保证 Router A 与公网连通（略）。

在接口 Serial2/0 上配置内部服务器功能，将 HTTP Server 的 IP 地址 192.168.1.2/24 映射为 200.1.1.100/24（和 Router A 的接口 Serial2/0 的 IP 地址在同一网段）。

```
<RouterA> system-view
[RouterA] interface Serial2/0
[RouterA-Serial2/0] nat server protocol tcp global 200.1.1.100 www inside 192.168.1.2 www
[RouterA-Serial2/0] quit
```

在接口 Serial2/1 上配置内部服务器功能，将 HTTP Server 的 IP 地址 192.168.1.2/24 映射为 200.2.1.100/24（和 Router A 的接口 Serial2/1 的 IP 地址在同一网段）。

```
[RouterA] interface Serial2/1
[RouterA-Serial2/1] nat server protocol tcp global 200.2.1.100 www inside 192.168.1.2 www
[RouterA-Serial2/1] quit
```

定义 10 号节点，使匹配反向入接口 Serial2/0 的报文的下一跳地址为 200.1.1.2/16。

```
[RouterA] policy-based-route test permit node 10
[RouterA-pbr-test-10] if-match reverse-input-interface Serial2/0
[RouterA-pbr-test-10] apply ip-address next-hop 200.1.1.2
[RouterA-pbr-test-10] quit
```

在以太网接口 GigabitEthernet0/1 上应用策略 test。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ip policy-based-route test
```

目 录

1 配置URPF 1-1
1.1 URPF简介 1-1
1.1.1 URPF概述 1-1
1.1.2 URPF处理流程 1-1
1.2 配置URPF 1-2

1 配置 URPF

说明

在以下的介绍中所指的路由器，代表了一般意义下的路由器以及运行了路由协议的以太网交换机。为提高可读性，在手册的描述中将不另行说明。

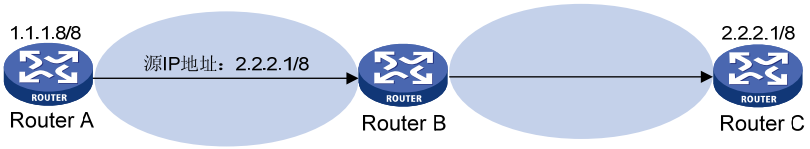
1.1 URPF 简介

1.1.1 URPF 概述

URPF（Unicast Reverse Path Forwarding，单播反向路径转发）的主要功能是用于防止基于源地址欺骗的网络攻击行为。

源地址欺骗攻击为入侵者构造出一系列带有伪造源地址的报文，对于使用基于 IP 地址验证的应用来说，此攻击方法可以导致未被授权用户以他人身份获得访问系统的权限，甚至是以管理员权限来访问。即使响应报文不能达到攻击者，同样也会造成对被攻击对象的破坏。

图1-1 源地址欺骗攻击示意图



如 图 1-1所示，在Router A上伪造源地址为 2.2.2.1/8 的报文，向服务器Router B发起请求，Router B响应请求时将向真正的“2.2.2.1/8”发送报文。这种非法报文对Router B和Router C都造成了攻击。

URPF 技术可以应用在上述环境中，阻止基于源地址欺骗的攻击。

1.1.2 URPF 处理流程

URPF 检查有严格（strict）型和松散（loose）型两种。此外，还可以支持 ACL 与链路层、缺省路由的检查。

URPF 的处理流程如下：

(1) 首先检查源地址合法性。

- 对于广播地址，直接予以丢弃。
- 对于全零地址，如果目的地址不是广播，则丢弃。（源地址为 0.0.0.0，目的地址为 255.255.255.255 的报文，可能是 DHCP 或者 BOOTP 报文，不做丢弃处理。）

(2) 如果报文的源地址在路由器的 FIB 表中存在

- 对于 strict 型检查，反向查找报文出接口，若其中至少有一个出接口和报文的入接口相匹配，则报文通过检查；否则报文将被拒绝。（反向查找是指查找以该报文源 IP 地址为目的 IP 地址的报文的出接口）
- 对于 loose 型检查，报文通过检查。

(3) 如果报文的源地址在路由器的 FIB 表中存在，且报文通过了检查，则检查链路层信息。

- 如果没有配置参数 link-check，则认为报文通过检查，进行正常的转发。
- 如果配置了参数 link-check，则根据 FIB 表中的下一跳查询 ARP 表，并比较 IP 报文源 MAC 地址与 ARP 表中的 MAC 地址是否一致，如果两者一致，则报文通过检查；如果查询失败或两者不一致，则报文将被拒绝。

(4) 如果报文的源地址在路由器的 FIB 表中不存在，则检查缺省路由及 URPF 的 allow-default-route 参数。

- 对于配置了缺省路由，但没有配置参数 allow-default-route 的情况，不管是 strict 型检查还是 loose 型检查，只要报文的源地址在路由器的 FIB 表中不存在，该报文都将被拒绝；
- 对于配置了缺省路由，同时又配置了参数 allow-default-route 的情况下，如果是 strict 型检查，只要缺省路由的出接口与报文的入接口一致，则报文将通过 URPF 的检查，进行正常的转发；如果缺省路由的出接口和报文的入接口不一致，则报文将拒绝。如果是 loose 型检查，报文都将通过 URPF 的检查，进行正常的转发。

(5) 当且仅当报文被拒绝后，才去匹配 ACL。如果被 ACL 允许通过，则报文继续进行正常的转发；如果被 ACL 拒绝，则报文被丢弃。

1.2 配置 URPF

表1-1 配置 URPF

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 URPF 检查	**ip urpf** { **loose** | **strict** } [**allow-default-route**] [**acl** *acl-number*]	必选 缺省情况下，禁止 URPF 检查

说明

- URPF 检查仅对接口收到的报文有效。
- 可以通过 **display ip interface** 命令查看 URPF 功能丢弃报文的统计信息。

目 录

1 IPv6 基础配置 1-1
1.1 IPv6 简介 1-1
1.1.1 IPv6 协议特点 1-1
1.1.2 IPv6 地址介绍 1-2
1.1.3 IPv6 邻居发现协议介绍 1-4
1.1.4 IPv6 PMTU发现 1-7
1.1.5 IPv6 DNS简介 1-7
1.1.6 协议规范 1-8
1.2 IPv6 基础配置任务简介 1-8
1.3 配置IPv6 基本功能 1-8
1.3.1 使能IPv6 报文转发功能 1-8
1.3.2 配置IPv6 单播地址 1-9
1.4 配置IPv6 邻居发现协议 1-10
1.4.1 配置静态邻居表项 1-10
1.4.2 配置接口上允许动态学习的邻居的最大个数 1-10
1.4.3 配置RA消息的相关参数 1-11
1.4.4 配置重复地址检测时发送邻居请求消息的次数 1-13
1.5 配置PMTU发现 1-13
1.5.1 配置接口MTU 1-13
1.5.2 配置指定地址的静态PMTU 1-13
1.5.3 配置PMTU老化时间 1-14
1.6 配置TCP6 1-14
1.7 配置IPv6 FIB转发功能 1-14
1.8 配置ICMPv6 报文发送 1-15
1.8.1 配置指定时间内发送ICMPv6 差错报文的最大个数 1-15
1.8.2 配置允许回复组播形式的Echo request报文 1-15
1.9 配置IPv6 DNS Client 1-16
1.9.1 配置静态IPv6 DNS 1-16
1.9.2 配置动态IPv6 DNS 1-16
1.10 IPv6 基础显示和维护 1-17
1.11 IPv6 基础典型配置举例 1-18
1.12 常见配置错误举例 1-21

1 IPv6 基础配置

1.1 IPv6 简介

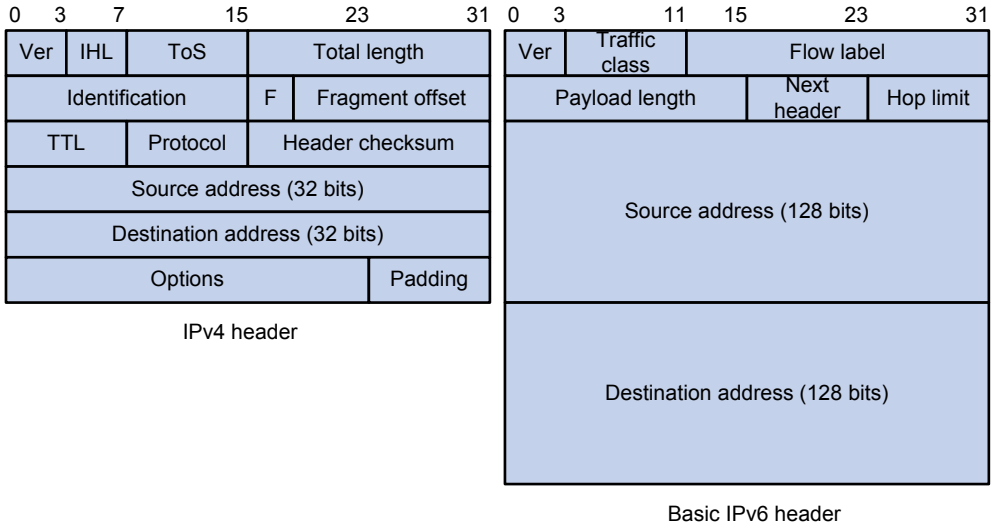
IPv6（Internet Protocol Version 6，因特网协议版本 6）是网络层协议的第二代标准协议，也被称为 IPng（IP Next Generation，下一代因特网），它是 IETF（Internet Engineering Task Force，Internet 工程任务组）设计的一套规范，是 IPv4 的升级版本。IPv6 和 IPv4 之间最显著的区别为：IP 地址的长度从 32 比特增加到 128 比特。

1.1.1 IPv6 协议特点

1. 简化的报文头格式

通过将 IPv4 报文头中的某些字段裁减或移入到扩展报文头，减小了 IPv6 基本报文头的长度。IPv6 使用固定长度的基本报文头，从而简化了转发设备对 IPv6 报文的处理，提高了转发效率。尽管 IPv6 地址长度是 IPv4 地址长度的四倍，但 IPv6 基本报文头的长度只有 40 字节，为 IPv4 报文头长度（不包括选项字段）的两倍。

图1-1 IPv4 报文头和 IPv6 基本报文头格式比较



2. 充足的地址空间

IPv6 的源地址与目的地址长度都是 128 比特（16 字节）。它可以提供超过 3.4×10^{38} 种可能的地址空间，完全可以满足多层次的地址划分需要，以及公有网络和机构内部私有网络的地址分配。

3. 层次化的地址结构

IPv6 的地址空间采用了层次化的地址结构，有利于路由快速查找，同时可以借助路由聚合，有效减少 IPv6 路由表占用的系统资源。

4. 地址自动配置

为了简化主机配置，IPv6 支持有状态地址配置和无状态地址配置：

- 有状态地址配置是指从服务器（如 DHCP 服务器）获取 IPv6 地址及相关信息；

- 无状态地址配置是指主机根据自己的链路层地址及路由器发布的前缀信息自动配置 IPv6 地址及相关信息。

同时，主机也可根据自己的链路层地址及默认前缀（FE80::/10）形成链路本地地址，实现与本链路上其他主机的通信。

5. 内置安全性

IPv6 将 IPSec 作为它的标准扩展头，可以提供端到端的安全特性。这一特性也为解决网络安全问题提供了标准，并提高了不同 IPv6 应用之间的互操作性。

6. 支持 QoS

IPv6 报文头的流标签（Flow Label）字段实现流量的标识，允许设备对某一流中的报文进行识别并提供特殊处理。

7. 增强的邻居发现机制

IPv6 的邻居发现协议是通过一组 ICMPv6（Internet Control Message Protocol for IPv6，IPv6 的因特网控制报文协议）消息实现的，管理着邻居节点间（即同一链路上的节点）信息的交互。它代替了 ARP（Address Resolution Protocol，地址解析协议）、ICMPv4 路由器发现和 ICMPv4 重定向消息，并提供了一系列其他功能。

8. 灵活的扩展报文头

IPv6 取消了 IPv4 报文头中的选项字段，并引入了多种扩展报文头，在提高处理效率的同时还大大增强了 IPv6 的灵活性，为 IP 协议提供了良好的扩展能力。IPv4 报文头中的选项字段最多只有 40 字节，而 IPv6 扩展报文头的大小只受到 IPv6 报文大小的限制。

1.1.2 IPv6 地址介绍

1. IPv6 地址表示方式

IPv6 地址被表示为以冒号（:）分隔的一连串 16 比特的十六进制数。每个 IPv6 地址被分为 8 组，每组的 16 比特用 4 个十六进制数来表示，组和组之间用冒号隔开，比如：2001:0000:130F:0000:0000:09C0:876A:130B。

为了简化 IPv6 地址的表示，对于 IPv6 地址中的“0”可以有下面的处理方式：

- 每组的前导“0”可省略，即上述地址可写为 2001:0:130F:0:0:9C0:876A:130B。
- 如果地址中包含连续两个或多个均为 0 的组，则可以用双冒号“::”来代替，即上述地址可写为 2001:0:130F::9C0:876A:130B。



注意

在一个 IPv6 地址中只能使用一次双冒号“::”，否则当设备将“::”转变为 0 以恢复 128 位地址时，将无法确定“::”所代表的 0 的个数。

IPv6 地址由两部分组成：地址前缀与接口标识。其中，地址前缀相当于 IPv4 地址中的网络号码字段部分，接口标识相当于 IPv4 地址中的主机号码部分。

地址前缀的表示方式为：IPv6 地址/前缀长度。其中，IPv6 地址是前面所列出的任一形式，而前缀长度是一个十进制数，表示 IPv6 地址最左边多少位为地址前缀。

2. IPv6 的地址分类

IPv6 主要有三种类型的地址：单播地址、组播地址和任播地址。

- 单播地址：用来唯一标识一个接口，类似于 IPv4 的单播地址。发送到单播地址的数据报文将被传送给此地址所标识的接口。
- 组播地址：用来标识一组接口（通常这组接口属于不同的节点），类似于 IPv4 的组播地址。发送到组播地址的数据报文被传送给此地址所标识的所有接口。
- 任播地址：用来标识一组接口（通常这组接口属于不同的节点）。发送到任播地址的数据报文被传送给此地址所标识的一组接口中距离源节点最近（根据使用的路由协议进行度量）的一个接口。



说明

IPv6 中没有广播地址，广播地址的功能通过组播地址来实现。

IPv6 地址类型是由地址前面几位（称为格式前缀）来指定的，主要地址类型与格式前缀的对应关系如 表 1-1所示。

表1-1 地址类型与格式前缀的对应关系

地址类型		格式前缀（二进制）	IPv6 前缀标识
单播地址	未指定地址	00...0 (128 bits)	::/128
	环回地址	00...1 (128 bits)	::1/128
	链路本地地址	1111111010	FE80::/10
	站点本地地址	1111111011	FEC0::/10
	全球单播地址	其他形式	-
组播地址		11111111	FF00::/8
任播地址		从单播地址空间中进行分配，使用单播地址的格式	

3. 单播地址的类型

IPv6 单播地址的类型可有多种，包括全球单播地址、链路本地地址和站点本地地址等。

- 全球单播地址等同于 IPv4 公网地址，提供给网络服务提供商。这种类型的地址允许路由前缀的聚合，从而限制了全球路由表项的数量。
- 链路本地地址用于邻居发现协议和无状态自动配置中链路本地上节点之间的通信。使用链路本地地址作为源或目的地址的数据报文不会被转发到其他链路上。
- 站点本地地址与 IPv4 中的私有地址类似。使用站点本地地址作为源或目的地址的数据报文不会被转发到本站点（相当于一个私有网络）外的其它站点。
- 环回地址：单播地址 0:0:0:0:0:0:0:1（简化表示为::1）称为环回地址，不能分配给任何物理接口。它的作用与在 IPv4 中的环回地址相同，即节点用来给自己发送 IPv6 报文。
- 未指定地址：地址“::”称为未指定地址，不能分配给任何节点。在节点获得有效的 IPv6 地址之前，可在发送的 IPv6 报文的源地址字段填入该地址，但不能作为 IPv6 报文中的目的地址。

4. 组播地址

表 1-2所示的组播地址，是预留的特殊用途的组播地址。

表1-2 预留的 IPv6 组播地址列表

地址	应用
FF01::1	节点本地范围所有节点组播地址
FF02::1	链路本地范围所有节点组播地址
FF01::2	节点本地范围所有路由器组播地址
FF02::2	链路本地范围所有路由器组播地址
FF05::2	站点本地范围所有路由器组播地址

另外，还有一类组播地址：被请求节点（Solicited-Node）地址。该地址主要用于获取同一链路上邻居节点的链路层地址及实现重复地址检测。每一个单播或任播 IPv6 地址都有一个对应的被请求节点地址。其格式为：

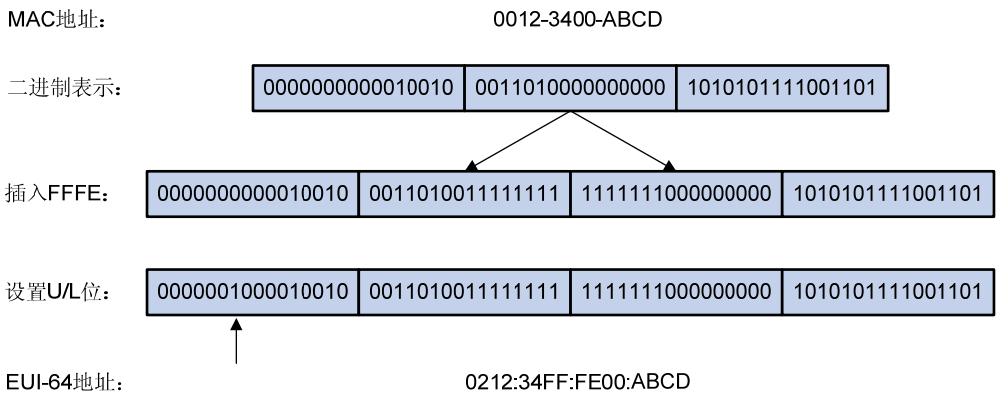
FF02:0:0:0:0:1:FFXX:XXXX

其中，FF02:0:0:0:0:1:FF 为 104 位固定格式；XX:XXXX 为单播或任播 IPv6 地址的后 24 位。

5. IEEE EUI-64 格式的接口标识符

IPv6 单播地址中的接口标识符用来标识链路上的一个唯一的接口。目前 IPv6 单播地址基本上都要求接口标识符为 64 位。IEEE EUI-64 格式的接口标识符是从接口的链路层地址（MAC 地址）变化而来的。IPv6 地址中的接口标识符是 64 位，而 MAC 地址是 48 位，因此需要在 MAC 地址的中间位置（从高位开始的第 24 位后）插入十六进制数 FFFE（1111111111111110）。为了确保这个从 MAC 地址得到的接口标识符是唯一的，还要将 Universal/Local (U/L)位（从高位开始的第 7 位）设置为“1”。最后得到的这组数就作为 EUI-64 格式的接口标识符。

图1-2 MAC 地址到 EUI-64 格式接口标识符的转换过程



1.1.3 IPv6 邻居发现协议介绍

IPv6 邻居发现协议使用五种类型的 ICMPv6 消息，实现下面一些功能：地址解析、验证邻居是否可达、重复地址检测、路由器发现/前缀发现、地址自动配置和重定向等功能。

邻居发现协议使用的ICMPv6 消息的类型及作用如 表 1-3所示。

表1-3 邻居发现协议使用的 ICMPv6 消息类型及作用

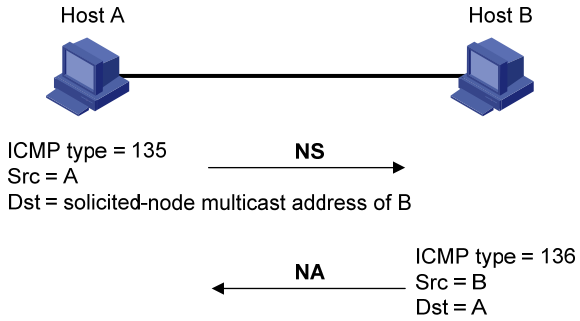
ICMPv6 消息	类型号	作用
邻居请求消息 NS（Neighbor Solicitation）	135	获取邻居的链路层地址
		验证邻居是否可达
		进行重复地址检测
邻居通告消息 NA（Neighbor Advertisement）	136	对 NS 消息进行响应
		节点在链路层变化时主动发送 NA 消息，向邻居节点通告本节点的变化信息
路由器请求消息 RS（Router Solicitation）	133	节点启动后，通过 RS 消息向路由器发出请求，请求前缀和其他配置信息，用于节点的自动配置
路由器通告消息 RA（Router Advertisement）	134	对 RS 消息进行响应
		在没有抑制 RA 消息发布的条件下，路由器会周期性地发布 RA 消息，其中包括前缀信息选项和一些标志位的信息
重定向消息（Redirect）	137	当满足一定的条件时，缺省网关通过向源主机发送重定向消息，使主机重新选择正确的下一跳地址进行后续报文的发送

邻居发现协议提供的主要功能如下：

1. 地址解析

获取同一链路上邻居节点的链路层地址（与IPv4 的ARP功能相同），通过邻居请求消息NS和邻居通告消息NA实现。如 图 1-3所示，节点A要获取节点B的链路层地址。

图1-3 地址解析示意图



(1) 节点 A 以组播方式发送 NS 消息。NS 消息的源地址是节点 A 的接口 IPv6 地址，目的地址是节点 B 的被请求节点组播地址，消息内容中包含了节点 A 的链路层地址。

(2) 节点 B 收到 NS 消息后，判断报文的目的地址是否为自己的 IPv6 地址对应的被请求节点组播地址。如果是，则节点 B 可以学习到节点 A 的链路层地址，并以单播方式返回 NA 消息，其中包含了自己的链路层地址。

(3) 节点 A 从收到的 NA 消息中就可获取到节点 B 的链路层地址。

2. 验证邻居是否可达

在获取到邻居节点的链路层地址后，通过邻居请求消息 NS 和邻居通告消息 NA 可以验证邻居节点是否可达。

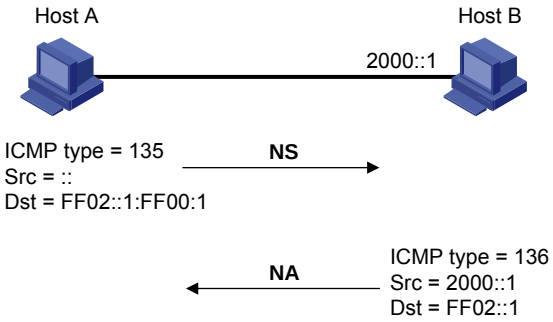
(1) 节点发送 NS 消息，其中目的地址是邻居节点的 IPv6 地址。

(2) 如果收到邻居节点的确认报文，则认为邻居可达；否则，认为邻居不可达。

3. 重复地址检测

当节点获取到一个IPv6 地址后，需要使用重复地址检测功能确定该地址是否已被其他节点使用（与IPv4 的免费ARP功能相似）。通过NS和NA可以实现重复地址检测，如 图 1-4所示。

图1-4 重复地址检测示意图



(1) 节点 A 发送 NS 消息，NS 消息的源地址是未指定地址::，目的地址是待检测的 IPv6 地址对应的被请求节点组播地址，消息内容中包含了待检测的 IPv6 地址。

(2) 如果节点 B 已经使用这个 IPv6 地址，则会返回 NA 消息。其中包含了自己的 IPv6 地址。

(3) 节点 A 收到节点 B 发来的 NA 消息，就知道该 IPv6 地址已被使用。反之，则说明该地址未被使用，节点 A 就可使用此 IPv6 地址。

4. 路由器发现/前缀发现及地址自动配置

路由器发现/前缀发现是指节点从收到的 RA 消息中获取邻居路由器及所在网络的前缀，以及其他配置参数。

地址无状态自动配置是指节点根据路由器发现/前缀发现所获取的信息，自动配置 IPv6 地址。

路由器发现/前缀发现通过路由器请求消息 RS 和路由器通告消息 RA 来实现，具体过程如下：

(1) 节点启动时，通过 RS 消息向路由器发出请求，请求前缀和其他配置信息，以便用于节点的配置。

(2) 路由器返回 RA 消息，其中包括前缀信息选项（路由器也会周期性地发布 RA 消息）。

(3) 节点利用路由器返回的 RA 消息中的地址前缀及其他配置参数，自动配置接口的 IPv6 地址及其他信息。

说明

- 前缀信息选项中不仅包括地址前缀的信息，还包括该地址前缀的首选生命期（preferred lifetime）和有效生命期（valid lifetime）。节点收到周期性发送的 RA 消息后，会根据该消息更新前缀的首选生命期和有效生命期。
- 在有效生命期内，自动生成的地址可以正常使用；有效生命期过期后，自动生成的地址将被删除。

---

5. 重定向功能

当主机启动时，它的路由表中可能只有一条到缺省网关的缺省路由。当满足一定的条件时，缺省网关会向源主机发送 ICMPv6 重定向消息，通知主机选择更好的下一跳进行后续报文的发送（与 IPv4 的 ICMP 重定向消息的功能相同）。

设备在满足下列条件时会发送对主机重定向的 ICMPv6 重定向报文：

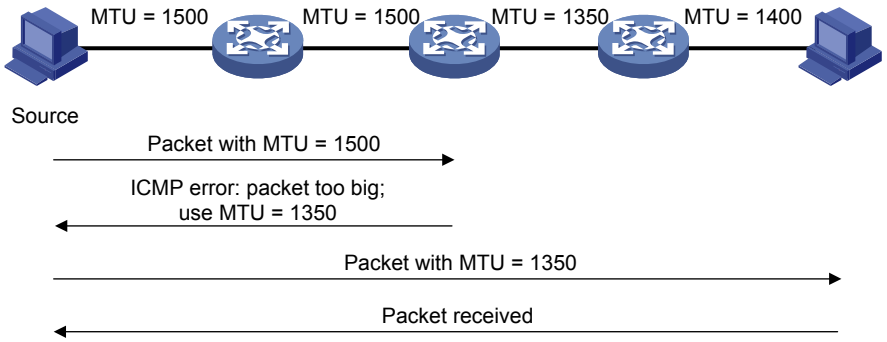
- 接收和转发数据报文的接口是同一接口；
- 被选择的路由本身没有被 ICMPv6 重定向报文创建或修改过；
- 被选择的路由不是缺省路由；
- 被转发的 IPv6 数据报文中不包含路由扩展头。

1.1.4 IPv6 PMTU 发现

报文从源端到目的端的传输路径中所经过的链路可能具有不同的 MTU。在 IPv6 中，当报文的长度大于链路的 MTU 时，报文的分片将在源端进行，从而减轻中间转发设备的处理压力，合理利用网络资源。

PMTU（Path MTU，路径MTU）发现机制的目的就是要找到从源端到目的端的路径上最小的MTU。PMTU的工作过程如 图 1-5所示。

图1-5 PMTU 发现工作过程



(1) 源端主机使用自己的 MTU 对报文进行分片，之后向目的主机发送报文。

(2) 中间转发设备接收到该报文进行转发时，如果发现转发报文的接口支持的 MTU 值小于报文长度，则会丢弃报文，并给源端返回一个 ICMPv6 差错报文，其中包含了转发失败的接口的 MTU。

(3) 源主机收到该差错报文后，将使用报文中所携带的 MTU 重新对报文进行分片并发送。

(4) 如此反复，直到目的端主机收到这个报文，从而确定报文从源端到目的端路径中的最小 MTU。

1.1.5 IPv6 DNS 简介

在 IPv6 网络中，通过支持 IPv6 的 DNS（Domain Name System，域名系统）实现域名与 IPv6 地址的转换。IPv6 DNS 与 IPv4 DNS 的区别只是将域名转换为 IPv6 地址，而非 IPv4 地址。

IPv6 DNS 与 IPv4 DNS 相同，也分为静态域名解析和动态域名解析。两种域名解析的作用和实现方式也与 IPv4 DNS 一样。具体描述请参见“IP 业务分册”中的“域名解析配置”。

通常，连接 IPv4 和 IPv6 网络的 DNS 服务器中既包含 A 记录（IPv4 地址）也包含 AAAA 记录（IPv6 地址），可以进行域名与 IPv4 地址的转换，也可以进行域名与 IPv6 地址的转换。此时，DNS 服务器兼具 IPv6 DNS 和 IPv4 DNS 的功能。

1.1.6 协议规范

与 IPv6 基础相关的协议规范有：

- RFC 1881：IPv6 Address Allocation Management
- RFC 1887：An Architecture for IPv6 Unicast Address Allocation
- RFC 1981：Path MTU Discovery for IP version 6
- RFC 2375：IPv6 Multicast Address Assignments
- RFC 2460：Internet Protocol, Version 6 (IPv6) Specification.
- RFC 2461：Neighbor Discovery for IP Version 6 (IPv6)
- RFC 2462：IPv6 Stateless Address Autoconfiguration
- RFC 2463：Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
- RFC 2464：Transmission of IPv6 Packets over Ethernet Networks
- RFC 2526：Reserved IPv6 Subnet Anycast Addresses
- RFC 3307：Allocation Guidelines for IPv6 Multicast Addresses
- RFC 3513：Internet Protocol Version 6 (IPv6) Addressing Architecture
- RFC 3596：DNS Extensions to Support IP Version 6

1.2 IPv6 基础配置任务简介

表1-4 IPv6 基础配置任务简介

配置任务	说明	详细配置
配置 IPv6 基本功能	必选	1.3
配置 IPv6 邻居发现协议	可选	1.4
配置 PMTU 发现	可选	1.5
配置 TCP6	可选	1.6
配置 IPv6 FIB 转发功能	可选	1.7
配置 ICMPv6 报文发送	可选	1.8
配置 IPv6 DNS Client	可选	1.9

1.3 配置 IPv6 基本功能

1.3.1 使能 IPv6 报文转发功能

在进行 IPv6 的相关配置以前，必须先使能 IPv6 报文转发功能。否则即使在接口上配置了 IPv6 地址，仍无法转发 IPv6 的报文，造成 IPv6 网络无法互通。

表1-5 使能 IPv6 报文转发功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，IPv6 报文转发功能处于关闭状态

1.3.2 配置 IPv6 单播地址

(1) IPv6 站点本地地址和全球单播地址可以通过下面三种方式配置：

- 采用 EUI-64 格式形成：当配置采用 EUI-64 格式形成 IPv6 地址时，接口的 IPv6 地址的前缀是所配置的前缀，而接口标识符则由接口的链路层地址转化而来。
- 手工配置：用户手工配置 IPv6 站点本地地址或全球单播地址。

(2) IPv6 的链路本地地址可以通过两种方式获得：

- 自动生成：设备根据链路本地地址前缀（FE80::/10）及接口的链路层地址，自动为接口生成链路本地地址；
- 手动指定：用户手工配置 IPv6 链路本地地址。

表1-6 配置 IPv6 单播地址

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图		**interface** *interface-type interface-number*	-
配置 IPv6 全球单播地址或站点本地地址	手工指定 IPv6 地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* }	二者必选其一 缺省情况下，接口上没有配置站点本地地址和全球单播地址
	采用 EUI-64 格式形成 IPv6 地址	**ipv6 address** *ipv6-address/prefix-length* **eui-64**	
配置 IPv6 链路本地地址	配置自动生成链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了 IPv6 站点本地地址或全球单播地址后，会自动生成链路本地地址
	手工指定接口的链路本地地址	**ipv6 address** *ipv6-address* **link-local**	

说明

- 当接口配置了 IPv6 站点本地地址或全球单播地址后，同时会自动生成链路本地地址。且与采用 **ipv6 address auto link-local** 命令生成的链路本地地址相同。此时如果手工指定接口的链路本地地址，则手工指定的有效。如果删除手工指定的链路本地地址，则接口的链路本地地址恢复为系统自动生成的地址。
- 配置链路本地地址时，手工指定方式的优先级高于自动生成方式。即如果先采用自动生成方式，之后手工指定，则手工指定的地址会覆盖自动生成的地址；如果先手工指定，之后采用自动生成的方式，则自动配置不生效，接口的链路本地地址仍是手工指定的。此时，如果删除手工指定的地址，则自动生成的链路本地地址会生效。
- **undo ipv6 address auto link-local** 命令只能删除使用 **ipv6 address auto link-local** 命令生成的链路本地地址。即如果此时已经配置了 IPv6 站点本地地址或全球单播地址，由于系统会自动生成链路本地地址，则接口仍有链路本地地址；如果此时没有配置 IPv6 站点本地地址或全球单播地址，则接口没有链路本地地址。
- 手工配置的全球单播地址的优先级高于自动生成的全球单播地址。如果在接口已经自动生成全球单播地址的情况下，手工配置前缀相同的全球单播地址，自动生成的地址将被覆盖。此后，即使删除手工配置的全球单播地址，已被覆盖的自动生成的全球单播地址也不会恢复。再次接收到 RA 报文后，设备根据报文携带的地址前缀信息，重新生成全球单播地址。
- 对于接口上的链路本地地址，只有当接口上不存在 IPv6 站点本地地址和全球单播地址，且没有配置 **ipv6 address auto link-local** 命令时，才会删除自动生成的链路本地地址。

1.4 配置 IPv6 邻居发现协议

1.4.1 配置静态邻居表项

将邻居节点的 IPv6 地址解析为链路层地址，可以通过邻居请求消息 NS 及邻居通告消息 NA 来动态实现，也可以通过手工配置静态邻居表项来实现。

设备根据 IPv6 地址及三层接口号来唯一标识一个静态邻居表项。目前，静态邻居表项有如下配置方式：

- 配置三层接口对应 IPv6 地址、链路层地址；

表1-7 配置静态邻居表项

操作	命令	说明
进入系统视图	**system-view**	-
配置静态邻居表项	**ipv6 neighbor** *ipv6-address mac-address* **interface** *interface-type interface-number*	必选

1.4.2 配置接口上允许动态学习的邻居的最大个数

设备可以通过 NS 消息和 NA 消息来动态获取邻居节点的链路层地址，并将其加入到邻居表中。如果动态获取的邻居表过大，将可能导致设备的转发性能下降。为此，可以通过设置接口上允许动态学习的邻居的最大个数来进行限制。当接口上动态学习的邻居个数达到所设置的最大值时，该接口将不再学习邻居信息。

表1-8 配置接口上允许动态学习的邻居的最大个数

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口上允许动态学习的邻居的最大个数	**ipv6 neighbors max-learning-num** *number*	可选 缺省情况下，接口上允许动态学习的邻居的最大个数为 1024

1.4.3 配置 RA 消息的相关参数

用户可以根据实际情况，配置接口是否发送RA消息及发送RA消息的时间间隔，同时可以配置RA消息中的相关参数以通告给主机。当主机接收到RA消息后，就可以采用这些参数进行相应操作。可以配置的RA消息中的参数及含义如 表 1-9所示。

表1-9 RA 消息中的参数及描述

参数	描述
跳数限制（Cur Hop Limit）	主机在发送 IPv6 报文时，将使用该参数值填充 IPv6 报文头中的 Hop Limit 字段。同时该参数值也作为设备应答报文中的 Hop Limit 字段值。
前缀信息（Prefix Information）	在同一链路上的主机收到设备发布的前缀信息后，可以进行无状态自动配置等操作。
被管理地址配置标志位（M flag）	用于确定主机是否采用有状态自动配置获取 IPv6 地址。 如果设置该标志位为 1，主机将通过有状态自动配置（例如 DHCP 服务器）来获取 IPv6 地址；否则，将通过无状态自动配置获取 IPv6 地址，即根据自己的链路层地址及路由器发布的前缀信息生成 IPv6 地址。
其他配置标志位（O flag）	用于确定主机是否采用有状态自动配置获取除 IPv6 地址外的其他信息。 如果设置其他配置标志位为 1，主机将通过有状态自动配置（例如 DHCP 服务器）来获取除 IPv6 地址外的其他信息；否则，将通过无状态自动配置获取其他信息。
路由器生存时间（Router Lifetime）	用于设置发布 RA 消息的路由器作为主机的默认路由器的时间。主机根据接收到的 RA 消息中的路由器生存时间参数值，就可以确定是否将发布该 RA 消息的路由器作为默认路由器。
邻居请求消息重传时间间隔（Retrans Timer）	设备发送 NS 消息后，如果未在指定的时间间隔内收到响应，则会重新发送 NS 消息。
保持邻居可达状态的时间（Reachable Time）	当通过邻居可达性检测确认邻居可达后，在所设置的可达时间内，设备认为邻居可达；超过设置的时间后，如果需要向邻居发送报文，会重新确认邻居是否可达。



说明

在接口上配置的邻居请求消息重传时间间隔及保持邻居可达状态的时间，既可作为 RA 消息中的信息发布给主机，也可作为本接口发送邻居请求消息的时间间隔及保持邻居可达状态的时间。

表1-10 配置 RA 消息的相关参数

操作	命令	说明
进入系统视图	**system-view**	-
配置跳数限制	**ipv6 nd hop-limit** *value*	可选 缺省情况下，路由器发布的跳数限制为 64 跳
进入接口视图	**interface** *interface-type interface-number*	-
取消对 RA 消息发布的抑制	**undo ipv6 nd ra halt**	必选 缺省情况下，抑制发布 RA 消息
配置 RA 消息发布的最大时间间隔和最小时间间隔	**ipv6 nd ra interval** *max-interval-value min-interval-value*	可选 缺省情况下，RA 消息发布的最大间隔时间为 600 秒，最小时间间隔为 200 秒 RA 消息周期性发布时，相邻两次的时间间隔是在最大时间间隔与最小时间间隔之间随机选取一个值作为周期性发布 RA 消息的时间间隔 配置的最小时间间隔应该小于等于最大时间间隔的 0.75 倍
配置 RA 消息中的前缀信息	**ipv6 nd ra prefix** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* } *valid-lifetime preferred-lifetime* [**no-autoconfig** | **off-link**] *	可选 缺省情况下，没有配置 RA 消息中的前缀信息，此时将使用发送 RA 消息的接口 IPv6 地址作为 RA 消息中的前缀信息
配置 RA 消息中不携带 MTU 选项	**ipv6 nd ra no-advlinkmtu**	可选 缺省情况下，RA 消息中携带 MTU 选项
设置被管理地址配置标志位为 1	**ipv6 nd autoconfig managed-address-flag**	可选 缺省情况下，被管理地址标志位为 0，即主机通过无状态自动配置获取 IPv6 地址
设置其他配置标志位为 1	**ipv6 nd autoconfig other-flag**	可选 缺省情况下，其他配置标志位为 0，即主机通过无状态自动配置获取其他信息
配置 RA 消息中路由器的生存时间	**ipv6 nd ra router-lifetime** *value*	可选 缺省情况下，RA 消息中路由器的生存时间为 1800 秒
配置邻居请求消息重传时间间隔	**ipv6 nd ns retrans-timer** *value*	可选 缺省情况下，接口发送 NS 消息的时间间隔为 1000 毫秒，接口发布的 RA 消息中 Retrans Timer 字段的值为 0
配置保持邻居可达状态的时间	**ipv6 nd nud reachable-time** *value*	可选 缺省情况下，接口保持邻居可达状态的时间为 30000 毫秒，接口发布的 RA 消息中 Reachable Timer 字段的值为 0



注意

RA 消息发布的最大间隔时间应该小于或等于 RA 消息中路由器的生存时间。

1.4.4 配置重复地址检测时发送邻居请求消息的次数

接口获得 IPv6 地址后，将发送邻居请求消息进行重复地址检测，如果在指定的时间内（通过 **ipv6 nd ns retrans-timer** 命令配置）没有收到响应，则继续发送邻居请求消息，当发送的次数达到所设置的次数后，仍未收到响应，则认为该地址可用。

表1-11 配置重复地址检测时发送邻居请求消息的次数

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置重复地址检测时发送邻居请求消息的次数	**ipv6 nd dad attempts** *value*	可选 缺省情况下，重复地址检测时发送邻居请求报文的次数为 1，当 *value* 值为 0 时，表示禁止重复地址检测

1.5 配置 PMTU 发现

1.5.1 配置接口 MTU

由于 IPv6 路由器不支持对报文进行分片，当路由器接口收到一个报文后，如果发现报文长度比转发接口的 MTU 值大，则会将其丢弃；同时将转发接口的 MTU 值通过 ICMPv6 报文的“Packet Too Big”消息发给源端主机，源端主机以该值重新发送 IPv6 报文。为减少报文被丢弃带来的额外流量开销，需要根据实际组网环境设置合适的接口 MTU 值。

表1-12 配置接口 MTU

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口 MTU	**ipv6 mtu** *mtu-size*	可选

1.5.2 配置指定地址的静态 PMTU

用户可以为指定的目的 IPv6 地址配置静态的 PMTU 值。当源端主机从接口发送报文时，将比较该接口的 MTU 与指定目的 IPv6 地址的静态 PMTU，如果报文长度大于二者中的最小值，则采用此最小值对报文进行分片。

表1-13 配置指定地址的静态 PMTU

操作	命令	说明
进入系统视图	**system-view**	-
配置指定 IPv6 地址对应的静态 PMTU 值	**ipv6 pathmtu** *ipv6-address* [*value*]	必选 缺省情况下，没有配置静态 PMTU 值

1.5.3 配置 PMTU 老化时间

通过“1.1.4 IPv6 PMTU发现”中的方法动态确定源端主机到目的端主机的PMTU后，源端主机将使用这个MTU值发送后续报文到目的端主机。当PMTU老化时间超时后，动态确定的PMTU值将会被删除，源端主机会通过PMTU机制重新确定发送报文的MTU值。

该配置对静态 PMTU 不起作用。

表1-14 配置 PMTU 老化时间

操作	命令	说明
进入系统视图	**system-view**	-
配置 PMTU 老化时间	**ipv6 pathmtu age** *age-time*	可选 缺省情况下，PMTU 的老化时间是 10 分钟

1.6 配置 TCP6

可以配置的 TCP6 属性包括：

- synwait 定时器：当发送 SYN 报文时，TCP6 启动 synwait 定时器，如果 synwait 定时器超时前未收到回应报文，则 TCP6 连接建立不成功。
- finwait 定时器：当 TCP6 的连接状态为 FIN_WAIT_2 时，启动 finwait 定时器，如果在定时器超时前没有收到报文，则 TCP6 连接终止；如果收到 FIN 报文，则 TCP6 连接状态变为 TIME_WAIT 状态；如果收到非 FIN 报文，则从收到的最后一个非 FIN 报文开始重新计时，在超时后中止连接。
- TCP6 的接收和发送缓冲区的大小。

表1-15 配置 TCP6

操作	命令	说明
进入系统视图	**system-view**	-
配置 TCP6 的 finwait 定时器	**tcp ipv6 timer fin-timeout** *wait-time*	可选 缺省情况下，finwait 定时器的值为 675 秒
配置 TCP6 的 synwait 定时器	**tcp ipv6 timer syn-timeout** *wait-time*	可选 缺省情况下，synwait 定时器的值 75 秒
配置 TCP6 的接收和发送缓冲区大小	**tcp ipv6 window** *size*	可选 缺省情况下，TCP6 的接收和发送缓冲区大小为 8KB

1.7 配置 IPv6 FIB 转发功能

在使能 IPv6 的 FIB 缓存功能后，当报文进行转发时将会查找 FIB 缓存，这样缩短了 IP 报文查找时间，从而提高了转发效率。

表1-16 配置 IPv6 FIB 转发功能

操作		命令	说明
进入系统视图		**system-view**	-
使能 IPv6 FIB 缓存功能	集中式设备	**ipv6 fibcache**	必选 缺省情况下，IPv6 FIB 缓存功能处于关闭状态
	分布式设备	**ipv6 fibcache** { *slot-number* | **all** }	



说明

多核数据平面不支持 fibcache，该特性仅为控制平面特有，当本机报文走控制平面 IPv6 转发时支持 fibcache

1.8 配置 ICMPv6 报文发送

1.8.1 配置指定时间内发送 ICMPv6 差错报文的最大个数

如果网络中短时间内发送的 ICMPv6 差错报文过多，将可能导致网络拥塞。为了避免这种情况，用户可以控制在指定时间内发送 ICMPv6 差错报文的最大个数，目前采用令牌桶算法来实现。

用户可以设置令牌桶的容量，即令牌桶中可以同时容纳的令牌数；同时可以设置令牌桶的刷新周期，即每隔多长时间将令牌桶内的令牌个数刷新为所配置的容量。一个令牌表示允许发送一个 ICMPv6 差错报文，每当发送一个 ICMPv6 差错报文，则令牌桶中减少一个令牌。如果连续发送的 ICMPv6 差错报文超过了令牌桶的容量，则后续的 ICMPv6 差错报文将不能被发送出去，直到按照所设置的刷新频率将新的令牌放入令牌桶中。

表1-17 配置指定时间内发送 ICMPv6 差错报文的最大个数

操作	命令	说明
进入系统视图	**system-view**	-
配置控制 ICMPv6 差错报文发送的令牌桶容量和刷新周期	**ipv6 icmp-error** { **bucket** *bucket-size* | **ratelimit** *interval* } *	可选 缺省情况下，令牌桶容量为 10，令牌桶的刷新周期为 100 毫秒，即每一个刷新周期内最多可以发送 10 个 ICMPv6 差错报文 刷新周期为 0 时，表示不限制 ICMPv6 差错报文的发送

1.8.2 配置允许回复组播形式的 Echo request 报文

如果允许主机回复组播形式的 Echo request 报文，则主机 A 可以构造目的地址为组播地址、源地址为主机 B 的 Echo request 报文，使该组播组中所有的主机都向主机 B 发送 Echo reply 报文，从而达到攻击主机 B 的目的。因此，为了避免主机利用设备达到攻击的目的，缺省情况下，不允许设备回复组播形式的 Echo request 报文。可以通过下面的命令，配置允许设备回复组播形式的 Echo request 报文。

表1-18 配置允许回复组播形式的 Echo request 报文

操作	命令	说明
进入系统视图	**system-view**	-

操作	命令	说明
配置允许回复组播形式的 Echo request 报文	**ipv6 icmpv6 multicast-echo-reply enable**	可选 缺省情况下，不允许回复组播形式的 Echo request 报文

1.9 配置 IPv6 DNS Client

1.9.1 配置静态 IPv6 DNS

配置静态 IPv6 DNS 就是配置将主机名与 IPv6 地址相对应。当使用 telnet 等应用时，可以直接使用主机名，由系统解析为 IPv6 地址。每个主机名只能对应 1 个 IPv6 地址。

表1-19 配置静态 IPv6 DNS

操作	命令	说明
进入系统视图	**system-view**	-
配置主机名和对应的 IPv6 地址	**ipv6 host** *hostname ipv6-address*	必选

1.9.2 配置动态 IPv6 DNS

如果用户需要使用动态域名解析功能，可以使用下面的命令使能动态域名解析功能，并配置域名服务器，这样才能将查询请求报文发送到正确的服务器进行解析。系统最多支持 6 个 DNS 服务器。

用户还可以配置域名后缀，以便实现只输入域名的部分字段，而由系统自动加上预先设置的后缀进行解析。系统最多支持 10 个域名后缀。

表1-20 配置动态 IPv6 DNS

操作	命令	说明
进入系统视图	**system-view**	-
使能动态域名解析功能	**dns resolve**	必选 缺省情况下，动态域名解析功能处于关闭状态
配置 IPv6 的 DNS 服务器	**dns server ipv6** *ipv6-address* [*interface-type interface-number*]	必选 当 DNS 服务器的 IPv6 地址为链路本地地址时，需要指定参数 *interface-type* 和 *interface-number*
配置域名后缀	**dns domain** *domain-name*	必选 缺省情况下，没有配置域名后缀，即只根据用户输入的域名信息进行解析



说明

dns resolve 和 **dns domain** 命令与 IPv4 DNS 的命令相同，详细信息请参见“IP 业务分册”中的“域名解析命令”。

1.10 IPv6 基础显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IPv6 配置后的运行情况，用户可以通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除相应的统计信息。

表1-21 IPv6 基础显示和维护

操作		命令
显示 DNS 域名后缀信息		**display dns domain** [**dynamic**]
显示 IPv6 动态域名缓存信息		**display dns ipv6 dynamic-host**
显示 IPv6 域名服务器的相关信息		**display dns ipv6 server** [**dynamic**]
显示 IPv6 FIB 转发信息表项	集中式设备	**display ipv6 fib** [*ipv6-address*]
	分布式设备	**display ipv6 fib** [*slot-number*] [*ipv6-address*]
显示 IPv6 FIB 缓存中的路由总数	集中式设备	**display ipv6 fibcache**
	分布式设备	**display ipv6 fibcache** *slot-number*
显示静态域名解析表中所有主机名与 IPv6 地址的对应关系		**display ipv6 host**
显示可以配置 IPv6 地址的接口的 IPv6 信息		**display ipv6 interface** [*interface-type* [*interface-number*]] [**verbose**]
显示邻居信息	集中式设备	**display ipv6 neighbors** { *ipv6-address* | **all** | **dynamic** | **interface** *interface-type interface-number* | **static** } [**|** { **begin** | **exclude** | **include** } *regular-expression*]
	分布式设备	**display ipv6 neighbors** { { *ipv6-address* | **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* } [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示符合指定条件的邻居表项的总个数	集中式设备	**display ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **static** } **count**
	分布式设备	**display ipv6 neighbors** { { **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* } **count**
显示 IPv6 的 PMTU 信息		**display ipv6 pathmtu** { *ipv6-address* | **all** | **dynamic** | **static** }
显示指定套接字的相关信息	集中式设备	**display ipv6 socket** [**socktype** *socket-type*] [*task-id socket-id*]
	分布式设备	**display ipv6 socket** [**socktype** *socket-type*] [*task-id socket-id*] [**slot** *slot-number*]
显示 TCP6 连接的统计信息		**display tcp ipv6 statistics**
显示 TCP6 连接的状态信息		**display tcp ipv6 status**
显示 UDP6 的统计信息		**display udp ipv6 statistics**
清除 IPv6 动态域名缓存信息		**reset dns ipv6 dynamic-host**
清除 FIB 缓存项	集中式设备	**reset ipv6 fibcache**
	分布式设备	**reset ipv6 fibcache** { *slot-number* | **all** }
清除 IPv6 邻居信息	集中式设备	**reset ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **static** }
	分布式设备	**reset ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **slot** *slot-number* | **static** }
清除 PMTU 值		**reset ipv6 pathmtu** { **all** | **static** | **dynamic**}

操作	命令
清除所有 TCP6 连接的统计信息	**reset tcp ipv6 statistics**
清除所有 UDP6 统计信息	**reset udp ipv6 statistics**

说明

display dns domain 命令与 IPv4 DNS 的命令相同，详细信息请参见“IP 业务分册”中的“域名解析命令”。

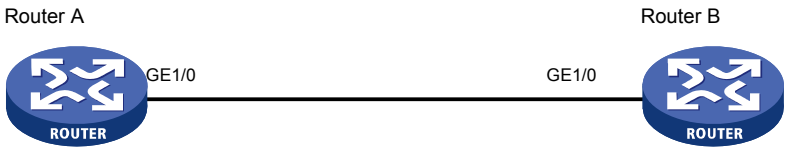
1.11 IPv6 基础典型配置举例

1. 组网需求

Router A 和 Router B 通过以太网接口直接相连，在接口上配置 IPv6 地址，验证它们之间的互通性。Router A 的接口 GE1/0 的全球单播地址为 3001::1/64，Router B 的接口 GE1/0 的全球单播地址为 3001::2/64。

2. 组网图

图1-6 IPv6 地址配置组网图



3. 配置步骤

(1) 配置 Router A

使能路由器的 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

手工指定接口 GigabitEthernet1/0 的全球单播地址。

```
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ipv6 address 3001::1/64
```

允许接口 GigabitEthernet1/0 发布 RA 消息。（缺省情况下，所有的接口不会发布 RA 消息）

```
[RouterA-GigabitEthernet1/0] undo ipv6 nd ra halt
```

(2) 配置 Router B

使能路由器的 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

手工指定接口 GigabitEthernet1/0 的全球单播地址。

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] ipv6 address 3001::2/64
```

4. 验证配置结果

显示 Router A 的接口信息。

```
[RouterA-GigabitEthernet1/0] display ipv6 interface GigabitEthernet 1/0 verbose
GigabitEthernet1/0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::2E0:FCFF:FE98:7654
  Global unicast address(es):
    2001::1, subnet is 2001::/64
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FF98:7654
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                    9
  InTooShorts:                   0
  InTruncatedPkts:               0
  InHopLimitExceeds:             0
  InBadHeaders:                  0
  InBadOptions:                  0
  ReasmReqds:                    0
  ReasmOKs:                      0
  InFragDrops:                   0
  InFragTimeouts:                0
  OutFragFails:                  0
  InUnknownProtos:               0
  InDelivers:                    7
  OutRequests:                   27
  OutForwDatagrams:              0
  InNoRoutes:                    0
  InTooBigErrors:                0
  OutFragOKs:                    0
  OutFragCreates:                0
  InMcastPkts:                   2
  InMcastNotMembers:             2
  OutMcastPkts:                  23
```

```
  InAddrErrors:                  0
  InDiscards:                    0
  OutDiscards:                   0
```

显示 Router B 的接口信息。

```
[RouterB-GigabitEthernet1/0] display ipv6 interface GigabitEthernet 1/0 verbose
GigabitEthernet1/0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:2
  Global unicast address(es):
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:2
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                        0
  InTooShorts:                       0
  InTruncatedPkts:             0
  InHopLimitExceeds:           0
  InBadHeaders:                      0
  InBadOptions:                      0
  ReasmReqds:                        0
  ReasmOKs:                          0
  InFragDrops:                       0
  InFragTimeouts:                    0
  OutFragFails:                      0
  InUnknownProtos:                   0
  InDelivers:                        0
  OutRequests:                       0
  OutForwDatagrams:                  0
  InNoRoutes:                        0
  InTooBigErrors:                    0
  OutFragOKs:                        0
  OutFragCreates:                    0
  InMcastPkts:                       0
  InMcastNotMembers:                 0
  OutMcastPkts:                      0
  InAddrErrors:                      0
  InDiscards:                        0
  OutDiscards:                       0
```

在路由器 Router A 上使用 Ping 测试和路由器 Router B 的互通性；在 Router B 上使用 Ping 测试和 Router A 的互通性。

 注意

在 Ping 链路本地地址时，需要使用**-i** 参数，来指定链路本地地址的接口。

```
[RouterA-GigabitEthernet1/0] ping ipv6 –c 1 3001::2
PING 3001::2 : 56  data bytes, press CTRL_C to break
    Reply from 3001::2
    bytes=56 Sequence=1 hop limit=64  time = 2 ms

  --- 3001::2 ping statistics ---
    1 packet(s) transmitted
    1 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 2/2/2 ms
[RouterB-GigabitEthernet1/0] ping ipv6 3001::1
  PING 3001::1 : 56  data bytes, press CTRL_C to break
    Reply from 3001::1
    bytes=56 Sequence=1 hop limit=255  time = 50 ms
    Reply from 3001::1
    bytes=56 Sequence=2 hop limit=255  time = 60 ms
    Reply from 3001::1
    bytes=56 Sequence=3 hop limit=255  time = 60 ms
    Reply from 3001::1
    bytes=56 Sequence=4 hop limit=255  time = 70 ms
    Reply from 3001::1
    bytes=56 Sequence=5 hop limit=255  time = 60 ms

  --- 3001::1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 50/60/70 ms
```

1.12 常见配置错误举例

1. 故障现象

无法 Ping 通对端的 IPv6 地址。

2. 故障排除

- 在任意视图使用 **display current-configuration** 命令或在系统视图下使用 **display this** 命令检查是否使能了 IPv6 报文转发功能。
- 在任意视图下使用 **display ipv6 interface** 命令检查接口配置的 IPv6 地址是否正确，接口状态是否为 up。

- 在用户视图下使用 **debugging ipv6 packet** 命令打开 IPv6 报文调试开关，根据调试信息进行判断。

目 录

1 邻接表配置 1-1
1.1 邻接表简介 1-1
1.1.1 邻接表介绍 1-1
1.1.2 邻接表中的术语解释 1-1
1.2 邻接表显示和维护 1-1

1 邻接表配置

说明

- 目前邻接表只用于配合硬件转发，尚不支持软件转发；
- 对于以太网，鉴于其属于广播网络，且在设备应用中的广泛性和特殊性，其邻居信息由专门的 ARP 表项进行存储与管理，不包含在邻接表的范围内。

1.1 邻接表简介

1.1.1 邻接表介绍

邻接表用于管理相连且处于激活态的邻居的信息。邻接表保存邻居的网络层地址（下一跳）、路由出接口、链路层业务类型、链路层地址（对 ATM 可以是 PVC，PPP 则没有这个信息）等。

邻居的概念实际是与网络层相关的，B 节点对于 A 节点可能是 IP 的邻居，但并不是 IPX 的邻居；一个邻居处于激活态实际上也是相对网络层而言的，目前邻接表只支持相对 IP 而言的邻居概念。

邻居与设备的连接可以通过各种链路协议，PPP、ATM、FR 等链路各有自己的邻居。

邻接表可以用于软件转发，配合转发表进行快速软件转发；对于支持硬件转发的产品，需要邻接表下发到驱动，用于硬件转发。

1.1.2 邻接表中的术语解释

- 路由接口：路由转发项中可见的出接口。
- 物理接口：实际存在的物理接口，如：串口、POS 接口等。
- 逻辑接口：没有对应物理实体，一般是在邻接表中为实现上方便而抽象出来的接口。如：Virtual-Template 接口等。
- 服务类型：邻接表对应的业务类型，目前只支持 PPP、FR 和 CHDLC。
- 报文处理类型：表示匹配上此表项的报文应该做的动作。如：FORWARDING 表示转发，DROPPKT 表示丢弃。
- 链路媒介类型：与路由出接口当前封装的链路层协议有关。如：P2P 表示点到点，代表有 PPP。
- 链路层头信息：IPv6 协议对应的链路层头信息。

1.2 邻接表显示和维护

在任意视图下执行 **display** 命令可以显示邻接表项的信息。

表1-1 邻接表显示和维护

操作	命令
显示 IPv4 邻接表项	**display adjacent-table** { **all** | **physical-interface** *interface-type interface-number* | **routing-interface** *interface-type interface-number* | **slot** *slotnum* } [**count** | **verbose**]
显示 IPv6 邻接表项	**display ipv6 adjacent-table** { **all** | **physical-interface** *interface-type interface-number* | **routing-interface** *interface-type interface-number* | **slot** *slotnum* } [**count** | **verbose**]

目 录

1 流分类配置 1-1
1.1 流分类简介 1-1
1.2 配置流分类策略 1-1
1.3 流分类显示和维护 1-1

1 流分类配置

1.1 流分类简介

流分类采用一定的规则识别符合某类特征的报文，是有区别地进行服务的前提和基础。

多核处理中，控制平面和数据平面分别运行在不同的内核和线程上，数据平面对于接收到的报文基于流进行处理，一条流用来标识具有相同特征并且处理流程相同的报文，通常可以用五元组来区分和划定一条流。报文接收后首先经过流分类模块的处理，给报文带上一个流标签送到数据处理平面。当数据平面收到一条流的首包时，通过相关业务模块的处理建立和流相关联的处理表项，后续报文就可以直接根据关联表项进行转发，提高性能。

1.2 配置流分类策略

表1-1 配置流分类策略

操作	命令	说明
进入系统视图	**system-view**	-
配置流分类策略	**forwarding policy { per-flow | per-packet }**	可选 缺省情况下，采用基于流调度的流分类策略

注意

当流分类策略配置为基于报文调度时，同一条流的报文会被发送到不同的数据平面进行处理，这样无法保证报文的处理顺序。当启动某些业务时，该业务可能无法处理同一条流的报文在不同的数据平面而将报文丢弃。

1.3 流分类显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后流分类的运行情况，通过查看显示信息验证配置的效果。

表1-2 流分类显示和维护

操作	命令
显示当前的流分类策略	**display forwarding policy**

目 录

1 IP虚拟分片重组配置 1-1
1.1 IP虚拟分片重组简介 1-1
1.2 配置IP虚拟分片重组功能 1-1
1.3 IP虚拟分片重组显示和维护 1-1
1.4 IP虚拟分片重组典型配置举例 1-2

1 IP 虚拟分片重组配置

1.1 IP 虚拟分片重组简介

为了避免每个业务模块（如：IPSec、NAT 和防火墙）单独处理后片先到（报文分片后）这种情况而导致复杂度过高，设备需要收到 IP 报文后就对分片报文进行虚拟分片重组。IP 虚拟分片重组功能可以对分片报文进行检验、排序和缓存，保证后续业务模块处理的都是顺序正确的分片报文。

同时，IP 虚拟分片重组功能还可以对下面几种分片攻击进行检测。如果检测到分片攻击，则设备会丢弃收到的分片报文，从而提高了设备的安全性。

- Tiny Fragment 攻击：如果设备收到分片报文的首片长度非常小，并且传输层协议（如：TCP、UDP）头字段放在第二个分片中，则认为是受到了 Tiny Fragment 攻击；
- Overlapping Fragment 攻击：如果设备收到了完全相同的分片报文，或者收到的分片报文与其前一分片或后一分片出现重叠时，则认为是受到了 Overlapping Fragment 攻击；
- Fragment-flood 攻击：如果设备收到的分片报文超过了指定的最大分片报文数或者设备上创建的分片队列个数超过了指定的最大分片队列个数，则认为是受到了 Fragment-flood 攻击。

1.2 配置 IP 虚拟分片重组功能

表1-1 配置 IP 虚拟分片重组功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 IP 虚拟分片重组功能	**ip virtual-reassembly** [**drop-fragments** | **max-fragments** *number* | **max-reassemblies** *number* | **timeout** *seconds*] *	必选 缺省情况下，关闭 IP 虚拟分片重组功能

注意

- 只能在报文的入接口配置 IP 虚拟分片重组特性。
- 不支持负载分担，即不支持同一个 IP 报文的不同分片从不同的接口到达。

1.3 IP 虚拟分片重组显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IP 虚拟分片重组的配置的情况及统计信息，通过查看显示信息验证配置的效果。

表1-2 IP 虚拟分片重组显示和维护

操作	命令
显示接口上的分片信息	**display ip virtual-reassembly** [**interface** *interface-type interface-number*]

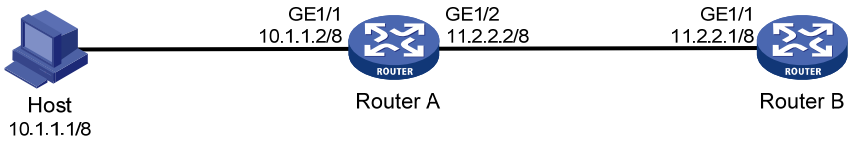
1.4 IP 虚拟分片重组典型配置举例

1. 组网需求

- Host 通过 Router A 与 Router B 进行通信。
- Router A 上接口 GigabitEthernet1/2 启用了 NAT 服务，为了保证安全和高效地运行 NAT，可以在 Router A 的接口 GigabitEthernet1/1 上使用 IP 虚拟分片重组功能对分片报文进行处理。

2. 组网图

图1-1 IP 虚拟分片重组配置举例组网图



3. 配置步骤

(1) 配置 Host

配置一条静态路由使 Host 可以与 Router B 通信。（略）

(2) 配置 Route A

配置 NAT 业务，并启用 IP 虚拟分片重组功能。

```
<RouterA> system-view
[RouterA] nat static 10.1.1.1 11.2.2.3
[RouterA] interface gigabitethernet 1/2
[RouterA-GigabitEthernet1/2] nat outbound static
[RouterA-GigabitEthernet1/2] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip virtual-reassembly
```

通过上面的配置，如果从接口 GigabitEthernet1/1 上收到乱序的分片报文，Router A 将会对其进行检验、重新排序，这时可以通过 **display ip virtual-reassembly** 命令查看相关的信息。

目 录

1 双协议栈配置 1-1
1.1 双协议栈简介 1-1
1.2 配置双协议栈 1-1

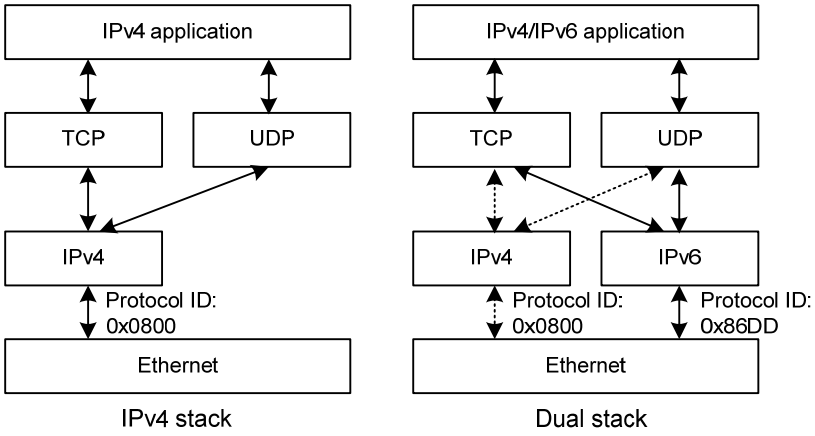
1 双协议栈配置

1.1 双协议栈简介

对于 IPv6 节点来说，兼容 IPv4 的最直接有效的办法就是保留一个完整的 IPv4 协议栈。同时支持 IPv4 协议和 IPv6 协议的网络节点即成为双协议栈节点。当双栈节点配置 IPv4 地址和 IPv6 地址后，就可以在相应接口上转发 IPv4 和 IPv6 报文。

当一个上层的应用支持IPv4 和IPv6 协议时，根据协议要求可以选用TCP或UDP作为传输层的协议，但在选择网络层协议时，它会优先选择IPv6 协议栈。图 1-1所示为IPv4 单协议栈和IPv4/IPv6 双协议栈的结构图。

图1-1 图 1-1 IPv4 单协议栈与 IPv4/IPv6 双协议栈结构（以太网）



1.2 配置双协议栈

为了启动双协议栈，必须使能 IPv6 报文转发能力。否则即使配置了接口的 IPv6 地址，仍无法转发 IPv6 的报文

表1-1 配置双协议栈

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 IPv4 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	必选 缺省情况下，没有为接口配置 IP 地址

操作			命令	说明
配置接口的IPv6地址	配置IPv6全球单播地址或站点本地地址	手工指定IPv6地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	二者必选其一 缺省情况下，接口没有配置站点本地地址和全球单播地址
		采用EUI-64格式形成IPv6地址	**ipv6 address** *ipv6-addressIprefix-length* **eui-64**	
	配置IPv6链路本地地址	配置自动生成链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了IPv6站点本地地址或全球单播地址后，同时会自动生成链路本地地址
		手工指定接口的链路本地地址	**ipv6 address** *ipv6-address* **link-local**	

说明

- IPv4地址的相关介绍请参见“IP业务分册”中的“IP地址配置”。
- IPv6地址的相关介绍请参见“IP业务分册”中的“IPv6基础配置”。
- 使能IPv6报文转发功能、配置接口IPv6地址的命令具体介绍请参见“IP业务分册”中的“IPv6基础命令”

目 录

1 隧道技术配置 1-1
1.1 隧道技术简介 1-1
1.1.1 IPv6 over IPv4 隧道 1-1
1.1.2 IPv4 over IPv4 隧道 1-4
1.1.3 IPv4 或IPv6 over IPv6 隧道 1-5
1.1.4 6PE概述 1-6
1.1.5 协议规范 1-6
1.2 隧道技术配置任务简介 1-7
1.3 配置IPv6 手动隧道 1-7
1.3.1 配置准备 1-7
1.3.2 配置IPv6 手动隧道 1-7
1.3.3 配置举例 1-9
1.4 配置IPv4 兼容IPv6 自动隧道 1-11
1.4.1 配置准备 1-11
1.4.2 配置IPv4 兼容IPv6 自动隧道 1-12
1.4.3 配置举例 1-13
1.5 配置 6to4 隧道 1-15
1.5.1 配置准备 1-15
1.5.2 配置 6to4 隧道 1-15
1.5.3 配置 6to4 隧道举例 1-17
1.5.4 配置 6to4 中继举例 1-19
1.6 配置ISATAP隧道 1-21
1.6.1 配置准备 1-21
1.6.2 配置ISATAP隧道 1-21
1.6.3 配置举例 1-22
1.7 配置IPv4 over IPv4 隧道 1-24
1.7.1 配置准备 1-24
1.7.2 配置IPv4 over IPv4 隧道 1-24
1.7.3 配置举例 1-25
1.8 配置IPv4 over IPv6 隧道 1-28
1.8.1 配置准备 1-28
1.8.2 配置IPv4 over IPv6 隧道 1-28
1.8.3 配置举例 1-29
1.9 配置IPv6 over IPv6 隧道 1-32
1.9.1 配置准备 1-32
1.9.2 配置IPv6 over IPv6 隧道 1-32
1.9.3 配置举例 1-34
1.10 隧道技术显示和维护 1-37
1.11 常见错误配置举例 1-37

1 隧道技术配置

1.1 隧道技术简介

随着 Internet 的日益膨胀，现有的 IPv4 地址已经十分紧缺，虽然使用分配临时 IPv4 地址或 NAT（Network Address Translator，网络地址转换）等技术，在一定程度上缓解了 IPv4 地址不足的状况，但也增加了地址解析和处理方面的开销，同时导致某些高层应用失效，而且仍然无法回避 IPv4 地址即将被分配殆尽这个问题。采用 128 位地址长度的 IPv6 协议，彻底解决了 IPv4 地址不足的难题，并且在地址容量、安全性、网络管理、移动性以及服务质量等方面有明显的改进，是下一代互联网络协议采用的核心标准之一。IPv6 与 IPv4 不兼容，但它同所有的 TCP / IP 协议族中的其他协议兼容，即 IPv6 完全可以取代 IPv4。

在 IPv6 成为主流协议之前，首先使用 IPv6 协议栈的网络希望能与当前仍被 IPv4 支撑着的 Internet 进行正常通信，因此必须开发出 IPv4 和 IPv6 互通技术以保证 IPv4 能够平稳过渡到 IPv6。此外，互通技术应该对信息传递做到高效无缝。国际上 IETF 组建了专门的 NGTRANS 工作组，开展对 IPv4 和 IPv6 过渡问题和高效无缝互通问题的研究。目前已经出现了多种过渡技术和互通方案，这些技术各有特点，用于解决不同过渡时期、不同环境的通信问题。

目前解决过渡问题的基本技术主要有 3 种：双协议栈（RFC2893），隧道技术（RFC2893）和 NAT-PT(RFC2766)。

隧道是一种封装技术，它利用一种网络协议来传输另一种网络协议，即利用一种网络传输协议，将其他协议产生的数据报文封装在它自己的报文中，然后在网络中传输。隧道（Tunnel）是一个虚拟的点对点的连接。在实际应用中仅支持点对点连接的虚拟接口为 Tunnel 接口。一个 Tunnel 提供了一条使封装的数据报文能够传输的通路，并且在一个 Tunnel 的两端可以分别对数据报文进行封装及解封装。隧道技术就是指包括数据封装、传输和解封装在内的全过程。

说明

- 双协议栈的相关配置请参见“IP 业务分册”中的“双协议栈配置”。
- NAT-PT 的相关配置请参见“IP 业务分册”中的“NAT-PT 配置”。

1.1.1 IPv6 over IPv4 隧道

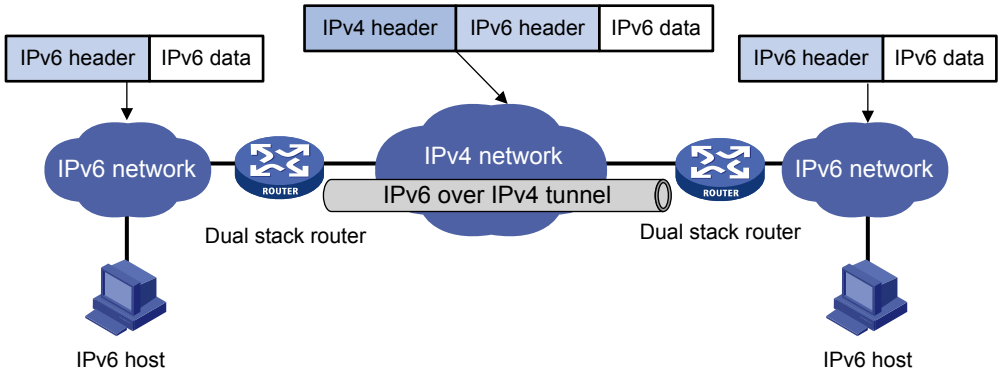
1. IPv6 over IPv4 隧道原理

IPv6 over IPv4 隧道机制是在IPv6 数据报文前封装上IPv4 的报文头，通过隧道（Tunnel）使IPv6 报文穿越IPv4 网络，实现隔离的IPv6 网络的互通，如 图 1-1所示。

注意

IPv6 over IPv4 隧道两端的设备必须支持 IPv4/IPv6 双协议栈。

图1-1 IPv6 over IPv4 隧道原理图



IPv6 over IPv4 隧道对报文的处理过程如下：

- IPv6 网络中的设备发送 IPv6 报文，到达隧道的源端设备。
- 隧道的源端设备根据路由表判定该报文要通过隧道进行转发，将会在 IPv6 报文前封装上 IPv4 的报文头，通过隧道的实际物理接口将报文转发出去。
- 封装报文通过隧道到达隧道目的端设备，目的端设备判断该封装报文的目的地是本设备后，将对报文进行解封装。
- 目的端设备根据解封装后的 IPv6 报文的目的地址将报文进行转发；如果目的地就是本设备，则将 IPv6 报文转给上层协议处理。

2. 配置隧道和自动隧道

IPv6 over IPv4 隧道可以建立在主机-主机、主机-设备、设备-主机、设备-设备之间。隧道的终点可能是 IPv6 报文的最终目的地，也可能需要进一步转发。

根据隧道终点的 IPv4 地址的获取方式不同，隧道分为“配置隧道”及“自动隧道”。

- 如果 IPv6 over IPv4 隧道的终点地址不能从 IPv6 报文的目的地址中自动获取，需要进行手工配置，这样的隧道即为“配置隧道”。
- 如果 IPv6 over IPv4 隧道的接口地址采用内嵌 IPv4 地址的特殊 IPv6 地址形式，即可以从 IPv6 报文的目的地址中自动获取隧道终点的 IPv4 地址，这样的隧道即为“自动隧道”。

3. IPv6 over IPv4 隧道模式

根据对 IPv6 报文的封装方式的不同，IPv6 over IPv4 隧道分为以下几种模式：

表1-1 IPv6 over IPv4 隧道模式

隧道类型	隧道模式
配置隧道	IPv6 手动隧道
	IPv6-over-IPv4 GRE （Generic Routing Encapsulation，通用路由封装）隧道
自动隧道	IPv4 兼容 IPv6 自动隧道
	6to4 隧道
	ISATAP（Intra-Site Automatic Tunnel Addressing Protocol，站点内自动隧道寻址协议）隧道

下面将对各隧道模式分别做详细介绍。首先，在 表 1-2中对各隧道模式的关键配置参数进行了简要列举。

表1-2 IPv6 over IPv4 隧道模式参数对比表

隧道模式	隧道源/目的地址	隧道接口地址
IPv6 手动隧道	源/目的地址为手动配置的 IPv4 地址	IPv6 地址
IPv6-over-IPv4GRE 隧道	源/目的地址为手动配置的 IPv4 地址	IPv6 地址
IPv4 兼容 IPv6 自动隧道	源地址为手动配置的 IPv4 地址，目的地址不需配置	IPv4 兼容 IPv6 地址，其格式为 ::IPv4-source-address/96
6to4 隧道	源地址为手动配置的 IPv4 地址，目的地址不需配置	6to4 地址，其格式为 2002:IPv4-source-address::/48
ISATAP 隧道	源地址为手动配置的 IPv4 地址，目的地址不需配置	ISATAP 地址，其格式为 Prefix:0:5EFE:IPv4-source-address/64

(1) IPv6 手动隧道

手动隧道是点到点之间的链路，一条链路就是一个单独的隧道。主要用于边缘路由器-边缘路由器或主机-边缘路由器之间定期安全通信的稳定连接，可实现与远端 IPv6 网络的连接。

(2) GRE 隧道

使用标准的 GRE 协议可对 IPv6 报文进行封装，使 IPv6 报文能通过隧道穿越 IPv4 网络。与 IPv6 手动隧道相同，GRE 隧道也是点到点之间的链路，每条链路都是一条单独的隧道。GRE 隧道主要用于边缘路由器-边缘路由器、主机-边缘路由器定期安全通信的稳定连接。相关配置请参见“VPN 分册”中的“GRE 配置”。

(3) IPv4 兼容 IPv6 自动隧道

IPv4 兼容 IPv6 自动隧道是点到多点的链路。隧道两端采用特殊的 IPv6 地址：IPv4 兼容 IPv6 地址，其格式为：0:0:0:0:0:0:a.b.c.d/96，其中 a.b.c.d 是 IPv4 地址。通过这个嵌入的 IPv4 地址可以自动确定隧道的终点，使 IPv6 隧道的建立非常方便。但由于它必须使用 IPv4 兼容 IPv6 地址，仍依赖于 IPv4 地址，在使用时有一定的局限性。

(4) 6to4 隧道

- 普通 6to4 隧道

6to4 隧道是点到多点的自动隧道，主要用于将多个 IPv6 孤岛通过 IPv4 网络连接到 IPv6 网络。6to4 隧道通过在 IPv6 报文的目的地址中嵌入 IPv4 地址，来实现自动获取隧道终点的 IPv4 地址。

6to4 隧道采用特殊的 6to4 地址，其格式为：2002:abcd:efgh:子网号::接口 ID/64，其中 2002 表示固定的 IPv6 地址前缀，abcd:efgh 表示该 6to4 隧道对应的 32 位全球唯一的 IPv4 源地址，用 16 进制表示（如 1.1.1.1 可以表示为 0101:0101）。2002:abcd:efgh 之后的部分唯一标识了一个主机在 6to4 网络内的位置。通过这个嵌入的 IPv4 地址可以自动确定隧道的终点，使隧道的建立非常方便。

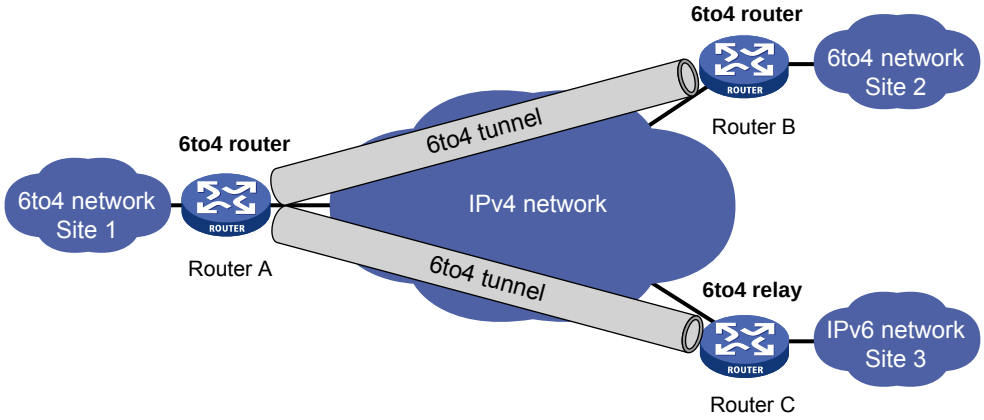
由于 6to4 地址的 64 位地址前缀中的 16 位子网号可以由用户自定义，前缀中的前 48 位已由固定数值、隧道起点或终点设备的 IPv4 地址确定，使 IPv6 报文通过隧道进行转发成为可能。6to4 隧道可以实现利用 IPv4 网络完成 IPv6 网络的互连，克服了 IPv4 兼容 IPv6 自动隧道使用的局限性。

- 6to4 中继

6to4 隧道只能用于前缀为 2002::/16 的 6to4 网络之间的通信，但在 IPv6 网络中也会使用像 2001::/16 这样的 IPv6 网络地址。为了实现 6to4 网络和其它 IPv6 网络的通信，必须有一台 6to4 路由器作为网关转发到 IPv6 网络的报文，这台路由器就叫做 6to4 中继（6to4 relay）路由器。

如下图所示，6to4 网络的边缘路由器 Router A 需配置一条静态路由，下一跳地址指向 6to4 中继路由器 Router C 的 6to4 地址，这样，所有去往 IPv6 网络的报文都会被转发到 6to4 中继路由器，之后再由 6to4 中继路由器转发到 IPv6 网络中，从而实现 6to4 网络（地址前缀以 2002 开始）与 IPv6 网络的互通。

图1-2 6to4 隧道和 6to4 中继原理图



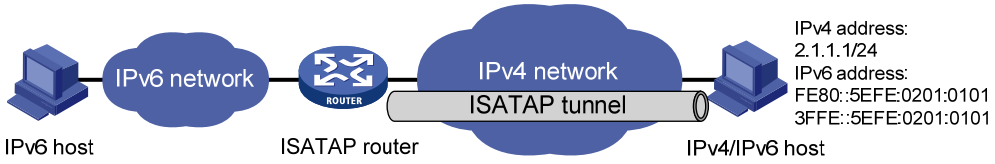
(5) ISATAP 隧道

随着 IPv6 技术的推广，现有的 IPv4 网络中将会出现越来越多的 IPv6 主机，ISATAP 隧道技术为这种应用提供了一个较好的解决方案。ISATAP 隧道是点到点的自动隧道技术，通过在 IPv6 报文的目的地址中嵌入的 IPv4 地址，可以自动获取隧道的终点。

使用 ISATAP 隧道时，IPv6 报文的目的地址和隧道接口的 IPv6 地址都要采用特殊的 ISATAP 地址。ISATAP 地址格式为：Prefix(64bit):0:5EFE:ip-address。其中，64 位的 Prefix 为任何合法的 IPv6 单播地址前缀，ip-address 为 32 位 IPv4 源地址，形式为 a.b.c.d 或者 abcd:efgh，且该 IPv4 地址不要求全球唯一。通过这个嵌入的 IPv4 地址就可以自动建立隧道，完成 IPv6 报文的传送。

ISATAP 隧道主要用于在 IPv4 网络中 IPv6 路由器-IPv6 路由器、IPv6 主机-IPv6 路由器的连接。

图1-3 ISATAP 隧道原理图



1.1.2 IPv4 over IPv4 隧道

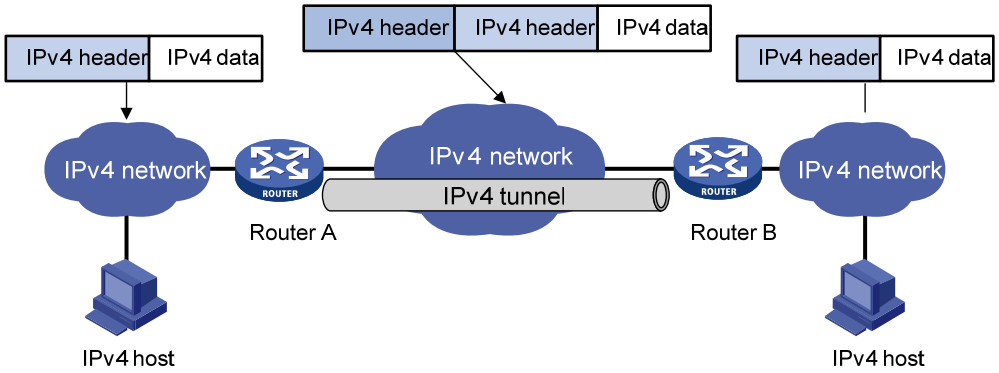
1. IPv4 over IPv4 隧道协议简介

IPv4 over IPv4 隧道（RFC1853）协议是对 IP 数据报进行封装，使得一个 IPv4 网络的数据能够在另一个 IPv4 网络中传输。

2. 报文封装及解封装

报文在 Tunnel 中传输经过封装与解封装两个过程，以下图为例说明这两个过程：

图1-4 IPv4 隧道原理图



- 封装流程

Router A 连接 IPv4 主机所在子网的接口收到 IP 数据报后，首先交由 IP 协议栈处理。IP 协议栈根据 IP 报头中的目的地址来确定如何路由此包。如果报文要路由到与 Router B 相连的 IPv4 主机的地址，则将此报文发给 Route A 上连接 Router B 的 Tunnel 接口。

Tunnel 接口收到此包后，进行 IPv4 over IPv4 的封装，封装完成后重新交给 IP 协议栈处理，IP 协议栈根据添加的 IP 报头确定出接口。

- 解封装流程

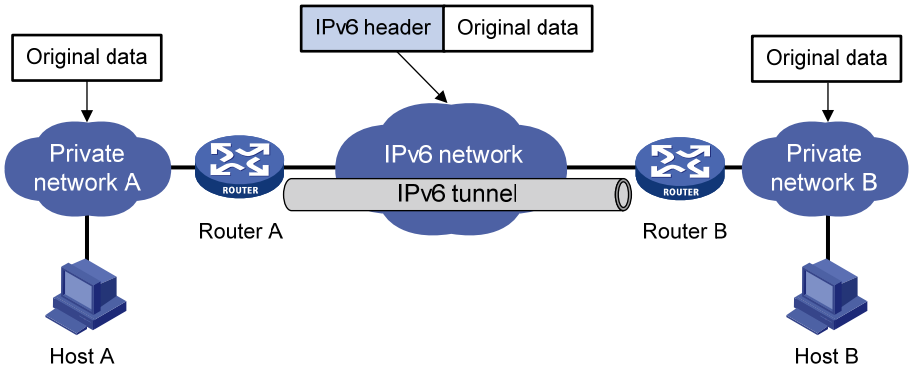
解封装过程和封装的过程相反。从网络接口收到的 IP 报文被送到 IP 协议栈，IP 协议栈检查 IP 报头的协议号；若发现此协议号为 4 即 IPv4，则将此 IP 数据包发送到隧道模块进行解封装处理；解封装之后的 IP 报文将重新被送到 IP 协议栈进行处理。

1.1.3 IPv4 或 IPv6 over IPv6 隧道

1. IPv4 或 IPv6 over IPv6 隧道协议简介

IPv4 或 IPv6 over IPv6 隧道（RFC2473）协议是对 IPv4 或者 IPv6 的数据报进行封装，使这些被封装的数据报能够在另一个 IPv6 网络中传输，封装后的数据报文即 IPv6 隧道报文。

图1-5 IPv6 隧道原理图



如 图 1-5所示，Original data指IPv4 或IPv6 报文。

2. 报文封装及解封装

- 封装流程

Router A 连接网络 A 的接口收到原始数据报后，首先交由对应的数据转发模块进行处理，并确定如何路由此数据包；若确定此数据包要路由到与 Router B 相连的 Host B 的地址，则将此报文发给 Router A 上连接 IPv6 网络的 Router B 的 Tunnel 接口。

Tunnel 口收到此包后添加 IPv6 报文头，封装完成后交给 IPv6 模块处理；IPv6 协议模块根据 IPv6 隧道头的目的地址重新决定路由。

- 解封装流程

解封装过程和封装的过程相反。从 IPv6 网络接口接收的数据包被送到 IPv6 协议模块进行处理；若乘客协议为 IPv4 或 IPv6，则数据包进入隧道处理模块进行解封装处理；解封装之后的数据包被送往相应的协议模块进行二次路由处理。

注意

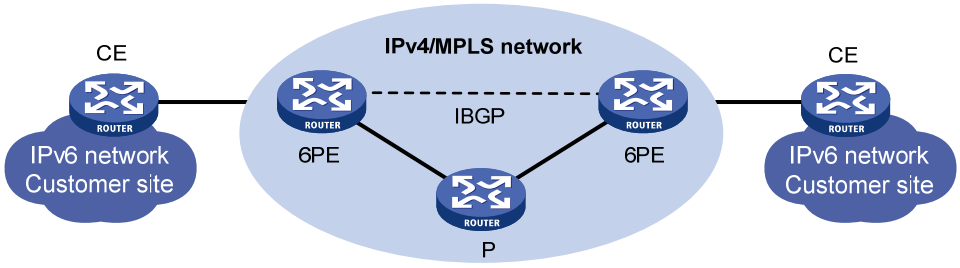
GRE 协议可以用来实现“IPv4 或 IPv6 over IPv6 隧道”的功能，相关配置请参见“VPN 分册”中的“GRE 配置”。

1.1.4 6PE 概述

6PE 是一种过渡技术，ISP 可以利用已有的 IPv4 骨干网为分散用户的 IPv6 网络提供接入能力。

6PE 的主要思想是：6PE（IPv6 Provider Edge，IPv6 供应商边缘）路由器将用户的 IPv6 路由信息转换为带有标签的 IPv6 路由信息，并且通过 IBGP（Internal Border Gateway Protocol，内部边界网关协议）会话扩散到 ISP 的 IPv4 骨干网中。6PE 路由器转发 IPv6 报文时，首先会将进入骨干网隧道的数据流打上标签。隧道可以是 GRE 隧道或者 MPLS LSP 等。

图1-6 6PE 组网图



当 ISP 想利用自己原有的 IPv4/MPLS 网络，使其通过 MPLS 具有 IPv6 流量交换能力时，只需要升级 PE 路由器就可以了。所以对于运营商来说，使用 6PE 技术作为 IPv6 过渡机制无疑是一个高效的解决方案，其操作风险也会小得多。

说明

有关 6PE 的详细介绍及配置请参考“IP 路由分册”中的“IPv6 BGP 配置”，下文不作介绍。

1.1.5 协议规范

与隧道技术相关的协议规范有：

- RFC1853：IP in IP Tunneling
- RFC2473：Generic Packet Tunneling in IPv6 Specification

- RFC2893：Transition Mechanisms for IPv6 Hosts and Routers
- RFC3056：Connection of IPv6 Domains via IPv4 Clouds
- RFC4214：Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)

1.2 隧道技术配置任务简介

表1-3 隧道技术配置任务简介

配置任务		说明	详细配置
配置 IPv6 over IPv4 隧道	配置 IPv6 手动隧道	可选 根据具体组网应用配置	1.3
	配置 IPv4 兼容 IPv6 自动隧道	可选 根据具体组网应用配置	1.4
	配置 6to4 隧道	可选 根据具体组网应用配置	1.5
	配置 ISATAP 隧道	可选 根据具体组网应用配置	1.6
配置 IPv4 over IPv4 隧道		可选 根据具体组网应用配置	1.7
配置 IPv4 over IPv6 隧道		可选 根据具体组网应用配置	1.8
配置 IPv6 over IPv6 隧道		可选 根据具体组网应用配置	1.9

1.3 配置 IPv6 手动隧道

1.3.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.3.2 配置 IPv6 手动隧道

表1-4 配置 IPv6 手动隧道

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建 Tunnel 接口并进入 Tunnel 接口视图	**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口

操作		命令	说明
设置 Tunnel 接口的 IPv6 地址	配置 IPv6 全球单播地址或站点本地地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	二者必选其一 缺省情况下，Tunnel 接口上没有设置 IPv6 全球单播地址或站点本地地址
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	配置 IPv6 链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了 IPv6 全球单播地址或站点本地地址后，会自动生成链路本地地址
		ipv6 address *ipv6-address* **link-local**	
配置为 IPv6 手动隧道模式		**tunnel-protocol ipv6-ipv4**	必选 缺省情况下，为 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口		**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置 Tunnel 接口的目的端地址		**destination** *ip-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址
设置隧道接口发送 IPv6 报文的 MTU 值		**ipv6 mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

- interface tunnel、**service slot** *slot-number* 和 tunnel-protocol 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。
- 隧道接口发送 IPv6 报文的 MTU 值的相关配置可参考“IP 业务分册/IPv6 基础配置命令”中的 ipv6 mtu。

注意

- 以上各项 Tunnel 接口下进行的功能特性配置，在删除 Tunnel 接口后，该接口上的所有配置也将被删除。
- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。在 Tunnel 的两端都要进行此项配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 配置静态路由时，需要手动配置到达目的地址（不是隧道的终点 IPv4 地址，而是封装前报文的目的 IPv6 地址）的路由，并配置出接口为本端 Tunnel 接口或下一跳为对端的 Tunnel 接口地址。在隧道的两端都要进行此项配置。
- 配置动态路由时，需要在隧道两端的 Tunnel 接口使能动态路由协议。在隧道的两端都要进行此项配置。相关配置请见“IP 路由分册”中的相关内容。

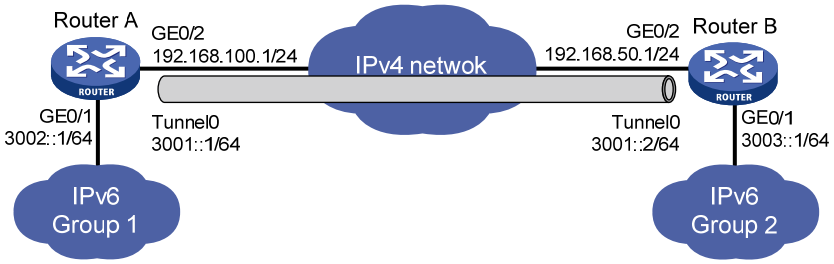
1.3.3 配置举例

1. 组网需求

如 图 1-7所示，两个IPv6 网络分别通过Router A和Router B与IPv4 网络连接，要求在Router A和Router B之间建立IPv6 手动隧道，使两个IPv6 网络可以互通。

2. 组网图

图1-7 IPv6 手动隧道组网图



3. 配置步骤

说明

已经配置 Router A 和 Router B 之间 IPv4 报文路由可达。

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ip address 192.168.100.1 255.255.255.0
[RouterA-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的 IPv6 地址。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ipv6 address 3002::1 64
[RouterA-GigabitEthernet0/1] quit
```

配置手动隧道。

```
[RouterA] interface tunnel 0
[RouterA-Tunnel0] ipv6 address 3001::1/64
[RouterA-Tunnel0] source GigabitEthernet 0/2
[RouterA-Tunnel0] destination 192.168.50.1
[RouterA-Tunnel0] tunnel-protocol ipv6-ipv4
[RouterA-Tunnel0] quit
```

配置从 Router A 经过 Tunnel0 接口到 Group 2 的静态路由。

```
[RouterA] ipv6 route-static 3003:: 64 tunnel 0
```

(2) 配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ip address 192.168.50.1 255.255.255.0
[RouterB-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的 IPv6 地址。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ipv6 address 3003::1 64
[RouterA-GigabitEthernet0/1] quit
```

配置手动隧道。

```
[RouterB] interface tunnel 0
[RouterB-Tunnel0] ipv6 address 3001::2/64
[RouterB-Tunnel0] source GigabitEthernet 0/2
[RouterB-Tunnel0] destination 192.168.100.1
[RouterB-Tunnel0] tunnel-protocol ipv6-ipv4
[RouterB-Tunnel0] quit
```

配置从 Router B 经过 Tunnel0 接口到 Group 1 的静态路由。

```
[RouterB] ipv6 route-static 3002:: 64 tunnel 0
```

4. 验证配置结果

完成以上配置之后，分别查看 Router A 和 Router B 的 Tunnel 接口状态如下：

```
[RouterA] display ipv6 interface tunnel 0 verbose
Tunnel0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:6401
  Global unicast address(es):
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FFA8:6401
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   55
……（略）
[RouterB] display ipv6 interface tunnel 0 verbose
Tunnel0 current state :UP
Line protocol current state :UP
```

```
IPv6 is enabled, link-local address is FE80::C0A8:3201
  Global unicast address(es):
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FFA8:3201
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   55
……（略）
```

从 Router A 上可以 Ping 通对端的 GigabitEthernet0/1 接口的 IPv6 地址：

```
[RouterA] ping ipv6 3003::1
  PING 3003::1 : 56  data bytes, press CTRL_C to break
    Reply from 3003::1
    bytes=56 Sequence=1 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=2 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=3 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=4 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=5 hop limit=64  time = 1 ms

  --- 3003::1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/1 ms
```

1.4 配置 IPv4 兼容 IPv6 自动隧道

1.4.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.4.2 配置 IPv4 兼容 IPv6 自动隧道

表1-5 配置 IPv4 兼容 IPv6 自动隧道

操作		命令	说明
进入系统视图		**system-view**	-
使能 IPv6 报文转发功能		**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建Tunnel接口并进入Tunnel接口视图		**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv6 地址	配置 IPv6 全球单播地址或站点本地地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	二者必选其一 缺省情况下，Tunnel 接口上没有设置 IPv6 全球单播地址或站点本地地址
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	配置 IPv6 链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了 IPv6 全球单播地址或站点本地地址后，会自动生成链路本地地址
		ipv6 address *ipv6-address* **link-local**	
配置为IPv4兼容IPv6自动隧道模式		**tunnel-protocol ipv6-ipv4 auto-tunnel**	必选 缺省情况下，为 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口		**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址或接口
设置隧道接口发送 IPv6 报文的 MTU 值		**ipv6 mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

- interface tunnel、**service slot** *slot-number* 和 tunnel-protocol 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。
- 隧道接口发送 IPv6 报文的 MTU 值的相关配置可参考“IP 业务分册/IPv6 基础配置命令”中的 ipv6 mtu。

注意

- IPv4 兼容 IPv6 自动隧道不需要配置目的地址。
- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。对于自动隧道，不支持动态路由。在 Tunnel 的两端都要进行转发路由的配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 对于自动隧道，使用同种封装协议的 Tunnel 接口不能同时配置完全相同的源地址。
- 配置静态路由时，需要手动配置到达目的地址（不是隧道的终点 IPv4 地址，而是封装前报文的目的 IPv6 地址）的路由，并配置出接口为本端 Tunnel 接口或下一跳为对端 Tunnel 接口地址。在隧道的两端都要进行此项配置。

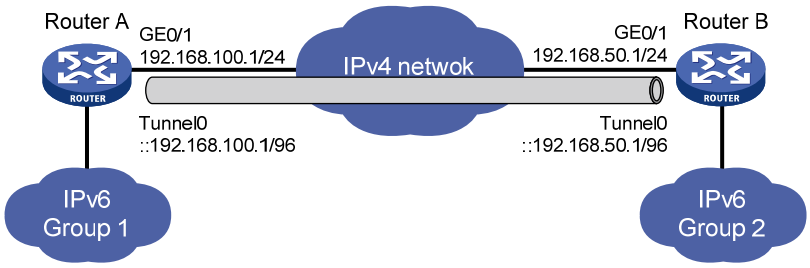
1.4.3 配置举例

1. 组网需求

如 图 1-8所示，两个IPv6 网络分别通过Router A和Router B与IPv4 网络连接，要求在Router A和Router B之间建立IPv6 自动隧道，使两个IPv6 网络可以互通。

2. 组网图

图1-8 IPv6 自动隧道组网图



3. 配置步骤

说明

已经配置 Router A 和 Router B 之间 IPv4 报文路由可达。

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ip address 192.168.100.1 255.255.255.0
[RouterA-GigabitEthernet0/1] quit
```

配置自动隧道。

```
[RouterA] interface tunnel 0
```

```
[RouterA-Tunnel0] ipv6 address ::192.168.100.1/96
[RouterA-Tunnel0] source GigabitEthernet 0/1
[RouterA-Tunnel0] tunnel-protocol ipv6-ipv4 auto-tunnel
```

(2) 配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ip address 192.168.50.1 255.255.255.0
[RouterB-GigabitEthernet0/1] quit
```

配置自动隧道。

```
[RouterB] interface tunnel 0
[RouterB-Tunnel0] ipv6 address ::192.168.50.1/96
[RouterB-Tunnel0] source GigabitEthernet 0/1
[RouterB-Tunnel0] tunnel-protocol ipv6-ipv4 auto-tunnel
```

4. 验证配置结果

完成以上配置之后，分别查看 Router A 和 Router B 的 Tunnel 接口状态如下：

```
[RouterA] display ipv6 interface tunnel 0 verbose
Tunnel0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:6401
  Global unicast address(es):
    ::192.168.100.1, subnet is ::/96
  Joined group address(es):
    FF02::1:FFA8:6401
    FF02::1:FF00:0
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   65
……（略）
[RouterB] display ipv6 interface tunnel 0 verbose
Tunnel0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:3201
  Global unicast address(es):
    ::192.168.50.1, subnet is ::/96
  Joined group address(es):
    FF02::1:FFA8:3201
    FF02::1:FF00:0
```

```
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   65
……（略）
```

从 Router A 上可以 Ping 通对端的 IPv4 兼容 IPv6 地址：

```
[RouterA] ping ipv6 ::192.168.50.1
  PING ::192.168.50.1 : 56  data bytes, press CTRL_C to break
    Reply from ::192.168.50.1
    bytes=56 Sequence=1 hop limit=64  time = 1 ms
    Reply from ::192.168.50.1
    bytes=56 Sequence=2 hop limit=64  time = 1 ms
    Reply from ::192.168.50.1
    bytes=56 Sequence=3 hop limit=64  time = 1 ms
    Reply from ::192.168.50.1
    bytes=56 Sequence=4 hop limit=64  time = 1 ms
    Reply from ::192.168.50.1
    bytes=56 Sequence=5 hop limit=64  time = 1 ms

  --- ::192.168.50.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/1 ms
```

1.5 配置 6to4 隧道

1.5.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.5.2 配置 6to4 隧道

表1-6 配置 6to4 隧道

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建 Tunnel 接口并进入 Tunnel 接口视图	**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口

操作		命令	说明
设置 Tunnel 接口的 IPv6 地址	配置 IPv6 全球单播地址或站点本地地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	二者必选其一 缺省情况下，Tunnel 接口上没有设置 IPv6 全球单播地址或站点本地地址
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	配置 IPv6 链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了 IPv6 全球单播地址或站点本地地址后，会自动生成链路本地地址
		ipv6 address *ipv6-address* **link-local**	
配置为 6to4 隧道模式		**tunnel-protocol ipv6-ipv4 6to4**	必选 缺省情况下，为 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口		**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置隧道接口发送 IPv6 报文的 MTU 值		**ipv6 mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

- interface tunnel、**service slot** *slot-number* 和 tunnel-protocol 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。
- 隧道接口发送 IPv6 报文的 MTU 值的相关配置可参考“IP 业务分册/IPv6 基础配置命令”中的 ipv6 mtu。

注意

- 6to4 隧道不需要配置目的地址，因为自动隧道的目的地址可以通过 6to4 IPv6 地址中嵌入的 IPv4 地址自动获得。
- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。对于自动隧道，不支持动态路由。在 Tunnel 的两端都要进行转发路由的配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 对于自动隧道，使用同种封装协议的 Tunnel 接口不能同时配置完全相同的源地址。
- 配置静态路由时，需要手动配置到达目的地址（不是隧道的终点 IPv4 地址，而是封装前报文的目的 IPv6 地址）的路由，并配置出接口为本端 Tunnel 接口或下一跳为对端 Tunnel 接口地址。在隧道的两端都要进行此项配置。

1.5.3 配置 6to4 隧道举例

1. 组网需求

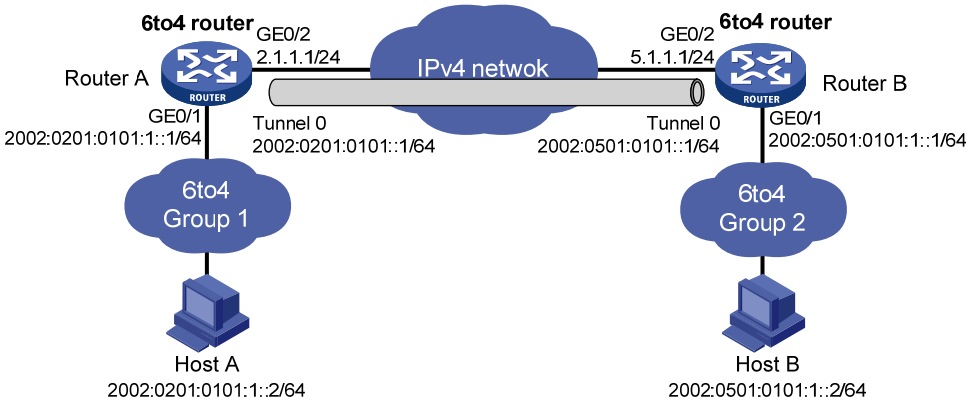
如 图 1-9所示，两个 6to4 网络通过网络边缘 6to4 router（Router A和Router B）与IPv4 网络相连，为了实现 6to4 网络中的主机Host A和Host B之间的互通，需要配置 6to4 隧道。

6to4 网络之间的互通需要为 6to4 网络内的主机及 6to4 router 配置 6to4 地址。

- Router A 上接口 GigabitEthernet0/2 的 IPv4 地址为 2.1.1.1/24，转换成 IPv6 地址后使用 6to4 前缀 2002:0201:0101::/48。对此前缀进行子网划分，Tunnel0 使用 2002:0201:0101::/64 子网，GigabitEthernet0/1 使用 2002:0201:0101:1::/64 子网。
- Router B 上接口 GigabitEthernet0/2 的 IPv4 地址为 5.1.1.1/24，转换成 IPv6 地址后使用 6to4 前缀 2002:0501:0101::/48。对此前缀进行子网划分，Tunnel0 使用 2002:0501:0101::/64 子网，GigabitEthernet0/1 使用 2002:0501:0101:1::/64 子网。

2. 组网图

图1-9 6to4 隧道组网图



3. 配置步骤

说明

已经配置 Router A 和 Router B 之间 IPv4 报文路由可达。

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ip address 2.1.1.1 24
[RouterA-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ipv6 address 2002:0201:0101:1::1/64
```

```
[RouterA-GigabitEthernet0/1] quit
```

配置 6to4 隧道。

```
[RouterA] interface tunnel 0
[RouterA-Tunnel0] ipv6 address 2002:201:101::1/64
[RouterA-Tunnel0] source GigabitEthernet 0/2
[RouterA-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[RouterA-Tunnel0] quit
```

配置到目的地址 2002::/16，下一跳为 Tunnel 接口的静态路由。

```
[RouterA] ipv6 route-static 2002:: 16 tunnel 0
```

(2) 配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ip address 5.1.1.1 24
[RouterB-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ipv6 address 2002:0501:0101:1::1/64
[RouterB-GigabitEthernet0/1] quit
```

配置 6to4 隧道。

```
[RouterB] interface tunnel 0
[RouterB-Tunnel0] ipv6 address 2002:0501:0101::1/64
[RouterB-Tunnel0] source GigabitEthernet 0/2
[RouterB-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[RouterB-Tunnel0] quit
```

配置到目的地址 2002::/16，下一跳为 Tunnel 接口的静态路由。

```
[RouterB] ipv6 route-static 2002:: 16 tunnel 0
```

4. 验证配置结果

完成以上配置之后，Host A 与 Host B 可以互相 Ping 通。

```
D:\>ping6 -s 2002:201:101:1::2 2002:501:101:1::2

Pinging 2002:501:101:1::2
from 2002:201:101:1::2 with 32 bytes of data:

Reply from 2002:501:101:1::2: bytes=32 time=13ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time<1ms

Ping statistics for 2002:501:101:1::2:
```

```
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

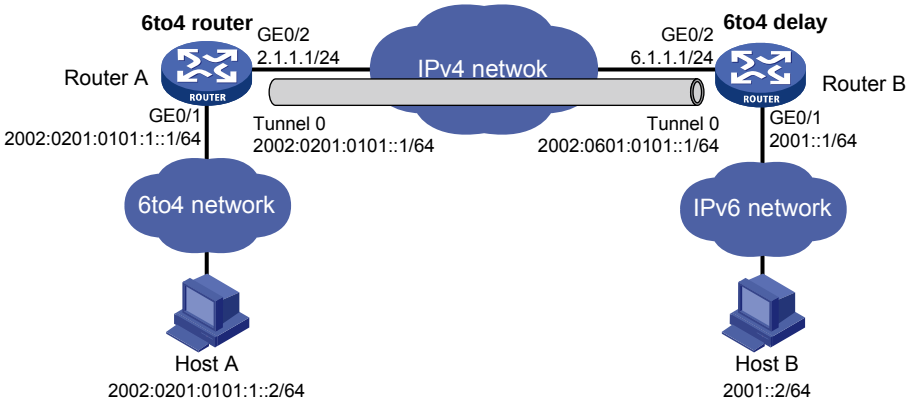
1.5.4 配置 6to4 中继举例

1. 组网需求

如 图 1-10所示，Router A为 6to4 路由器，其IPv6 侧的网络使用 6to4 地址。Router B作为 6to4 中继路由器，它和IPv6 网络（2001::/16）相连。要求在Router A和Router B之间配置 6to4 隧道，使得 6to4 网络中的主机与IPv6 网络中的主机互通。

2. 组网图

图1-10 6to4 中继组网图



3. 配置步骤

说明

- 已经配置 Router A 和 Router B 之间 IPv4 报文路由可达。
- 6to4 中继路由器的配置与 6to4 路由器的配置相同，但为实现 6to4 网络与 IPv6 网络的互通，需要在 6to4 路由器上配置到 IPv6 网络的路由。

---

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ip address 2.1.1.1 255.255.255.0
[RouterA-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ipv6 address 2002:0201:0101:1::1/64
[RouterA-GigabitEthernet0/1] quit
```

配置 6to4 隧道。

```
[RouterA] interface tunnel 0
[RouterA-Tunnel0] ipv6 address 2002:0201:0101::1/64
[RouterA-Tunnel0] source GigabitEthernet 0/2
[RouterA-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[RouterA-Tunnel0] quit
```

配置到纯 IPv6 网络的缺省路由和到目的地址 2002::/16，下一跳为 Tunnel 接口的静态路由。

```
[RouterA] ipv6 route-static :: 0 2002:0601:0101::1
[RouterA] ipv6 route-static 2002:: 16 tunnel 0
```

(2) 配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/2 的地址。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ip address 6.1.1.1 255.255.255.0
[RouterB-GigabitEthernet0/2] quit
```

配置接口 GigabitEthernet0/1 的地址。

```
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ipv6 address 2001::1/16
[RouterB-GigabitEthernet0/1] quit
```

配置 6to4 隧道。

```
[RouterB] interface tunnel 0
[RouterB-Tunnel0] ipv6 address 2002:0601:0101::1/64
[RouterB-Tunnel0] source GigabitEthernet 0/2
[RouterB-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[RouterB-Tunnel0] quit
```

配置到目的地址 2002::/16，下一跳为 Tunnel 接口的静态路由。

```
[RouterB] ipv6 route-static 2002:: 16 tunnel 0
```

4. 验证配置结果

完成以上配置之后，Host A 可以 Ping 通 Host B。

```
D:\>ping6 -s 2002:201:101:1::2 2001::2

Pinging 2001::2
from 2002:201:101:1::2 with 32 bytes of data:

Reply from 2001::2: bytes=32 time=13ms
Reply from 2001::2: bytes=32 time=1ms
Reply from 2001::2: bytes=32 time=1ms
Reply from 2001::2: bytes=32 time<1ms

Ping statistics for 2001::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
```

```
Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

1.6 配置 ISATAP 隧道

1.6.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.6.2 配置 ISATAP 隧道

表1-7 配置 ISATAP 隧道

操作		命令	说明
进入系统视图		**system-view**	-
使能 IPv6 报文转发功能		**ipv6**	必选 缺省情况下，IPv6 报文转发功能处于关闭状态
创建 Tunnel 接口并进入 Tunnel 接口视图		**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv6 地址	配置 IPv6 全球单播地址或站点本地地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	二者必选其一 缺省情况下，Tunnel 接口上没有设置 IPv6 全球单播地址或站点本地地址
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	配置 IPv6 链路本地地址	**ipv6 address auto link-local**	可选 缺省情况下，当接口配置了 IPv6 全球单播地址或站点本地地址后，会自动生成链路本地地址
		ipv6 address *ipv6-address* **link-local**	
配置为 ISATAP 隧道模式		**tunnel-protocol ipv6-ipv4 isatap**	必选 缺省情况下，为 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口		**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置隧道接口发送 IPv6 报文的 MTU 值		**ipv6 mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

- interface tunnel、**service slot** *slot-number* 和 tunnel-protocol 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。
- 隧道接口发送 IPv6 报文的 MTU 值的相关配置可参考“IP 业务分册/IPv6 基础配置命令”中的 ipv6 mtu。

注意

- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。对于自动隧道，不支持动态路由。在 Tunnel 的两端都要进行转发路由的配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置，配置命令的详细解释请参见相应的命令手册。
- 对于自动隧道，使用同种封装协议的 Tunnel 接口不能同时配置完全相同的源地址。
- 配置静态路由时，需要手动配置到达目的地址（不是隧道的终点 IPv4 地址，而是封装前报文的目的 IPv6 地址）的路由，并配置出接口为本端 Tunnel 接口或下一跳为对端 Tunnel 接口地址。在隧道的两端都要进行此项配置。

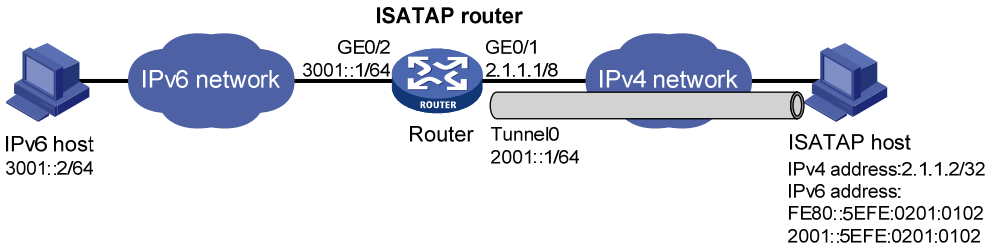
1.6.3 配置举例

1. 组网需求

如 图 1-11所示，IPv6 网络和IPv4 网络通过ISATAP路由器相连，要求将IPv4 网络中的IPv6 主机通过ISATAP隧道接入到IPv6 网络。

2. 组网图

图1-11 ISATAP 隧道组网图



3. 配置步骤

说明

已经配置 Router 的 GigabitEthernet0/1 和 ISATAP host 之间 IPv4 报文路由可达。

(1) 配置 Router

使能 IPv6 转发功能。

```
<Router> system-view
[Router] ipv6
```

配置各接口地址。

```
[Router] interface GigabitEthernet 0/2
[Router-GigabitEthernet0/2] ipv6 address 3001::1/64
[Router-GigabitEthernet0/2] quit
[Router] interface GigabitEthernet 0/1
[Router-GigabitEthernet0/1] ip address 2.1.1.1 255.0.0.0
[Router-GigabitEthernet0/1] quit
```

配置 ISATAP 隧道。

```
[Router] interface tunnel 0
[Router-Tunnel0] ipv6 address 2001::1/64 eui-64
[Router-Tunnel0] source GigabitEthernet 0/1
[Router-Tunnel0] tunnel-protocol ipv6-ipv4 isatap
```

取消对 RA 消息发布的抑制，使主机可以通过路由器发布的 RA 消息获取地址前缀等信息。

```
[Router-Tunnel0] undo ipv6 nd ra halt
[Router-Tunnel0] quit
```

配置到 ISATAP 主机的静态路由。

```
[Router] ipv6 route-static 2001:: 16 tunnel 0
```

(2) 配置 ISATAP 主机

ISATAP 主机上的具体配置与主机的操作系统有关，下面仅以 Windows XP 操作系统为例进行说明。

在 Windows XP 上，ISATAP 接口通常为接口 2，只要在该接口上配置 ISATAP 路由器的 IPv4 地址即可完成主机侧的配置。先看看这个 ISATAP 接口的信息：

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  does not use Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 0.0.0.0
  router link-layer address: 0.0.0.0
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1280 (true link MTU 65515)
  current hop limit 128
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
  DAD transmits 0
  default site prefix length 48
```

它自动生成了一个 ISATAP 格式的 link-local 地址（fe80::5efe:2.1.1.2）。我们需要设置这个接口上的 ISATAP 路由器的 IPv4 地址：

```
C:\>ipv6 rlu 2 2.1.1.1
```

只需要这么一个命令，这就完成了主机的配置，我们再来看看这个 ISATAP 接口的信息：

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  uses Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 2.1.1.2
  router link-layer address: 2.1.1.1
    preferred global 2001::5efe:2.1.1.2, life 29d23h59m46s/6d23h59m46s (public)
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1500 (true link MTU 65515)
```

```
  current hop limit 255
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
  DAD transmits 0
  default site prefix length 48
```

对比前后的区别，我们可以看到主机获取了 2001::/64 的前缀，自动生成地址 2001::5efe:2.1.1.2，同时还会发现这么一行“uses Router Discovery”表明主机启用了路由器发现，这时 Ping 一下路由器上隧道接口的 IPv6 地址，可以 Ping 通，这时候表明 ISATAP 隧道已经成功建立。

```
C:\>ping 2001::5efe:2.1.1.1

Pinging 2001::5efe:2.1.1.1 with 32 bytes of data:

Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms

Ping statistics for 2001::5efe:2.1.1.1:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

4. 验证配置结果

完成以上配置之后，ISATAP 主机就可访问 IPv6 网络中的主机。

1.7 配置 IPv4 over IPv4 隧道

1.7.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.7.2 配置 IPv4 over IPv4 隧道

表1-8 配置 IPv4 over IPv4 隧道

操作	命令	说明
进入系统视图	**system-view**	-
创建 Tunnel 接口并进入 Tunnel 接口视图	**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv4 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	必选 缺省情况下，Tunnel 接口上没有设置 IPv4 地址
配置隧道模式	**tunnel-protocol ipv4-ipv4**	可选 缺省情况下，采用 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则会造成报文传输失败
设置 Tunnel 接口的源端地址或接口	**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口

操作	命令	说明
设置 Tunnel 接口的目的端地址	**destination** *ip-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址
设置隧道接口的 MTU 值	**mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）	**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

interface tunnel、tunnel-protocol、**service slot** *slot-number* 和 mtu 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。

注意

- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。在 Tunnel 的两端都要进行此项配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 本端隧道接口的 IPv4 地址与隧道的目的地址不能在同一个网段内。
- 配置经过隧道接口的路由时，路由的目的地址不能与该隧道的目的地址在同一个网段内。
- 对两个或两个以上使用同种封装协议的 Tunnel 接口，不能同时配置完全相同的源地址和目的地址。
- 配置 Tunnel 接口的源端地址时，若采用配置源接口形式，则 Tunnel 的源地址取的是源接口的主 IP 地址。
- 配置动态路由时，需要在隧道两端的 Tunnel 接口使能动态路由协议。在隧道的两端都要进行此项配置。相关配置请见“IP 路由分册”中的相关内容。

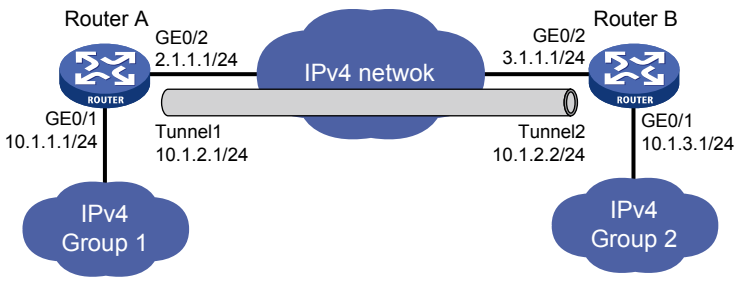
1.7.3 配置举例

1. 组网需求

运行 IP 协议的两个子网 Group 1 和 Group 2，通过在路由器 Router A 和路由器 Router B 之间使用三层隧道协议 IPv4 over IPv4 实现互联。

2. 组网图

图1-12 IPv4 over IPv4 隧道组网图



3. 配置步骤

说明

已经配置 Router A 和 Router B 之间 IPv4 报文路由可达。

(1) 配置 Router A

配置接口 GigabitEthernet0/1。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ip address 10.1.1.1 255.255.255.0
[RouterA-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ip address 2.1.1.1 255.255.255.0
[RouterA-GigabitEthernet0/2] quit
```

创建 Tunnel1 接口。

```
[RouterA] interface tunnel 1
```

配置 Tunnel1 接口的 IP 地址。

```
[RouterA-Tunnel1] ip address 10.1.2.1 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterA-Tunnel1] tunnel-protocol ipv4-ipv4
```

配置 Tunnel1 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] source 2.1.1.1
```

配置 Tunnel1 接口的目的地址（RouterB 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] destination 3.1.1.1
[RouterA-Tunnel1] quit
```

配置从 Router A 经过 Tunnel1 接口到 Group 2 的静态路由。

```
[RouterA] ip route-static 10.1.3.0 255.255.255.0 tunnel 1
```

(2) 配置 Router B

配置接口 GigabitEthernet0/1。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ip address 10.1.3.1 255.255.255.0
[RouterB-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ip address 3.1.1.1 255.255.255.0
[RouterB-GigabitEthernet0/2] quit
```

创建 Tunnel2 接口。

```
[RouterB] interface tunnel 2
```

配置 Tunnel2 接口的 IP 地址。

```
[RouterB-Tunnel2] ip address 10.1.2.2 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterB-Tunnel2] tunnel-protocol ipv4-ipv4
```

配置 Tunnel2 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] source 3.1.1.1
```

配置 Tunnel2 接口的目的地址（Router A 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] destination 2.1.1.1
[RouterB-Tunnel2] quit
```

配置从 Router B 经过 Tunnel2 接口到 Group 1 的静态路由。

```
[RouterB] ip route-static 10.1.1.0 255.255.255.0 tunnel 2
```

4. 验证配置结果

完成以上配置之后，分别查看 Router A 和 Router B 的 Tunnel 接口状态如下：

```
<RouterA> display interface tunnel 1
Tunnel1 current state: UP
Line protocol current state: UP
Description: Tunnel1 Interface
The Maximum Transmit Unit is 64000
Internet Address is 10.1.2.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source 2.1.1.1, destination 3.1.1.1
Tunnel protocol/transport IP/IP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  2 bytes/sec, 0 packets/sec
    4 packets input,  256 bytes
    0 input error
    12 packets output,  768 bytes
    0 output error

<RouterB> display interface tunnel 2
Tunnel2 current state: UP
Line protocol current state: UP
Description: Tunnel2 Interface
The Maximum Transmit Unit is 64000
Internet Address is 10.1.2.2/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source 3.1.1.1, destination 2.1.1.1
Tunnel protocol/transport IP/IP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
```

```
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    5 packets input,  320 bytes
    0 input error
    9 packets output,  576 bytes
    0 output error
```

从 Router A 可以 Ping 通对端的 GigabitEthernet0/1 接口的 IPv4 地址：

```
[RouterA] ping 10.1.3.1
  PING 10.1.3.1: 56  data bytes, press CTRL_C to break
    Reply from 10.1.3.1: bytes=56 Sequence=1 ttl=255 time=15 ms
    Reply from 10.1.3.1: bytes=56 Sequence=2 ttl=255 time=15 ms
    Reply from 10.1.3.1: bytes=56 Sequence=3 ttl=255 time=16 ms
    Reply from 10.1.3.1: bytes=56 Sequence=4 ttl=255 time=16 ms
    Reply from 10.1.3.1: bytes=56 Sequence=5 ttl=255 time=15 ms

  --- 10.1.3.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 15/15/16 ms
```

1.8 配置 IPv4 over IPv6 隧道

1.8.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IPv6 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.8.2 配置 IPv4 over IPv6 隧道

表1-9 配置 IPv4 over IPv6 隧道

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建 Tunnel 接口并进入 Tunnel 接口视图	**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv4 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	必选 缺省情况下，Tunnel 接口上没有设置 IPv4 地址
配置隧道模式	**tunnel-protocol ipv4-ipv6**	可选 缺省情况下，采用 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口	**source** { *ipv6-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口

操作	命令	说明
设置 Tunnel 接口的目的端地址	**destination** *ipv6-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址
设置隧道接口发送 IPv4 报文的 MTU 值	**mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）	**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

interface tunnel、**service slot** *slot-number*、tunnel-protocol 和 mtu 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。

注意

- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。在 Tunnel 的两端都要进行此项配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 对两个或两个以上使用同种封装协议的 Tunnel 接口，不能同时配置完全相同的源地址和目的地址。
- 配置 Tunnel 接口的源端地址时，若采用配置源接口形式，则 Tunnel 的源地址取的是源接口的主 IP 地址。
- 配置动态路由时，需要在隧道两端的 Tunnel 接口使能动态路由协议。在隧道的两端都要进行此项配置。相关配置请见“IP 路由分册”中的相关内容。

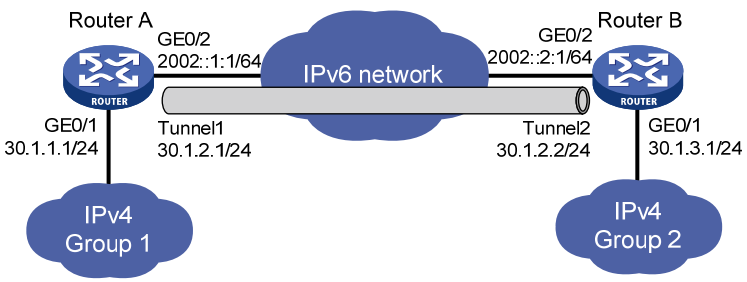
1.8.3 配置举例

1. 组网需求

运行 IP 协议的两个子网 Group 1 和 Group 2，通过在路由器 Router A 和路由器 Router B 之间使用 IPv4 over IPv6 隧道，穿越 IPv6 网络实现互联。

2. 组网图

图1-13 IPv4 over IPv6 隧道组网图



3. 配置步骤


说明

已经配置 Router A 和 Router B 之间 IPv6 报文路由可达。

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/1。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ip address 30.1.1.1 255.255.255.0
[RouterA-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ipv6 address 2002::1:1 64
[RouterA-GigabitEthernet0/2] quit
```

创建 Tunnel1 接口。

```
[RouterA] interface tunnel 1
```

配置 Tunnel1 接口的 IP 地址。

```
[RouterA-Tunnel1] ip address 30.1.2.1 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterA-Tunnel1] tunnel-protocol ipv4-ipv6
```

配置 Tunnel1 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] source 2002::1:1
```

配置 Tunnel1 接口的目的地址（Router B 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] destination 2002::2:1
[RouterA-Tunnel1] quit
```

配置从 Router A 经过 Tunnel1 接口到 Group 2 的静态路由。

```
[RouterA] ip route-static 30.1.3.0 255.255.255.0 tunnel 1
```

(2) 配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/1。

```
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ip address 30.1.3.1 255.255.255.0
[RouterB-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ipv6 address 2002::2:1 64
[RouterB-GigabitEthernet0/2] quit
```

创建 Tunnel2 接口。

```
[RouterB] interface tunnel 2
```

配置 Tunnel2 接口的 IP 地址。

```
[RouterB-Tunnel2] ip address 30.1.2.2 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterB-Tunnel2] tunnel-protocol ipv4-ipv6
```

配置 Tunnel2 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] source 2002::2:1
```

配置 Tunnel2 接口的目的地址（Router A 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] destination 2002::1:1
[RouterB-Tunnel2] quit
```

配置从 Router B 经过 Tunnel2 接口到 Group 1 的静态路由。

```
[RouterB] ip route-static 30.1.1.0 255.255.255.0 tunnel 2
```

4. 验证配置结果

完成以上配置之后，分别查看 Router A 和 Router B 的 Tunnel 接口状态如下：

```
<RouterA> display interface tunnel 1
Tunnel1 current state: UP
Line protocol current state: UP
Description: Tunnel1 Interface
The Maximum Transmit Unit is 64000
Internet Address is 30.1.2.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source 2002::0001:0001, destination 2002::0002:0001
Tunnel protocol/transport IP/IPv6
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    152 packets input,  9728 bytes
    0 input error
    168 packets output,  10752 bytes
    0 output error

<RouterB> display interface tunnel 2
Tunnel2 current state: UP
Line protocol current state: UP
Description: Tunnel2 Interface
The Maximum Transmit Unit is 64000
Internet Address is 30.1.2.2/24 Primary
```

```
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source 2002::0002:0001, destination 2002::0001:0001
Tunnel protocol/transport IP/IPv6
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  1 bytes/sec, 0 packets/sec
    Last 300 seconds output:  1 bytes/sec, 0 packets/sec
    167 packets input,  10688 bytes
    0 input error
    170 packets output,  10880 bytes
    0 output error
```

从 Router A 可以 Ping 通对端的 GigabitEthernet0/1 接口的 IPv4 地址：

```
[RouterA] ping 30.1.3.1
  PING 30.1.3.1: 56  data bytes, press CTRL_C to break
    Reply from 30.1.3.1: bytes=56 Sequence=1 ttl=255 time=46 ms
    Reply from 30.1.3.1: bytes=56 Sequence=2 ttl=255 time=15 ms
    Reply from 30.1.3.1: bytes=56 Sequence=3 ttl=255 time=16 ms
    Reply from 30.1.3.1: bytes=56 Sequence=4 ttl=255 time=15 ms
    Reply from 30.1.3.1: bytes=56 Sequence=5 ttl=255 time=16 ms

  --- 30.1.3.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 15/21/46 ms
```

1.9 配置 IPv6 over IPv6 隧道

1.9.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IPv6 地址，能够进行正常通讯。这些接口将作为 Tunnel 接口的源接口，以保证隧道目的地址路由可达。

1.9.2 配置 IPv6 over IPv6 隧道

表1-10 配置 IPv6 over IPv6 隧道

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建 Tunnel 接口并进入 Tunnel 接口视图	**interface tunnel** *number*	必选 缺省情况下，设备上无 Tunnel 接口

操作		命令	说明
设置 Tunnel 接口的 IPv6 地址	配置 IPv6 全球单播地址或站点本地地址	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	必选之一 缺省情况下，Tunnel 接口上没有设置 IPv6 地址
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	配置 IPv6 链路本地地址	**ipv6 address auto link-local**	
		ipv6 address *ipv6-address* **link-local**	
配置隧道模式		**tunnel-protocol ipv6-ipv6**	可选 缺省情况下，采用 GRE 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口		**source** { *ipv6-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置 Tunnel 接口的目的端地址		**destination** *ipv6-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址
设置隧道接口发送 IPv6 报文的 MTU 值		**ipv6 mtu** *mtu-size*	可选
配置当前接口的转发流量业务处理板（分布式设备）		**service slot** *slot-number*	可选 缺省情况下，接口没有设置转发流量业务处理板

说明

interface tunnel、**service slot** *slot-number* 和 tunnel-protocol 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。

隧道接口发送 IPv6 报文的 MTU 值的相关配置可参考“IP 业务分册/IPv6 基础配置命令”中的 ipv6 mtu。

注意

- 如果隧道两端 Tunnel 接口的地址不在同一个网段，则必须配置通过隧道到达对端的转发路由，以便需要进行封装的报文能正常转发。用户可以配置静态路由，也可以配置动态路由。在 Tunnel 的两端都要进行此项配置，配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置。
- 本端隧道接口的 IPv6 地址与隧道的目的地址不能在同一个网段内。
- 配置经过隧道接口的路由时，路由的目的地址不能与该隧道的目的地址在同一个网段内。
- 对两个或两个以上使用同种封装协议的 Tunnel 接口，不能同时配置完全相同的源地址和目的地址。
- 配置 Tunnel 接口的源端地址时，若采用配置源接口形式，则 Tunnel 的源地址取的是源接口的主 IP 地址。
- 配置动态路由时，需要在隧道两端的 Tunnel 接口使能动态路由协议。在隧道的两端都要进行此项配置。相关配置请见“IP 路由分册”中的相关内容。
- 只有 IPv6 over IPv6 隧道有最大嵌套封装次数限制。

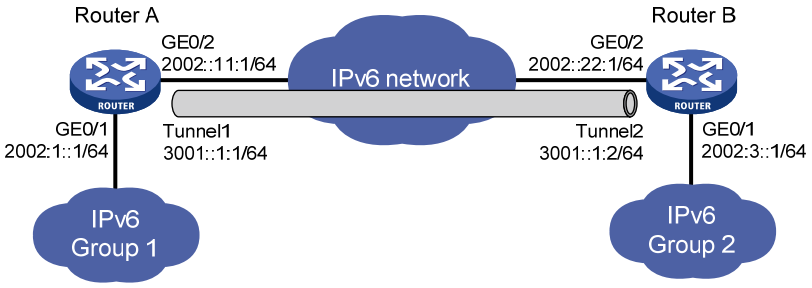
1.9.3 配置举例

1. 组网需求

运行 IPv6 协议的两个子网 Group 1 和 Group 2，通过在路由器 Router A 和路由器 Router B 之间使用三层隧道协议 IPv6 over IPv6 实现互联。

2. 组网图

图1-14 IPv6 over IPv6 隧道组网图



3. 配置步骤

说明

已经配置 Router A 和 Router B 之间 IPv6 报文路由可达。

(1) 配置 Router A

使能 IPv6 转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet0/1。

```
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ipv6 address 2002:1::1 64
[RouterA-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] ipv6 address 2002::11:1 64
[RouterA-GigabitEthernet0/2] quit
```

创建 Tunnel1 接口。

```
[RouterA] interface tunnel 1
```

配置 Tunnel1 接口的 IP 地址。

```
[RouterA-Tunnel1] ipv6 address 3001::1:1 64
```

配置 Tunnel 封装模式。

```
[RouterA-Tunnel1] tunnel-protocol ipv6-ipv6
```

配置 Tunnel1 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] source 2002::11:1
```

配置 Tunnel1 接口的目的地址（Router B 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterA-Tunnel1] destination 2002::22:1
[RouterA-Tunnel1] quit
```

配置从 Router A 经过 Tunnel1 接口到 Group 2 的静态路由。

```
[RouterA] ipv6 route-static 2002:3:: 64 tunnel 1
```

(2)　配置 Router B

使能 IPv6 转发功能。

```
<RouterB> system-view
[RouterB] ipv6
```

配置接口 GigabitEthernet0/1。

```
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] ipv6 address 2002:3::1 64
[RouterB-GigabitEthernet0/1] quit
```

配置接口 GigabitEthernet0/2（隧道的实际物理接口）。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] ipv6 address 2002::22:1 64
[RouterB-GigabitEthernet0/2] quit
```

创建 Tunnel2 接口。

```
[RouterB] interface tunnel 2
```

配置 Tunnel2 接口的 IP 地址。

```
[RouterB-Tunnel2] ipv6 address 3001::1:2 64
```

配置 Tunnel 封装模式。

```
[RouterB-Tunnel2] tunnel-protocol ipv6-ipv6
```

配置 Tunnel2 接口的源地址（GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] source 2002::22:1
```

配置 Tunnel2 接口的目的地址（Router A 的 GigabitEthernet0/2 的 IP 地址）。

```
[RouterB-Tunnel2] destination 2002::11:1
[RouterB-Tunnel2] quit
```

配置从 Router B 经过 Tunnel2 接口到 Group 1 的静态路由。

```
[RouterB] ipv6 route-static 2002:1:: 64 tunnel 2
```

4. 验证配置结果

完成以上配置之后，分别查看 Router A 和 Router B 的 Tunnel 接口状态如下：

```
<RouterA> display ipv6 interface tunnel 1 verbose
Tunnel1 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::2013:1
  Global unicast address(es):
    3001::1:1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
```

```
    FF02::1:FF01:1
    FF02::1:FF13:1
    FF02::2
    FF02::1
  MTU is 1460 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
……（略）
<RouterB> display ipv6 interface tunnel 2 verbose
Tunnel2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::2024:1
  Global unicast address(es):
    3001::1:2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF01:2
    FF02::1:FF24:1
    FF02::2
    FF02::1
  MTU is 1460 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
……（略）
```

从 Router A 可以 Ping 通对端的 GigabitEthernet0/1 接口的 IPv6 地址：

```
<RouterA> ping ipv6 2002:3::1
  PING 2002:3::1 : 56  data bytes, press CTRL_C to break
    Reply from 2002:3::1
    bytes=56 Sequence=1 hop limit=64  time = 31 ms
    Reply from 2002:3::1
    bytes=56 Sequence=2 hop limit=64  time = 1 ms
    Reply from 2002:3::1
    bytes=56 Sequence=3 hop limit=64  time = 16 ms
    Reply from 2002:3::1
    bytes=56 Sequence=4 hop limit=64  time = 16 ms
    Reply from 2002:3::1
    bytes=56 Sequence=5 hop limit=64  time = 31 ms

  --- 2002:3::1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/19/31 ms
```

1.10 隧道技术显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IPv6 隧道配置后的运行情况，通过查看显示信息验证配置的效果。

表1-11 隧道技术显示和维护

操作	命令
显示指定 Tunnel 接口的相关信息	**display interface tunnel** [*number*]
显示指定 Tunnel 接口的 IPv6 相关信息	**display ipv6 interface tunnel** [*number*] [**verbose**]



说明

display interface tunnel 和 **display ipv6 interface tunnel** 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。

1.11 常见错误配置举例

故障现象：在 Tunnel 接口上配置了相关的参数后（例如隧道的起点、终点地址和隧道模式）仍未处于 up 状态。

故障排除：可以按照如下步骤进行。

(1) Tunnel 接口未处于 up 状态的最常见原因是隧道起点的物理接口没有处于 up 状态。使用 **display interface tunnel** 和 **display ipv6 interface tunnel** 命令查看隧道起点的物理接口状态为 up 还是 down。如果隧道状态是 down 的，可以通过用户视图下的 **debugging tunnel event** 调试命令查看隧道 down 的原因。

(2) Tunnel 接口未处于 up 状态的另一个可能的原因是隧道的终点地址不可达。使用 **display ipv6 routing-table** 和 **display ip routing-table** 命令查看是否终点地址通过路由可达。如果路由表中没有保证隧道通讯的路由项，请配置相关路由。

目 录

1 UDP Helper配置 1-1
1.1 UDP Helper简介 1-1
1.2 配置UDP Helper 1-1
1.3 UDP Helper显示和维护 1-2
1.4 UDP Helper典型配置举例 1-2
1.4.1 UDP Helper配置举例 1-2

1 UDP Helper 配置

说明

UDP Helper 功能中对于接口的相关配置，目前只能在 VLAN 接口、三层以太网接口（包括子接口）上进行。

1.1 UDP Helper 简介

网络中的主机有时需要通过发送广播报文，来获得网络配置或查询网络中其他设备的名称。但是，当主机与服务器或待查询的设备不在同一个广播域时，主机就无法获得所需要的信息。

为解决上述问题，设备提供了 UDP Helper 功能。通过该功能可以实现对指定 UDP 端口的 IP 广播报文进行中继转发，即将指定 UDP 端口的广播报文转换为单播报文发送给指定的目的服务器，起到中继的作用。

使能 UDP Helper 功能后，如果设备接收到广播报文，将根据报文的 UDP 目的端口号来判断是否要对其中继转发，并进行相应的处理：

- 如果报文的 UDP 目的端口号与配置的需要中继转发的 UDP 端口号匹配，则修改 IP 报文头的目的 IP 地址，将报文发给指定的目的服务器；
- 否则，直接将报文送给上层协议处理。

1.2 配置 UDP Helper

表1-1 配置 UDP Helper

操作	命令	说明
进入系统视图	**system-view**	-
使能 UDP Helper 功能	**udp-helper enable**	必选 缺省情况下，UDP Helper 功能处于关闭状态
配置需要中继转发的 UDP 端口	**udp-helper port** { *port-number* | **dns** | **netbios-ds** | **netbios-ns** | **tacacs** | **tftp** | **time** }	必选 缺省情况下，没有配置中继转发的 UDP 端口
进入接口视图	**interface** *interface-type interface-number*	-
配置中继转发的目的服务器	**udp-helper server** *ip-address*	必选 缺省情况下，没有配置中继转发的目的服务器

注意

- UDP Helper 功能不能中继转发 DHCP 广播报文，即中继转发的 UDP 端口不能配置为 67 和 68。
- 参数 **dns**、**netbios-ds**、**netbios-ns**、**tacacs**、**tftp**、**time** 有两种配置方法：指定端口号配置和指定参数配置。例如：**udp-helper port** 53 和 **udp-helper port dns** 的效果是一样的。
- 关闭 UDP Helper 功能后，所有已配置的 UDP 端口都被取消。
- 设备上，最多可以配置 256 个需要中继转发的 UDP 端口。
- 一个接口上，最多可以配置 20 个中继转发的目的服务器。

1.3 UDP Helper 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 UDP 中继转发的目的服务器信息以及中继转发到目的服务器的报文数目，通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除 UDP 中继转发的报文统计数目。

表1-2 UDP Helper 显示和维护

操作	命令
显示 UDP 中继转发的相关信息	**display udp-helper server** [**interface** *interface-type interface-number*]
清除 UDP 中继转发的报文统计数目	**reset udp-helper packet**

1.4 UDP Helper 典型配置举例

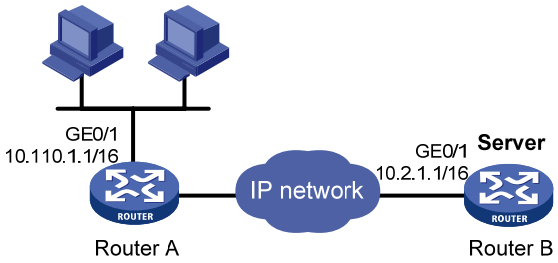
1.4.1 UDP Helper 配置举例

1. 组网需求

Router A 的 GigabitEthernet0/1 接口的 IP 地址为 10.110.1.1/16，连接到网段 10.110.0.0/16。配置将目的 UDP 端口号为 55，目的 IP 地址为 255.255.255.255 的广播 UDP 报文和目的地址为此网段子网广播地址 10.110.255.255 的广播 UDP 报文中继转发到指定的目的服务器 10.2.1.1/16。

2. 组网图

图1-1 UDP Helper 配置举例一组网图（路由应用）



3. 配置步骤

说明

在开始下面的配置之前，假设 Router A 到网段 10.2.0.0/16 有路由可达。

使能 UDP Helper 功能。

```
<RouterA> system-view
[RouterA] udp-helper enable
```

配置将目的 UDP 端口号为 55 的广播报文进行中继转发。

```
[RouterA] udp-helper port 55
```

配置中继转发的目的服务器为 10.2.1.1。

```
[RouterA] interface gigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] ip address 10.110.1.1 16
[RouterA-GigabitEthernet0/1] udp-helper server 10.2.1.1
```

目 录

1 DHCPv6 配置 1-1
1.1 DHCPv6 简介 1-1
1.1.1 基本概念 1-1
1.1.2 DHCPv6 的典型组网应用 1-2
1.1.3 DHCPv6 中继的工作过程 1-2
1.1.4 协议规范 1-3
1.2 配置DHCPv6 中继 1-3
1.2.1 配置准备 1-3
1.2.2 配置步骤 1-3
1.3 DHCPv6 显示和维护 1-3
1.4 DHCPv6 典型配置举例 1-4
1.4.1 DHCPv6 中继典型配置举例 1-4

1 DHCPv6 配置

1.1 DHCPv6 简介

DHCPv6（Dynamic Host Configuration Protocol for IPv6，支持 IPv6 的动态主机配置协议）是针对 IPv6 编址方案设计的、为主机分配 IPv6 地址和其他网络配置参数的协议。

与其他 IPv6 地址分配方式（手工配置、通过路由器公告消息中的网络前缀无状态自动配置等）相比，DHCPv6 具有以下优点：

- 更好地控制地址的分配。通过 DHCPv6 不仅可以记录为主机分配的地址，还可以为特定主机分配特定的地址，以便于网络管理。
- 除了 IPv6 地址外，还可以为主机提供 DNS 服务器、域名等网络配置参数。

1.1.1 基本概念

1. 所有 DHCPv6 服务器和中继的组播地址

DHCPv6 中采用组播地址 FF02::1:2 来表示链路范围内所有的 DHCPv6 服务器和中继。

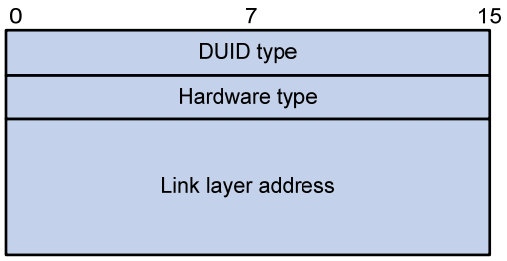
2. DUID

DUID（DHCP Unique Identifier，DHCP 唯一标识符）是唯一标识一台 DHCPv6 设备（包括客户端、中继和服务器）的标识符，用于 DHCPv6 设备之间的相互验证。

目前，设备采用RFC 3315 规定的DUID-LL（DUID Based on Link-layer Address，基于链路层地址的DUID）作为DHCPv6 设备的标识。DUID-LL的结构如 图 1-1所示：

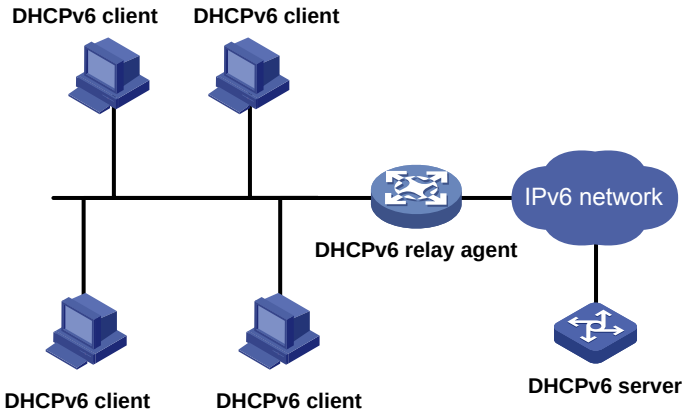
- DUID type：DUID 类型。DUID-LL 类型取值为 0x0003。
- Hardware type：硬件类型。设备支持的硬件类型为以太网，取值为 0x0001。
- Link layer address：链路层地址。取值为设备的桥 MAC 地址。

图1-1 DUID-LL 结构



1.1.2 DHCPv6 的典型组网应用

图1-2 DHCPv6 的典型组网

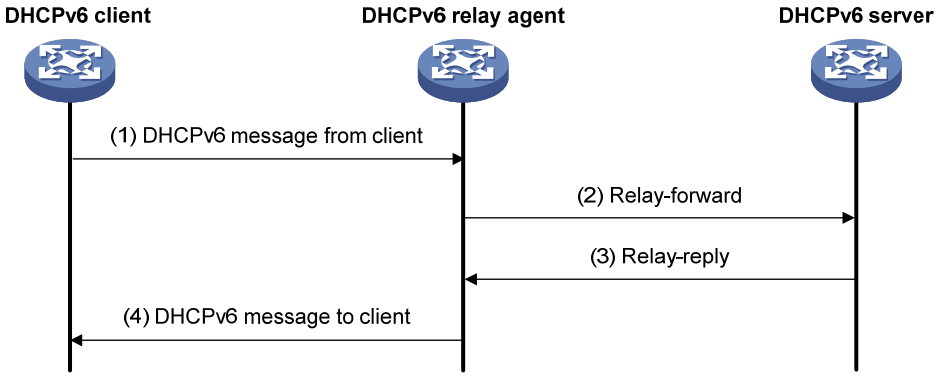


DHCPv6 的典型组网如 图 1-2所示。DHCPv6 客户端通过链路范围的组播地址与DHCPv6 服务器通信，以获取IPv6 地址和其他网络配置参数。如果服务器和客户端不在同一个链路范围内，则需要通过DHCPv6 中继来转发报文，这样可以避免在每个链路范围内都部署DHCPv6 服务器，既节省了成本，又便于进行集中管理。

1.1.3 DHCPv6 中继的工作过程

DHCPv6 中继和 DHCPv6 服务器之间交互两种报文：Relay-forward 和 Relay-reply。

图1-3 DHCPv6 中继的工作过程



如 图 1-3所示，DHCPv6 client通过DHCPv6 中继转发报文，获取IPv6 地址和其他网络配置参数的过程为：

(1) DHCPv6 客户端向所有 DHCPv6 服务器和中继的组播地址 FF02::1:2 发送请求；

(2) DHCPv6 中继接收到请求后，将其封装在 Relay-forward 报文的中继消息选项（Relay Message Option）中，并将 Relay-forward 报文发送给 DHCPv6 服务器；

(3) DHCPv6 服务器从 Relay-forward 报文中解析出客户端的请求，为客户端选取 IPv6 地址和其他参数，构造应答消息，将应答消息封装在 Relay-reply 报文的中继消息选项中，并将 Relay-reply 报文发送给 DHCPv6 中继；

(4) DHCPv6 中继从 Relay-reply 报文中解析出服务器的应答，转发给 DHCPv6 客户端；

DHCPv6 客户端根据 DHCPv6 服务器分配的 IPv6 地址和其他参数进行网络配置。

1.1.4 协议规范

与 DHCPv6 相关的协议规范有：

- RFC 3736：Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6
- RFC 3315：Dynamic Host Configuration Protocol for IPv6 (DHCPv6)
- RFC 2462：IPv6 Stateless Address Autoconfiguration

1.2 配置 DHCPv6 中继

工作在 DHCPv6 中继模式的接口接收到 DHCPv6 客户端发来的报文后，将其封装在 Relay-forward 报文中，并发送给指定的 DHCPv6 服务器，由 DHCPv6 服务器为客户端分配 IPv6 地址和其他网络配置参数。

1.2.1 配置准备

配置 DHCPv6 中继之前，需要通过 **ipv6** 命令使能 IPv6 报文的转发功能。

1.2.2 配置步骤

表1-1 配置 DHCPv6 中继

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口工作在 DHCPv6 中继模式，并指定 DHCPv6 服务器的地址	**ipv6 dhcp relay server-address** *ipv6-address* [**interface** *interface-type interface-number*]	必选 缺省情况下，接口没有工作在 DHCPv6 中继模式，也未指定任何 DHCPv6 服务器



说明

- 通过多次执行 **ipv6 dhcp relay server-address** 命令可以指定多个 DHCPv6 服务器，一个接口下最多可以指定 8 个 DHCPv6 服务器。DHCPv6 中继接收到 DHCPv6 客户端报文后，将其转发给所有的 DHCPv6 服务器。
- 如果指定的 DHCPv6 服务器地址为链路本地地址或链路范围的组播地址，则必须通过 **ipv6 dhcp relay server-address** 命令的 **interface** 参数指定出接口，否则报文可能会无法到达服务器。
- 接口上指定的 DHCPv6 服务器地址全部删除后，该接口不再工作在 DHCPv6 中继模式。

1.3 DHCPv6 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 DHCPv6 中继的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 DHCPv6 中继的统计信息。

表1-2 DHCPv6 显示和维护

操作	命令
显示本设备 DUID	**display ipv6 dhcp duid**
显示 DHCPv6 中继指定的 DHCPv6 服务器地址信息	**display ipv6 dhcp relay server-address** { **all** | **interface** *interface-type interface-number* }
显示 DHCPv6 中继的报文统计信息	**display ipv6 dhcp relay statistics**
清除 DHCPv6 中继的报文统计信息	**reset ipv6 dhcp relay statistics**

1.4 DHCPv6 典型配置举例

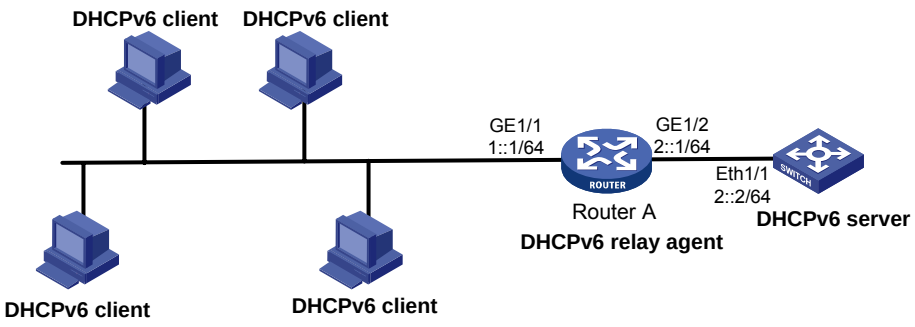
1.4.1 DHCPv6 中继典型配置举例

1. 组网需求

- DHCPv6 客户端所在网络地址为 1::/64，DHCPv6 服务器的地址为 2::2/64。客户端和服务器不在同一个链路，需要通过 DHCPv6 中继转发报文。
- Router A 作为 DHCPv6 中继，为客户端和服务器转发报文。
- Router A 同时作为 1::/64 网络的网关设备，通过 RA 消息中的 M 标志位和 O 标志位指定该网络中的主机通过 DHCPv6 获取 IPv6 地址和其他网络配置参数。

2. 组网图

图1-4 DHCPv6 中继组网图



3. 配置步骤

(1) 配置 Router A 作为 DHCPv6 中继

使能 IPv6 报文转发功能。

```
<RouterA> system-view
[RouterA] ipv6
```

配置接口 GigabitEthernet1/1 和 GigabitEthernet1/2 的 IPv6 地址。

```
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ipv6 address 2::1 64
[RouterA-GigabitEthernet1/2] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ipv6 address 1::1 64
```

在接口 GigabitEthernet1/1 上配置 DHCPv6 服务器地址。

```
[RouterA-GigabitEthernet1/1] ipv6 dhcp relay server-address 2::2
```

(2) 配置 Router A 作为网关

配置发布 RA 消息，并配置 M 和 O 标志位。

```
[RouterA-GigabitEthernet1/1] undo ipv6 nd ra halt
[RouterA-GigabitEthernet1/1] ipv6 nd autoconfig managed-address-flag
[RouterA-GigabitEthernet1/1] ipv6 nd autoconfig other-flag
```

(3) 验证配置结果

完成上述配置后，查看 DHCPv6 服务器的地址信息。

```
[RouterA-GigabitEthernet1/1] display ipv6 dhcp relay server-address all
Interface: GE1/1
Server address(es)                              Output Interface
2::2
```

查看 DHCPv6 中继转发报文的统计信息。

```
[RouterA-GigabitEthernet1/1] display ipv6 dhcp relay statistics
Packets dropped               :  0
    Error                     :  0
    Excess of rate limit      :  0
Packets received              :  14
    SOLICIT                   :  0
    REQUEST                   :  0
    CONFIRM                   :  0
    RENEW                     :  0
    REBIND                    :  0
    RELEASE                   :  0
    DECLINE                   :  0
    INFORMATION-REQUEST       :  7
    RELAY-FORWARD             :  0
    RELAY-REPLY               :  7
Packets sent                  :  14
    ADVERTISE                 :  0
    RECONFIGURE               :  0
    REPLY                     :  7
    RELAY-FORWARD             :  7
    RELAY-REPLY               :  0
```

目 录

1 IPv6 单播策略路由配置 1-1
1.1 IPv6 单播策略路由简介 1-1
1.2 配置IPv6 单播策略路由 1-1
1.2.1 配置IPv6 策略 1-1
1.2.2 配置IPv6 系统策略路由 1-3
1.2.3 配置IPv6 接口策略路由 1-3
1.3 IPv6 单播策略路由显示和维护 1-4
1.4 IPv6 单播策略路由典型配置举例 1-4
1.4.1 基于源地址的策略路由配置举例 1-4
1.4.2 基于报文大小的策略路由配置举例 1-6

1 IPv6 单播策略路由配置

1.1 IPv6 单播策略路由简介

策略路由（policy-based-route）是一种依据用户制定的策略进行路由选择的机制。与单纯依照 IPv6 报文的目的地址查找路由表进行转发不同，策略路由可以提供基于到达报文的源地址、地址长度等信息而灵活地进行路由选择。

策略路由可分为系统策略路由和接口策略路由：

- 系统策略路由对本地产生的报文进行策略路由，它只对本地产生的报文起作用，对转发的报文不起作用。对于一般的转发和安全等方面的使用需求，大多数情况下使用接口策略路由；
- 接口策略路由作用于到达该接口的报文，它只对转发的报文起作用，对本地产生的报文（比如本地的 ping 报文）不起作用。

一般来讲，策略路由的优先级要高于普通路由，即报文先按照策略路由进行转发。如果无法匹配所有的策略路由条件，再按照普通路由进行转发。但对于配置了缺省出接口（下一跳）的情况，则是先进行普通路由的转发，如果无法匹配，再进行策略路由转发。

1.2 配置 IPv6 单播策略路由

1.2.1 配置 IPv6 策略

IPv6 策略可以包含若干节点，每个策略节点由 *node-number* 来指定，*node-number* 的值越小优先级越高，优先级高的路由策略优先被执行。一个策略用来引入一条路由，对 IPv6 报文转发进行路由选择。策略的具体内容由 **if-match** 和 **apply** 子句来指定。

if-match 子句定义该节点的匹配规则，**apply** 子句定义通过该节点过滤后进行的动作。**if-match** 子句定义了那些需要使用策略路由的报文，当报文满足 **if-match** 子句时，则执行策略路由对应节点的 **apply** 子句，以完成报文的转发。

IPv6 单播策略路由提供了两种 **if-match** 子句，分别为 **if-match packet-length** 和 **if-match acl6**，同一类型的 **if-match** 子句只能有一条；IPv6 单播策略路由同时提供六种 **apply** 子句：**apply ipv6-precedence**、**apply output-interface**、**apply ipv6-address next-hop**、**apply default output-interface** 、**apply ipv6-address default next-hop** 、**apply destination-based-forwarding**，同一类型的 **apply** 子句最多只能配置一个。在满足所有 **if-match** 子句的情况下，**apply** 子句执行的优先级情况如下：

- 配置报文的优先级：**apply ipv6-precedence**，只要配置了该子句，该子句就一定会执行；
- 配置策略路由出接口和下一跳：**apply output-interface** 和 **apply ipv6-address next-hop**，其中 **apply output-interface** 命令的优先级高于 **apply ipv6-address next-hop**。当两条命令同时配置并且都有效时，系统只会执行 **apply output-interface** 命令；
- 使能基于目的地址的 IPv6 报文转发：**apply destination-based-forwarding**。当使能基于目的地址的 IPv6 报文转发时，则允许策略执行失败的报文按照正常转发；当关闭该使能开关时，则丢弃报文，直接跳出转发流程。如果不配置该子句，则 **apply default output-interface** 和 **apply ipv6-address default next-hop** 子句不会生效。

- 配置策略路由缺省出接口和下一跳：**apply default output-interface** 和 **apply ipv6-address default next-hop**，同样，**apply default output-interface** 命令的优先级高于 **apply ipv6-address default next-hop**。当两条命令同时配置并且都有效时，系统只会执行 **apply default output-interface** 命令。执行缺省出接口和下一跳命令的前提是，在策略路由中报文没有配置出接口或者下一跳，或者配置的出接口和下一跳无效，并且使能了基于目的地址的 IPv6 报文转发，但报文的目的 IP 地址在路由表中没有查到相应的路由，这时才会使用策略路由配置的缺省下一跳或者出接口。

IPv6 策略的每个节点可以配置多个 **if-match** 子句，而且这些子句之间是“与”的关系，如果某一节点不配置 **if-match** 子句，则所有报文都会通过该节点的过滤。IPv6 策略的不同节点间是“或”的关系，即，如果通过了 IPv6 策略路由的一个节点，就认为通过该 IPv6 策略的过滤，不再进行下一节点的测试。

在配置 IPv6 策略的节点时，可以指定节点的匹配模式为 **permit** 或者 **deny**：

- **permit** 指定节点的匹配模式为允许。当路由项通过该节点的过滤后，将执行该节点的 **apply** 子句，不进入下一个节点的测试；如果路由项没有通过该节点过滤，将进入下一个节点继续测试。
- **deny** 指定节点的匹配模式为拒绝，这时 **apply** 子句不会被执行。当路由项满足该节点的所有 **if-match** 子句时，将被拒绝通过该节点，不进入下一个节点；如果路由项不满足该节点的 **if-match** 子句，将进入下一个节点继续测试。

通过一个节点所定义的策略的报文将不再参与其他策略的过滤和处理。如果报文不能通过一个策略所有节点的过滤，则认为没有通过该条策略。该报文按正常转发流程处理。

在一个 IPv6 策略中最多可以指定五个下一跳或者五个出接口，并且可以起到按流进行负载分担的作用。

表1-1 配置 IPv6 策略

操作	命令	说明
进入系统视图	**system-view**	-
创建策略，并进入该策略视图	**ipv6 policy-based-route** *policy-name* [**deny** | **permit**] **node** *sequence-num*	必选 缺省情况下，没有定义策略或策略节点
配置 IPv6 报文的长度匹配条件	**if-match packet-length** *min-len max-len*	可选
配置IPv6报文的ACL匹配条件	**if-match acl6** *acl6-number*	可选
配置报文的优先级	**apply ipv6-precedence** { *type* | *value* }	可选
配置报文的发送接口	**apply output-interface** *interface-type interface-number*	可选
配置报文的下一跳	**apply ipv6-address next-hop** *ipv6-address*	可选
配置基于目的地址的 IPv6 报文转发	apply destination-based-forwarding	可选

操作	命令	说明
配置报文的缺省发送接口	**apply default output-interface** *interface-type interface-number*	可选 只有配置了 **apply destination-based-forwarding**，本功能才会生效
配置报文的缺省下一跳	**apply ipv6-address default next-hop** *ipv6-address*	可选 只有配置了 **apply destination-based-forwarding**，本功能才会生效

说明

- 如果只添加一个节点而没有任何匹配项和设置操作项，则所有报文都匹配，不再继续往下匹配。但是 IPv6 单播策略路由的统计数字不改变。
- 如果添加一个节点，只有匹配项，没有设置操作项，则进行匹配，但不执行相应的操作，不再继续往下匹配。但是 IPv6 单播策略路由的统计数字不改变。
- 如果添加一个节点，没有匹配项，有设置操作项，则所有报文都匹配，根据 **permit/deny** 执行相应的操作，不再继续往下匹配。但是 IPv6 单播策略路由的统计数字改变。
- 如果匹配项中使用的 ACL 根本不存在，则缺省是不匹配任何报文。
- 当直接出接口指定为本地的以太网接口或子接口时，虽然从指定接口转发，但不能正常通信，因为这几个接口是广播域，不能确定下一跳，因此必须指定下一跳。
- 当配置节点的匹配模式为 **deny** 时，不执行节点的 **apply** 命令，数据包将走正常路由表转发，因此没有 **deny** 的调试信息以及相应的统计信息。

1.2.2 配置 IPv6 系统策略路由

系统策略路由只对本地产生的报文有效，最多只能配置一条系统策略路由。

表1-2 配置 IPv6 系统策略路由

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 系统策略路由	**ipv6 local policy-based-route** *policy-name*	必选 缺省情况下，禁止 IPv6 系统策略路由

1.2.3 配置 IPv6 接口策略路由

接口策略路由影响到达本接口的报文。一个接口上最多只能配置一条接口策略路由。

表1-3 配置 IPv6 接口策略路由

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 IPv6 接口策略路由	**ipv6 policy-based-route** *policy-name*	必选 缺省情况下，禁止 IPv6 接口策略路由

1.3 IPv6 单播策略路由显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IPv6 单播策略路由配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令重置 IPv6 策略路由的统计信息。

表1-4 IPv6 单播策略路由显示和维护

操作	命令
显示系统和接口配置并使能的 IPv6 策略路由的信息	**display ipv6 policy-based-route**
显示已经使能的 IPv6 策略路由的配置情况	**display ipv6 policy-based-route setup** { *policy-name* | **interface** *interface-type interface-number* | **local** }
显示已经使能的 IPv6 策略路由的统计信息	**display ipv6 policy-based-route statistics** { **interface** *interface-type interface-number* | **local** }
显示已经配置的 IPv6 策略路由（集中式设备）	**display ipv6 config policy-based-route** [*policy-name*]
显示已经配置的 IPv6 策略路由（分布式设备）	**display ipv6 config policy-based-route** [*policy-name* [**slot** *slot-number*]]
重置 IPv6 策略路由的统计信息	**reset ipv6 policy-based-route statistics**

1.4 IPv6 单播策略路由典型配置举例

1.4.1 基于源地址的策略路由配置举例

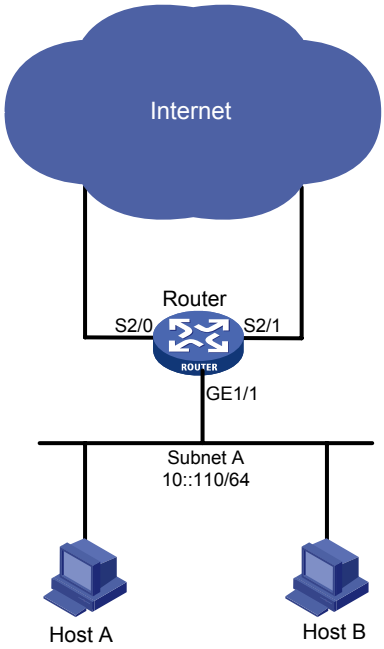
1. 组网需求

定义策略 aaa 的策略路由，控制所有从以太网接口 GigabitEthernet1/1 接收的 TCP 报文，通过串口 Serial2/0 发送。对其它 IPv6 报文，按照查找路由表的方式转发。

- 5 号节点，表示匹配 ACL 3001 的报文将被发往串口 Serial2/0；
- 10 号节点，表示匹配 ACL 3002 的报文不做策略路由处理；

2. 组网图

图1-1 基于源地址的策略路由配置举例组网图



3. 配置步骤

定义访问控制列表，ACL 3001 匹配 TCP 报文，ACL 3002 匹配 IPv6 报文。

```
<Router> system-view
[Router] ipv6
[Router] acl ipv6 number 3001
[Router-acl6-adv-3001] rule permit tcp
[Router-acl6-adv-3001] quit
[Router] acl ipv6 number 3002
[Router-acl6-adv-3002] rule permit ipv6
[Router-acl6-adv-3002] quit
```

定义 5 号节点，使 TCP 报文被发往串口 Serial2/0。

```
[Router] ipv6 policy-based-route aaa permit node 5
[Router-pbr6-aaa-5] if-match acl6 3001
[Router-pbr6-aaa-5] apply output-interface serial 2/0
[Router-pbr6-aaa-5] quit
```

定义 10 号节点，对于其它 IPv6 报文不进行策略路由。

```
[Router] ipv6 policy-based-route aaa deny node 10
[Router-pbr6-aaa-10] if-match acl6 3002
[Router-pbr6-aaa-10] quit
```

在以太网口 GigabitEthernet1/1 上应用定义的策略 aaa，处理此接口接收的报文。

```
[Router] interface GigabitEthernet 1/1
[Router-Ethernet1/0] ipv6 address 10::110 64
[Router-Ethernet1/0] ipv6 policy-based-route aaa
```

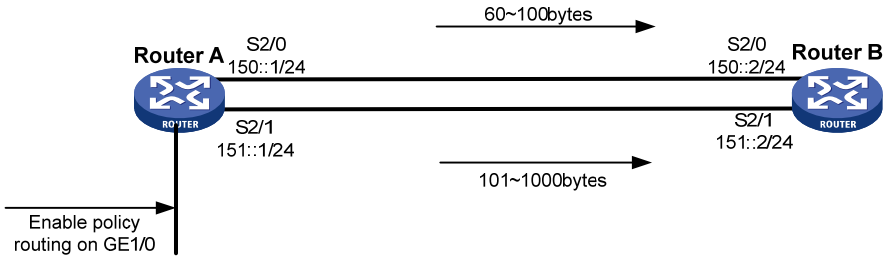
1.4.2 基于报文大小的策略路由配置举例

1. 组网需求

在 Router A 的 GigabitEthernet1/1 接口上应用 IPv6 策略 lab1；这个策略将大小为 64～100 字节的 IPv6 报文设置 150::2/64 作为下一跳 IPv6 地址；而将大小为 101～1000 字节的 IPv6 报文设置 151::2/64 作为下一跳 IPv6 地址。所有其它长度的 IPv6 报文都按照查找路由表的方式转发。

2. 组网图

图1-2 基于报文大小的策略路由配置举例组网图



3. 配置步骤

(1) 配置 Router A

配置动态路由协议 RIPng。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] ripng 1
[RouterA-ripng-1] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ipv6 address 150::1 64
[RouterA-Serial2/0] ripng 1 enable
[RouterA-Serial2/0] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ipv6 address 151::1 64
[RouterA-Serial2/1] ripng 1 enable
[RouterA-Serial2/1] quit
```

在以太网口 GigabitEthernet1/1 上应用定义的策略 lab1，处理此接口接收的报文。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ipv6 address 192::1 64
[RouterA-GigabitEthernet1/1] ripng 1 enable
[RouterA-GigabitEthernet1/1] ipv6 policy-based-route lab1
[RouterA-GigabitEthernet1/1] quit
```

将大小为 64～100 字节的 IPv6 报文转发到下一跳 150::2/64；而将大小为 101～1000 字节的 IPv6 报文转发到下一跳 151::2/64。

```
[RouterA] ipv6 policy-based-route lab1 permit node 10
[RouterA-pbr6-lab1-10] if-match packet-length 64 100
[RouterA-pbr6-lab1-10] apply ipv6-address next-hop 150::2
[RouterA-pbr6-lab1-10] quit
[RouterA] ipv6 policy-based-route lab1 permit node 20
```

```
[RouterA-pbr6-lab1-20] if-match packet-length 101 1000
[RouterA-pbr6-lab1-20] apply ipv6-address next-hop 151::2
```

(2) 配置 Router B

配置动态路由协议 RIPng。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] ripng 1
[RouterB-ripng-1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ipv6 address 150::2 64
[RouterB-Serial2/0] ripng 1 enable
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ipv6 address 151::2 64
[RouterB-Serial2/1] ripng 1 enable
[RouterB-Serial2/1] quit
```

目 录

1 NAT-PT配置 1-1
1.1 NAT-PT简介 1-1
1.1.1 NAT-PT机制 1-1
1.1.2 NAT-PT实现过程 1-2
1.1.3 协议规范 1-2
1.2 NAT-PT配置任务简介 1-3
1.3 配置NAT-PT 1-3
1.3.1 配置准备 1-3
1.3.2 使能NAT-PT功能 1-3
1.3.3 配置NAT-PT前缀 1-3
1.3.4 配置IPv4 侧报文的映射 1-4
1.3.5 配置IPv6 侧报文的映射 1-4
1.3.6 配置NAT-PT转换后报文的ToS/Traffic Class字段值 1-6
1.4 NAT-PT显示和维护 1-6
1.5 NAT-PT典型配置举例 1-6
1.5.1 配置IPv6 侧动态映射 1-6
1.5.2 配置IPv4 侧静态映射和IPv6 侧静态映射 1-8
1.6 常见配置错误举例 1-10

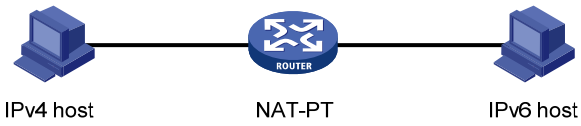
1 NAT-PT 配置

1.1 NAT-PT 简介

IPv6 的应用是个循序渐进的过程，在很长时间内，IPv4 网络和 IPv6 网络会同时存在且需要相互通信。在 IPv4 网络完全过渡到 IPv6 网络之前，两个网络之间直接的网络通信可以通过 NAT-PT（Network Address Translation-Protocol Translation，附带协议转换的网络地址转换）所提供的 IPv4 和 IPv6 地址之间的相互转换功能来实现。例如，使用此技术可以使 IPv6 网络中的主机直接访问 IPv4 网络中的 FTP 服务器。

如 图 1-1所示，NAT-PT作用于IPv4 和IPv6 网络边缘的设备上，所有的地址转换过程都在该设备上实现，对IPv4 和IPv6 网络来说是透明的，即用户不必改变目前的IPv4 网络中主机得配置就可实现 IPv6 网络与IPv4 网络的通信。

图1-1 NAT-PT 示意图



但由于 NAT-PT 具有下面一些局限性，所以在一些场合不推荐使用（如 IPv6 网络中主机跨越 IPv4 网络与另一 IPv6 网络中主机通信推荐使用隧道技术）：

- 属于同一会话的请求和响应都必须通过同一 NAT-PT 设备才能进行 NAT-PT 转换。
- 不能转换 IPv4 报文头的可选项部分。
- 缺少端到端的安全性。

目前，设备支持 NAT-PT 转换的应用协议包括 ICMP、DNS、FTP 以及与网络层协议相关但协议字段不涉及地址信息的相关协议。

1.1.1 NAT-PT 机制

有三种 NAT-PT 机制可实现 IPv4 和 IPv6 地址之间的相互转换：

1. 静态映射的 NAT-PT 机制

静态映射的 NAT-PT 机制是指采用手工配置的 IPv6 地址与 IPv4 地址的一一对应关系来实现 IPv6 地址与 IPv4 地址的转换。

2. 动态映射的 NAT-PT 机制

和静态映射不同，动态映射没有 IPv6 和 IPv4 地址之间的一一对应关系。动态映射要求先创建一个地址池，然后根据需要从地址池中选取空闲地址来完成 IPv6 地址与 IPv4 地址的映射。

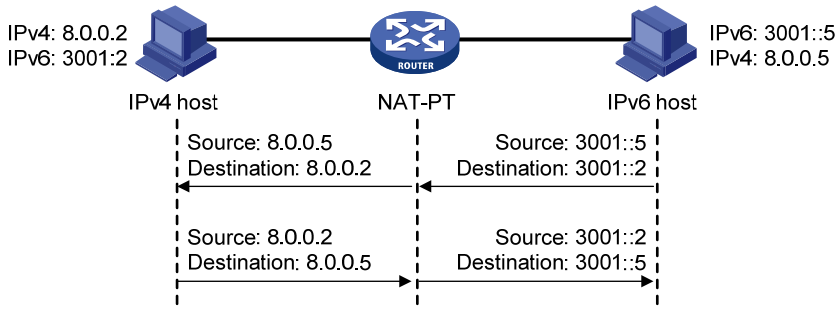
3. NAPT-PT 机制

NAPT-PT（Network Address Port Translation-Protocol Translation，附带协议转换的网络地址端口转换）是在 IP 地址动态转换的基础上对 TCP、UDP 的端口号也进行 IPv6 到 IPv4 的转换。采用这种“地址＋端口号”的映射方式，不同的 IPv6 地址转换时，可以对应同一个 IPv4 地址，通过端口号来区分不同的 IPv6 主机，从而使多个 IPv6 主机能共享一个 IPv4 地址完成转换。

1.1.2 NAT-PT 实现过程

1. IPv6 侧发起会话

图1-2 NAT-PT 的实现过程（IPv6 侧发起会话）



(1) 当 IPv6 网络中的设备发送给 IPv4 网络的报文到达 NAT-PT 设备后，如果设备判断该报文是要转发到 IPv4 网络，则会利用在 IPv6 侧已经配置的静态或者动态的映射关系，进行 IPv6 地址到 IPv4 地址的转换，将报文的源 IPv6 地址转换为 IPv4 地址。

(2) 根据 IPv4 侧配置的 IPv4 与 IPv6 的映射关系将目的地址转换为 IPv4 地址。如果没有配置映射关系，那么，如果报文中的目的 IPv6 地址的低 32 位可以直接转换为合法的 IPv4 地址，则直接转换为目的 IPv4 地址；否则，转换不成功。

(3) 报文的源 IPv6 地址和目的 IPv6 地址都转换为 IPv4 地址后，设备按照正常的转发流程将报文转发到 IPv4 网络中的设备。同时，将 IPv6 地址与转换后的 IPv4 地址的映射关系保存在设备中。

(4) IPv4 网络中的设备发送给 IPv6 网络中的报文到达 NAT-PT 设备后，将根据存储的映射关系进行相反的转换，从而将报文回送到 IPv6 网络中的设备。

2. IPv4 侧发起会话

(1) 当 IPv4 网络中的设备发送给 IPv6 网络的报文到达 NAT-PT 设备后，如果设备判断该报文是要转发到 IPv6 网络，则会利用在 IPv4 侧已经配置的静态或者动态的映射关系，进行 IPv4 地址到 IPv6 地址的转换，将报文的源 IPv4 地址转换为 IPv6 地址。

(2) 根据 IPv6 侧配置的 IPv6 与 IPv4 的映射关系将目的 IPv4 地址转换为 IPv6 地址。

(3) 报文的源 IPv4 地址和目的 IPv4 地址都转换为 IPv6 地址后，设备按照正常的转发流程将报文转发到 IPv6 网络中的设备。同时，将 IPv4 地址与转换后的 IPv6 地址的映射关系保存在设备中。

(4) IPv6 网络中的设备发送给 IPv4 网络中的报文到达 NAT-PT 设备后，将根据存储的映射关系进行相反的转换，从而将报文回送到 IPv4 网络中的设备。

1.1.3 协议规范

与 NAT-PT 相关的协议规范有：

- RFC 2765：Stateless IP/ICMP Translation Algorithm
- RFC 2766：Network Address Translation - Protocol Translation (NAT-PT)

1.2 NAT-PT 配置任务简介

表1-1 NAT-PT 配置任务简介

配置任务	说明	详细配置
使能 NAT-PT 功能	必选	1.3.2
配置 NAT-PT 前缀	必选	1.3.3
配置 IPv4 侧报文的映射	必选	1.3.4
配置 IPv6 侧报文的映射	必选	1.3.5
配置 NAT-PT 转换后报文的 TOS/Traffic Class 字段值	可选	1.3.6

1.3 配置 NAT-PT

1.3.1 配置准备

在配置 NAT-PT 之前，设备需要先使能 IPv6 转发功能，并根据实际情况在需要进行 NAT-PT 地址转换的接口上配置 IPv4 或 IPv6 地址。

1.3.2 使能 NAT-PT 功能

表1-2 使能 NAT-PT 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 NAT-PT 功能	**natpt enable**	必选 缺省情况下，NAT-PT 功能处于关闭状态

1.3.3 配置 NAT-PT 前缀

当报文要从 IPv6 网络发送到 IPv4 网络时，接收到该报文的具有 NAT-PT 功能的设备会检测报文 IPv6 目的地址的前缀，只有与所配置的 NAT-PT 前缀相同的报文才允许进行 IPv6 到 IPv4 的转换。

对于从 IPv4 侧到 IPv6 侧的报文，转换后的 IPv6 源地址前缀为配置的 NAT-PT 前缀。

表1-3 配置 NAT-PT 前缀

操作	命令	说明
进入系统视图	**system-view**	-
配置 NAT-PT 前缀	**natpt prefix** *natpt-prefix* [**interface** *interface-type interface-number*]	必选

注意

- NAT-PT 前缀不允许与使能 NAT-PT 功能的 IPv6 侧接口的网络地址相同。
- 被引用的 NAT-PT 前缀不能直接删除，必需取消引用配置后才能删除。

1.3.4 配置 IPv4 侧报文的映射

IPv4 侧报文映射是指从 IPv4 到 IPv6 的报文转换。当报文从 IPv4 网络发送到 IPv6 网络时，源 IPv4 地址将会按照配置的映射关系转换为 IPv6 地址。

IPv4 侧报文映射分为静态映射和动态映射两种：

- 静态映射是指设置 IPv4 地址与 IPv6 地址的一一对应关系，把源 IPv4 地址转换为相应的 IPv6 地址。
- 动态映射是指对于从 IPv4 侧发送到 IPv6 侧的报文，如果其 IPv4 源地址符合指定的 ACL 规则，则会在源 IPv4 地址前面添加 NAT-PT 前缀，转换为 IPv6 地址。

表1-4 配置 IPv4 侧报文的映射

操作		命令	说明
进入系统视图		**system-view**	-
配置 IPv4 侧报文的映射	配置 IPv4 侧报文的静态映射	**natpt v4bound static** { *ipv4-address ipv6-address* | **v6server protocol** *protocol-type ipv4-address ipv4-port-number ipv6-address ipv6-port-number* }	二者必选其一
	配置 IPv4 侧报文的动态映射	**natpt v4bound dynamic acl number** *acl-number* **prefix** *natpt-prefix*	

说明

- 在配置 IPv4 侧报文的动态映射时，命令中 *natpt-prefix* 参数必须通过 **natpt prefix** 命令事先配置。
- 关于 ACL 规则的详细配置请参见“安全分册”中的“ACL 配置”。

1.3.5 配置 IPv6 侧报文的映射

IPv6 侧报文映射是指从 IPv6 到 IPv4 的报文转换。当报文从 IPv6 网络发送到 IPv4 网络时，源 IPv6 地址将会按照配置的映射关系转换为 IPv4 地址。

IPv6 侧报文映射分为静态映射和动态映射两种：

- 静态映射是指设置 IPv6 地址与 IPv4 地址的一一对应关系，把源 IPv6 地址转换为相应的 IPv4 地址。
- 动态映射是指对于从 IPv6 侧到 IPv4 侧的报文，如果源 IPv6 地址符合指定的 IPv6 ACL 规则或者目的地址 IPv6 地址符合 NAT-PT 前缀，则会把源 IPv6 地址转换为指定 NAT-PT 地址池中的 IPv4 地址或者指定接口的 IPv4 地址。

设备支持四种动态映射方式：

(1) 组合 1：IPv6 ACL 与地址池的组合

如果 IPv6 报文的源 IPv6 地址符合指定的 IPv6 ACL 规则，源 IPv6 地址就会转换为配置命令中指定地址池中的 IPv4 地址。

(2) 组合 2：IPv6 ACL 与接口地址的组合

如果 IPv6 报文的源 IPv6 地址符合指定的 IPv6 ACL 规则，源 IPv6 地址就会转换为配置命令中指定接口的 IPv4 地址。

(3) 组合 3：NAT-PT 前缀与地址池的组合

如果 IPv6 报文的目的 IPv6 地址包含 NAT-PT 前缀，源 IPv6 地址就会转换为配置命令中指定地址池中的 IPv4 地址。

(4) 组合 4：NAT-PT 前缀与接口地址的组合

如果 IPv6 报文的目的 IPv6 地址包含 NAT-PT 前缀，源 IPv6 地址就会转换为配置命令中指定接口的 IPv4 地址。

如果在配置 IPv6 侧报文的动态映射时采用组合 1 或者组合 3，则首先需要配置 NAT-PT 地址池。

地址池是一组连续的 IPv4 地址，用于 IPv6 到 IPv4 报文的动态转换。当报文从 IPv6 网络发送到 IPv4 网络时，如果设置了组合 1 或者组合 3 的 NAT-PT 动态转换，则设备就会从地址池中选择一个 IPv4 地址作为 IPv6 报文转换后的源地址。

表1-5 配置 IPv6 侧报文的映射

操作		命令	说明
进入系统视图		**system-view**	-
配置IPv6侧报文的映射	配置 IPv6 侧报文的静态映射	**natpt v6bound static** *ipv6-address ipv4-address*	二者必选其一
	配置 IPv6 侧报文的动态映射	请参见 表 1-6配置IPv6 侧报文的动态映射	

表1-6 配置 IPv6 侧报文的动态映射

操作	命令	说明
配置 NAT-PT 地址池，定义 IPv6 侧报文的动态映射为：如果 IPv6 报文的源 IPv6 地址符合指定的 IPv6 ACL 规则，源 IPv6 地址就会转换为指定地址池中的 IPv4 地址	**natpt address-group** *group-number start-ipv4-address end-ipv4-address* **natpt v6bound dynamic acl6 number** *acl-number* **address-group** *address-group* [**no-pat**]	四者必选其一
定义 IPv6 侧报文的动态映射为：如果 IPv6 报文的源 IPv6 地址符合指定的 IPv6 ACL 规则，源 IPv6 地址就会转换为指定接口的 IPv4 地址	**natpt v6bound dynamic acl6 number** *acl-number* **interface** *interface-type interface-number*	
配置 NAT-PT 地址池，定义 IPv6 侧报文的动态映射为：如果 IPv6 报文的目的 IPv6 地址包含 NAT-PT 前缀，源 IPv6 地址就会转换为指定地址池中的 IPv4 地址	**natpt address-group** *group-number start-ipv4-address end-ipv4-address* **natpt v6bound dynamic prefix** *natpt-prefix* **address-group** *address-group* [**no-pat**]	
定义 IPv6 侧报文的动态映射为：如果 IPv6 报文的目的 IPv6 地址包含 NAT-PT 前缀，源 IPv6 地址就会转换为指定接口的 IPv4 地址	**natpt v6bound dynamic prefix** *natpt-prefix* **interface** *interface-type interface-number*	

说明

- 在配置 IPv6 侧报文的动态映射时，命令中 *natpt-prefix* 参数必须通过 **natpt prefix** 命令事先配置。
- 关于 ACL 规则的详细配置请参见“安全分册”中的“ACL 配置”。

1.3.6 配置 NAT-PT 转换后报文的 ToS/Traffic Class 字段值

用户可以设置在 NAT-PT 转换后，报文中的 ToS/Traffic Class 字段值的取值为 0 还是与转换前对应的 Traffic Class/ToS 字段值相同。

表1-7 配置 NAT-PT 转换后报文的 ToS/Traffic Class 字段值

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv4 报文转换为 IPv6 报文后，IPv6 报文中的 Traffic Class 字段值为 0	**natpt turn-off traffic-class**	必选 缺省情况下，当 IPv4 报文转换为 IPv6 报文后，IPv6 报文中的 Traffic Class 字段与转换前的 IPv4 报文的 ToS 字段值相同
配置 IPv6 报文转换为 IPv4 报文后，IPv4 报文中的 ToS 字段值为 0	**natpt turn-off tos**	必选 缺省情况下，当 IPv6 报文转换为 IPv4 报文后，IPv4 报文中的 ToS 字段与转换前的 IPv6 报文的 Traffic Class 字段值相同

1.4 NAT-PT 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 NAT-PT 的运行情况，用户可以通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 NAT-PT 动态地址映射信息及统计信息。

表1-8 NAT-PT 显示和维护

操作	命令
显示所有 NAT-PT 配置信息	**display natpt all**
显示 NAT-PT 地址池配置信息	**display natpt address-group**
显示 NAT-PT 的静态和动态映射信息	**display natpt address-mapping**
显示 NAT-PT 的统计信息	**display natpt statistics**
清除所有的 NAT-PT 统计信息（集中式设备）	**reset natpt statistics**
清除所有的 NAT-PT 统计信息（分布式设备）	**reset natpt statistics** [**slot** *slot-number*]

1.5 NAT-PT 典型配置举例

1.5.1 配置 IPv6 侧动态映射

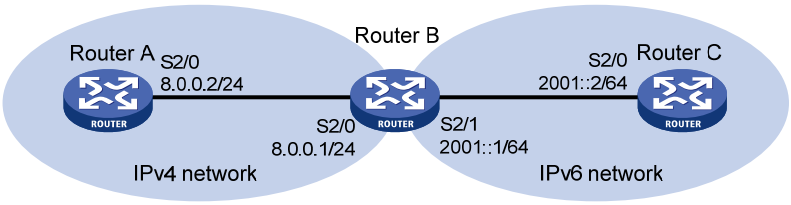
1. 组网需求

- IPv4 网络与 IPv6 网络通过 NAT-PT 设备 Router B 相连；

- Router B 上配置 IPv6 侧报文动态映射，使 IPv6 网络中的主机可以主动访问 IPv4 网络中的主机，而 IPv4 网络中的主机不能主动访问 IPv6 网络中的主机。

2. 组网图

图1-3 配置 IPv6 侧报文动态映射组网图



3. 配置步骤

(1) 配置 IPv4 侧设备 Router A

```
<RouterA> system-view
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 8.0.0.2 255.255.255.0
[RouterA-Serial2/0] quit
```

(2) 配置 IPv6 侧设备 Router C

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] interface serial 2/0
[RouterC-Serial2/0] ipv6 address 2001::2/64
[RouterC-Serial2/0] quit
```

配置缺省路由。

```
[RouterC] ipv6 route-static 3001:: 16 2001::1
```

(3) 配置 Router B

配置接口地址，使能接口 NAT-PT 能力。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address 8.0.0.1 255.255.255.0
[RouterB-Serial2/0] natpt enable
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ipv6 address 2001::1/64
[RouterB-Serial2/1] natpt enable
[RouterB-Serial2/1] quit
```

配置 NAT-PT 前缀。

```
[RouterB] natpt prefix 3001::
```

配置 NAT-PT 地址池。

```
[RouterB] natpt address-group 1 8.0.0.10 8.0.0.19
```

配置 IPv6 侧报文的动态映射。

```
[RouterB] natpt v6bound dynamic prefix 3001:: address-group 1
```

4. 验证配置结果

按以上配置完成后，在 Router C 上执行命令 **ping ipv6** 3001::0800:0002 应该可以收到响应报文。此时可以在设备 Router B 上看到建立的 NAT-PT 会话。

```
[RouterB] display session table verbose
Initiator:
  Source IP/Port : 2001::2/32768
  Dest IP/Port   : 3001::0800:0002/0
  VPN-Instance/VLAN ID/VLL ID:
Responder:
  Source IP/Port : 8.0.0.2/0
  Dest IP/Port   : 8.0.0.19/0
  VPN-Instance/VLAN ID/VLL ID:
Pro: ICMP(58)   App: unknown           State: ICMP-OPEN
Start time: 2008-12-25 10:04:02  TTL: 38s
Received packet(s)(Init): 5 packet(s) 400 byte(s)
Received packet(s)(Reply): 5 packet(s) 300 byte(s)
```

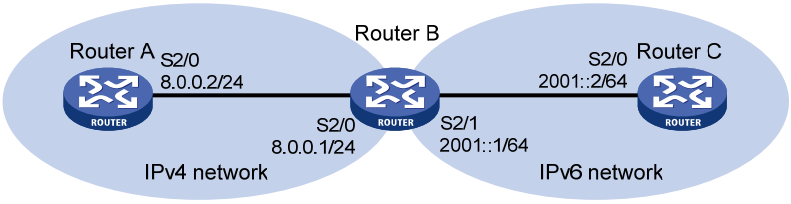
1.5.2 配置 IPv4 侧静态映射和 IPv6 侧静态映射

1. 组网需求

- IPv4 网络与 IPv6 网络通过 NAT-PT 设备 Router B 相连；
- Router B 上配置 IPv4 侧报文静态映射和 IPv6 侧报文静态映射，使 IPv4 网络和 IPv6 网络之间可以互相访问。

2. 组网图

图1-4 NAT-PT 组网图（IPv4/IPv6 侧为静态映射）



3. 配置步骤

(1) 配置 IPv4 侧设备 Router A

```
<RouterA> system-view
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 8.0.0.2 255.255.255.0
[RouterA-Serial2/0] quit
[RouterA] ip route-static 0.0.0.0 0 serial2/0
```

(2) 配置 IPv6 侧设备 Router C

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] interface serial 2/0
```

```
[RouterC-Serial2/0] ipv6 address 2001::2/64
[RouterC-Serial2/0] quit
[RouterC] ipv6 route-static :: 0 serial 2/0
```

(3) 配置 Router B

配置接口地址，使能接口 NAT-PT 能力。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address 8.0.0.1 255.255.255.0
[RouterB-Serial2/0] natpt enable
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ipv6 address 2001::1/64
[RouterB-Serial2/1] natpt enable
[RouterB-Serial2/1] quit
```

配置 NAT-PT 前缀。

```
[RouterB] natpt prefix 3001::
```

配置 IPv4 侧报文的静态映射。

```
[RouterB] natpt v4bound static 8.0.0.2 3001::5
```

配置 IPv6 侧报文的静态映射。

```
[RouterB] natpt v6bound static 2001::2 8.0.0.5
```

4. 验证配置结果

按以上配置完成后，在 Router A 上执行命令 **ping** 8.0.0.5 可以收到响应报文，并且在设备 Router B 上通过 **display** 命令可以看到建立如下 NAT-PT 会话。

```
[RouterB] display session table verbose
Initiator:
  Source IP/Port : 2001::2/32768
  Dest IP/Port   : 3001::0005/0
  VPN-Instance/VLAN ID/VLL ID:
Responder:
  Source IP/Port : 8.0.0.2/0
  Dest IP/Port   : 8.0.0.5/0
  VPN-Instance/VLAN ID/VLL ID:
Pro: ICMP(58)   App: unknown           State: ICMP-OPEN
Start time: 2008-12-25 10:04:02  TTL: 38s
Received packet(s)(Init): 5 packet(s) 400 byte(s)
Received packet(s)(Reply): 5 packet(s) 300 byte(s)
```

在 Router C 上执行命令 **ping ipv6** 3001::5 命令可以收到响应报文，并且在设备 Router B 上通过 **display** 命令可以看到建立如下 NAT-PT 会话。

```
[RouterB] display session table verbose
Initiator:
  Source IP/Port : 2001::2/32768
  Dest IP/Port   : 3001::0005/0
```

```
  VPN-Instance/VLAN ID/VLL ID:
Responder:
  Source IP/Port : 8.0.0.2/0
  Dest IP/Port   : 8.0.0.5/0
  VPN-Instance/VLAN ID/VLL ID:
Pro: ICMP(58)   App: unknown           State: ICMP-OPEN
Start time: 2008-12-25 10:04:02  TTL: 38s
Received packet(s)(Init): 5 packet(s) 400 byte(s)
Received packet(s)(Reply): 5 packet(s) 300 byte(s)
```

1.6 常见配置错误举例

1. 故障现象

NAT-PT 不正常。

2. 故障排除

- 打开 NAT-PT 调试开关。
- 首先根据设备的调试信息定位错误，再参考其他命令进一步判断。在调试过程中，检查转换后的源地址是否正确。如果不正确，问题可能出在地址池配置错误。

目 录

1 IP路由概述 1-1
1.1 IP路由和路由表 1-1
1.1.1 路由 1-1
1.1.2 路由表 1-1
1.2 路由协议概述 1-2
1.2.1 静态路由与动态路由 1-2
1.2.2 动态路由协议分类 1-2
1.2.3 路由协议及路由优先级 1-3
1.2.4 负载分担与路由备份 1-4
1.2.5 路由迭代 1-4
1.2.6 路由信息共享 1-5
1.3 配置全局路由器ID号 1-5
1.4 路由表显示和维护 1-5

1 IP 路由概述

1.1 IP 路由和路由表

1.1.1 路由

在因特网中进行路由选择要使用路由器，路由器根据所收到的报文的目的地址选择一条合适的路由（通过某一网络），并将报文传送到下一个路由器。路径中最后的路由器负责将报文送交目的主机。

1.1.2 路由表

1. 路由表的作用

路由器转发分组的关键是路由表。每个路由器中都保存着一张路由表，表中每条路由项都指明了要到达某子网或某主机的分组应通过路由器的哪个物理接口发送，就可到达该路径的下一个路由器，或者不需再经过别的路由器便可传送到直接相连的网络中的目的主机。

根据来源不同，路由表中的路由通常可分为以下三类：

- 链路层协议发现的路由（也称为接口路由或直连路由）
- 由网络管理员手工配置的静态路由
- 动态路由协议发现的路由

2. 路由表中内容

路由表中包含了下列关键项：

- 目的地址：用来标识 IP 数据报的目的地址或目的网络。
- 网络掩码：与目的地址一起来标识目的主机或路由器所在的网段的地址。将目的地址和网络掩码“逻辑与”后可得到目的主机或路由器所在网段的地址。例如：目的地址为 129.102.8.10、掩码为 255.255.0.0 的主机或路由器所在网段的地址为 129.102.0.0。掩码由若干个连续“1”构成，既可以用点分十进制法表示，也可以用掩码中连续“1”的个数来表示。
- 出接口：指明 IP 报文将从该路由器哪个接口转发。
- 下一跳 IP 地址：更接近目的网络的下一个路由器地址。如果只配置了出接口，下一跳 IP 地址是出接口的地址。
- 本条路由加入 IP 路由表的优先级：对于同一目的地，可能存在若干条不同下一跳的路由，这些不同的路由可能是由不同的路由协议发现的，也可能是手工配置的静态路由。优先级高（数值小）的路由将成为当前的最优路由。

根据路由目的地的不同，可划分为：

- 子网路由：目的地为子网
- 主机路由：目的地为主机

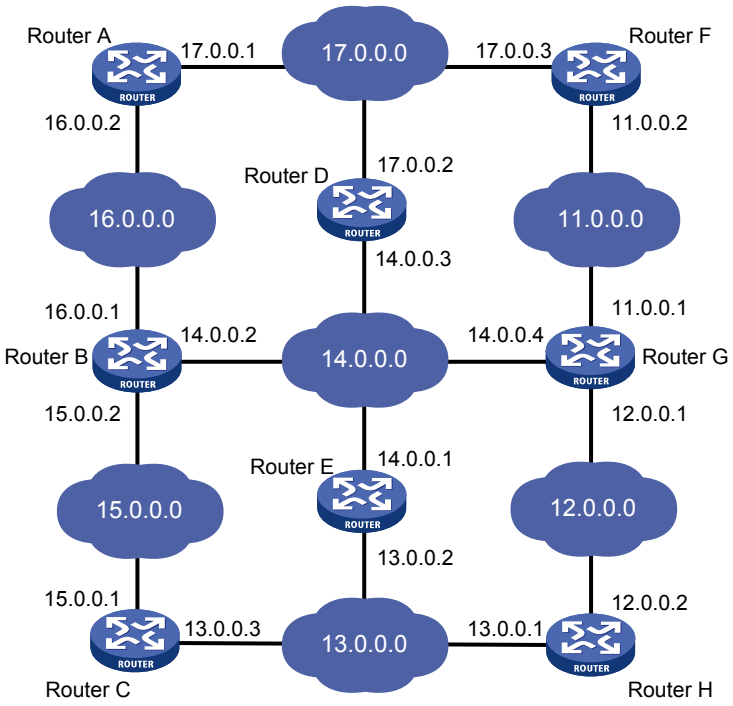
另外，根据目的地与该路由器是否直接相连，又可分为：

- 直接路由：目的地所在网络与路由器直接相连
- 间接路由：目的地所在网络与路由器非直接相连

为了不使路由表过于庞大，可以设置一条缺省路由。凡数据报文查找路由表失败，便根据缺省路由转发。

在 图 1-1的因特网中，各网络中的数字是该网络的网络地址。路由器Router G与三个网络相连，因此有三个IP地址和三个物理接口，其路由表如图所示。

图1-1 路由表示意图



Destiantion Network	Nexthop	Interface
11.0.0.0	11.0.0.1	2
12.0.0.0	12.0.0.1	1
13.0.0.0	12.0.0.2	1
14.0.0.0	14.0.0.4	3
15.0.0.0	14.0.0.2	3
16.0.0.0	14.0.0.2	3
17.0.0.0	11.0.0.2	2

1.2 路由协议概述

1.2.1 静态路由与动态路由

静态路由配置方便，对系统要求低，适用于拓扑结构简单并且稳定的小型网络。其缺点是每当网络拓扑结构发生变化，都需要人工重新配置，不能自动适应。

动态路由协议有自己的路由算法，能够自动适应网络拓扑的变化，适用于具有一定规模的网络拓扑。其缺点是配置比较复杂，对系统的要求高于静态路由，并将占用一定的网络资源。

1.2.2 动态路由协议分类

对动态路由协议的分类可采用以下不同标准：

1. 根据作用范围

根据作用的范围，路由协议可分为：

- 内部网关协议（Interior Gateway Protocol，简称 IGP）：在一个自治系统内部运行，常见的 IGP 协议包括 RIP、OSPF 和 IS-IS。

- 外部网关协议（Exterior Gateway Protocol，简称 EGP）：运行于不同自治系统之间，BGP 是目前最常用的 EGP。

说明

自治系统（Autonomous System）是拥有同一选路策略，并在同一技术管理部门下运行的一组路由器。

2. 根据使用算法

根据使用的算法，路由协议可分为：

- 距离矢量协议（Distance-Vector）：包括 RIP 和 BGP。其中，BGP 也被称为路径矢量协议（Path-Vector）。
- 链路状态协议（Link-State）：包括 OSPF 和 IS-IS。

以上两种算法的主要区别在于发现和计算路由的方法不同。

3. 根据目的地址类型

根据目的地址的类型，路由协议可分成：

- 单播路由协议（Unicast Routing Protocol）：包括 RIP、OSPF、BGP 和 IS-IS 等。
- 组播路由协议（Multicast Routing Protocol）：包括 PIM-SM、PIM-DM 等。

本部分手册主要介绍单播路由协议，组播路由协议请参见“IP 组播分册”。

4. 根据 IP 协议版本

根据 IP 协议的版本，路由协议可分成：

- IPv4 路由协议：包括 RIP、OSPFv2、BGP4 和 IS-IS 等。
- IPv6 路由协议：包括 RIPng、OSPFv3、BGP4+和支持 IPv6 的 IS-IS 等。

1.2.3 路由协议及路由优先级

对于相同的目的地，不同的路由协议（包括静态路由）可能会发现不同的路由，但这些路由并不都是最优的。事实上，在某一时刻，到某一目的地的当前路由仅能由唯一的路由协议来决定。为了判断最优路由，各路由协议（包括静态路由）都被赋予了一个优先级，当存在多个路由信息源时，具有较高优先级的路由协议发现的路由将成为当前路由。各种路由协议及其发现路由的缺省优先级如表 1-1所示。

其中：0 表示直接连接的路由，256 表示任何来自不可信源端的路由。数值越小表明优先级越高。

表1-1 路由协议及缺省时的路由优先级

路由协议或路由种类	相应路由的优先级
DIRECT	0
OSPF	10
IS-IS	15
STATIC	60
RIP	100

路由协议或路由种类	相应路由的优先级
OSPF ASE	150
OSPF NSSA	150
IBGP	255
EBGP	255
UNKNOWN	256

除直连路由（DIRECT）外，各种路由的优先级都可由用户手工进行配置。另外，每条静态路由的优先级都可以不相同。

说明

IPv4 路由和 IPv6 路由有独立的路由表，两者互不影响。

1.2.4 负载分担与路由备份

1. 负载分担

多路由模式是指对同一路由协议来说，允许配置多条目的地相同且开销也相同的路由。当到同一目的地的路由中，没有更高优先级的路由时，这几条路由都被采纳，在转发去往该目的地的报文时，依次通过各条路径发送，从而实现网络的负载分担。对于同一目的地，特定的路由协议也可能会发现几条等值的路由，如果该路由协议在所有活跃的路由协议中优先级最高，那么这几条不同的路由都被看作当前有效的路由。这样，在路由协议层面上，保证了 IP 流量的负载分担。

说明

- 系统所允许的负载分担的具体路由条数，请以设备的实际情况为准。
- 目前实现支持负载分担的路由协议为静态路由、RIP、OSPF、BGP 和 IS-IS。

2. 路由备份

使用路由备份可以提高网络的可靠性。用户可根据实际情况，配置到同一目的地的多条路由，其中优先级最高的一条路由作为主路由，其余优先级较低的路由作为备份路由。

正常情况下，路由器采用主路由转发数据。当线路出现故障时，该路由变为非激活状态，路由器选择备份路由中优先级最高的转发数据。这样，也就实现了从主路由到备份路由的切换。当主路由恢复正常时，路由器也恢复相应的路由，并重新选择路由。由于该路由的优先级最高，路由器选择主路由来发送数据。这就是从备份路由到主路由的切换。

1.2.5 路由迭代

对于 BGP 路由（直连 EBGP 路由除外）和静态路由（配置了下一跳）以及多跳 RIP 路由而言，其所携带的下一跳信息可能并不是直接可达，从指导转发的角度而言，它需要找到到达下一跳的直连出接口。路由迭代的过程就是通过路由的下一跳信息来找到直连出接口的过程。而对于 OSPF 和 IS-IS 等链路状态路由协议而言，其下一跳是直接在路由计算的时就得到，因此，不需要进行路由迭代。

1.2.6 路由信息共享

由于各路由协议采用的路由算法不同，不同的路由协议可能会发现不同的路由。如果网络规模较大，当使用多种路由协议时，往往需要在不同的路由协议间能够共享各自发现的路由。每种路由协议都有相应的路由引入机制，具体内容请参见“IP 路由分册”中各路由协议模块有关引入外部路由的描述。

1.3 配置全局路由器 ID 号

一些动态路由协议要求使用 Router ID，如果在启动这些路由协议时没有显示指定 Router ID，则缺省使用路由管理的 Router ID。

表1-2 配置全局路由器 ID 号

配置任务	说明	详细配置
进入系统视图	**system-view**	-
配置全局路由器 ID	**router id** *router-id*	可选 缺省情况下，未配置路由管理的 Router ID

1.4 路由表显示和维护

查看路由表的信息是定位路由问题的基本方法。在任意视图下执行 **display** 命令可以显示路由表信息。在用户视图下执行 **reset** 命令可以清除路由表的统计信息。

表1-3 路由表显示和维护

操作	命令
查看路由表中当前激活路由的摘要信息	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] [**verbose** | | { **begin | exclude | include** } *regular-expression*]
查看指定目的地址的路由	**display ip routing-table** *ip-address* [*mask-length* | *mask*] [**longer-match**] [**verbose**]
查看指定目的地址范围内的路由	**display ip routing-table** *ip-address1* { *mask-length* | *mask* } *ip-address2* { *mask-length* | *mask* } [**verbose**]
查看通过指定基本访问控制列表过滤的路由	**display ip routing-table acl** *acl-number* [**verbose**]
查看通过指定前缀列表过滤的路由	**display ip routing-table ip-prefix** *ip-prefix-name* [**verbose**]
查看指定协议发现的路由	**display ip routing-table protocol** *protocol* [**inactive** | **verbose**]
查看全局路由管理 ID	**display router id**
查看路由表或 VPN 路由表中的综合路由统计信息	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **statistics**
清除路由表或 VPN 路由表中的综合路由统计信息	**reset ip routing-table statistics protocol** [**vpn-instance** *vpn-instance-name*] { **all** | *protocol* }
查看 IPv6 路由表摘要信息	**display ipv6 routing-table**
查看 IPv6 路由表详细信息	**display ipv6 routing-table verbose**
查看指定 IPv6 路由的信息	**display ipv6 routing-table** *ipv6-address prefix-length* [**longer-match**] [**verbose**]

操作	命令
查看经过指定的基本 IPv6 ACL（访问控制列表）过滤的路由	**display ipv6 routing-table acl** *acl6-number* [**verbose**]
查看经过指定 IPv6 前缀列表过滤的路由信息	**display ipv6 routing-table ipv6-prefix** *ipv6-prefix-name* [**verbose**]
查看指定协议发现的 IPv6 路由信息	**display ipv6 routing-table protocol** *protocol* [**inactive** | **verbose**]
查看 IPv6 路由统计信息	**display ipv6 routing-table statistics**
查看在指定地址范围内的 IPv6 路由信息	**display ipv6 routing-table** *ipv6-address1 prefix-length1 ipv6-address2 prefix-length2* [**verbose**]
清除 IPv6 路由表中的统计信息	**reset ipv6 routing-table statistics protocol** { **all** | *protocol* }

目 录

1 BGP配置 1-1
1.1 BGP简介 1-1
1.1.1 BGP概述 1-1
1.1.2 BGP的消息类型 1-2
1.1.3 BGP的路由属性 1-4
1.1.4 BGP的选路规则 1-8
1.1.5 IBGP和IGP同步 1-10
1.1.6 大规模BGP网络所遇到的问题 1-10
1.1.7 BGP GR 1-14
1.1.8 MP-BGP 1-14
1.1.9 协议规范 1-15
1.2 BGP配置任务简介 1-15
1.3 配置BGP连接 1-16
1.3.1 配置准备 1-16
1.3.2 创建BGP连接 1-16
1.3.3 配置建立TCP连接使用的源接口 1-17
1.3.4 配置允许同非直连邻居建立EBGP连接 1-18
1.4 控制路由信息的生成 1-18
1.4.1 配置准备 1-18
1.4.2 配置BGP发布本地路由 1-19
1.4.3 配置BGP引入其它路由 1-19
1.4.4 配置BGP引入其它协议缺省路由 1-19
1.5 控制路由信息的发布与接收 1-20
1.5.1 配置准备 1-20
1.5.2 配置BGP路由聚合 1-20
1.5.3 配置向对等体/对等体组发送缺省路由 1-20
1.5.4 配置BGP路由信息的发布/接收策略 1-21
1.5.5 配置BGP与IGP路由同步 1-22
1.5.6 限制从BGP对等体/对等体组接收的路由前缀数量 1-23
1.5.7 配置BGP路由衰减 1-23
1.5.8 配置shortcut路由 1-23
1.6 配置BGP的路由属性 1-24
1.6.1 配置准备 1-24
1.6.2 为接收路由分配首选值 1-24
1.6.3 配置BGP的路由优先级 1-24
1.6.4 配置Local-Preference缺省值 1-24
1.6.5 配置MED属性 1-25
1.6.6 配置下一跳属性 1-27
1.6.7 配置AS_PATH属性 1-28
1.7 调整和优化BGP网络 1-30
1.7.1 配置准备 1-30
1.7.2 配置BGP存活时间间隔与保持时间 1-30

1.7.3 配置发送路由更新报文的时间间隔 1-31
1.7.4 配置BGP软复位 1-31
1.7.5 使能EBGP连接快速复位功能 1-32
1.7.6 配置BGP建立TCP连接时进行MD5 认证 1-32
1.7.7 配置BGP负载分担 1-33
1.7.8 禁止与对等体/对等体组建立会话 1-33
1.8 配置大型BGP网络 1-33
1.8.1 配置准备 1-33
1.8.2 配置BGP对等体组 1-33
1.8.3 配置BGP团体 1-35
1.8.4 配置BGP路由反射器 1-35
1.8.5 配置BGP联盟 1-36
1.9 配置BGP GR 1-37
1.10 开启Trap功能 1-38
1.11 使能BGP日志功能 1-38
1.12 BGP显示和维护 1-38
1.12.1 显示BGP 1-38
1.12.2 复位BGP连接 1-39
1.12.3 清除BGP信息 1-40
1.13 BGP典型配置举例 1-40
1.13.1 BGP基本配置 1-40
1.13.2 BGP与IGP交互配置 1-44
1.13.3 BGP负载分担配置 1-46
1.13.4 BGP团体配置 1-48
1.13.5 BGP路由反射器配置 1-50
1.13.6 BGP联盟配置 1-52
1.13.7 BGP路径选择配置 1-56
1.14 BGP常见错误配置举例 1-59
1.14.1 BGP对等体关系不能建立 1-59

1 BGP 配置

1.1 BGP 简介

1.1.1 BGP 概述

BGP（Border Gateway Protocol，边界网关协议）是一种用于 AS（Autonomous System，自治系统）之间的动态路由协议。AS 是拥有同一选路策略，在同一技术管理部门下运行的一组路由器。

早期发布的三个版本分别是 BGP-1（RFC 1105）、BGP-2（RFC 1163）和 BGP-3（RFC 1267），当前使用的版本是 BGP-4（RFC 1771，已更新至 RFC 4271）。BGP-4 做为事实上的 Internet 外部路由协议标准，被广泛应用于 ISP（Internet Service Provider，因特网服务提供商）之间。

说明

下文中若不做特殊说明，所指的 BGP 均为 BGP-4。

BGP 特性描述如下：

- BGP 是一种外部网关协议（Exterior Gateway Protocol，EGP），与 OSPF、RIP 等内部网关协议（Interior Gateway Protocol，IGP）不同，其着眼点不在于发现和计算路由，而在于控制路由的传播和选择最佳路由。
- BGP 使用 TCP 作为其传输层协议（端口号 179），提高了协议的可靠性。
- BGP 支持 CIDR（Classless Inter-Domain Routing，无类别域间路由）。
- 路由更新时，BGP 只发送更新的路由，大大减少了 BGP 传播路由所占用的带宽，适用于在 Internet 上传播大量的路由信息。
- BGP 路由通过携带 AS 路径信息彻底解决路由环路问题。
- BGP 提供了丰富的路由策略，能够对路由实现灵活的过滤和选择。
- BGP 易于扩展，能够适应网络新的发展。

发送 BGP 消息的路由器称为 BGP 发言者（BGP Speaker），它接收或产生新的路由信息，并发布（Advertise）给其它 BGP 发言者。当 BGP 发言者收到来自其它自治系统的新路由时，如果该路由比当前已知路由更优、或者当前还没有该路由，它就把这条路由发布给自治系统内所有其它 BGP 发言者。

相互交换消息的 BGP 发言者之间互称对等体（Peer），若干相关的对等体可以构成对等体组（Peer group）。

BGP 在路由器上以下列两种方式运行：

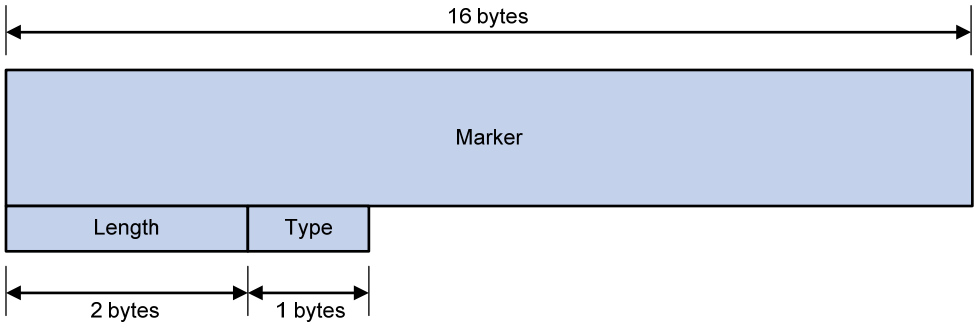
- IBGP（Internal BGP）：当 BGP 运行于同一自治系统内部时，被称为 IBGP；
- EBGP（External BGP）：当 BGP 运行于不同自治系统之间时，称为 EBGP。

1.1.2 BGP 的消息类型

1. 消息头格式

BGP有 5 种消息类型：Open、Update、Notification、Keepalive和Route-refresh。这些消息有相同的报文头，其格式如 图 1-1所示。

图1-1 BGP 消息的报文头格式



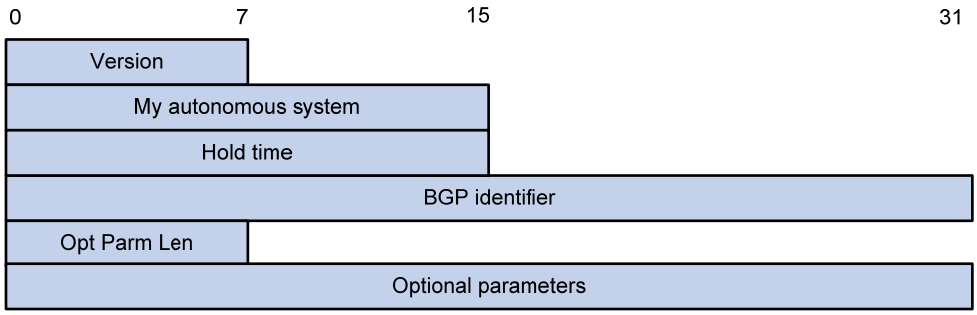
主要字段的解释如下：

- Marker：16 字节，用于标明 BGP 报文边界，所有比特均为“1”。
- Length：2 字节，BGP 消息总长度（包括报文头在内），以字节为单位。
- Type：1 字节，BGP 消息的类型。其取值从 1 到 5，分别表示 Open、Update、Notification、Keepalive 和 Route-refresh 消息。其中，前四种消息是在 RFC 1771 中定义，而 Type 为 5 的消息则是在 RFC 2918 中定义的。

2. Open

Open消息是TCP连接建立后发送的第一个消息，用于建立BGP对等体之间的连接关系。其消息格式如 图 1-2所示。

图1-2 BGP Open 消息格式



主要字段的解释如下：

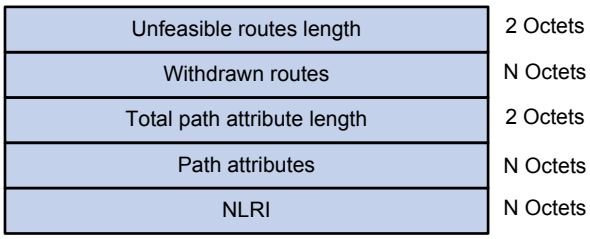
- Version：BGP 的版本号。对于 BGP-4 来说，其值为 4。
- My autonomous system：本地 AS 号。通过比较两端的 AS 号可以确定是 EBGP 连接还是 IBGP 连接。
- Hold time：保持时间。在建立对等体关系时两端要协商 Hold Time，并保持一致。如果在这个时间内未收到对端发来的 Keepalive 消息或 Update 消息，则认为 BGP 连接中断。
- BGP identifier：BGP 标识符。以 IP 地址的形式表示，用来识别 BGP 路由器。

- Opt Parm Len（Optional Parameters Length）：可选参数的长度。如果为 0 则没有可选参数。
- Optional parameters：可选参数。用于多协议扩展（Multiprotocol Extensions）等功能。

3. Update

Update消息用于在对等体之间交换路由信息。它既可以发布可达路由信息，也可以撤销不可达路由信息。其消息格式如 图 1-3所示。

图1-3 BGP Update 消息格式



一条 Update 报文可以通告一类具有相同路径属性的可达路由，这些路由放在 NLRI（Network Layer Reachable Information，网络层可达信息）字段中，Path Attributes 字段携带了这些路由的属性，BGP 根据这些属性进行路由的选择；同时 Update 报文还可以携带多条不可达路由，被撤销的路由放在 Withdrawn Routes 字段中。

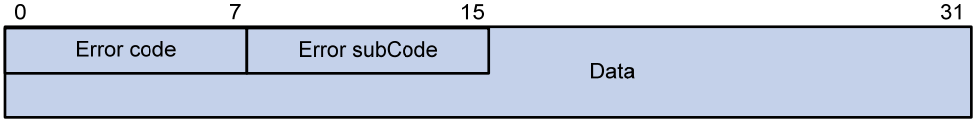
主要字段的解释如下：

- Unfeasible routes length：不可达路由字段的长度，以字节为单位。如果为 0 则说明没有 Withdrawn Routes 字段。
- Withdrawn routes：不可达路由的列表。
- Total path attribute length：路径属性字段的长度，以字节为单位。如果为 0 则说明没有 Path Attributes 字段。
- Path attributes：与 NLRI 相关的所有路径属性列表，每个路径属性由一个 TLV（Type-Length-Value）三元组构成。BGP 正是根据这些属性值来避免环路，进行选路，协议扩展等。
- NLRI（Network Layer Reachability Information）：可达路由的前缀和前缀长度二元组。

4. Notification

当BGP检测到错误状态时，就向对等体发出Notification消息，之后BGP连接会立即中断。其消息格式如 图 1-4所示。

图1-4 BGP Notification 消息格式



主要字段的解释如下：

- Error code：差错码，指定错误类型。
- Error subcode：差错子码，错误类型的详细信息。

- Data：用于辅助发现错误的原因，它的内容依赖于具体的差错码和差错子码，记录的是出错部分的数据，长度不固定。

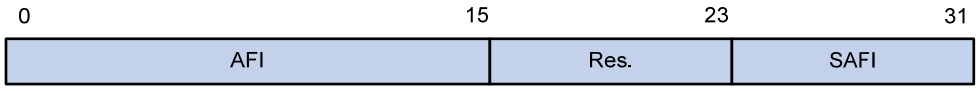
5. Keepalive

BGP 会周期性地向对等体发出 Keepalive 消息，用来保持连接的有效性。其消息格式中只包含报文头，没有附加其他任何字段。

6. Route-refresh

Route-refresh消息用来要求对等体重新发送指定地址族的路由信息。其消息格式如 图 1-5所示。

图1-5 BGP Route-refresh 消息格式



主要的字段解释如下：

- AFI：Address Family Identifier，地址族标识。
- Res.：保留，必须置 0。
- SAFI：Subsequent Address Family Identifier，子地址族标识。

1.1.3 BGP 的路由属性

1. 路由属性的分类

BGP 路由属性是一组参数，它对特定的路由进行了进一步的描述，使得 BGP 能够对路由进行过滤和选择。

事实上，所有的 BGP 路由属性都可以分为以下四类：

- 公认必须遵循（Well-known mandatory）：所有 BGP 路由器都必须能够识别这种属性，且必须存在于 Update 消息中。如果缺少这种属性，路由信息就会出错。
- 公认可选（Well-known discretionary）：所有 BGP 路由器都可以识别，但不要求必须存在于 Update 消息中，可以根据具体情况来选择。
- 可选过渡（Optional transitive）：在 AS 之间具有可传递性的属性。BGP 路由器可以不支持此属性，但它仍然会接收带有此属性的路由，并通告给其他对等体。
- 可选非过渡（Optional non-transitive）：如果 BGP 路由器不支持此属性，该属性被忽略，且不会通告给其他对等体。

BGP路由几种基本属性和对应的类别如 表 1-1所示。

表1-1 路由属性和类别

属性名称	类别
ORIGIN	公认必须遵循
AS_PATH	公认必须遵循
NEXT_HOP	公认必须遵循
LOCAL_PREF	公认可选
ATOMIC_AGGREGATE	公认可选

属性名称	类别
AGGREGATOR	可选过渡
COMMUNITY	可选过渡
MULTI_EXIT_DISC (MED)	可选非过渡
ORIGINATOR_ID	可选非过渡
CLUSTER_LIST	可选非过渡

2. 几种主要的路由属性

(1) 源（ORIGIN）属性

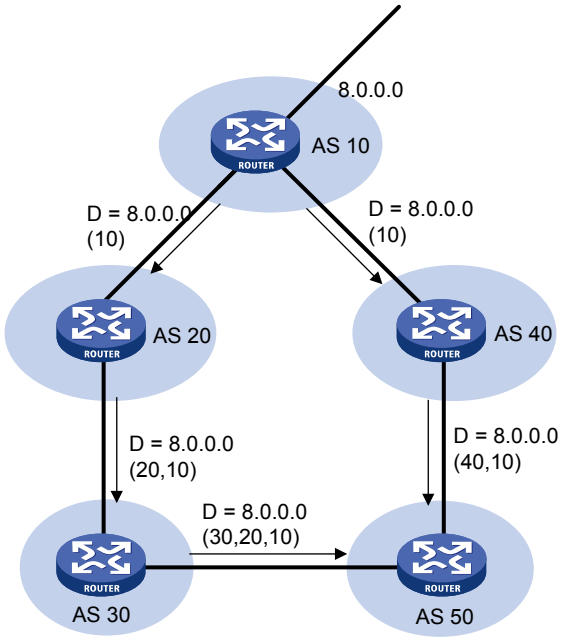
ORIGIN 属性定义路由信息的来源，标记一条路由是怎么成为 BGP 路由的。它有以下三种类型：

- IGP：优先级最高，说明路由产生于本 AS 内。
- EGP：优先级次之，说明路由通过 EGP 学到。
- incomplete：优先级最低，它并不是说明路由不可达，而是表示路由的来源无法确定。例如，引入的其它路由协议的路由信息。

(2) AS 路径（AS_PATH）属性

AS_PATH属性按一定次序记录了某条路由从本地到目的地址所要经过的所有AS号。当BGP将一条路由通告到其他AS时，便会把本地AS号添加在AS_PATH列表的最前面。收到此路由的BGP路由器根据AS_PATH属性就可以知道去目的地址所要经过的AS。离本地AS最近的相邻AS号排在前面，其他AS号按顺序依次排列。如 图 1-6所示。

图1-6 AS_PATH 属性



通常情况下，BGP 不会接受 AS_PATH 中已包含本地 AS 号的路由，从而避免了形成路由环路的可能。

说明

在目前的实现中，根据某些特殊的具体应用需求，可以通过配置 **peer allow-as-loop** 命令允许 AS 号重复。

同时，AS_PATH 属性也可用于路由的选择和过滤。在其他因素相同的情况下，BGP 会优先选择路径较短的路由。比如在上图中，AS 50 中的 BGP 路由器会选择经过 AS 40 的路径作为到目的地址 8.0.0.0 的最优路由。

在某些应用中，可以使用路由策略来人为的增加 AS 路径的长度，以便更为灵活地控制 BGP 路径的选择。

通过配置 AS 路径过滤列表，还可以针对 AS_PATH 属性中所包含的 AS 号来对路由进行过滤。

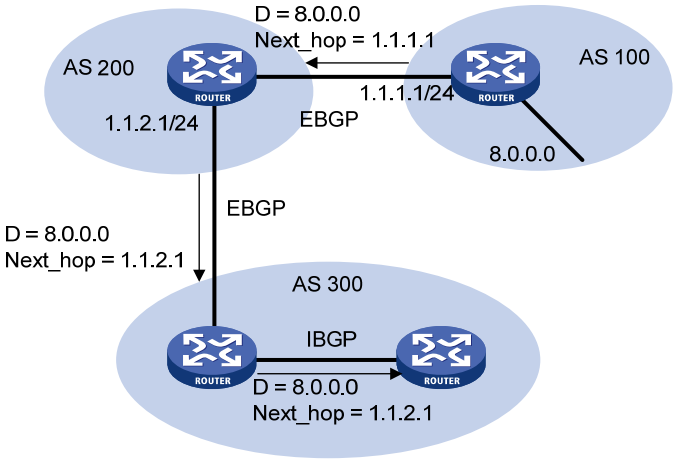
(3) 下一跳（NEXT_HOP）属性

BGP 的下一跳属性和 IGP 的有所不同，不一定就是邻居路由器的 IP 地址。

下一跳属性取值情况分为三种，如 图 1-7所示。

- BGP 发言者把自己产生的路由发给所有邻居时，将把该路由信息的下一跳属性设置为自己与对端连接的接口地址；
- BGP 发言者把接收到的路由发送给 EBGP 对等体时，将把该路由信息的下一跳属性设置为本地与对端连接的接口地址；
- BGP发言者把从EBGP邻居得到的路由发给IBGP邻居时，并不改变该路由信息的下一跳属性。如果配置了负载分担，路由被发给IBGP邻居时则会修改下一跳属性。关于“负载分担”的概念请参见“1.1.4 BGP的选路规则”。

图1-7 下一跳属性



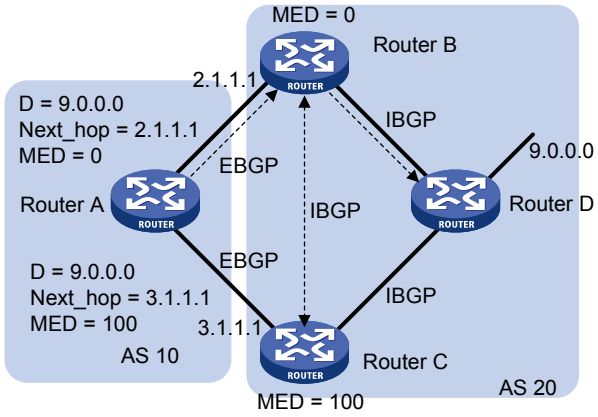
(4) MED（MULTI_EXIT_DISC）

MED 属性仅在相邻两个 AS 之间交换，收到此属性的 AS 一方不会再将其通告给任何其他第三方 AS。

MED属性相当于IGP使用的度量值（metrics），它用于判断流量进入AS时的最佳路由。当一个运行BGP的路由器通过不同的EBGP对等体得到目的地址相同但下一跳不同的多条路由时，在其它条

件相同的情况下，将优先选择MED值较小者作为最佳路由。如 图 1-8所示，从AS 10 到AS 20 的流量将选择Router B作为入口。

图1-8 MED 属性



通常情况下，BGP 只比较来自同一个 AS 的路由的 MED 属性值。

说明

在目前的实现中，通过配置 **compare-different-as-med** 命令，可以强制 BGP 比较来自不同 AS 的路由的 MED 属性值。

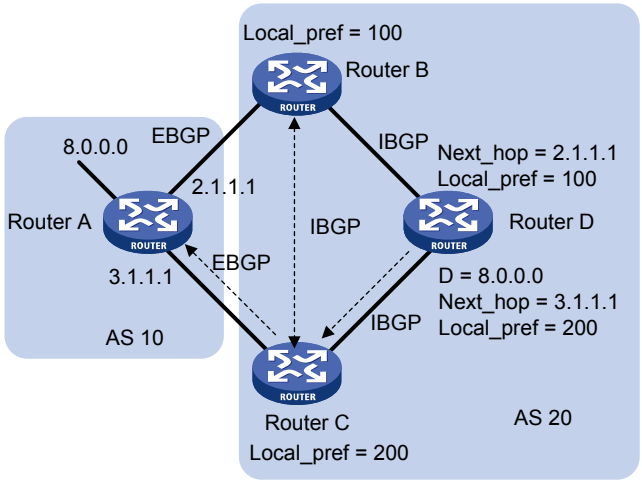
---

(5) 本地优先（LOCAL_PREF）属性

LOCAL_PREF 属性仅在 IBGP 对等体之间交换，不通告给其他 AS。它表明 BGP 路由器的优先级。

LOCAL_PREF属性用于判断流量离开AS时的最佳路由。当BGP的路由器通过不同的IBGP对等体得到目的地址相同但下一跳不同的多条路由时，将优先选择LOCAL_PREF属性值较高的路由。如 图 1-9所示，从AS 20 到AS 10 的流量将选择Router C作为出口。

图1-9 LOCAL_PREF 属性



(6) 团体（COMMUNITY）属性

团体属性用来简化路由策略的应用和降低维护管理的难度。它是一组有 相同特征的目的地址的集合，没有物理上的边界，与其所在的 AS 无关。公认的团体属性有：

- INTERNET：缺省情况下，所有的路由都属于 INTERNET 团体。具有此属性的路由可以被通告给所有的 BGP 对等体。
- NO_EXPORT：具有此属性的路由在收到后，不能被发布到本地AS之外。如果使用了联盟，则不能被发布到联盟之外，但可以发布给联盟中的其他子AS（关于联盟的定义请参见“1.1.6 大规模BGP网络所遇到的问题”）。
- NO_ADVERTISE：具有此属性的路由被接收后，不能被通告给任何其他的 BGP 对等体。
- NO_EXPORT_SUBCONFED：具有此属性的路由被接收后，不能被发布到本地 AS 之外，也不能发布到联盟中的其他子 AS。

1.1.4 BGP 的选路规则

1. BGP 选择路由的策略

在目前的实现中，BGP 选择路由时采取如下策略：

- 首先丢弃下一跳（NEXT_HOP）不可达的路由；
- 优选 Preferred-value 值最大的路由；
- 优选本地优先级（LOCAL_PREF）最高的路由；
- 优选聚合路由；
- 优选 AS 路径（AS_PATH）最短的路由；
- 依次选择 ORIGIN 类型为 IGP、EGP、Incomplete 的路由；
- 优选 MED 值最低的路由；
- 依次选择从 EBGP、联盟、IBGP 学来的路由；
- 优选下一跳 Cost 值最低的路由；
- 优选 CLUSTER_LIST 长度最短的路由；
- 优选 ORIGINATOR_ID 最小的路由；
- 优选 Router ID 最小的路由器发布的路由；
- 优选地址最小的对等体发布的路由。



说明

- CLUSTER_ID 为路由反射器的集群 ID，CLUSTER_LIST 由 CLUSTER_ID 序列组成，反射器将自己的 CLUSTER_ID 加入 CLUSTER_LIST 中，若反射器收到路由中 CLUSTER_LIST 中包含有自己的 CLUSTER_ID，则丢弃该路由，从而避免群内环路的发生。
- 如果配置了负载分担，并且有多条到达同一目的地的路由，则根据配置的路由条数选择多条路由进行负载分担。

---

2. 应用 BGP 负载分担时的选路

在 BGP 中，由于协议本身的特殊性，它产生的路由的下一跳地址可能不是当前路由器直接相连的邻居。常见的一个原因是：IBGP 之间发布路由信息时不改变下一跳。这种情况下，为了能够将报文正确转发出去，路由器必须先找到一个直接可达的地址（查找 IGP 建立的路由表项），通过这个地址到达路由表中指示的下一跳。在上述过程中，去往直接可达地址的路由被称为依赖路由，BGP 路由依赖于这些路由指导报文转发。根据下一跳地址找到依赖路由的过程就是路由迭代（recursion）。

目前系统支持基于迭代的 BGP 负载分担，即如果依赖路由本身是负载分担的（假设有三个下一跳地址），则 BGP 也会生成相同数量的下一跳地址来指导报文转发。需要说明的是，基于迭代的 BGP 负载分担并不需要命令配置，这一特性在系统上始终启用。

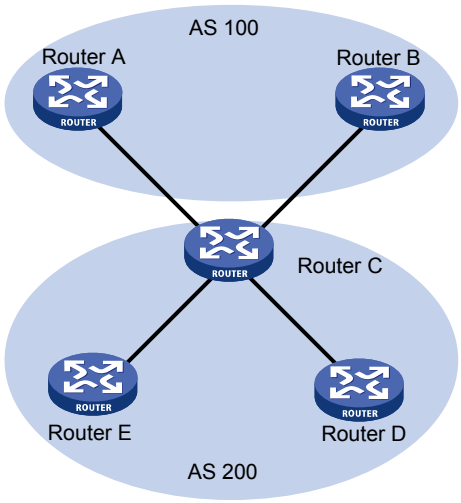
在实现方法上，BGP 的负载分担与 IGP 的负载分担有所不同：

- IGP 是通过协议定义的路由算法，对到达同一目的地址的不同路由，根据计算结果，将度量值（metric）相等的（如 RIP、OSPF）路由进行负载分担，选择的标准很明确（按 metric）。
- BGP 本身并没有路由计算的算法，它只是一个选路的路由协议，因此，不能根据一个明确的度量值决定是否对路由进行负载分担，但 BGP 有丰富的选路规则，可以在对路由进行一定的选择后，有条件地进行负载分担，也就是将负载分担加入到 BGP 的选路规则中去。

说明

- BGP 只对 AS_PATH 属性、ORIGIN 属性、LOCAL_PREF 和 MED 值完全相同的路由进行负载分担。
- BGP 负载分担特性适用于 EBGP、IBGP 以及联盟之间。
- 如果有多条到达同一目的地的路由，则根据配置的路由条数选择多条路由进行负载分担。

图1-10 BGP 负载分担示意图



在 图 1-10中，Router D和Router E是Router C的IBGP对等体。当Router A和Router B同时向Router C通告到达同一目的地的路由时，如果用户在Router C配置了负载分担（如balance 2），则当满足一定的选路规则后，并且两条路由具有相同的AS_PATH属性、ORIGIN属性、LOCAL_PREF和MED值时，Router C就把接收的两条路由同时加入到转发表中，实现BGP路由的负载分担。Router C只向Router D和Router E转发一次该路由，AS_PATH不变，但NEXT_HOP属性改变为Router C的地址，而不是原来的EBGP对等体地址。其它的BGP过渡属性将按最佳路由的属性传递。

3. BGP 发布路由的策略

在目前的实现中，BGP 发布路由时采用如下策略：

- 存在多条有效路由时，BGP 发言者只将最优路由发布给对等体；
- BGP 发言者只把自己使用的路由发布给对等体；

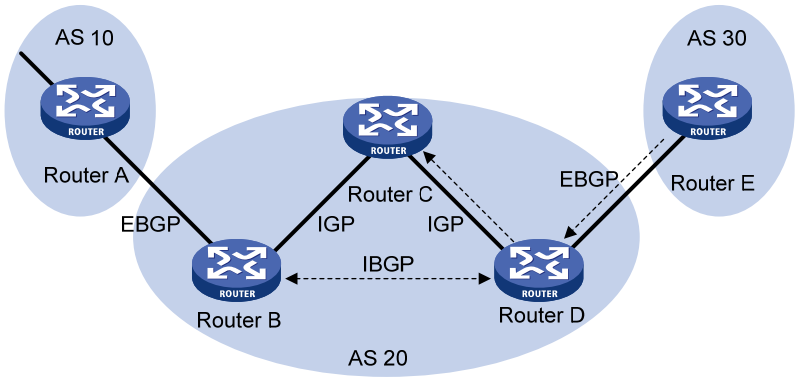
- BGP 发言者从 EBGP 获得的路由会向它所有 BGP 对等体发布（包括 EBGP 对等体和 IBGP 对等体）；
- BGP 发言者从 IBGP 获得的路由不向它的 IBGP 对等体发布；
- BGP 发言者从 IBGP 获得的路由发布给它的 EBGP 对等体（关闭 BGP 与 IGP 同步的情况下，IBGP 路由被直接发布；开启 BGP 与 IGP 同步的情况下，该 IBGP 路由只有在 IGP 也发布了这条路由时才会被同步并发布给 EBGP 对等体）；
- 连接一旦建立，BGP 发言者将把自己所有的 BGP 路由发布给新对等体。

1.1.5 IBGP 和 IGP 同步

同步是指 IBGP 和 IGP 之间的同步，其目的是为了避免出现误导外部 AS 路由器的现象发生。

如果一个AS中有非BGP路由器提供转发服务，经该AS转发的IP报文将可能因为目的地址不可达而被丢弃。如 图 1-11所示，Router E通过BGP从Router D可以学到Router A的一条路由 8.0.0.0/8，于是将到这个目的地址的报文转发给Router D，Router D查询路由表，发现下一跳是Router B（通过**peer next-hop-local**命令手动设置）。由于Router D从IGP学到了到Router B的路由，所以通过路由迭代，Router D将报文转发给Router C。但Router C并不知道去 8.0.0.0/8 的路由，于是将报文丢弃。

图1-11 IBGP 和 IGP 同步



如果设置了同步特性，在 IBGP 路由加入路由表并发布给 EBGP 对等体之前，会先检查 IGP 路由表。只有在 IGP 也知道这条 IBGP 路由时，它才会被加入到路由表，并发布给 EBGP 对等体。

在下面的情况中，可以关闭同步特性。

- 本 AS 不是过渡 AS（上图中的 AS20 就属于一个过渡 AS）
- 本 AS 内所有路由器建立 IBGP 全连接

1.1.6 大规模 BGP 网络所遇到的问题

1. 路由聚合

在大规模的网络中，BGP 路由表十分庞大，使用路由聚合（Routes Aggregation）可以大大减小路由表的规模。

路由聚合实际上是将多条路由合并的过程。这样 BGP 在向对等体通告路由时，可以只通告聚合后的路由，而不是将所有的具体路由都通告出去。

目前系统支持自动聚合和手动聚合方式。使用后者还可以控制聚合路由的属性，以及决定是否发布具体路由。

2. BGP 路由衰减

BGP 路由衰减（Route Dampening）用来解决路由不稳定的问题。路由不稳定的主要表现形式是路由振荡（Route flaps），即路由表中的某条路由反复消失和重现。

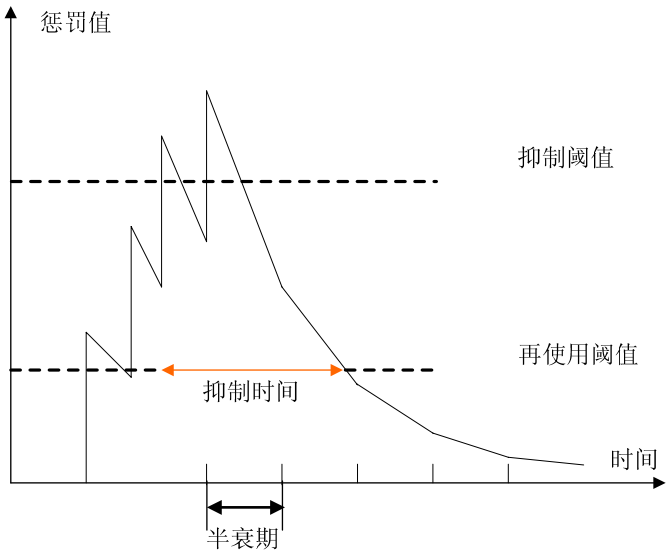
发生路由振荡时，路由协议就会向邻居发布路由更新，收到更新报文的路由器需要重新计算路由并修改路由表。所以频繁的路由振荡会消耗大量的带宽资源和 CPU 资源，严重时会影响到网络的正常工作。

在多数情况下，BGP 协议都应用于复杂的网络环境中，路由变化十分频繁。为了防止持续的路由振荡带来的不利影响，BGP 使用衰减来抑制不稳定的路由。

BGP 衰减使用惩罚值来衡量一条路由的稳定性，惩罚值越高则说明路由越不稳定。路由每发生一次振荡（路由从激活状态变为未激活状态，称为一次路由振荡），BGP 便会给此路由增加一定的惩罚值（1000，此数值为系统固定，不可修改）。当惩罚值超过抑制阈值时，此路由被抑制，不加入到路由表中，也不再向其他 BGP 对等体发布更新报文。

被抑制的路由每经过一段时间，惩罚值便会减少一半，这个时间称为半衰期（Half-life）。当惩罚值降到再使用阈值时，此路由变为可用并被加入到路由表中，同时向其他 BGP 对等体发布更新报文。

图1-12 BGP 衰减示意图



3. 对等体组

对等体组（Peer Group）是一些具有某些相同属性的对等体的集合。当一个对等体加入对等体组中时，此对等体将获得与所在对等体组相同的配置。当对等体组的配置改变时，组内成员的配置也相应改变。

在大型 BGP 网络中，对等体的数量会很多，其中很多对等体具有相同的策略，在配置时会重复使用一些命令，利用对等体组在很多情况下可以简化配置。

将对等体加入对等体组中，对等体与对等体组具有相同的路由更新策略，提高了路由发布效率。

注意

如果对等体和对等体组都对某个选项做了配置，配置以最后一次的修改为准。

4. 团体

对等体组可以使一组对等体共享相同的策略，而利用团体可以使多个 AS 中的一组 BGP 路由器共享相同的策略。团体是一个路由属性，在 BGP 对等体之间传播，它并不受到 AS 范围的限制。

BGP 路由器在将带有团体属性的路由发布给其它对等体之前，可以改变此路由原有的团体属性。

除了使用公认的团体属性外，用户还可以使用团体属性列表自定义扩展团体属性，以便更为灵活地控制路由策略。

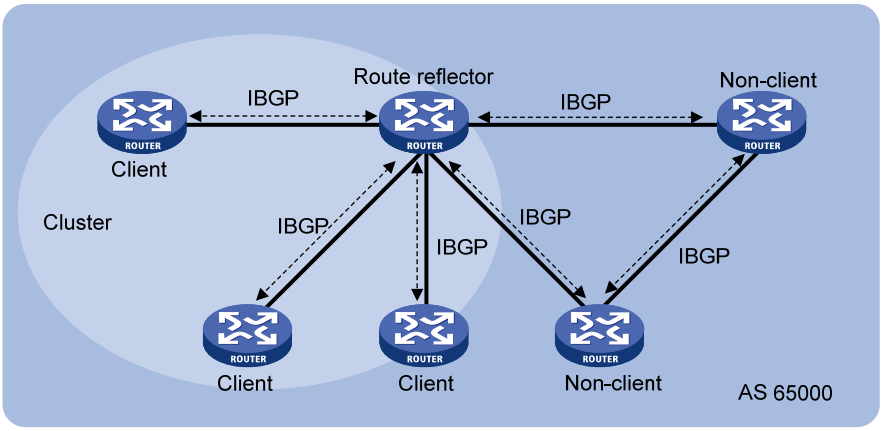
5. 路由反射器

为保证 IBGP 对等体之间的连通性，需要在 IBGP 对等体之间建立全连接关系。假设在一个 AS 内部有 n 台路由器，那么应该建立的 IBGP 连接数就为 n(n-1)/2。当 IBGP 对等体数目很多时，对网络资源和 CPU 资源的消耗都很大。

利用路由反射可以解决这一问题。在一个 AS 内，其中一台路由器作为路由反射器 RR（Route Reflector），其它路由器做为客户机（Client）与路由反射器之间建立 IBGP 连接。路由反射器在客户机之间传递（反射）路由信息，而客户机之间不需要建立 BGP 连接。

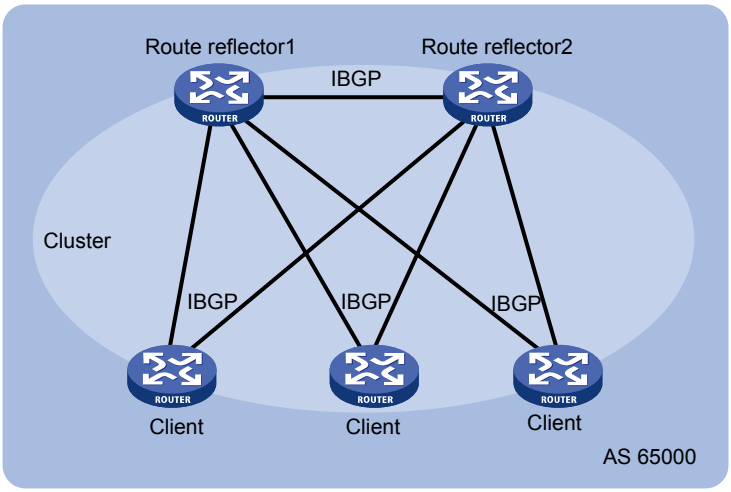
既不是反射器也不是客户机的BGP路由器被称为非客户机（Non-Client）。非客户机与路由反射器之间，以及所有的非客户机之间仍然必须建立全连接关系。其示意图如 图 1-13所示。

图1-13 路由反射器示意图



路由反射器和它的客户机组成了一个集群（Cluster）。某些情况下，为了增加网络的可靠性和防止单点故障，可以在一个集群中配置一个以上的路由反射器。这时，位于相同集群中的每个路由反射器都要配置相同的Cluster_ID，以避免路由循环。如 图 1-14所示。

图1-14 多路由反射器



在某些网络中，路由反射器的客户机之间已经建立了全连接，它们可以直接交换路由信息，此时客户机到客户机之间的路由反射是没有必要的，而且还占用带宽资源。目前，系统支持配置相关命令来禁止在客户机之间反射路由。

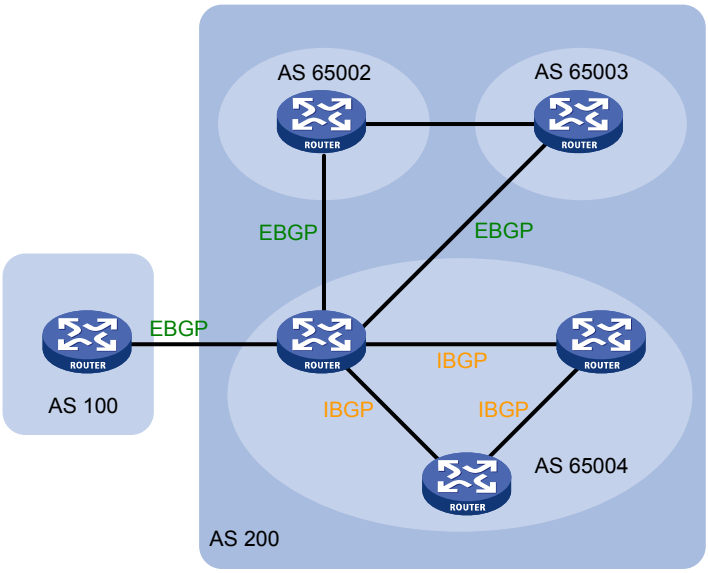
说明

禁止客户机之间的路由反射后，客户机到非客户机之间的路由仍然可以被反射。

6. 联盟

联盟（Confederation）是处理AS内部的IBGP网络连接激增的另一种方法，它将一个自治系统划分为若干个子自治系统，每个子自治系统内部的IBGP对等体建立全连接关系，子自治系统之间建立联盟内部EBGP连接关系。其示意图如 图 1-15所示。

图1-15 联盟示意图



在不属于联盟的 BGP 发言者看来，属于同一个联盟的多个子自治系统是一个整体，外界不需要了解内部的子自治系统情况，联盟 ID 就是标识联盟这一整体的自治系统号，如上图中的 AS 200 就是联盟 ID。

联盟的缺陷是：从非联盟方案向联盟方案转变时，要求路由器重新进行配置，逻辑拓扑也要改变。

在大型 BGP 网络中，路由反射器和联盟可以被同时使用。

1.1.7 BGP GR



说明

有关 GR（Graceful Restart，平滑重启）的原理介绍请参见“系统分册”中的“GR 概述”。

基于 BGP 的 GR Restarter 为了与 BGP 对等体建立一个 BGP 会话连接，首先要发送一个包含了 GR 能力的 OPEN 消息到对端，BGP 对等体收到该消息后，得知发送方已具有 GR 能力。这样，通过 OPEN 消息交互 GR 能力，GR Restarter 与其 BGP 对等体之间协商建立起 GR Session 连接。如果双方都没有交换 GR 能力的信息，建立起的会话也就不具备 GR 能力。

对于分布式设备，当进行主备倒换时，会话项将丢失，此时具备 GR 感知能力的 BGP 对等体会将所有与该 GR Restarter 有关的路由进行失效标记。但在 GR Time 内仍按照这些路由进行报文转发，这样确保了在从 BGP 对等体重新收集路由信息的过程中没有报文丢失。

对于分布式设备，主备倒换完成，GR Restarter 会重新与 BGP 对等体建立 GR Session 连接，同时发送新的 GR 消息以宣告其重启完毕。此时两个 BGP 对等体间进行路由信息交换。交换完成后，GR Restarter 根据新的路由转发信息更新路由表和转发表，删除失效的路由，完成 BGP 协议收敛。

1.1.8 MP-BGP

1. MP-BGP 概述

传统的 BGP-4 只能管理 IPv4 单播的路由信息，对于使用其它网络层协议（如 IPv6 等）的应用，在跨自治系统传播时就受到一定限制。

为了提供对多种网络层协议的支持，IETF 对 BGP-4 进行了扩展，形成 MP-BGP，目前的 MP-BGP 标准是 RFC 4760（Multiprotocol Extensions for BGP-4，BGP-4 的多协议扩展）。

支持 BGP 扩展的路由器与不支持 BGP 扩展的路由器可以互通。

2. MP-BGP 的扩展属性

BGP-4 使用的报文中，与 IPv4 地址格式相关的三条信息都由 Update 报文携带，这三条信息分别是：NLRI、路径属性中的 NEXT_HOP、路径属性中的 AGGREGATOR（该属性中包含形成聚合路由的 BGP 发言者的 IP 地址）。

为实现对多种网络层协议的支持，BGP-4 需要将网络层协议的信息反映到 NLRI 及 NEXT_HOP。MP-BGP 中引入了两个新的路径属性：

- MP_REACH_NLRI：Multiprotocol Reachable NLRI，多协议可达 NLRI。用于发布可达路由及下一跳信息。
- MP_UNREACH_NLRI：Multiprotocol Unreachable NLRI，多协议不可达 NLRI。用于撤销不可达路由。

这两种属性都是可选非过渡（Optional non-transitive）的，因此，不提供多协议能力的 BGP 发言者将忽略这两个属性的信息，不把它们传递给其它邻居。

3. 地址族

MP-BGP 采用地址族（Address Family）来区分不同的网络层协议，关于地址族的一些取值可以参

考 RFC 1700（Assigned Numbers）。目前，系统实现了多种 MP-BGP 扩展应用，包括对 VPN 的扩展、对 IPv6 的扩展等，不同的扩展应 1 用在各自的地址族视图下配置。

说明

- 有关 VPN 的扩展应用，请参见“MPLS 分册”中的“MPLS L3VPN 配置”。
- 有关 IPv6 的扩展应用，请参见“IP 路由分册”中的“IPv6 BGP 配置”。
- 本章不对 MP-BGP 地址族视图下的、与特定应用相关的命令作详细介绍。

1.1.9 协议规范

与 BGP 相关的协议规范有：

- RFC1771：A Border Gateway Protocol 4 (BGP-4)
- RFC2858：Multiprotocol Extensions for BGP-4
- RFC3392：Capabilities Advertisement with BGP-4
- RFC2918：Route Refresh Capability for BGP-4
- RFC2439：BGP Route Flap Damping
- RFC1997：BGP Communities Attribute
- RFC2796：BGP Route Reflection
- RFC3065：Autonomous System Confederations for BGP
- draft-ietf-idr-restart-08：Graceful Restart Mechanism for BGP

1.2 BGP 配置任务简介

表1-2 BGP 配置任务简介

配置任务		说明	详细配置
配置 BGP 连接	创建 BGP 连接	必选	1.3.2
	配置建立 TCP 连接使用的源接口	可选	1.3.3
	配置允许同非直连邻居建立 EBGP 连接	可选	1.3.4
控制路由信息的生成	配置 BGP 发布本地路由	二者必选其一	1.4
	配置 BGP 引入其他路由		
	配置 BGP 引入其他协议缺省路由	可选	
控制路由信息的发布与接收	配置 BGP 路由聚合	可选	1.5.2
	配置向对等体/对等体组发送缺省路由		1.5.3
	配置 BGP 路由信息的发布/接受策略		1.5.4
	配置 BGP 与 IGP 路由同步		1.5.5
	限制从 BGP 对等体/对等体组接收的路由前缀数量		1.5.6
	配置 BGP 路由衰减		1.5.7
	配置 shortcut 路由		1.5.8

配置任务		说明	详细配置
配置 BGP 的路由属性	为接收路由分配首选值	可选	1.6.2
	配置 BGP 的路由优先级		1.6.3
	配置 Local-Preference 缺省值		1.6.4
	配置 MED 属性		1.6.5
	配置下一跳属性		1.6.6
	配置 AS_PATH 属性		1.6.7
调整和优化 BGP 网络	配置 BGP 存活时间间隔与保持时间	可选	1.7.2
	配置发送路由更新报文的时间间隔		1.7.3
	配置 BGP 软复位		1.7.4
	使能 EBGP 连接快速复位功能		1.7.5
	配置 BGP 建立 TCP 连接时进行 MD5 认证		1.7.6
	配置 BGP 负载分担		1.7.7
	禁止与对等体/对等体组建立会话		1.7.8
配置 BGP 大型网络	配置 BGP 对等体组	可选	1.8.2
	配置 BGP 团体		1.8.3
	配置 BGP 路由反射器		1.8.4
	配置 BGP 联盟		1.8.5
配置 BGP GR		可选	1.9
开启 Trap 功能		可选	1.10
使能 BGP 日志功能		可选	1.11

1.3 配置 BGP 连接

本节讲述最基本的 BGP 网络配置过程。

说明

在本节中，不对 BGP 和 MP-BGP 进行严格的区分，命令的适用情况请参考所在的视图。

1.3.1 配置准备

在配置 BGP 连接之前，须保证相邻节点在网络层互通。

1.3.2 创建 BGP 连接

要在路由器上创建 BGP 连接，必须配置该路由器所在的 AS 编号即启动 BGP、配置对等体（对等体组）的 AS 编号。

Router ID 用来在一个自治系统中唯一的标识一台路由器，一台路由器如果要运行 BGP 协议，则必须存在 Router ID。

- 用户可以在启动 BGP 进入 BGP 视图后指定 Router ID，配置时，必须保证自治系统中任意两台路由器的 ID 都不相同。通常的做法是将路由器的 ID 配置为与该路由器某个接口的 IP 地址一致，为了增加网络的可靠性，建议将 Router ID 配置为 Loopback 接口的 IP 地址。
- 如果没有在 BGP 视图下配置 Router ID，则缺省使用全局 Router ID。关于全局 Router ID 的详细说明，请参考“IP 路由分册”中的“IP 路由概述”。
- 如果 BGP 使用的是全局 Router ID，当 Router ID 所在接口被删除时会重新选择路由器的 Router ID。
- 如果是在 BGP 视图下手工配置的 Router ID，只有使用 **undo router-id** 命令才能重新选择路由器的 Router ID。

表1-3 创建 BGP 连接

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
配置 Router ID	**router-id** *router-id*	可选 缺省情况下，与全局 Router ID 相同
指定对等体/对等体组及其 AS 号	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选 缺省情况下，没有指定对等体/对等体组 AS 号
配置使用 **peer as-number** 命令创建邻居时默认使能 IPv4 单播地址族	**default ipv4-unicast**	可选 缺省情况下，IPv4 单播地址族处于使能状态
激活指定对等体	**peer** *ip-address* **enable**	可选 缺省情况下，BGP 对等体是激活的
配置对等体/对等体组的描述信息	**peer** { *group-name* | *ip-address* } **description** *description-text*	可选 缺省情况下，对等体/对等体组没有描述信息

注意

- 一个路由器只能位于一个 AS 内，因此一台路由器只能运行一个 BGP 进程。
- 必须首先创建对等体组，才能进行对等体组的相关配置。关于对等体组的创建可以参见“1.8.2 配置BGP对等体组”。

1.3.3 配置建立 TCP 连接使用的源接口

BGP 使用 TCP 作为其传输层协议，缺省情况下，BGP 使用到达对等体最佳路由的出接口作为与对等体/对等体组建立 TCP 连接的源接口。

当建立 BGP 连接的路由器之间存在冗余链路时，如果路由器上的一个接口发生故障 down 掉，建立 TCP 连接的源接口可能会随之发生变化，导致 BGP 需要重新建立 TCP 连接，造成网络振荡。

BGP 对等体利用逻辑链路也可以建立对等体关系，并且 Loopback 接口通常比较稳定，建议网络管理员将建立 TCP 连接使用的源接口配置为 Loopback 接口，在网络中存在冗余链路时不会因为其中某个接口或链路的故障而使 BGP 连接中断，从而提高 BGP 连接的可靠性和稳定性。

表1-4 配置建立 TCP 连接使用的源接口

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置与对等体/对等体组创建 BGP 会话时建立 TCP 连接使用的源接口	**peer** { *group-name* | *ip-address* } **connect-interface** *interface-type interface-number*	必选 缺省情况下，BGP 使用到达 BGP 对等体的最佳路由的出接口作为与对等体/对等体组创建 BGP 会话时建立 TCP 连接的源接口

注意

当两个设备之间建立多条 BGP 连接时，如果没有为每个连接明确指定建立 TCP 连接的源接口，可能会由于无法根据到达 BGP 对等体的最优路由确定 TCP 连接源接口从而导致无法建立 TCP 连接，在此情况下建议用户配置 BGP 对等体时明确配置 BGP 会话建立 TCP 连接的源接口。

1.3.4 配置允许同非直连邻居建立 EBGP 连接

当前路由器要与另外一个路由器建立 EBGP 连接，它们必须具有直连的物理链路，如果不满足这一要求，则必须使用 **peer ebgp-max-hop** 命令允许它们之间经过多跳建立 TCP 连接。

表1-5 配置允许同非直连邻居建立 EBGP 连接

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置允许同非直接相连网络上的邻居建立 EBGP 连接	**peer** { *group-name* | *ip-address* } **ebgp-max-hop** [*hop-count*]	必选 缺省情况下，不允许同非直接相连网络上的邻居建立 EBGP 连接

说明

直连 EBGP 对等体不需要配置 **peer ebgp-max-hop** 命令。

1.4 控制路由信息的生成

BGP 与 IGP 不同，其着眼点主要在于控制路由的传播和选择最佳路由。

要生成 BGP 路由，主要有以下两种途径：

- 配置 BGP 发布本地路由
- 引入其它路由

网络管理员也可以根据需要配置引入其它协议的缺省路由。

1.4.1 配置准备

在控制 BGP 路由信息的生成之前，需要先创建 BGP 连接。

1.4.2 配置 BGP 发布本地路由

用户可以在 BGP 视图下配置发布某个网段的路由，从而将该路由发布给对等体，通过该种方式发布的路由的 Origin 属性为 IGP。网络管理员还可以通过使用路由策略更为灵活地控制所发布的路由。

需要注意的是，要发布的本地路由必须存在于 IP 路由表中。

表1-6 配置 BGP 发布本地路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
将网段路由发布到 BGP 路由表中	**network** *ip-address* [*mask* | *mask-length*] **route-policy** *route-policy-name*	必选 缺省情况下，BGP 不发布任何网段路由

1.4.3 配置 BGP 引入其它路由

BGP 可以向邻居 AS 发送本地 AS 内部网络的路由信息，但 BGP 不是自己去发现 AS 内部的路由信息，而是将 IGP 的路由信息引入到 BGP 路由表中，并发布给对等体。在引入 IGP 路由时，可以针对不同的路由协议来对路由信息进行过滤。

通过引入方式发布的路由的 Origin 属性为 incomplete。

表1-7 配置 BGP 引入其它路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
引入其它协议路由信息并通告	**import-route** *protocol* [*process-id* | **all-processes**] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 缺省情况下，BGP 不引入其它协议的路由

1.4.4 配置 BGP 引入其它协议缺省路由

BGP 不能通过 **import-route** 命令从其它协议引入缺省路由，如果要引入其它协议缺省路由，必须要使用下面命令进行配置。

表1-8 配置 BGP 引入其它协议缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
引入其它协议路由信息并通告	**import-route** *protocol* [*process-id* | **all-processes**] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 缺省情况下，BGP 不引入其它协议的路由
允许将缺省路由引入到 BGP 路由表中	**default-route imported**	必选 缺省情况下，BGP 不允许将缺省路由引入到 BGP 路由表中

1.5 控制路由信息的发布与接收

1.5.1 配置准备

在控制 BGP 路由信息的发布与接收之前，需要先配置 BGP 连接。

1.5.2 配置 BGP 路由聚合

在中型或大型 BGP 网络中，在向对等体发布路由信息时，需要配置路由聚合，减小对等体路由表中的路由数量。BGP 支持自动聚合和手动聚合两种聚合方式，同时配置时，手动聚合的优先级高于自动聚合的优先级。

1. 配置路由自动聚合

配置自动聚合功能后，BGP 将对引入的 IGP 子网路由进行聚合，不再发布子网路由，而是发布聚合后的自然网段的路由。

需要注意的是，用 **network** 命令发布的路由不能进行自动聚合。

表1-9 配置 BGP 引入其他路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置对引入的 IGP 子网路由进行自动聚合	**summary automatic**	必选 缺省情况下，不对引入的 IGP 子网路由进行自动聚合

2. 配置路由手动聚合

自动聚合是按照自然网段进行聚合，而且只能对 IGP 引入的子网路由进行聚合。

通过配置手动聚合，用户可以同时对 IGP 引入的子网路由和用 **network** 命令发布的路由进行聚合，而且还可以根据需要定义聚合路由的子网掩码长度。

表1-10 配置路由手动聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置路由手动聚合	**aggregate** *ip-address* { *mask* | *mask-length* } [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	必选 缺省情况下，不进行路由聚合

1.5.3 配置向对等体/对等体组发送缺省路由

通过配置向对等体/对等体组发送缺省路由，不论本地路由表中是否存在缺省路由，都将向指定对等体/对等体组发布一条下一跳地址为本地地址的缺省路由。

表1-11 配置向对等体/对等体组发送缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
向对等体/对等体组发送缺省路由	**peer** { *group-name* | *ip-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	必选 缺省情况下，不向对等体/对等体组发送缺省路由

1.5.4 配置 BGP 路由信息的发布/接收策略

1. 配置准备

配置 BGP 路由信息的发布/接收策略前，根据采取的策略，需要配置下列过滤器：

- 访问控制列表
- IP 地址前缀列表
- 路由策略
- AS 路径过滤列表

关于访问控制列表的详细配置过程，请参考“安全分册”中的“ACL 配置”。

关于 IP 地址前缀列表、路由策略和 AS 路径过滤列表的详细配置过程，请参考“IP 路由分册”中的“路由策略配置”。

2. 配置 BGP 路由信息的发布策略

用户可以使用路由策略、访问控制列表、AS 路径过滤列表或 IP 地址前缀列表对发布给对等体的路由信息进行过滤，可以根据需求选择过滤策略。

表1-12 配置 BGP 路由信息的发布策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置引入的路由在发布时进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct** | **isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | | **static**]	必选其一 缺省情况下，不对引入的路由在发布时进行过滤 可以根据需求选择过滤策略 同时配置几种过滤策略时，按照如下顺序执行： **filter-policy export** **peer filter-policy export** **peer as-path-acl export** **peer ip-prefix export** **peer route-policy export** 只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被发布
对发布给对等体/对等体组的路由指定路由策略	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	
为对等体/对等体组设置基于 ACL 的过滤策略	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **export**	
为对等体/对等体组设置基于 AS 路径过滤列表的 BGP 路由过滤策略	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **export**	
为对等体/对等体组设置基于 IP 前缀列表的路由过滤策略	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **export**	

3. 配置 BGP 路由信息的接收策略

用户可以使用路由策略、访问控制列表、AS 路径过滤列表或 IP 地址前缀列表对从等体接收的路由信息进行过滤，只有满足某些条件的路由才能被 BGP 接收，并加到路由表中。

网络管理员可以根据需求选择过滤策略，可以对从所有对等体接收的路由信息进行过滤，也可以从指定对等体接收的路由信息进行过滤。

需要注意的是，对等体组的成员可以与所在的组使用不同的入方向路由策略，即接收路由时，各对等体可以选择自己的策略。

表1-13 配置 BGP 路由信息的接收策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	必选其一 缺省情况下，不对接收的路由信息进行过滤 可以根据需求选择过滤策略 同时配置几种过滤策略时，按照如下顺序执行： **filter-policy import** **peer filter-policy import** **peer as-path-acl import** **peer ip-prefix import** **peer route-policy import** 只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被接收
对来自对等体/对等体组的路由指定路由策略	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **import**	
为对等体/对等体组设置基于 ACL 的过滤策略	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **import**	
为对等体/对等体组设置基于 AS 路径过滤列表的 BGP 路由过滤策略	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **import**	
为对等体/对等体组设置基于 IP 前缀列表的路由过滤策略	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **import**	

1.5.5 配置 BGP 与 IGP 路由同步

BGP 路由器收到一条 IBGP 路由，缺省只检查该路由的下一跳是否可达。使能同步特性后，当 BGP 路由器收到一条 IBGP 路由时，如果在其 IGP 路由表中没有同样的路由信息（即 IGP 没有与 BGP 同步），那么它将不会再把此路由信息发送给它的 EBGP 对等体。

表1-14 配置 BGP 与 IGP 路由同步

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 BGP 与 IGP 路由同步	**synchronization**	必选 缺省情况下，BGP 和 IGP 路由不同步

1.5.6 限制从 BGP 对等体/对等体组接收的路由前缀数量

表1-15 限制从 BGP 对等体/对等体组接收的路由前缀数量

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置当从指定对等体/对等体组收到的路由前缀数量超过最大值时，路由器中断 BGP 连接	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* [*percentage-value*]	三者必选其一 缺省情况下，没有限制从 BGP 对等体/对等体组接收的路由前缀数量
配置当从指定对等体/对等体组收到的路由前缀数量超过最大值时，仅打印告警信息，路由器保持与指定对等体/对等体组的连接	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* **alert-only** [*percentage-value*]	
配置当从指定对等体/对等体组收到的路由前缀数量超过允许的最大值时，路由器中断 BGP 连接，但会自动与对等体重建连接	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* **reconnect** *reconnect-time* [*percentage-value*]	

1.5.7 配置 BGP 路由衰减

通过配置 BGP 衰减，可以抑制不稳定的路由信息，不将这类路由加入到路由表中，也不将这类路由向其他 BGP 对等体发布。

表1-16 配置 BGP 路由衰减

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 BGP 路由衰减	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	必选 缺省情况下，没有配置 BGP 路由衰减

1.5.8 配置 shortcut 路由

缺省情况下，接收到的 EBGP 路由的路由管理优先级为 255，比本地生成的路由的管理优先级低。

通过将一条 EBGP 路由被配置成 short-cut，那么这条路由在加入到路由表时将使用本地生成路由的路由管理值，而不是 EBGP 路由的路由管理值，此时，这条路由成为最佳路由的可能性被提高了很多。

表1-17 配置 shortcut 路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 **short-cut** 路由	**network** *ip-address* [*mask* | *mask-length*] **short-cut**	可选 缺省情况下，接收到的 EBGP 路由的路由管理优先级为 255

1.6 配置 BGP 的路由属性

BGP 具有很多路由属性，利用这些属性可以改变 BGP 的选路策略。

1.6.1 配置准备

在配置 BGP 的路由属性之前，需要先配置 BGP 连接。

1.6.2 为接收路由分配首选值

BGP 选择路由时将首先丢弃下一跳不可达的路由，其次再优选 Preferred-value 值最大的路由。

缺省情况下，从对等体/对等体组学到的路由的首选值为 0，网络管理员可以为从某个对等体/对等体组接收的路由配置首选值，从而提高从指定对等体/对等体学到的路由的优先级。

表1-18 为接收路由分配首选值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
为从对等体/对等体组接收的路由分配首选值	**peer** { *group-name* | *ip-address* } **preferred-value** *value*	可选 缺省情况下，从对等体/对等体组接收的路由的首选值为 0

1.6.3 配置 BGP 的路由优先级

由于路由器上可能同时运行多个动态路由协议，就存在各个路由协议之间路由信息共享和选择的问题。系统为每一种路由协议设置一个优先级，在不同协议发现同一条路由时，优先级高的路由将被优先选择。

用户可以修改 EBGP 路由、IBGP 路由以及用 **network** 命令发布的本地路由的路由优先级；还可以应用路由策略为匹配过滤条件的特定路由配置优先级，对于那些没有匹配的路由，使用缺省优先级。

表1-19 配置 BGP 的路由优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 BGP 路由的管理优先级	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	可选 缺省情况下，EBGP 路由的管理优先级为 255，IBGP 路由的管理优先级为 255，本地产生的 BGP 路由的管理优先级为 130

1.6.4 配置 Local-Preference 缺省值

Local-Preference 用来判断流量离开 AS 时的最佳路由。当 BGP 的路由器通过不同的 IBGP 对等体得到目的地址相同但下一跳不同的多条路由时，将优先选择 Local-Preference 值较高的路由。

用户可以通过 **default local-preference** 命令改变 BGP 路由器向 IBGP 对等体发送的路由 Local-Preference 的缺省值。

表1-20 配置 Local-Preference 缺省值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置本地优先级的缺省值	**default local-preference** *value*	可选 缺省情况下，本地优先级的缺省值为 100

1.6.5 配置 MED 属性

MED 用来判断流量进入 AS 时的最佳路由。当一个运行 BGP 的路由器通过不同的 EBGP 对等体得到目的地址相同但下一跳不同的多条路由时，在其它条件相同的情况下，将优先选择 MED 值较小者作为最佳路由。

1. 配置 MED 缺省值

表1-21 配置 MED 缺省值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置系统 MED 的缺省值	**default med** *med-value*	可选 缺省情况下，MED 的缺省值为 0

2. 配置允许比较来自不同 AS 邻居路由的 MED 属性值

缺省情况下，BGP 只比较来自同一个 AS 的路由的 MED 属性值。

通过配置 **compare-different-as-med** 命令，可以强制 BGP 比较来自不同 AS 的路由的 MED 属性值。

表1-22 配置允许比较来自不同 AS 邻居路由的 MED 属性值

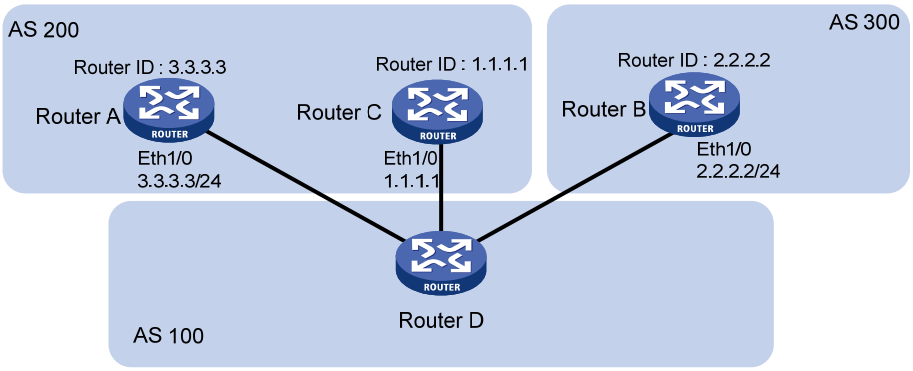
操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置允许比较来自不同 AS 邻居的路由路径的 MED 属性值	**compare-different-as-med**	必选 缺省情况下，不允许比较来自不同 AS 邻居的路由路径的 MED 属性值

3. 配置对来自同一 AS 的路由进行 MED 排序优选

缺省情况下，系统不会对来自同一 AS 的路由进行 MED 排序优选。

BGP 选择最优路由时是将新的路由和当前 BGP 路由表中的最优路由进行比较，只要新的路由比当前 BGP 路由表中的最优路由更优，新的路由将成为最优路由，路由学习的顺序有可能会影响最优路由的选择结果。

图1-16 MED 排序优选示意图



如上图所示所示，Router D 分别从 Router A 和 Router B 学习到了到达网段 10.0.0.0 的路由，由于 Router B 的 Router ID 值较小，因此，从 Router B 学来的路由被选为最优路由：

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    3.3.3.3         50                      0       200e
```

当路由器再从 Router C 学习到到达 10.0.0.0 网段的路由时，由于它只和当前路由表的最优路由进行比较，由于 Router C 的 Router ID 值更小，相对更优，它将成为最优路由。

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          1.1.1.1         60                      0       200e
 * i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    3.3.3.3         50                      0       200e
```

但是如果将这条路由与从 Router A 学习到的路由进行比较，那么由于来自同一 AS 200 且 MED 值更大，应该视为无效路由。

通过在 Router D 上配置 **bestroute compare-med** 命令，Router D 对于学习到的路由会首先按照路由来自的 AS 分组，对来自同一 AS 的路由根据 MED 值的大小进行优选，首先选出来 MED 值最小的路由，然后再对来自不同 AS 的路由进行优选，从而避免路由优选结果的不确定性。配置对来自同一 AS 的路由进行 MED 排序优选后，Router D 上的 BGP 路由表如下所示，从 Router C 学习到的到达 10.0.0.0 网段的路由将成为最优路由。

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          3.3.3.3         50                      0       200e
 * i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    1.1.1.1         60                      0       200e
```

需要注意的是，BGP 将无法进行负载分担，因为 BGP 要求进行负载分担的路由必须 AS_PATH 相同。

表1-23 配置对来自同一 AS 的路由进行 MED 排序优选

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置根据路由来自的 AS 进行分组，并根据 MED 排序优选	**bestroute compare-med**	必选 缺省情况下，不根据路由来自的 AS 进行分组对 MED 排序优选

4. 配置允许比较来自同一联盟不同子自治系统邻居路由的 MED 属性值

表1-24 配置允许比较来自同一联盟不同子自治系统邻居路由的 MED 属性值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置允许比较来自同一联盟不同子自治系统邻居路由的 MED 属性值	**bestroute med-confederation**	必选 缺省情况下，不比较来自同一联盟不同子自治系统邻居路由的 MED 属性值

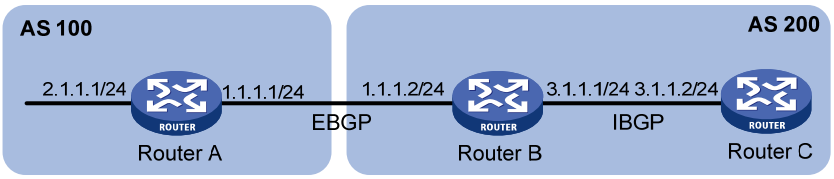
说明

只有 AS_PATH 里不包含联盟体外的子自治系统编号时，才会比较来自同一联盟不同子自治系统邻居路由的 MED 属性值，例如有三条路由，它们的 AS-PATH 值分别为 65006 65009、65007 65009 和 65008 65009，MED 值分别为 2、3、1，由于第三条路由包含了联盟体外的子自治系统编号，因此在选择最优路由时第一条路由将成为最优路由。

1.6.6 配置下一跳属性

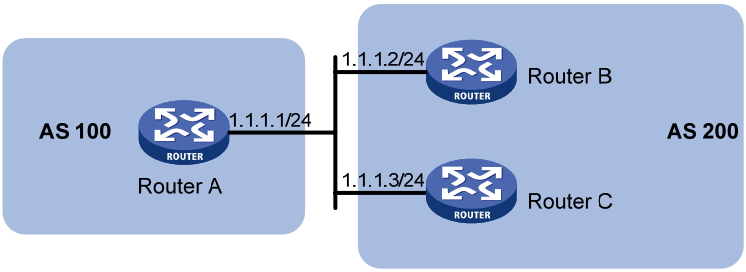
缺省情况下，路由器向 IBGP 对等体/对等体组发布路由时，不将自身地址作为下一跳，但有的时候为了保证 IBGP 邻居能够找到下一跳，可以配置将自身地址作为下一跳。以下图为例，Router A 与 Router B 建立 EBGP 邻居关系，Router B 与 Router C 建立 IBGP 邻居关系，Router B 在向 Router C 发布从 Router A 学到的 BGP 路由时，如果 Router C 上没有到达 1.1.1.1/24 的路由，可以在 Router B 上配置 **peer next-hop-local** 命令将 3.1.1.1/24 做为下一跳，这样，Router C 就能找到下一跳。

图1-17 配置 BGP 下一跳属性应用组网图一



在一些比较特殊的组网环境中（即两个 BGP 连接在同一网段的广播网），路由器向 EBGP 对等体/对等体组发布路由时不会将自身地址作为下一跳，以下图为例：Router A 与 Router B 建立 EBGP 邻居关系，Router B 与 Router C 建立 IBGP 邻居关系，两个 BGP 连接都位于同一个广播网 1.1.1.0/24 中，Router B 向 Router A 发布 EBGP 路由时不会将自身地址 1.1.1.2/24 作为下一跳，但如果用户有需要，也可以通过通过配置 **peer next-hop-local** 命令实现将自身地址 1.1.1.2/24 作为下一跳。

图1-18 配置 BGP 下一跳属性应用组网图二



需要注意的是，如果配置了 BGP 负载分担，则不论是否配置了 **peer next-hop-local** 命令，本地路由器向 IBGP 对等体/对等体组发布路由时都先将下一跳地址改变为自身地址。

表1-25 配置 BGP 的下一跳属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置发布路由时将自身地址作为下一跳	**peer** { *group-name* | *ip-address* } **next-hop-local**	可选 缺省情况下，向 EBGP 对等体/对等体组发布路由时，将自身地址作为下一跳；向 IBGP 对等体/对等体组发布路由时，不将自身地址作为下一跳

1.6.7 配置 AS_PATH 属性

1. 配置允许本地 AS 号重复出现的次数

通常情况下，BGP 会检查对等体发来的路由的 AS_PATH 属性，如果其中已存在本地 AS 号，则 BGP 会忽略此路由，以免形成路由环路。

通过配置 **peer allow-as-loop** 命令，可以允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数。

表1-26 配置允许本地 AS 号重复出现的次数

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置允许本地 AS 号重复出现的次数	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	必选 缺省情况下，不允许本地 AS 重复出现

2. 禁止路由器将 AS_PATH 当作选路算法中的一个因素

路由器在选择最优路由时会优选 AS 路径最短的路由，通过如下配置可以禁止路由器将 AS_PATH 当作选路算法中的一个因素。

表1-27 禁止路由器将 AS_PATH 当作选路算法中的一个因素

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
禁止路由器将 AS_PATH 当作选路算法中的一个因素	**bestroute as-path-neglect**	必选 缺省情况下，路由器将 AS_PATH 当作选路算法中的一个因素

3. 为对等体/对等体组定制一个虚拟的自治系统号

当网络管理员需要进行系统移植时，例如，Router A 原来位于 AS 2，现在需要将它移植到 AS 3 里，可以在 Router A 上为已经创建的 EBGP 对等体（组）配置一个虚拟的本地自治系统号 2，将本地

真实的 AS 号 3 隐藏起来；由于位于其他 AS 内的 EBGP 对等体只能看到这个虚拟 AS 号 2，因此它们不需要改变现有配置，从而在系统升级调试过程中可以不中断现有的 BGP 服务。

表1-28 为对等体/对等体组定制一个虚拟的自治系统号

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
为对等体/对等体组定制一个虚拟的自治系统号	**peer** { *group-name* | *ip-address* } **fake-as** *as-number*	必选 缺省情况下，没有为对等体/对等体组配置虚拟的本地自治系统号

说明

peer fake-as 命令只适用于 EBGP 对等体/对等体组。

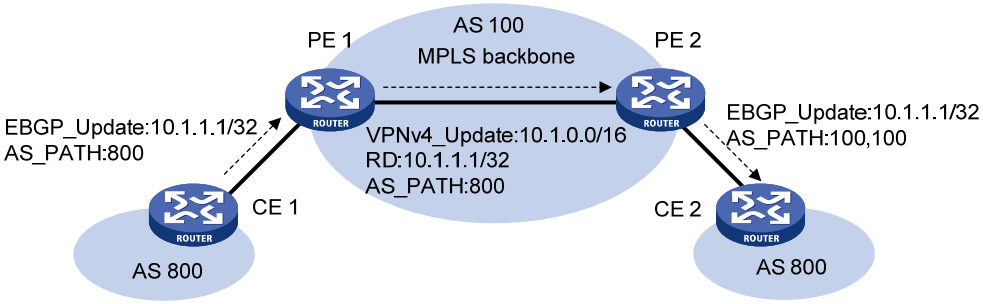
4. 配置 AS 号替换功能

在 MPLS L3VPN 中，如果 PE 和 CE 之间运行 EBGP，由于 BGP 使用 AS 号检测路由环路，为保证路由信息的正确发送，需要为物理位置不同的站点分配不同的 AS 号。

如果物理分散的 CE 复用相同的 AS 号，就应该在 PE 上配置 BGP 的 AS 号替换功能。此功能是 BGP 的出口策略，在发布路由时有效。

使能了 BGP 的 AS 号替换功能后，当 PE 向指定对等体中的 CE 发布路由时，如果路由的 AS_PATH 中有与 CE 相同的 AS 号，将被替换成 PE 的 AS 号后再发布。

图1-19 BGP AS 号替换应用示意图



如 图 1-19所示，CE 1 和CE 2 都使用AS号 800，在PE 2 上使能针对CE 2 的AS号替换功能。当CE 1 发来的Update信息从PE 2 发布给CE 2 时，PE 2 发现AS_PATH中存在与CE 2 相同的AS号 800，就把它替换为自己的AS号 100，这需要注意的是，如果需要完全的连接性，PE 1 上也需要做类似的配置。

表1-29 配置 AS 号替换功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-

操作	命令	说明
配置用本地 AS 号替换 AS_PATH 属性中指定对等体/对等体组的 AS 号	**peer** { *group-name* | *ip-address* } **substitute-as**	必选 缺省情况下，没有用本地 AS 号替换 AS_PATH 属性中指定对等体/对等体组的 AS 号

注意

替换 AS_PATH 属性中的 AS 号命令仅在特定组网环境下使用，错误的配置会引起路由环路。

5. 配置发送 BGP 更新报文时 AS_PATH 属性中不携带私有 AS 号

表1-30 配置发送 BGP 更新报文时 AS_PATH 属性中不携带私有 AS 号

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置发送 BGP 更新报文时 AS_PATH 属性中不携带私有 AS 号	**peer** { *group-name* | *ip-address* } **public-as-only**	必选 缺省情况下，发送 BGP 更新报文时，携带私有自治系统号

1.7 调整和优化 BGP 网络

1.7.1 配置准备

在调整和优化 BGP 网络之前，需要先配置 BGP 连接。

1.7.2 配置 BGP 存活时间间隔与保持时间

当对等体间建立了 BGP 连接后，它们定时向对端发送 Keepalive 消息，以防止路由器认为 BGP 连接已中断。若路由器在设定的连接保持时间（Holdtime）内未收到对端的 Keepalive 消息或 Update 报文，则认为此 BGP 连接已中断，从而断开此 BGP 连接。

用户可以配置当前路由的 BGP 存活时间间隔与保持时间，也可以配置与指定对等体/对等体组建立连接的存活时间间隔和保持时间，如果当前路由器的存活时间间隔（保持时间）与对等体/对等体组上配置的不一致，数值较小者做为协商后的保持时间。

表1-31 配置 BGP 存活时间间隔与保持时间

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 BGP 的存活时间间隔与保持时间	**timer keepalive** *keepalive* **hold** *holdtime*	可选 缺省情况下，存活时间间隔为 60 秒，保持时间为 180 秒
配置指定对等体/对等体组的存活时间间隔和保持时间	**peer** { *group-name* | *ip-address* } **timer keepalive** *keepalive* **hold** *holdtime*	

注意

- 合理的最大 Keepalive 消息发送间隔为保持时间的三分之一，且该发送间隔不能小于 1 秒，因此，保持时间如果不为 0，则最小为 3 秒。
- 使用 **timer** 命令配置的定时器比使用 **peer timer** 命令配置的定时器优先级要低。
- 如果与对等体的邻居关系已经建立，必须要复位 BGP 连接才能使新配置的存活时间间隔与保持时间生效。

1.7.3 配置发送路由更新报文的时间间隔

表1-32 配置发送路由更新报文的时间间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置对等体/对等体组发送同一路由更新报文的时间间隔	**peer** { *group-name* | *ip-address* } **route-update-interval** *interval*	可选 缺省情况下，向 IBGP 对等体发送同一路由更新报文的时间间隔为 15 秒，向 EBGP 对等体发送同一路由更新报文的时间间隔为 30 秒

1.7.4 配置 BGP 软复位

BGP 的选路策略改变后，为了使新的策略生效，必须复位 BGP 连接，但这样会造成短暂的 BGP 连接中断。

如果当前路由器和对等体都支持 Route-Refresh 功能，通过使能 Route-Refresh 功能，在策略改变时，系统可以在不中断 BGP 连接的情况下，自动对 BGP 路由表进行动态刷新。

如果对等体不支持 Route-Refresh 功能，则需要将从对等体接收的所有路由更新保存在本地，当选路策略发生改变后，在不中断连接的情况下重新刷新 BGP 路由表，并应用新的策略。

1. 通过使能 Route-Refresh 实现 BGP 软复位

在对等体支持并使能 Route-Refresh 功能的情况下，如果 BGP 的路由策略发生了变化，本地路由器会向 BGP 对等体发布 Router-Refresh 消息，收到此消息的对等体会将其路由信息重新发给本地路由器。这样，在不中断 BGP 连接的情况下，就可以对 BGP 路由表进行动态更新，并应用新的选路策略。

表1-33 通过使能 Route-Refresh 实现 BGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
使能 BGP 路由刷新功能	**peer** { *group-name* | *ip-address* } **capability-advertise route-refresh**	可选 缺省情况下，BGP 路由刷新功能处于使能状态

2. 通过将所有路由更新保存在本地实现 BGP 软复位

当对等体不支持 Route-Refresh 功能时，可通过配置 **peer keep-all-routes** 命令将从对等体接收的所有路由更新保存在本地用户，然后再执行 **refresh bgp** 命令对保存在本地的所有路由使用路由策略重新过滤一遍。

表1-34 通过将所有路由更新保存在本地实现 BGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
禁止 BGP 路由刷新和多协议扩展功能	**peer** { *group-name* | *ip-address* } **capability-advertise conventional**	必选 缺省情况下，BGP 路由刷新和多协议扩展功能处于使能状态
保存所有来自对等体/对等体组的原始路由信息	**peer** { *group-name* | *ip-address* } **keep-all-routes**	必选 缺省情况下，不保存所有来自对等体/对等体组的原始路由信息
退回用户视图	**return**	-
手工对 BGP 连接进行软复位	**refresh bgp** { **all** | *ip-address* | **group** *group-name* | **external** | **internal** } { **export** | **import** }	必选

1.7.5 使能 EBGP 连接快速复位功能

路由器在设定的连接保持时间（Holdtime）内未收到对端的 Keepalive 消息或 Update 报文，则认为此 BGP 连接已中断，从而断开此 BGP 连接。

通过使能 EBGP 连接快速复位功能，如果当前路由器与直连的 EBGP 邻居链路变为 down 状态，与该邻居建立的 EBGP 会话将重新建立。

表1-35 使能 EBGP 连接快速复位功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
使能 EBGP 连接快速复位功能	**ebgp-interface-sensitive**	可选 缺省情况下，使能 EBGP 连接快速复位功能处于关闭状态

1.7.6 配置 BGP 建立 TCP 连接时进行 MD5 认证

BGP 使用 TCP 做为传输层协议，为提高 BGP 的安全性，可以配置 BGP 在建立 TCP 连接时进行 MD5 认证，即两台路由器必须配置相同的密码，才能建立 TCP 连接，

BGP 的 MD5 认证并不能对 BGP 报文认证，它只是为 TCP 连接设置 MD5 认证密码，由 TCP 完成认证。如果认证失败，则不建立 TCP 连接。

表1-36 配置 BGP 建立 TCP 连接时进行 MD5 认证

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置 BGP 建立 TCP 连接时进行 MD5 认证	**peer** { *group-name* | *ip-address* } **password** { **cipher** | **simple** } *password*	可选 缺省情况下，BGP 在建立 TCP 连接时不进行 MD5 认证

1.7.7 配置 BGP 负载分担

如果到一个目的地有几条开销相同的路径，可以实现等价路由负载分担，IP 报文在这几个链路上负载分担，以提高链路利用率。

该配置用以设置 BGP 协议的最大等价路由条数。

表1-37 配置 BGP 负载分担

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置进行 BGP 负载分担的路由条数	**balance** *number*	可选 缺省情况下，BGP 不进行路由负载分担

1.7.8 禁止与对等体/对等体组建立会话

表1-38 禁止与对等体/对等体组建立会话

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
禁止与对等体/对等体组建立会话	**peer** { *group-name* | *ip-address* } **ignore**	可选 缺省情况下，允许与 BGP 对等体/对等体组建立会话

1.8 配置大型 BGP 网络

在大型 BGP 网络中，对等体的数目众多，配置和维护极为不便，可以根据组网需要，配置对等体组、团体、路由反射器或联盟，以降低管理难度和提高路由发布效率。

1.8.1 配置准备

在配置大型 BGP 网络之前，需完成以下任务：相邻节点的网络层互通。

1.8.2 配置 BGP 对等体组

对等体组是具有相同更新策略的对等体的集合。

在大型 BGP 网络中，对等体的数量会很多，其中，很多对等体需要配置相同的策略，通过配置对等体组并将对等体加入到对等体组，可以使对等体获得与所在对等体组相同的配置，而且当对等体组的配置改变时，组内成员的配置也相应改变，从而简化配置。

根据对等体所在的 AS，对等体组可分为：

- IBGP 对等体组：对等体组中的对等体与当前路由器位于同一 AS。
- EBGP 对等体组：对等体组中的对等体与当前路由器位于不同 AS。

需要注意的是，如果对等体组中已经存在对等体，则不能改变该对等体组的 AS 号，也不能使用 **undo** 命令删除已指定的 AS 号。

1. 配置 IBGP 对等体组

创建 IBGP 对等体组后，系统在将对等体加入 IBGP 对等体组时，会自动在 BGP 视图下创建该对等体，并设置其 AS 号为本地 AS 号。

表1-39 配置 BGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	- 缺省情况下，系统没有运行 BGP
创建 IBGP 对等体组	**group** *group-name* [**internal**]	可选
向对等体组中加入对等体	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	可向组中加入多个对等体。系统会自动在 BGP 视图下创建该对等体，并设置其 AS 号为本地 AS 号

2. 配置 EBGP 对等体组

根据对等体组中的对等体是否属于同一个外部 AS，EBGP 对等体组又可以分为纯 EBGP 对等体组和混合 EBGP 对等体组；如果对等体组中的对等体属于同一个外部 AS，该对等体组就是纯 EBGP 对等体组；如果对等体组中的对等体属于不同外部 AS，该对等体组就是混合 EBGP 对等体组。

用户有两种方式配置 EBGP 对等体组：

- 第一种方式是创建对等体组后，先指定 AS 号，再将对等体加入到对等体组中；
- 第二种方式是创建对等体组完毕后，将已经创建的对等体加入对等体组中，或将对等体加入对等体中时指定 AS 号。

表1-40 配置 EBGP 对等体组（方式一）

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
创建 EBGP 对等体组	**group** *group-name* **external**	必选
设置对等体组的 AS 号	**peer** *group-name* **as-number** *as-number*	必选
向对等体组中加入对等体	**peer** *ip-address* **group** *group-name*	必选

表1-41 配置 EBGP 对等体组（方式二）

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
创建 EBGP 对等体组	**group** *group-name* **external**	必选
向对等体组中加入对等体	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	必选

1.8.3 配置 BGP 团体

团体是一组有相同特征的目的地址的集合，没有物理上的边界，与其所在的 AS 无关。

用户可以先通过路由策略定义一组目的地址属于某个指定团体，并配置向对等体/对等体组发布团体属性：

- 在向对等体/对等体组发布路由时可以指定发布策略，在发布策略中通过团体属性对发布给对等体/对等体组的路由信息进行过滤；
- 对等体/对等体组在接收路由信息时也可以指定接收策略，在接收策略中通过团体属性对从等体/对等体组接收的路由信息进行过滤。

在路由策略中通过团体来进行路由过滤，可以提高路由策略配置的灵活度，并简化路由策略的管理，从而降低维护管理的难度。

关于路由策略的详细配置，请参见“IP 路由分册”中的“路由策略配置”。

表1-42 配置 BGP 团体

操作		命令	说明
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	-
配置向对等体/对等体组发布团体属性	配置向对等体/对等体组发布团体属性	**peer** { *group-name* | *ip-address* } **advertise-community**	必选 缺省情况下，不将团体属性和扩展团体属性发布给任何对等体/对等体组
	配置向对等体/对等体组发布扩展团体属性	**peer** { *group-name* | *ip-address* } **advertise-ext-community**	
对发布给对等体/对等体组的路由指定路由策略		**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，不指定对等体/对等体组的路由策略

1.8.4 配置 BGP 路由反射器

如果同一个 AS 内有多个 BGP 路由器，为了减少在同一 AS 内建立的 IBGP 连接数，可以把几个 BGP 路由器划分为一个集群，将其中的一台路由器配置为反射器，其它路由器作为客户机。

为了增加网络的可靠性和防止单点故障，可以在一个集群中配置一个以上的路由反射器，这时，网络管理员必须给位于相同集群中的每个路由反射器配置相同的集群 ID，以避免路由循环。

表1-43 配置 BGP 路由反射器

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	- 缺省情况下，系统不运行 BGP
配置将本机作为路由反射器，并将对等体/对等体组作为路由反射器的客户	**peer** { *group-name* | *ip-address* } **reflect-client**	必选 缺省情况下，没有配置路由反射器及其客户
配置允许客户到客户的路由反射	**reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
配置路由反射器的集群 ID	**reflector cluster-id** *cluster-id*	可选 缺省情况下，每个路由反射器是使用自己的 Router ID 作为集群 ID

注意

- 通常情况下，路由反射器的客户之间不要求是全连接的，路由缺省通过反射器从一个客户反射到其它客户；如果客户之间是全连接的，可以禁止客户间的反射，以便减少开销。
- 通常，一个集群里只有一个路由反射器。此时是由反射器的路由器 ID 来识别该集群的。设置多个路由反射器可提高网络的稳定性。如果一个集群中配有多个路由反射器，请使用相关命令为所有的路由反射器配置同样的集群 ID，以避免路由循环。

1.8.5 配置 BGP 联盟

联盟（Confederation）是处理 AS 内部的 IBGP 网络连接激增的另一种方法，它将一个自治系统划分为若干个子自治系统，每个子自治系统内部的 IBGP 对等体建立全连接关系，子自治系统之间建立 EBGP 连接关系。

1. BGP 联盟基本配置

网络管理员将一个自治系统划分为若干个子自治系统后，如果路由器位于联盟中的某个子自治系统中，需要在路由器上做如下配置：

(1) 首先，启动 BGP，并指定该路由器所属的子自治系统号；

(2) 其次，配置联盟 ID。在不属于联盟的 BGP 发言者看来，属于同一个联盟的多个子自治系统是一个整体，联盟 ID 就是标识联盟这一整体的自治系统号；

(3) 如果该路由器与该联盟的其它子自治系统建立 EBGP 邻居关系，需要在该路由器上指定该联盟体中除了自己还包含哪些子自治系统。

一个联盟最多可包括 32 个子自治系统，配置属于联盟的子自治系统时使用的 *as-numbei* 仅在联盟内部有效。

表1-44 配置 BGP 联盟

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-

操作	命令	说明
配置联盟 ID	**confederation id** *as-number*	必选 缺省情况下，未配置联盟的 ID
指定一个联盟体中包含了哪些子自治系统	**confederation peer-as** *as-number-list*	必选 缺省情况下，未指定一个联盟体中包含了哪些子自治系统

2. 配置联盟兼容性

如果其他路由器的联盟实现机制不同于 RFC 3065 标准，可以通过如下配置与未采用 RFC 3065 配置的 AS 联盟兼容。

表1-45 配置联盟兼容性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置与未采用 RFC 3065 配置的 AS 联盟兼容	**confederation nonstandard**	可选 缺省情况下，配置的联盟与 RFC 3065 一致

1.9 配置 BGP GR

分别在作为 GR Restarter 和 GR Helper 的设备上进行以下配置。

说明

一台设备可以同时充当 GR Restarter 和 GR Helper。

表1-46 配置 BGP GR

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	- 缺省情况下，系统没有运行 BGP
使能 BGP 协议的 GR 能力	**graceful-restart**	必选 缺省情况下，BGP 协议的 GR 能力处于关闭状态
配置对端重建 BGP 会话连接的最大时间	**graceful-restart timer restart** *timer*	可选 缺省情况下，对端重建 BGP 会话连接的最大时间为 150 秒
配置等待 End-Of-RIB 标记的时间	**graceful-restart timer wait-for-rib** *timer*	可选 缺省情况下，等待 End-Of-RIB 标记的时间为 180 秒

说明

- 通常情况下，对端重建会话连接的最大时间应小于 OPEN 消息中的 Holdtime 时间。接收侧（Receiving Speaker）检测到会话中断后，将等待指定时间以便重新建立会话连接。
- End-Of-RIB（End of Router-Information-Base，路由信息库结束）标记用来标定路由更新发送的结束。

1.10 开启 Trap 功能

开启 BGP 模块的 Trap 功能后，该模块会生成级别为 level4 的警告 Trap 报文，用于报告该模块的重要事件。生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-47 开启 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 BGP 模块的 Trap 功能	**snmp-agent trap enable bgp**	可选 缺省情况下，BGP 模块的 Trap 功能处于开启状态

1.11 使能 BGP 日志功能

表1-48 使能 BGP 日志功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	- 缺省情况下，系统没有运行 BGP
全局使能 BGP 日志功能	**log-peer-change**	可选 缺省情况下，全局 BGP 日志功能处于使能状态
记录指定对等体/对等体组的会话状态和事件信息	**peer** { *group-name* | *ip-address* } **log-change**	可选 记录对等体/对等体组的会话状态和事件信息

1.12 BGP 显示和维护

1.12.1 显示 BGP

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 BGP 的运行情况，通过查看显示信息验证配置的效果。

表1-49 BGP 配置显示

操作	命令
显示对等体组信息	**display bgp group** [*group-name*]
显示 BGP 发布的路由信息	**display bgp network**
显示 AS 路径信息	**display bgp paths** [*as-regular-expression*]
显示 BGP 对等体/对等体组的信息	**display bgp peer** [*ip-address* { **log-info** | **verbose** } | *group-name* **log-info** | **verbose**]
显示 BGP 路由信息	**display bgp routing-table** [*ip-address* [{ *mask* | *mask-length* } [**longer-prefixes**]]]
显示匹配指定 AS 路径过滤列表的 BGP 路由信息	**display bgp routing-table as-path-acl** *as-path-acl-number*
显示 CIDR 的 BGP 路由信息	**display bgp routing-table cidr**
显示匹配指定 BGP 团体的 BGP 路由信息	**display bgp routing-table community** [*aa:nn*&<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**]* [**whole-match**]
显示匹配指定 BGP 团体列表的 BGP 路由	**display bgp routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>
显示 BGP 衰减的 BGP 路由信息	**display bgp routing-table dampened**
显示 BGP 衰减的配置参数	**display bgp routing-table dampening parameter**
显示源 AS 不一致的路由	**display bgp routing-table different-origin-as**
显示 BGP 路由振荡统计信息	**display bgp routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *ip-address* [{ *mask* | *mask-length* } [**longer-match**]]]
显示 BGP 标签路由信息	display bgp routing-table label
显示向指定的 BGP 对等体发送或者从 BGP 对等体收到的路由信息	**display bgp routing-table peer** *ip-address* { **advertised-routes** | **received-routes** } [*ip-address* [*mask* | *mask-length*] | **statistic**]
显示匹配 AS 正则表达式的路由信息	**display bgp routing-table regular-expression** *as-regular-expression*
显示 BGP 的路由统计信息	**display bgp routing-table statistic**

1.12.2 复位 BGP 连接

当 BGP 路由策略或协议发生变化后，如果需要通过复位 BGP 连接使新的配置生效，请在用户视图下进行下列配置。

表1-50 复位 BGP 连接

操作	命令
复位所有 BGP 连接	**reset bgp all**
复位与指定 AS 之间的 BGP 连接	**reset bgp** *as-number*
复位与指定对等体的 BGP 连接	**reset bgp** *ip-address* [**flap-info**]
复位所有 EBGP 连接	**reset bgp external**
复位与指定对等体组的 BGP 连接	**reset bgp group** *group-name*

操作	命令
复位所有 IBGP 连接	**reset bgp internal**
复位 IPv4 单播地址族下的所有 BGP 连接	**reset bgp ipv4 all**

1.12.3 清除 BGP 信息

在用户视图下，执行 **reset** 命令可以清除 BGP 相关统计信息。

表1-51 清除 BGP 信息

操作	命令
清除路由的衰减信息并释放被抑制的路由	**reset bgp dampening** [*ip-address* [*mask* | *mask-length*]]
清除路由的振荡统计信息	**reset bgp flap-info** [**regexp** *as-path-regular-expression* | **as-path-acl** *as-path-acl-number* | *ip-address* [*mask* | *mask-length*]]

1.13 BGP 典型配置举例

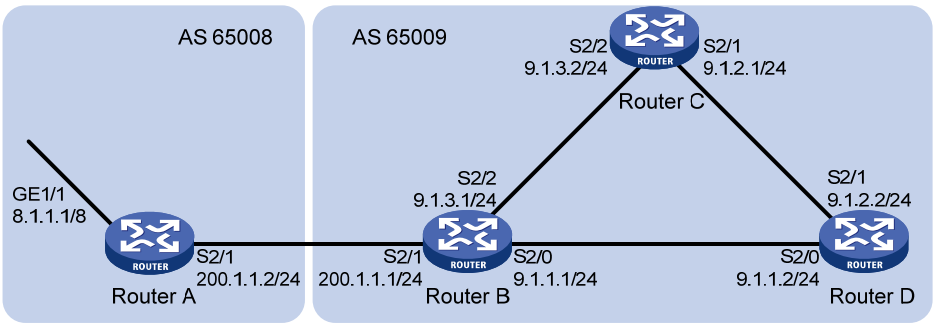
1.13.1 BGP 基本配置

1. 组网需求

如 图 1-20所示，所有路由器均运行BGP协议，Router A和Router B之间建立EBGP连接，Router B、Router C和Router D之间建立IBGP全连接。

2. 组网图

图1-20 BGP 基本配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IBGP 连接

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 65009
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 9.1.1.2 as-number 65009
[RouterB-bgp] peer 9.1.3.2 as-number 65009
```

```
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 65009
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] peer 9.1.3.1 as-number 65009
[RouterC-bgp] peer 9.1.2.2 as-number 65009
[RouterC-bgp] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] bgp 65009
[RouterD-bgp] router-id 4.4.4.4
[RouterD-bgp] peer 9.1.1.1 as-number 65009
[RouterD-bgp] peer 9.1.2.1 as-number 65009
[RouterD-bgp] quit
```

(3) 配置 EBGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 65008
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] peer 200.1.1.1 as-number 65009
```

将 8.0.0.0/8 网段路由通告到 BGP 路由表中。

```
[RouterA-bgp] network 8.0.0.0
[RouterA-bgp] quit
```

配置 Router B。

```
[RouterB] bgp 65009
[RouterB-bgp] peer 200.1.1.2 as-number 65008
[RouterB-bgp] quit
```

查看 Router B 的 BGP 对等体的连接状态。

```
[RouterB] display bgp peer

 BGP local router ID : 2.2.2.2
 Local AS number : 65009
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  9.1.1.2     4 65009       56       56     0       0 00:40:54 Established
  9.1.3.2     4 65009       49       62     0       0 00:44:58 Established
  200.1.1.2   4 65008       49       65     0       1 00:44:03 Established
```

可以看出，Router B 到其他路由器的 BGP 连接均已建立。

查看 Router A 路由表信息。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0           0.0.0.0       0                  0        i
```

显示 Router B 的路由表。

```
[RouterB] display bgp routing-table
 Total Number of Routes: 1

 BGP Local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0          200.1.1.2     0                  0       65008i
```

显示 Router C 的路由表。

```
[RouterC] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED       LocPrf    PrefVal Path/Ogn

   i  8.0.0.0          200.1.1.2     0         100       0       65008i
```



说明

从路由表可以看出，Router A 没有学到 AS 65009 内部的任何路由，Router C 虽然学到了 AS 65008 中的 8.0.0.0 的路由，但因为下一跳 200.1.1.2 不可达，所以也不是有效路由。

(4) 配置 BGP 引入直连路由

配置 Router B。

```
[RouterB] bgp 65009
[RouterB-bgp] import-route direct
```

显示 Router A 的 BGP 路由表。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 4

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0           0.0.0.0       0                  0       i
 *>   9.1.1.0/24        200.1.1.1     0                  0       65009?
 *>   9.1.3.0/24        200.1.1.1     0                  0       65009?
 *    200.1.1.0         200.1.1.1     0                  0       65009?
```

显示 Router C 的路由表。

```
[RouterC] display bgp routing-table

 Total Number of Routes: 4

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop       MED        LocPrf    PrefVal Path/Ogn

 *>i  8.0.0.0           200.1.1.2     0          100       0       65008i
 *>i  9.1.1.0/24        9.1.3.1       0          100       0       ?
 * i  9.1.3.0/24        9.1.3.1       0          100       0       ?
 *>i  200.1.1.0         9.1.3.1       0          100       0       ?
```

可以看出，到 8.0.0.0 的路由变为有效路由，下一跳为 Router A 的地址。

使用 Ping 进行验证。

```
[RouterC] ping 8.1.1.1
  PING 8.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 8.1.1.1: bytes=56 Sequence=1 ttl=254 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=2 ttl=254 time=47 ms
    Reply from 8.1.1.1: bytes=56 Sequence=3 ttl=254 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=4 ttl=254 time=16 ms
    Reply from 8.1.1.1: bytes=56 Sequence=5 ttl=254 time=31 ms

  --- 8.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 16/31/47 ms
```

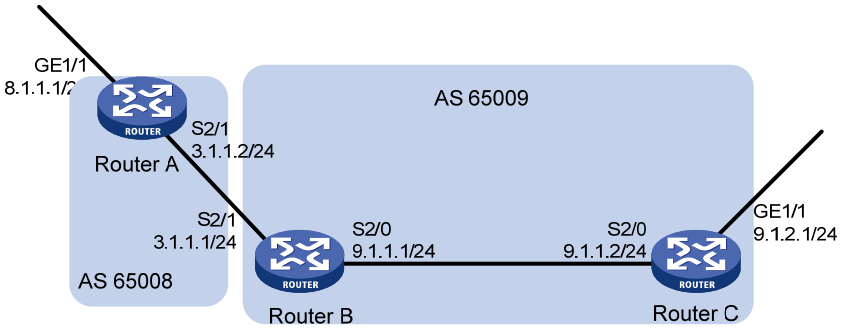
1.13.2 BGP 与 IGP 交互配置

1. 组网需求

- 在 AS 65009 内使用 OSPF 作为 IGP 协议；
- Router A 和 Router B 建立 EBGP 连接，Router C 为 AS 65009 内部的一台非 BGP 路由器。

2. 组网图

图1-21 BGP 与 IGP 交互配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 OSPF（略）

(3) 配置 EBGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 65008
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] peer 3.1.1.1 as-number 65009
```

将 8.1.1.0/24 网段通告到 BGP 路由表中。

```
[RouterA-bgp] network 8.1.1.0 24
[RouterA-bgp] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 65009
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 3.1.1.2 as-number 65008
[RouterB-bgp] quit
```

(4) 配置 BGP 与 IGP 交互

在 Router B 上配置 BGP 引入 OSPF 路由。

```
[RouterB] bgp 65009
[RouterB-bgp] import-route ospf 1
[RouterB-bgp] quit
```

查看 Router A 的路由表。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 3
 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.1.1.0/24        0.0.0.0       0                  0       i
 *>   9.1.2.0/24        3.1.1.1       1563               0       65009?
```

在 Router B 上配置 OSPF 引入 BGP 路由。

```
[RouterB] ospf
[RouterB-ospf-1] import-route bgp
[RouterB-ospf-1] quit
```

查看 RouterC 的路由表。

```
<RouterC> display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8

Destination/Mask    Proto  Pre  Cost       NextHop         Interface

8.1.1.0/24          O_ASE  150  1          9.1.1.1         S2/0
9.1.1.0/24          Direct 0    0          9.1.1.2         S2/0
9.1.1.1/32          Direct 0    0          9.1.1.1         S2/0
9.1.1.2/32          Direct 0    0          127.0.0.1       InLoop0
9.1.2.0/24          Direct 0    0          9.1.2.1         GE1/1
9.1.2.1/32          Direct 0    0          127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0          127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0          127.0.0.1       InLoop0
```

(5) 配置路由自动聚合

配置 Router B。

```
[RouterB] bgp 65009
[RouterB-bgp] summary automatic
```

显示 Router A 的 BGP 路由表。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
```

```
     Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>  8.1.1.0/24         0.0.0.0        0                    0       i
 *>  9.0.0.0            3.1.1.1        0                    0       65009?
```

使用 Ping 进行验证。

```
[RouterA] ping -a 8.1.1.1 9.1.2.1
  PING 9.1.2.1: 56  data bytes, press CTRL_C to break
    Reply from 9.1.2.1: bytes=56 Sequence=1 ttl=254 time=15 ms
    Reply from 9.1.2.1: bytes=56 Sequence=2 ttl=254 time=31 ms
    Reply from 9.1.2.1: bytes=56 Sequence=3 ttl=254 time=47 ms
    Reply from 9.1.2.1: bytes=56 Sequence=4 ttl=254 time=46 ms
    Reply from 9.1.2.1: bytes=56 Sequence=5 ttl=254 time=47 ms

  --- 9.1.2.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 15/37/47 ms
```

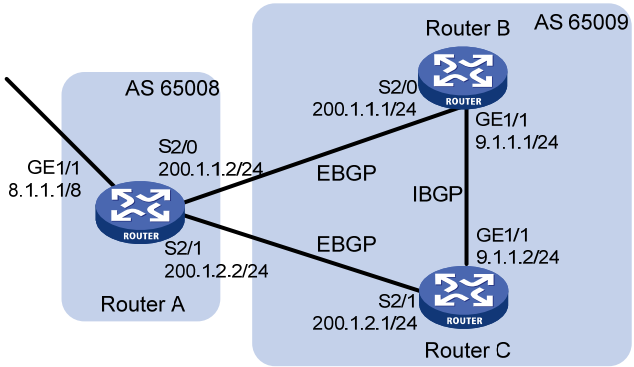
1.13.3 BGP 负载分担配置

1. 组网需求

- 所有路由器都配置 BGP，Router A 在 AS 65008 中，Router B 和 Router C 在 AS 65009 中。
- Router A 与 Router B、Router C 之间运行 EBGP，Router B 和 Router C 之间运行 IBGP。
- 在 Router A 上配置负载分担的路由条数为 2，以提高链路利用率。

2. 组网图

图1-22 BGP 负载分担配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 BGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 65008
[RouterA-bgp] router-id 1.1.1.1
```

```
[RouterA-bgp] peer 200.1.1.1 as-number 65009
[RouterA-bgp] peer 200.1.2.1 as-number 65009
[RouterA-bgp] network 8.0.0.0 255.0.0.0
[RouterA-bgp] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 65009
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 200.1.1.2 as-number 65008
[RouterB-bgp] peer 9.1.1.2 as-number 65009
[RouterB-bgp] network 9.1.1.0 255.255.255.0
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 65009
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] peer 200.1.2.2 as-number 65008
[RouterC-bgp] peer 9.1.1.1 as-number 65009
[RouterC-bgp] network 9.1.1.0 255.255.255.0
[RouterC-bgp] quit
```

查看 Router A 的路由表。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0           0.0.0.0         0                     0       i
 *>   9.1.1.0/24        200.1.1.1       0                     0       65009i
 *                      200.1.2.1       0                     0       65009i
```

从路由表中可以看出，到目的地址 9.1.1.0/24 有两条有效路由，其中下一跳为 200.1.1.1 的路由前有标志“*>”，表示它是当前有效的最优路由（因为 Router B 的路由器 ID 要小一些）；而下一跳为 200.1.2.1 的路由前只有标志“*”，表示它是当前有效的路由，但不是最优的。

(3) 配置负载分担

配置 Router A。

```
[RouterA] bgp 65008
[RouterA-bgp] balance 2
[RouterA-bgp] quit
```

查看 Router A 的路由表。

```
[RouterA] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0            0.0.0.0        0                    0       i
 *>   9.1.1.0/24         200.1.1.1      0                    0       65009i
 *>                      200.1.2.1      0                    0       65009i
```

从路由表中可以看到，BGP 路由 9.1.1.0/24 存在两个下一跳，分别是 200.1.1.1 和 200.1.2.1，两条路由前都有标志“*>”，表明它们都是当前有效的最优路由。

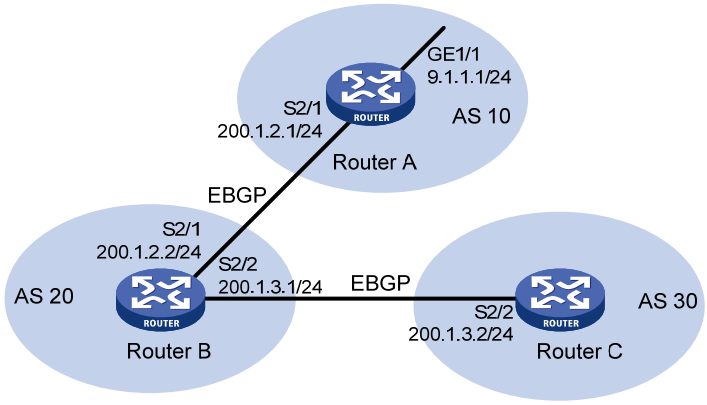
1.13.4 BGP 团体配置

1. 组网需求

- Router B 分别与 Router A、Router C 之间建立 EBGP 连接。
- 通过在 Router A 上配置 NO_EXPORT 团体属性，使得 AS 10 发布到 AS 20 中的路由，不再被 AS 20 向其他 AS 发布。

2. 组网图

图1-23 BGP 团体组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 EBGP

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 10
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] peer 200.1.2.2 as-number 20
```

```
[RouterA-bgp] network 9.1.1.0 255.255.255.0
[RouterA-bgp] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 20
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 200.1.2.1 as-number 10
[RouterB-bgp] peer 200.1.3.2 as-number 30
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 30
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] peer 200.1.3.1 as-number 20
[RouterC-bgp] quit
```

查看 Router B 的路由表。

```
[RouterB] display bgp routing-table 9.1.1.0

 BGP local router ID : 2.2.2.2
 Local AS number : 20
 Paths:   1 available, 1 best

BGP routing table entry information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Original nexthop: 200.1.2.1
 AS-path         : 10
 Origin          : igp
 Attribute value : MED 0, pref-val 0, pre 255
 State           : valid, external, best,
 Advertised to such 1 peers:
    200.1.3.2
```

可以看出，Router B 把收到的路由发布给了位于 AS 30 内的 Router C。

查看 Router C 的路由表。

```
[RouterC] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   9.1.1.0/24        200.1.3.1       0                     0       20 10i
```

从路由表可以确认，Router C 从 Router B 那里学到了目的地址为 9.1.1.0/24 的路由。

(3) 配置 BGP 团体属性

配置路由策略。

```
[RouterA] route-policy comm_policy permit node 0
[RouterA-route-policy] apply community no-export
[RouterA-route-policy] quit
```

应用路由策略。

```
[RouterA] bgp 10
[RouterA-bgp] peer 200.1.2.2 route-policy comm_policy export
[RouterA-bgp] peer 200.1.2.2 advertise-community
```

查看 Router B 的路由表。

```
[RouterB] display bgp routing-table 9.1.1.0
 BGP local router ID : 2.2.2.2
 Local AS number : 20
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Original nexthop: 200.1.2.1
 Community       : No-Export
 AS-path         : 10
 Origin          : igp
 Attribute value : MED 0, pref-val 0, pre 255
 State           : valid, external, best,
 Not advertised to any peers yet
```

在 Router B 的 BGP 路由表中可以看到配置的团体属性。此时在 Router C 的 BGP 路由表中已经没有到目的地址 9.1.1.0/24 的路由。

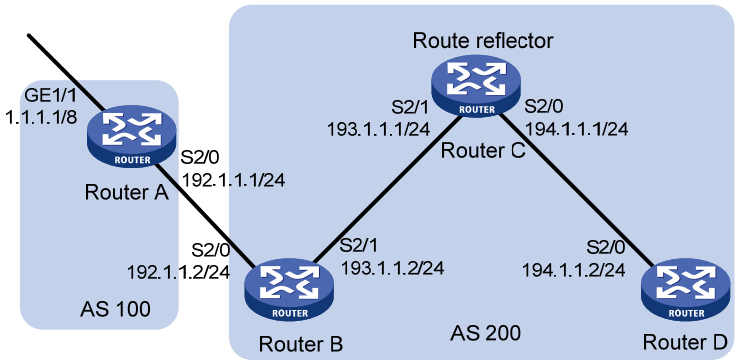
1.13.5 BGP 路由反射器配置

1. 组网需求

- 所有路由器运行 BGP 协议，Router A 与 Router B 建立 EBGP 连接，Router C 与 Router B 和 Router D 之间建立 IBGP 连接。
- Router C 作为路由反射器，Router B 和 Router D 为 Router C 的客户机。
- Router D 能够通过 Router C 学到路由 1.0.0.0/8。

2. 组网图

图1-24 配置 BGP 路由反射器的组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 BGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 100
[RouterA-bgp] peer 192.1.1.2 as-number 200
```

通告 1.0.0.0/8 网段路由到 BGP 路由表中。

```
[RouterA-bgp] network 1.0.0.0
[RouterA-bgp] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 200
[RouterB-bgp] peer 192.1.1.1 as-number 100
[RouterB-bgp] peer 193.1.1.1 as-number 200
[RouterB-bgp] peer 193.1.1.1 next-hop-local
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 200
[RouterC-bgp] peer 193.1.1.2 as-number 200
[RouterC-bgp] peer 194.1.1.2 as-number 200
[RouterC-bgp] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] bgp 200
[RouterD-bgp] peer 194.1.1.1 as-number 200
[RouterD-bgp] quit
```

(3) 配置路由反射器

配置 Router C。

```
[RouterC] bgp 200
[RouterC-bgp] peer 193.1.1.2 reflect-client
[RouterC-bgp] peer 194.1.1.2 reflect-client
[RouterC-bgp] quit
```

(4) 验证配置效果

查看 Router B 的 BGP 路由表。

```
[RouterB] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 200.1.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network        NextHop         MED        LocPrf     PrefVal Path/Ogn

 *>  1.0.0.0        192.1.1.1       0                     0       100i
```

查看 Router D 的 BGP 路由表。

```
[RouterD] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 200.1.2.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network        NextHop         MED        LocPrf     PrefVal Path/Ogn

   i 1.0.0.0        193.1.1.2       0          100        0       100i
```

可以看出，Router D 从 Router C 已经学到了 1.0.0.0/8 路由。

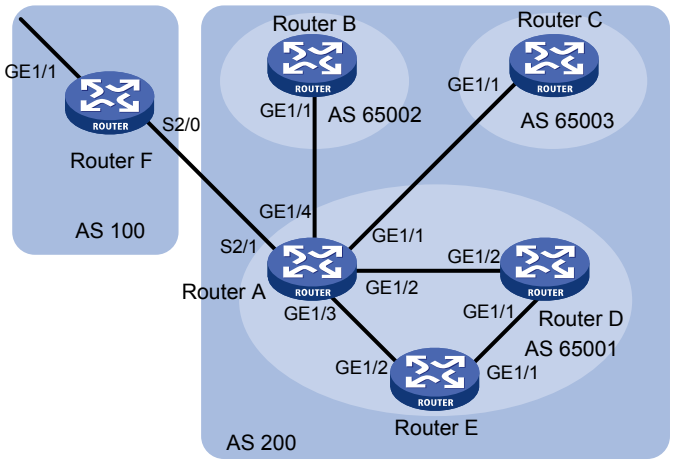
1.13.6 BGP 联盟配置

1. 组网需求

AS 200 中有多台 BGP 路由器，为了减少 IBGP 的连接数，现将他们划分为 3 个子自治系统：AS 65001、AS 65002 和 AS 65003。其中 AS 65001 内的三台路由器建立 IBGP 全连接。

2. 组网图

图1-25 配置联盟组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	S2/1	200.1.1.1/24	Router D	GE1/1	10.1.5.1/24
	GE1/1	10.1.2.1/24		GE1/2	10.1.3.2/24
	GE1/2	10.1.3.1/24	Router E	GE1/1	10.1.5.2/24
	GE1/3	10.1.4.1/24		GE1/2	10.1.4.2/24
	GE1/4	10.1.1.1/24	Router F	GE1/1	9.1.1.1/24
Router B	GE1/1	10.1.1.2/24		S2/0	200.1.1.2/24
Router C	GE1/1	10.1.2.2/24			

3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 BGP 联盟

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 65001
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] confederation id 200
[RouterA-bgp] confederation peer-as 65002 65003
[RouterA-bgp] peer 10.1.1.2 as-number 65002
[RouterA-bgp] peer 10.1.1.2 next-hop-local
[RouterA-bgp] peer 10.1.2.2 as-number 65003
[RouterA-bgp] peer 10.1.2.2 next-hop-local
[RouterA-bgp] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 65002
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] confederation id 200
[RouterB-bgp] confederation peer-as 65001 65003
[RouterB-bgp] peer 10.1.1.1 as-number 65001
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
```

```
[RouterC] bgp 65003
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] confederation id 200
[RouterC-bgp] confederation peer-as 65001 65002
[RouterC-bgp] peer 10.1.2.1 as-number 65001
[RouterC-bgp] quit
```

(3) 配置 AS 65001 内的 IBGP 连接

配置 Router A。

```
[RouterA] bgp 65001
[RouterA-bgp] peer 10.1.3.2 as-number 65001
[RouterA-bgp] peer 10.1.3.2 next-hop-local
[RouterA-bgp] peer 10.1.4.2 as-number 65001
[RouterA-bgp] peer 10.1.4.2 next-hop-local
[RouterA-bgp] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] bgp 65001
[RouterD-bgp] router-id 4.4.4.4
[RouterD-bgp] confederation id 200
[RouterD-bgp] peer 10.1.3.1 as-number 65001
[RouterD-bgp] peer 10.1.5.2 as-number 65001
[RouterD-bgp] quit
```

配置 Router E。

```
<RouterE> system-view
[RouterE] bgp 65001
[RouterE-bgp] router-id 5.5.5.5
[RouterE-bgp] confederation id 200
[RouterE-bgp] peer 10.1.4.1 as-number 65001
[RouterE-bgp] peer 10.1.5.1 as-number 65001
[RouterE-bgp] quit
```

(4) 配置 AS 100 和 AS 200 之间的 EBGP 连接

配置 Router A。

```
[RouterA] bgp 65001
[RouterA-bgp] peer 200.1.1.2 as-number 100
[RouterA-bgp] quit
```

配置 Router F。

```
<RouterF> system-view
[RouterF] bgp 100
[RouterF-bgp] router-id 6.6.6.6
[RouterF-bgp] peer 200.1.1.1 as-number 200
[RouterF-bgp] network 9.1.1.0 255.255.255.0
[RouterF-bgp] quit
```

(5) 查看配置结果

查看 RouterB 的 BGP 路由表。

```
[RouterB] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED      LocPrf   PrefVal Path/Ogn

 *>i  9.1.1.0/24        10.1.1.1       0         100       0       (65001) 100i
[RouterB] display bgp routing-table 9.1.1.0

 BGP local router ID : 2.2.2.2
 Local AS number : 65002
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 10.1.1.1 (1.1.1.1)
 Relay Nexthop   : 0.0.0.0
 Original nexthop: 10.1.1.1
 AS-path         : (65001) 100
 Origin          : igp
 Attribute value : MED 0, localpref 100, pref-val 0, pre 255
 State           : valid, external-confed, best,
 Not advertised to any peers yet
```

查看 Router D 的 BGP 路由表。

```
[RouterD] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>i  9.1.1.0/24        10.1.3.1      0          100      0       100i
[RouterD] display bgp routing-table 9.1.1.0

 BGP local router ID : 4.4.4.4
 Local AS number : 65001
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 10.1.3.1 (1.1.1.1)
 Relay Nexthop   : 0.0.0.0
 Original nexthop: 10.1.3.1
```

```
AS-path         : 100
Origin          : igp
Attribute value : MED 0, localpref 100, pref-val 0, pre 255
State           : valid, internal, best,
Not advertised to any peers yet
```

通过以上显示信息可以看出：

- Router F 只需要和 Router A 建立 EBGP 连接，而不需要和 Router B、Router C 建立连接，同样可以通过联盟将路由信息传递给 Router B 和 Router C。
- Router B 和 Router D 在同一个联盟里，但是属于不同的子自治域，它们都是通过 Router A 来获取外部路由信息，生成的 BGP 路由表项也是一致的，等效于在同一个自治域内，但是又不需要物理上全连接。

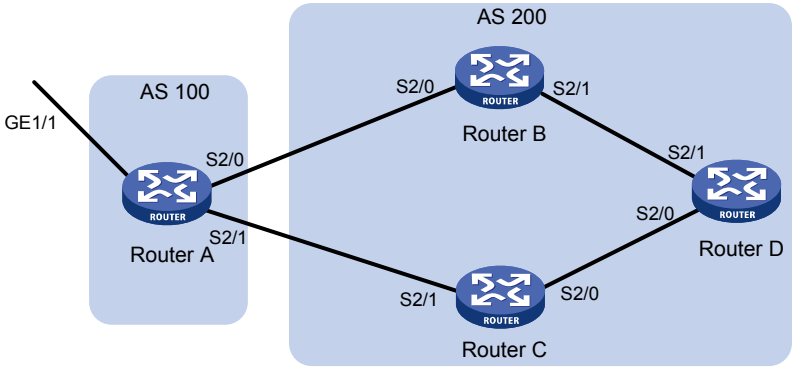
1.13.7 BGP 路径选择配置

1. 组网需求

- 所有路由器都运行 BGP 协议。Router A 与 Router B 和 Router C 之间运行 EBGP；Router D 与 Router B 和 Router C 之间运行 IBGP。
- AS 200 中运行 OSPF 协议。
- 配置不同的路由策略，使得 Router D 优选 Router C 学到的 1.0.0.0/8 路由。

2. 组网图

图1-26 配置 BGP 路径选择的组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE1/1	1.0.0.0/8	Router D	S2/0	195.1.1.1/24
	S2/0	192.1.1.1/24		S2/1	194.1.1.1/24
	S2/1	193.1.1.1/24	Router C	S2/0	195.1.1.2/24
Router B	S2/0	192.1.1.2/24		S2/1	193.1.1.2/24
	S2/1	194.1.1.2/24			

3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 Router B、Router C 和 Router D 之间运行 OSPF 协议

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
```

```
[RouterB-ospf] area 0
[RouterB-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf
[RouterC-ospf] area 0
[RouterC-ospf-1-area-0.0.0.0] network 193.1.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] quit
[RouterC-ospf-1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ospf
[RouterD-ospf] area 0
[RouterD-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.0] quit
[RouterD-ospf-1] quit
```

(3) 配置 BGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 100
[RouterA-bgp] peer 192.1.1.2 as-number 200
[RouterA-bgp] peer 193.1.1.2 as-number 200
```

将 1.0.0.0/8 网段通告到 Router A 的 BGP 路由表中。

```
[RouterA-bgp] network 1.0.0.0 8
[RouterA-bgp] quit
```

配置 Router B。

```
[RouterB] bgp 200
[RouterB-bgp] peer 192.1.1.1 as-number 100
[RouterB-bgp] peer 194.1.1.1 as-number 200
[RouterB-bgp] quit
```

配置 Router C。

```
[RouterC] bgp 200
[RouterC-bgp] peer 193.1.1.1 as-number 100
[RouterC-bgp] peer 195.1.1.1 as-number 200
[RouterC-bgp] quit
```

配置 Router D。

```
[RouterD] bgp 200
[RouterD-bgp] peer 194.1.1.2 as-number 200
[RouterD-bgp] peer 195.1.1.2 as-number 200
```

```
[RouterD-bgp] quit
```

(4) 通过配置 1.0.0.0/8 路由的不同属性值，使得 Router D 优选 Router C 学到的路由。

- 在 Router A 上对发布给对等体 192.1.1.2 的 1.0.0.0/8 路由配置较高的 MED 属性值，使得 Router D 优选 Router C 学到的路由。

定义编号为 2000 的 ACL，允许路由 1.0.0.0/8 通过。

```
[RouterA] acl number 2000
[RouterA-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[RouterA-acl-basic-2000] quit
```

定义两个 Route-policy，一个名为 apply_med_50，为路由 1.0.0.0/8 设置 MED 属性值为 50；另一个名为 apply_med_100，为路由 1.0.0.0/8 设置 MED 属性值为 100。

```
[RouterA] route-policy apply_med_50 permit node 10
[RouterA-route-policy] if-match acl 2000
[RouterA-route-policy] apply cost 50
[RouterA-route-policy] quit
[RouterA] route-policy apply_med_100 permit node 10
[RouterA-route-policy] if-match acl 2000
[RouterA-route-policy] apply cost 100
[RouterA-route-policy] quit
```

对发布给对等体 193.1.1.2（Router C）的路由应用名为 apply_med_50 的 Route-policy，对发布给对等体 192.1.1.2（Router B）的路由应用名为 apply_med_100 的 Route-policy。

```
[RouterA] bgp 100
[RouterA-bgp] peer 193.1.1.2 route-policy apply_med_50 export
[RouterA-bgp] peer 192.1.1.2 route-policy apply_med_100 export
[RouterA-bgp] quit
```

查看 Router D 的 BGP 路由表。

```
[RouterD] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 194.1.1.1
 Status codes: * - valid, > - best, d – damped,
               h – history,  i – internal, s – suppressed, S – Stale
               Origin : i – IGP, e – EGP, ? – incomplete
     Network         NextHop        MED        LocPrf     PrefVal Path/Ogn

 *>i 1.0.0.0         193.1.1.1      50         100        0       100i
 * i                 192.1.1.1      100        100        0       100i
```

可以看到，Router D 从 Router C 学到 1.0.0.0/8 的路由是最优的。

- 在 Router B 和 Router C 上分别对 1.0.0.0/8 路由配置不同的本地优先级，使得 Router D 优选 Router C 学到的路由。

在 Router C 上定义编号为 2000 的 ACL，允许 1.0.0.0/8 路由通过。

```
[RouterC] acl number 2000
[RouterC-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
```

```
[RouterC-acl-basic-2000] quit
```

在 Router C 上定义名为 localpref 的 Route-policy，设置路由 1.0.0.0/8 的本地优先级为 200（缺省的本地优先级为 100）。

```
[RouterC] route-policy localpref permit node 10
[RouterC-route-policy] if-match acl 2000
[RouterC-route-policy] apply local-preference 200
[RouterC-route-policy] quit
```

为从 BGP 对等体 193.1.1.1 的路由应用名为 localpref 的 Router-policy。

```
[RouterC] bgp 200
[RouterC-bgp] peer 193.1.1.1 route-policy localpref import
[RouterC-bgp] quit
```

查看 Router D 的 BGP 路由表。

```
[RouterD] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 194.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network          NextHop       MED        LocPrf     PrefVal Path/Ogn

 *>i 1.0.0.0          193.1.1.1     0          200        0       100i
 * i                  192.1.1.1     0          100        0       100i
```

可以看到，Router D 从 Router C 学到 1.0.0.0/8 的路由是最优的。

1.14 BGP 常见错误配置举例

1.14.1 BGP 对等体关系不能建立

1. 故障现象

使用 **display bgp peer** 命令查看 BGP 对等体的信息，发现与对端的连接无法进入 Established 状态。

2. 分析

BGP 邻居的建立需要能够使用 179 端口建立 TCP 会话，以及能够正确交换 Open 消息。

3. 处理过程

(1) 用 **display current-configuration** 命令检查邻居的 AS 号配置是否正确。

(2) 用 **display bgp peer** 命令检查邻居的 IP 地址是否正确。

(3) 如果使用 Loopback 接口，检查是否配置了 **peer connect-interface** 命令。

(4) 如果是物理上非直连的 EBGP 邻居，检查是否配置了 **peer ebgp-max-hop** 命令。

(5) 检查路由表中是否存在到邻居的可用路由。

(6) 使用 **ping** 命令检查链路是否畅通。

(7) 使用 **display tcp status** 命令检查 TCP 连接是否正常。

(8) 检查是否配置了禁止 TCP 端口 179 的 ACL。

目 录

1 IS-IS配置 1-1
1.1 IS-IS简介 1-1
1.1.1 基本概念 1-1
1.1.2 IS-IS区域 1-3
1.1.3 IS-IS的网络类型 1-5
1.1.4 IS-IS的PDU格式 1-6
1.1.5 支持的IS-IS特性 1-12
1.1.6 协议规范 1-14
1.2 IS-IS配置任务简介 1-15
1.3 配置IS-IS基本功能 1-16
1.3.1 配置准备 1-16
1.3.2 使能IS-IS功能 1-16
1.3.3 配置路由级别 1-16
1.3.4 配置接口网络类型 1-17
1.4 配置IS-IS路由信息控制 1-18
1.4.1 配置准备 1-18
1.4.2 配置IS-IS链路度量值 1-18
1.4.3 配置IS-IS路由优先级 1-19
1.4.4 配置IS-IS最大等价路由条数 1-19
1.4.5 配置IS-IS路由聚合 1-20
1.4.6 配置IS-IS发布缺省路由 1-20
1.4.7 配置IS-IS引入外部路由 1-21
1.4.8 配置IS-IS路由过滤 1-21
1.4.9 配置IS-IS路由渗透 1-22
1.5 配置IS-IS调整和优化 1-23
1.5.1 配置准备 1-23
1.5.2 配置Hello/CSNP报文发送时间间隔 1-23
1.5.3 配置Hello报文失效数目 1-23
1.5.4 配置DIS优先级 1-24
1.5.5 禁止接口发送和接收IS-IS报文 1-24
1.5.6 配置在PPP接口上取消建立邻接关系必须在同一网段的限制 1-24
1.5.7 配置接口发送小型Hello报文 1-25
1.5.8 配置LSP参数 1-25
1.5.9 配置SPF参数 1-29
1.5.10 配置LSDB过载标志位 1-29
1.6 配置IS-IS验证功能 1-30
1.6.1 配置准备 1-30
1.6.2 配置邻居关系验证 1-30
1.6.3 配置区域验证 1-30
1.6.4 配置路由域验证 1-31
1.7 配置IS-IS主机名映射 1-31
1.7.1 配置IS-IS静态主机名映射 1-31

1.7.2 配置IS-IS动态主机名映射 1-32
1.8 配置IS-IS GR 1-32
1.9 配置邻接状态输出 1-33
1.10 使能IS-IS的TRAP功能 1-34
1.11 配置IS-IS和MIB绑定 1-34
1.12 IS-IS显示和维护 1-34
1.13 IS-IS典型配置举例 1-35
1.13.1 IS-IS基本配置举例 1-35
1.13.2 配置IS-IS的DIS选择 1-39
1.13.3 配置IS-IS引入外部路由 1-43
1.13.4 IS-IS GR配置举例 1-47
1.13.5 IS-IS验证配置举例 1-49

1 IS-IS 配置

1.1 IS-IS 简介

IS-IS（Intermediate System-to-Intermediate System intra-domain routing information exchange protocol，中间系统到中间系统的域内路由信息交换协议）最初是国际标准化组织（International Organization for Standardization，ISO）为它的无连接网络协议（ConnectionLess Network Protocol，CLNP）设计的一种动态路由协议。

为了提供对 IP 的路由支持，IETF（Internet Engineering Task Force，因特网工程任务组）在 RFC 1195 中对 IS-IS 进行了扩充和修改，使它能够同时应用在 TCP/IP 和 OSI 环境中，称为集成化 IS-IS（Integrated IS-IS 或 Dual IS-IS）。

IS-IS 属于内部网关协议（Interior Gateway Protocol，IGP），用于自治系统内部。IS-IS 是一种链路状态协议，使用最短路径优先（Shortest Path First，SPF）算法进行路由计算。

1.1.1 基本概念

1. IS-IS 路由协议的基本术语

- IS（Intermediate System）：中间系统。相当于 TCP/IP 中的路由器，是 IS-IS 协议中生成路由和传播路由信息的基本单元。在下文中 IS 和路由器具有相同的含义。
- ES（End System）：终端系统。相当于 TCP/IP 中的主机系统。ES 不参与 IS-IS 路由协议的处理，ISO 使用专门的 ES-IS 协议定义终端系统与中间系统间的通信。
- RD（Routing Domain）：路由域。在一个路由域中多个 IS 通过相同的路由协议来交换路由信息。
- Area：区域，路由域的细分单元，IS-IS 允许将整个路由域分为多个区域。
- LSDB（Link State DataBase）：链路状态数据库。网络内所有链路的状态组成了链路状态数据库，在每一个 IS 中都至少有一个 LSDB。IS 使用 SPF 算法，利用 LSDB 来生成自己的路由。
- LSPDU（Link State Protocol Data Unit）：链路状态协议数据单元，简称 LSP。在 IS-IS 中，每一个 IS 都会生成 LSP，此 LSP 包含了本 IS 的所有链路状态信息。
- NPDU（Network Protocol Data Unit）：网络协议数据单元，是 OSI 中的网络层协议报文，相当于 TCP/IP 中的 IP 报文。
- DIS（Designated IS）：广播网络上选举的指定中间系统，也可以称为指定 IS。
- NSAP（Network Service Access Point）：网络服务接入点，即 OSI 中网络层的地址，用来标识一个抽象的网络服务访问点，描述 OSI 模型的网络地址结构。

2. IS-IS 地址结构

(1) NSAP

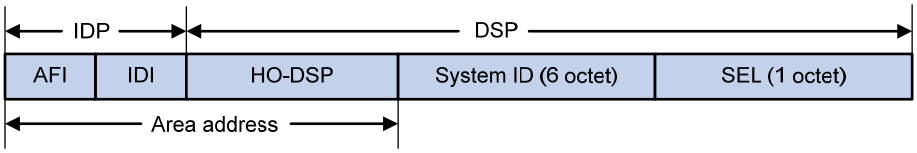
如 图 1-1所示，NSAP由IDP（Initial Domain Part）和DSP（Domain Specific Part）组成。IDP相当于IP地址中的主网络号，DSP相当于IP地址中的子网号和主机地址。

IDP 部分是 ISO 规定的，它由 AFI（Authority and Format Identifier）与 IDI（Initial Domain Identifier）组成，AFI 表示地址分配机构和地址格式，IDI 用来标识域。

DSP 由 HO-DSP（High Order Part of DSP）、SystemID 和 SEL 三个部分组成。HO-DSP 用来分割区域，SystemID 用来区分主机，SEL 指示服务类型。

IDP 和 DSP 的长度都是可变的，NSAP 总长最多是 20 个字节，最少 8 个字节。

图1-1 IS-IS 协议的地址结构示意图



(2) 区域地址

IDP 和 DSP 中的 HODSP 一起，既能够标识路由域，也能够标识路由域中的区域，因此，它们一起被称为区域地址（Area Address）。两个不同的路由域中不允许有相同的区域地址。

一般情况下，一台路由器只需要配置一个区域地址，且同一区域中所有节点的区域地址都要相同。为了支持区域的平滑合并、分割及转换，一台路由器最多可配置 3 个区域地址。

(3) System ID

System ID 用来在区域内唯一标识主机或路由器。它的长度固定为 48 比特。

在实际应用中，一般使用 Router ID 与 System ID 进行对应。假设一台路由器使用接口 Loopback0 的 IP 地址 168.10.1.1 作为 Router ID，则它在 IS-IS 使用的 System ID 可通过如下方法转换得到：

- 将 IP 地址 168.10.1.1 的每一部分都扩展为 3 位，不足 3 位的在前面补 0；
- 将扩展后的地址 168.010.001.001 重新划分为 3 部分，每部分由 4 位数字组成，得到的 1680.1000.1001 就是 System ID。

实际 System ID 的指定可以有不同的方法，但要保证能够唯一标识主机或路由器。

(4) SEL

SEL（NSAP Selector，有时也写成 N-SEL）的作用类似 IP 中的“协议标识符”，不同的传输协议对应不同的 SEL。在 IP 中，SEL 均为 00。

(5) 路由方式

由于这种地址结构明确的定义了区域，Level-1 路由器很容易识别出发往它所在的区域之外的报文，这些报文是需要转交给 Level-1-2 路由器的。

- Level-1 路由器利用 System ID 进行区域内的路由，如果发现报文的目的地址不属于自己所在的区域，就将报文转发给最近的 Level-1-2 路由器。
- Level-2 路由器根据区域地址进行区域间的路由。

3. NET

NET（Network Entity Title，网络实体名称）指示的是 IS 本身的网络层信息，不包括传输层信息，可以看作是一类特殊的 NSAP，即 SEL 为 0 的 NSAP 地址。因此，NET 的长度与 NSAP 的相同，最多为 20 个字节，最少为 8 个字节。

通常情况下，一台路由器配置一个 NET 即可，当区域需要重新划分时，例如将多个区域合并，或者将一个区域划分为多个区域，这种情况下配置多个 NET 可以在重新配置时仍然能够保证路由的正确性。由于一台路由器最多可配置 3 个区域地址，所以最多也只能配置 3 个 NET。在配置多个 NET 时，必须保证它们的 System ID 都相同。

例如 NET 为：ab.cdef.1234.5678.9abc.00，则其中 Area 为 ab.cdef，System ID 为 1234.5678.9abc，SEL 为 00。

1.1.2 IS-IS 区域

1. 两级结构

为了支持大规模的路由网络，IS-IS 在路由域内采用两级的分层结构。一个大的路由域通常被分成多个区域（Areas）。一般来说，我们将 Level-1 路由器部署在区域内，Level-2 路由器部署在区域间，Level-1-2 路由器部署在 Level-1 路由器和 Level-2 路由器的中间。

2. Level-1 与 Level-2

(1) Level-1 路由器

Level-1 路由器负责区域内的路由，它只与属于同一区域的 Level-1 和 Level-1-2 路由器形成邻居关系，维护一个 Level-1 的 LSDB，该 LSDB 包含本区域的路由信息，到区域外的报文转发给最近的 Level-1-2 路由器。

(2) Level-2 路由器

Level-2 路由器负责区域间的路由，可以与同一区域或者其它区域的 Level-2 和 Level-1-2 路由器形成邻居关系，维护一个 Level-2 的 LSDB，该 LSDB 包含区域间的路由信息。所有 Level-2 路由器和 Level-1-2 路由器组成路由域的骨干网，负责在不同区域间通信，骨干网必须是物理连续的。

(3) Level-1-2 路由器

同时属于 Level-1 和 Level-2 的路由器称为 Level-1-2 路由器，可以与同一区域的 Level-1 和 Level-1-2 路由器形成 Level-1 邻居关系，也可以与同一区域或者其他区域的 Level-2 和 Level-1-2 路由器形成 Level-2 的邻居关系。Level-1 路由器必须通过 Level-1-2 路由器才能连接至其他区域。Level-1-2 路由器维护两个 LSDB，Level-1 的 LSDB 用于区域内路由，Level-2 的 LSDB 用于区域间路由。



说明

- 属于不同区域的 Level-1 路由器不能形成邻居关系。
- Level-2 路由器是否形成邻居关系则与区域无关。

图 1-2为一个运行IS-IS协议的网络，其中Area 1 是骨干区域，该区域中的所有路由器均是Level-2 路由器。另外 4 个区域为非骨干区域，它们都通过Level-1-2 路由器与骨干路由器相连。

图1-2 IS-IS 拓扑结构图之一

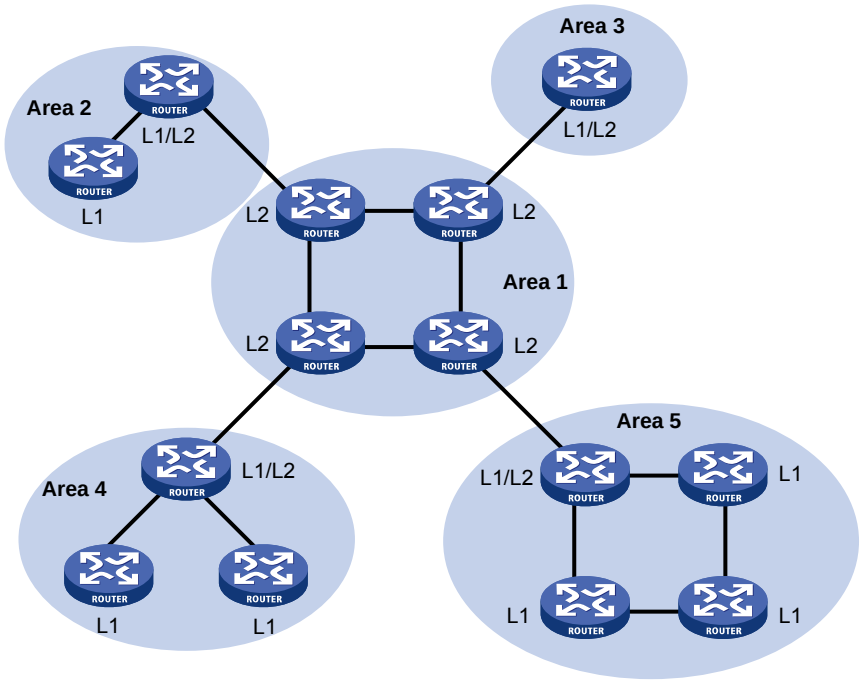
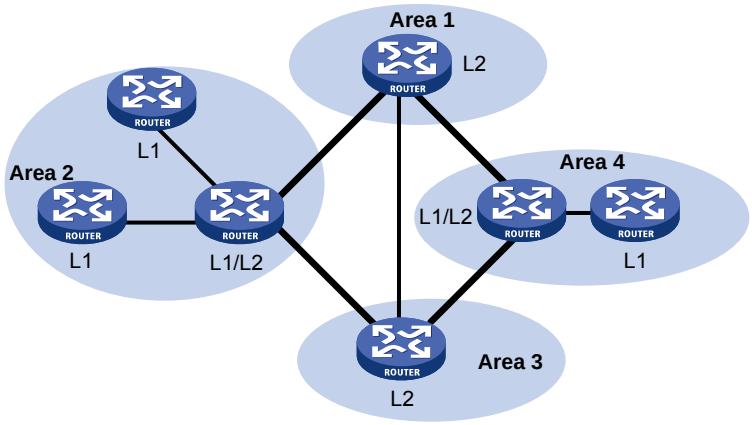


图 1-3是IS-IS的另外一种拓扑结构图。其中Level-1-2 路由器不仅仅用来连接Level-1 和Level-2 路由器，而且还与其它Level-2 路由器一起构成了IS-IS的骨干网。在这个拓扑中，并没有规定哪个区域是骨干区域。所有Level-2 路由器和Level-1-2 路由器构成了IS-IS的骨干网，他们可以属于不同的区域，但必须是物理连续的。

图1-3 IS-IS 拓扑结构图之二



说明

IS-IS 的骨干网（Backbone）指的不是一个特定的区域。

IS-IS 不论是 Level-1 还是 Level-2 路由，都采用 SPF 算法，分别生成最短路径树（Shortest Path Tree，SPT）。

3. 路由渗透

通常情况下，区域内的路由通过 Level-1 的路由器进行管理。所有的 Level-2 路由器和 Level-1-2 路由器构成一个 Level-2 区域。因此，一个 IS-IS 的路由域可以包含多个 Level-1 区域，但只有一个 Level-2 区域。

Level-1 区域必须且只能与 Level-2 区域相连，不同的 Level-1 区域之间并不相连。

Level-1 区域内的路由信息通过 Level-1-2 路由器发布到 Level-2 区域，因此，Level-2 路由器知道整个 IS-IS 路由域的路由信息。但是，在缺省情况下，Level-2 路由器并不将自己知道的其它 Level-1 区域以及 Level-2 区域的路由信息发布到 Level-1 区域。这样，Level-1 路由器将不了解本区域以外的路由信息，Level-1 路由器只将去往其它区域的报文发送到最近的 Level-1-2 路由器，所以可能导致对本区域之外的目的地址无法选择最佳的路由。

为解决上述问题，IS-IS 提供了路由渗透功能，使 Level-1-2 路由器可以将已知的其它 Level-1 区域以及 Level-2 区域的路由信息发布到指定的 Level-1 区域。

1.1.3 IS-IS 的网络类型

1. 网络类型

IS-IS 只支持两种类型的网络，根据物理链路不同可分为：

- 广播链路：如 Ethernet、Token-Ring 等。
- 点到点链路：如 PPP、HDLC 等。



说明

对于 NBMA（Non-Broadcast Multi-Access）网络，如 ATM，需对其配置子接口，并将子接口类型配置为点到点网络或广播网络。IS-IS 不能在点到多点（Point to MultiPoint，P2MP）链路上运行。

---

2. DIS 和伪节点

在广播网络中，IS-IS 需要在所有的路由器中选举一个路由器作为 DIS。

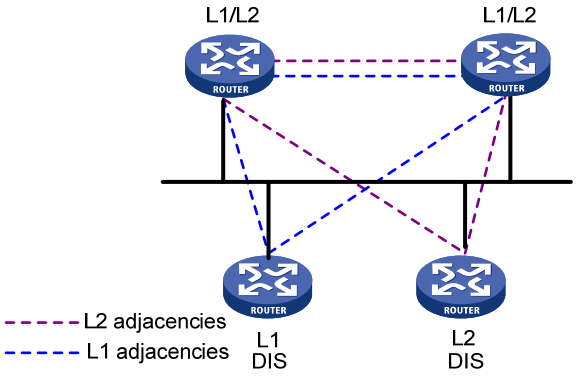
Level-1 和 Level-2 的 DIS 是分别选举的，用户可以为不同级别的 DIS 选举设置不同的优先级。DIS 优先级数值越高，被选中的可能性就越大。如果优先级最高的路由器有多台，则其中 SNPA（Subnetwork Point of Attachment，子网连接点）地址（广播网络中的 SNPA 地址是 MAC 地址）最大的路由器会被选中。不同级别的 DIS 可以是同一台路由器，也可以是不同的路由器。

与 OSPF 的不同点：

- 优先级为 0 的路由器也参与 DIS 的选举；
- 当有新的路由器加入，并符合成为 DIS 的条件时，这个路由器会被选中成为新的 DIS，此更改会引起一组新的 LSP 泛洪。

在IS-IS广播网中，同一网段上的同一级别的路由器之间都会形成邻接关系，包括所有的非DIS路由器之间也会形成邻接关系。如 图 1-4所示。

图1-4 IS-IS 广播网的 DIS 和邻接关系



DIS 用来创建和更新伪节点（Pseudonodes），并负责生成伪节点的 LSP，用来描述这个网络上有哪些路由器。

伪节点是用来模拟广播网络的一个虚拟节点，并非真实的路由器。在 IS-IS 中，伪节点用 DIS 的 System ID 和一个字节的 Circuit ID（非 0 值）标识。

使用伪节点可以简化网络拓扑，减少 SPF 的资源消耗。

说明

IS-IS 广播网络上所有的路由器之间都形成邻接关系，但 LSDB 的同步仍然依靠 DIS 来保证。

1.1.4 IS-IS 的 PDU 格式

1. PDU 头格式

IS-IS报文是直接封装在数据链路层的帧结构中的。PDU（Protocol Data Unit，协议数据单元）可以分为两个部分，报文头和变长字段部分。其中报文头又可分为通用报头和专用报头。对于所有PDU来说，通用报头都是相同的，但专用报头根据PDU类型不同而有所差别，如 图 1-5所示。

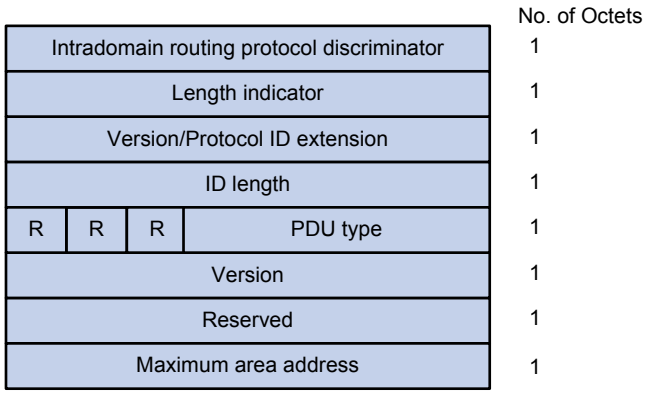
图1-5 PDU 格式

PDU common header	PDU specific header	Variable length fields (CLV)

2. 通用报头格式

所有的PDU都有相同的通用报头格式，如 图 1-6所示。

图1-6 PDU 头格式



主要字段的解释如下：

- Intradomain Routing Protocol Discriminator（域内路由协议鉴别符）：设置为 0x83。
- Length Indicator（长度标识符）：PDU 头部的长度（包括通用报头和专用报头），以字节为单位。
- Version/Protocol ID Extension（版本/协议标识扩展）：设置为 1（0x01）。
- ID Length（标识长度）：NSAP 地址和 NET 的 ID 长度。
- R（Reserved，保留）：设置为 0。
- PDU Type（PDU类型）：详细信息请参考 表 1-1。
- Version（版本）：设置为 1（0x01）。
- Maximum Area Address（最大区域地址数）：支持的最大区域个数。

表1-1 PDU 类型对应关系表

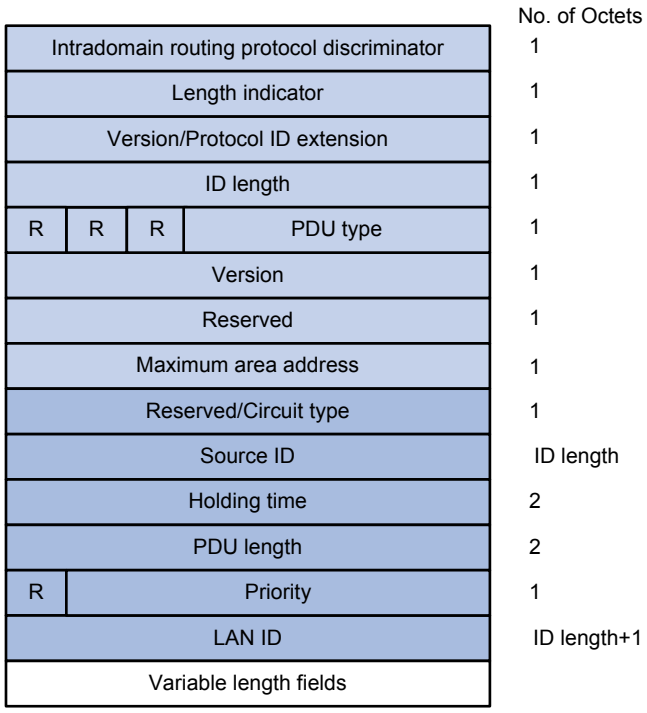
类型值	PDU 类型	简称
15	Level-1 LAN IS-IS Hello PDU	L1 LAN IIH
16	Level-2 LAN IS-IS Hello PDU	L2 LAN IIH
17	Point-to-Point IS-IS Hello PDU	P2P IIH
18	Level-1 Link State PDU	L1 LSP
20	Level-2 Link State PDU	L2 LSP
24	Level-1 Complete Sequence Numbers PDU	L1 CSNP
25	Level-2 Complete Sequence Numbers PDU	L2 CSNP
26	Level-1 Partial Sequence Numbers PDU	L1 PSNP
27	Level-2 Partial Sequence Numbers PDU	L2 PSNP

3. Hello

Hello 报文用于建立和维持邻居关系，也称为 IIH（IS-to-IS Hello PDUs）。其中，广播网中的 Level-1 路由器使用 Level-1 LAN IIH，广播网中的 Level-2 路由器使用 Level-2 LAN IIH，点到点网络中的路由器则使用 P2P IIH。

它们的报文格式有所不同。广播网中的Hello报文格式如 图 1-7所示（浅蓝色部分是通用报文头）。

图1-7 L1/L2 LAN IIH 格式

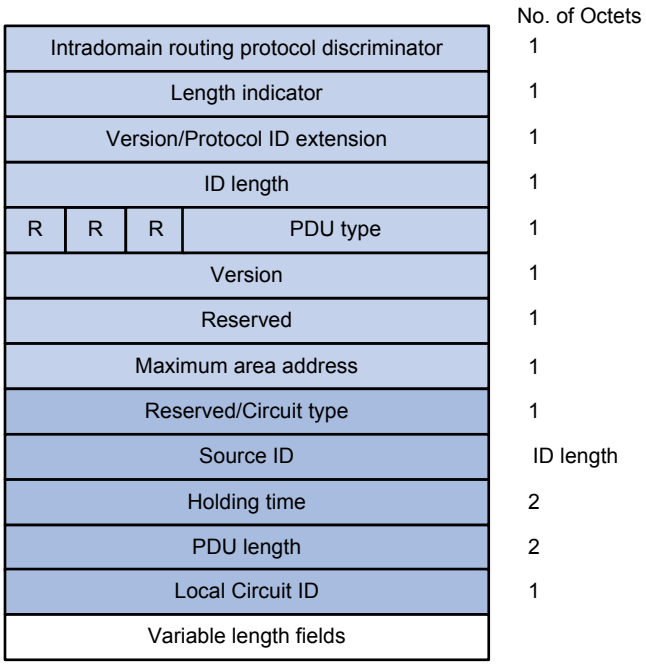


主要字段的解释如下：

- Reserved/Circuit Type：高位的 6 比特保留，值为 0。低位的 2 比特表示路由器的类型（00 保留，01 表示 L1，10 表示 L2，11 表示 L1/2）。
- Source ID：发送 Hello 报文的路由器的 System ID。
- Holding Time：保持时间。在此时间内如果没有收到邻居发来的 Hello 报文，则中止已建立的邻居关系。
- PDU Length：PDU 的总长度，以字节为单位。
- Priority：选举 DIS 的优先级。
- LAN ID：包括 System ID 和一字节的伪节点 ID。

点到点网络中的Hello报文格式如 图 1-8所示（浅蓝色部分是通用报文头）。

图1-8 P2P IIH 格式



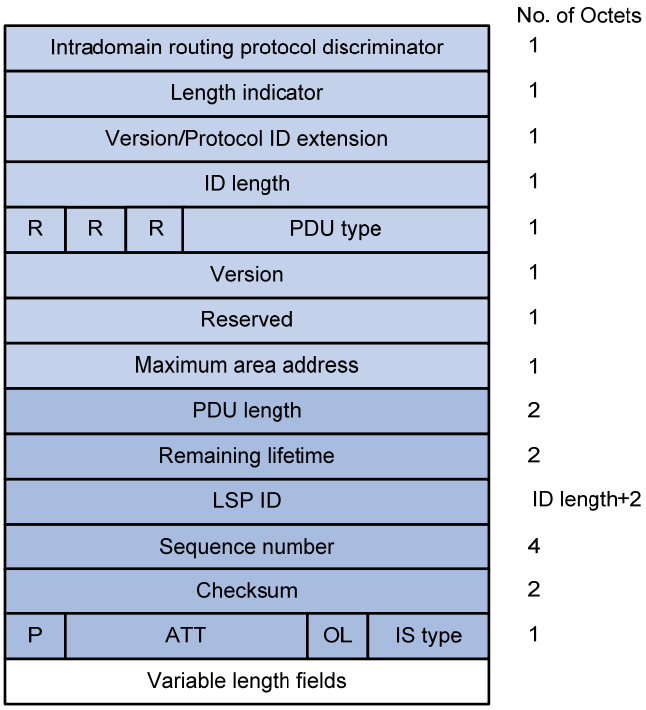
从图中可以看出，P2P IIH 中的多数字段与 LAN IIH 相同。不同的是没有 Priority 和 LAN ID 字段，而多了一个 Local Circuit ID 字段，表示本地链路 ID。

4. LSP 报文格式

LSP 用于交换链路状态信息。LSP 分为两种：Level-1 LSP 和 Level-2 LSP。Level-1 路由器传送 Level-1 LSP，Level-2 路由器传送 Level-2 LSP，Level-1-2 路由器则可传送以上两种 LSP。

两类LSP有相同的报文格式，如 图 1-9所示（浅蓝色部分是通用报文头）。

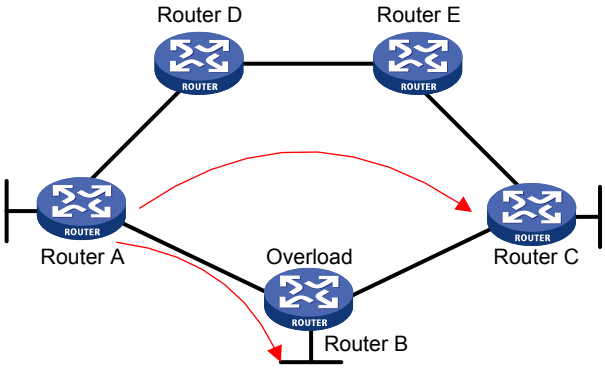
图1-9 L1/L2 LSP 格式



主要字段的解释如下：

- PDU Length：PDU 的总长度，以字节为单位。
- Remaining Lifetime：LSP 的存活时间，以秒为单位。
- LSP ID：由三部分组成，System ID、伪节点 ID（一字节）和 LSP 的分片号（一字节）。
- Sequence Number：LSP 的序列号。
- Checksum：LSP 的校验和。
- P（Partition Repair）：仅与 L2 LSP 有关，表示路由器是否支持自动修复区域分割。
- ATT（Attachment）：由 L1/L2 路由器产生，但仅与 L1 LSP 有关，表示产生此 LSP 的路由器（L1/L2 路由器）与多个区域相连接。
- OL（LSDB Overload）：表示本路由器因内存不足而导致LSDB不完整。其它路由器在得知这一信息后，就不会再利用这台路由器转发需要经过它传送的数据流，但到此路由器直连地址的报文仍然可以被转发。如 图 1-10所示，假设正常情况下Router A到Router C的报文都是经过Router B转发，但如果Router B的OL位置 1，则Router A会认为Router B的路由不完整，从而将报文通过Router D、Router E转发给Router C，但到Router B直连地址的报文不受影响。

图1-10 LSDB Overload 示意图



- IS Type：生成 LSP 的路由器的类型。

5. SNP 格式

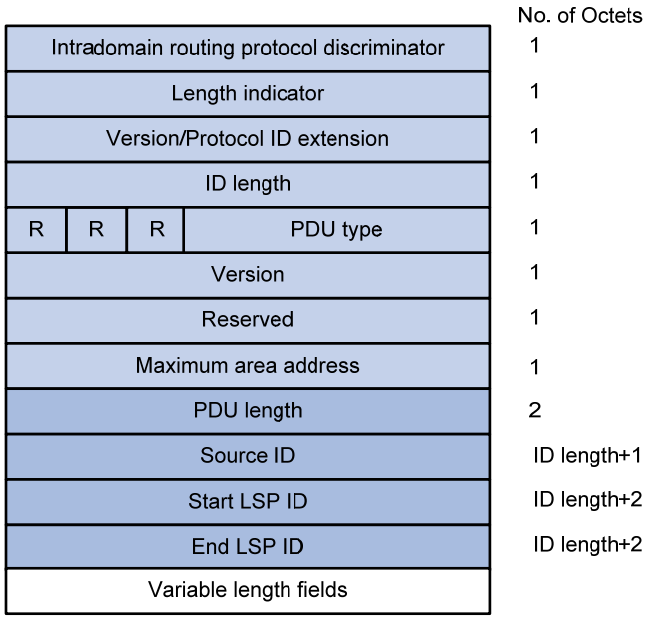
时序报文 SNP（Sequence Number PDUs）用于确认邻居之间最新接收的 LSP，作用类似于确认（Acknowledge）报文，但更有效。

SNP 包括 CSNP（Complete SNP，全时序报文）和 PSNP（Partial SNP，部分时序报文），进一步又可分为 Level-1 CSNP、Level-2 CSNP、Level-1 PSNP 和 Level-2 PSNP。

CSNP 包括 LSDB 中所有 LSP 的摘要信息，从而可以在相邻路由器间保持 LSDB 的同步。在广播网络上，CSNP 由 DIS 定期发送（缺省的发送周期为 10 秒）；在点到点链路上，CSNP 只在第一次建立邻接关系时发送。

CSNP的报文格式如 图 1-11所示（浅蓝色部分是通用报文头）。

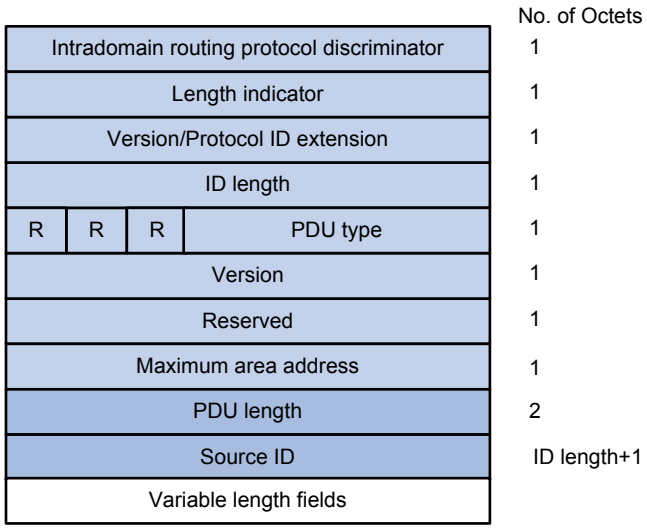
图1-11 L1/L2 CSNP 格式



PSNP 只列举最近收到的一个或多个 LSP 的序号，它能够一次对多个 LSP 进行确认。当发现 LSDB 不同步时，也用 PSNP 来请求邻居发送新的 LSP。

PSNP的报文格式如 图 1-12所示：

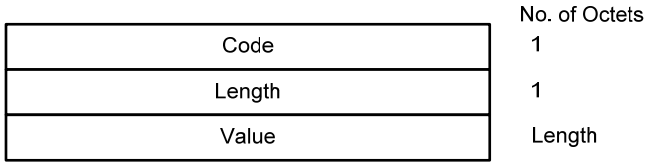
图1-12 L1/L2 PSNP 格式



6. CLV

PDU中的变长字段部分是多个CLV（Code-Length-Value）三元组。其格式如 图 1-13所示：

图1-13 CLV 格式



不同PDU类型所包含的CLV是不同的，如 表 1-2所示。

表1-2 PDU 类型和包含的 CLV 名称

CLV Code	名称	所应用的 PDU 类型
1	Area Addresses	IIH、LSP
2	IS Neighbors（LSP）	LSP
4	Partition Designated Level-2 IS	L2 LSP
6	IS Neighbors（MAC Address）	LAN IIH
7	IS Neighbors（SNPA Address）	LAN IIH
8	Padding	IIH
9	LSP Entries	SNP
10	Authentication Information	IIH、LSP、SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH、LSP
130	IP External Reachability Information	L2 LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH、LSP

其中，Code 值从 1 到 10 的 CLV 在 ISO 10589 中定义（有 2 类未在上表中列出），其它几种 CLV 在 RFC 1195 中定义。

1.1.5 支持的 IS-IS 特性

1. 多实例和多进程

为了方便管理，提高控制效率，IS-IS 支持多进程和多实例特性。多进程允许为一个指定的 IS-IS 进程关联一组接口，从而保证该进程进行的所有协议操作都仅限于这一组接口。这样，就可以实现一台路由器有多个 IS-IS 协议进程，每个进程负责唯一的一组接口。

对于支持 VPN 的路由器，每个 IS-IS 进程都与一个指定的 VPN 实例相关联。这样，所有附加到该进程的接口都应与该进程相关联的 VPN 实例相关联。

2. 热备份

具有分布式结构的路由器可支持 IS-IS 热备份（Hot Standby，HSB）特性。IS-IS 将需要备份的数据从主用主控板（Active Main Board，AMB）备份到备用主控板（Standby Main Board，SMB）。无论何时主用主控板出现故障，备用主控板都会变成激活状态，接替工作，从而保证 IS-IS 不受影响，保持正常工作。

IS-IS 支持两种类型的 HSB。一种 HSB 的所有 IS-IS 数据都是同步备份的，AMB 和 SMB 切换后，IS-IS 可以立即工作，保持平稳运行。另一种 HSB 是在 AMB 和 SMB 的切换过程中只备份 IS-IS 配置信息，IS-IS 进行 GR，重新向邻居发送邻接请求，同步 LSDB 数据库。

3. IS-IS GR

说明

有关 GR（Graceful Restart，平滑重启）的原理介绍请参见“系统分册”中的“GR 概述”。

基于 IS-IS 的 GR Restarter 进行协议重启后，为了与其 IS-IS 邻居重新同步链路状态数据库，它必须完成下列两项任务：

- 在不改变邻接关系的前提下，重新获取网络中的有效 IS-IS 邻居信息；
- 重新获取网络链路状态数据库的内容。

IS-IS 协议重启完毕后，GR Restarter 会立即向邻接的 GR Helper 发送一个 IS-IS GR 信号。这样，IS-IS 邻居就不会复位与其的邻居关系。在收到其 IS-IS 邻居的响应后，GR Restarter 会重新恢复与其的邻居关系列表。

邻居关系重新建立后，GR Restarter 与其所有具备 GR 感知能力的 IS-IS 邻居之间同步数据库，并交换路由信息。交换完成后，GR Restarter 根据新的路由转发信息更新路由表和转发表，删除失效的路由，完成 IS-IS 协议收敛。

4. IS-IS TE

IS-IS TE（Traffic Engineering，流量工程）支持建立和维护 TE 的 LSP（Label Switched Path，标签交换路径）。

MPLS 在构建 CR LSP（Constraint-based Routed LSP，基于约束路由的 LSP）时，需要了解本区域中所有链路的流量属性信息。它可以通过 IS-IS 来获取链路的流量工程信息。



说明

IS-IS TE 的详细配置请参见“MPLS 分册”中的“MPLS TE 配置”。

---

5. 管理标记

管理标记简化了对路由信息的管理，该标记用来携带关于 IP 地址前缀的管理信息，利用它可以控制对各种外部路由协议的引入，以及承载 BGP 的标准或扩展团体属性。

6. LSP 分片扩展

IS-IS 通过泛洪 LSP 来宣告链路状态信息，由于一个 LSP 能够承载的信息量有限，IS-IS 将对 LSP 进行分片。每个 LSP 分片由产生该 LSP 的结点或伪结点的 SystemID、PseudnodeID（普通 LSP 为 0，Pseudonode LSP 为非 0）、LSPNumber（LSP 分片号）组合起来唯一标识，由于 LSPNumber 字段的长度是 1 字节，因此，IS-IS 路由器可产生的分片数最大为 256，限制了 IS-IS 路由器可以发布的链路信息量。

IS-IS LSP 分片扩展特性可使 IS-IS 路由器生成更多的 LSP 分片，通过为路由器配置附加的虚拟系统、每个虚拟系统都可生成 256 个 LSP 分片（最多可配置 50 个虚拟系统），使得 IS-IS 路由器可最多生成 13056 个 LSP 分片。

(1) 常用术语

- 初始系统（Originating System）：实际运行 IS-IS 协议的路由器。使能 IS-IS LSP 分片扩展功能后，可以为路由器配置附加的虚拟系统，“Originating System”指的是原来运行的、“真正”的 IS-IS 进程。
- 系统 ID（System ID）：初始系统的系统 ID。

- 附加系统 ID（Additional System ID）：使能 IS-IS LSP 分片扩展功能后，为 IS-IS 路由器配置的附加的虚拟 SystemID。每个附加系统 ID 都可以生成 256 个额外扩展的 LSP 分片。附加系统 ID 和系统 ID 一样，在整个路由域中必须唯一。
- 虚拟系统（Virtual System）：由附加系统 ID 标识的系统，用来生成扩展 LSP 分片。
- 原始 LSP（Original LSP）：由初始系统生成的 LSP，在其 LSP ID 中指定的 System ID 为初始系统的系统 ID。
- 扩展 LSP（Extended LSP）：由虚拟系统生成的 LSP，在其 LSP ID 中指定的 System ID 为附加系统 ID。

通过配置附加系统 ID，IS-IS 路由器可以在扩展分片 LSP 中宣告更多的链路状态信息，可以把虚拟系统看作一个虚拟的路由器，扩展分片 LSP 就是由附加系统 ID 标识的虚拟系统发布的。

(2) 操作模式

IS-IS 路由器运行 LSP 分片扩展特性的操作模式有两种：

- Mode-1：当网络中的部分路由器不支持 LSP 分片扩展特性时，使用工作模式 1。在该模式下，初始系统与每个虚拟系统都建立邻居关系，且初始系统到虚拟系统的链路开销为 0，因此，虚拟系统就好象网络中与初始系统相连的路由器，但只能通过初始系统才能到达虚拟系统，因此，不支持分片扩展特性的 IS-IS 路由器收到扩展分片 LSP 时无需做任何修改也能够正常工作，但是对通过虚拟系统发布的扩展 LSP 分片中的链路状态信息有所限制。
- Mode-2：当网络中所有路由器都支持 LSP 分片扩展特性时，推荐使用工作模式 2。在该模式下，网络中所有 IS-IS 路由器都知道虚拟系统生成的 LSP 属于的初始系统，对虚拟系统发布的 LSP 扩展分片的链路状态信息没有限制。

LSP 分片扩展特性的操作模式基于区域和路由层次配置，mode-1 可以向前兼容，支持分片扩展特性的路由器能够与不支持分片扩展特性的路由器协同工作；mode-2 要求处在同一区域且 Level 相同的路由器都支持分片扩展特性。

7. 动态主机名交换机制

动态主机名交换机制是为了方便对 IS-IS 网络的维护和管理而引入的，它为 IS-IS 路由器提供了一种从主机名到 System ID 映射的服务。这个动态的主机名信息在 LSP 中以一个动态主机名 CLV 的形式发布。

这个机制同时还提供将主机名与广播网中的 DIS 相关联的服务，并将此信息通过伪节点的 LSP 以动态主机名 CLV 的形式发布出去。

在维护和管理中，使用主机名比使用 System ID 会更直观，也更容易记忆。配置此功能后，当在路由器上使用 **display** 命令显示 IS-IS 相关信息时，看到的是路由器的主机名，而不再是 System ID。

1.1.6 协议规范

与 IS-IS 相关的协议规范有：

- ISO 10589：ISO IS-IS Routing Protocol
- ISO 9542：ES-IS Routing Protocol
- ISO 8348：Ad2 Network Services Access Points
- RFC 1195：Use of OSI IS-IS for Routing in TCP/IP and Dual Environments

- RFC 2763：Dynamic Hostname Exchange Mechanism for IS-IS
- RFC 2966：Domain-wide Prefix Distribution with Two-Level IS-IS
- RFC 2973：IS-IS Mesh Groups
- RFC 3277：IS-IS Transient Blackhole Avoidance
- RFC 3358：Optional Checksums in ISIS
- RFC 3373：Three-Way Handshake for IS-IS Point-to-Point Adjacencies
- RFC 3567：Intermediate System to Intermediate System (IS-IS) Cryptographic Authentication
- RFC 3719：Recommendations for Interoperable Networks using IS-IS
- RFC 3786：Extending the Number of IS-IS LSP Fragments Beyond the 256 Limit
- RFC 3787：Recommendations for Interoperable IP Networks using IS-IS
- RFC 3784：IS-IS extensions for Traffic Engineering
- RFC 3847：Restart signaling for IS-IS

1.2 IS-IS 配置任务简介

表1-3 IS-IS 配置任务简介

配置任务		说明	详细配置
配置 IS-IS 基本功能	使能 IS-IS 功能	必选	1.3.2
	配置路由级别	可选	1.3.3
	配置接口网络类型	可选	1.3.4
配置IS-IS路由信息控制	配置 IS-IS 链路度量值	可选	1.4.2
	配置 IS-IS 路由优先级	可选	1.4.3
	配置 IS-IS 最大等价路由条数	可选	1.4.4
	配置 IS-IS 路由聚合	可选	1.4.5
	配置 IS-IS 发布缺省路由	可选	1.4.6
	配置 IS-IS 引入外部路由	可选	1.4.7
	配置 IS-IS 路由过滤	可选	1.4.8
	配置 IS-IS 路由渗透	可选	1.4.9
配置 IS-IS 调整和优化	配置 Hello/CSNP 报文发送时间间隔	可选	1.5.2
	配置 Hello 报文失效数目	可选	1.5.3
	配置 DIS 优先级	可选	1.5.4
	禁止接口发送和接收 IS-IS 报文	可选	1.5.5
	配置在 PPP 接口上取消建立邻接关系必须在同一网段的限制	可选	1.5.6
	配置接口发送小型 Hello 报文	可选	1.5.7
	配置 LSP 参数	可选	1.5.8
	配置 SPF 参数	可选	1.5.9
	配置 LSDB 过载标志位	可选	1.5.10

配置任务		说明	详细配置
配置 IS-IS 验证功能	配置邻居关系验证	可选	1.6.2
	配置区域验证	可选	1.6.3
	配置路由域验证	可选	1.6.4
配置 IS-IS 主机名 映射	配置 IS-IS 静态主机名映射	可选	1.7.1
	配置 IS-IS 动态主机名映射	可选	1.7.2
配置 IS-IS GR		可选	1.8
配置邻接状态输出		可选	1.9
使能 IS-IS 的 Trap 功能		可选	1.10
配置 IS-IS 和 MIB 绑定		可选	1.11

1.3 配置 IS-IS 基本功能

1.3.1 配置准备

在配置 IS-IS 基本功能之前，需完成以下任务：

- 配置链路层协议
- 配置接口的网络层地址，使相邻节点网络层可达

1.3.2 使能 IS-IS 功能

表1-4 使能 IS-IS 功能

操作	命令	说明
进入系统视图	**system-view**	-
启动 IS-IS 路由进程，进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，系统没有运行任何 IS-IS 进程
配置网络实体名称	**network-entity** *net*	必选 缺省情况下，没有配置 NET
退回至系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置指定接口上使能 IS-IS 路由进程	**isis enable** [*process-id*]	必选 缺省情况下，IS-IS 功能在接口上处于关闭状态，且没有任何 IS-IS 进程与其关联

1.3.3 配置路由级别

建议用户在配置 IS-IS 时配置路由器类型：

- 如果只有一个区域，建议用户将所有路由器设置为 Level-1 或者 Level-2，因为没有必要让所有路由器同时维护两个完全相同的 LSDB。
- 在 IP 网络中使用时，建议将所有的路由器都设置为 Level-2，这样有利于以后的扩展。

当路由器类型是 Level-1（Level-2）时，接口的链路类型只能为 Level-1（Level-2），当路由器类型是 Level-1-2 时，接口的链路类型缺省为 Level-1-2，当路由器只需要与对端建立 Level-1（Level-2)的邻接关系时，可以将接口的链路类型配置为 Level-1（Level-2)来限制接口上所能建立的邻接关系，如 Level-1 的接口只能建立 Level-1 的邻接关系，Level-2 的接口只能建立 Level-2 的邻接关系，让接口只发送和接收 Level-1（Level-2)类型的 Hello 报文，既减少了路由器的处理时间又节省了带宽。

表1-5 配置路由级别

操作	命令	说明
进入系统视图	**system-view**	-
启动 IS-IS 路由进程，进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置路由器类型	**is-level** { **level-1** | **level-1-2** | **level-2** }	可选 缺省情况下，路由器类型为 **level-1-2**
退回至系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的链路类型	**isis circuit-level** [**level-1** | **level-1-2** | **level-2**]	可选 缺省情况下，接口的链路类型为 Level-1-2

1.3.4 配置接口网络类型

接口网络类型不同，其工作机制也略微不同，如：当网络类型为广播网时，需要选举 DIS、通过泛洪 CSNP 报文来实现 LSDB 同步；当网络类型为 P2P 时，不需要选举 DIS，LSDB 同步机制也不同。

当只有两台路由器接入到同一个广播网时，通过将接口网络类型配置为 P2P 可以使 IS-IS 按照 P2P 而不是广播网的工作机制运行，避免 DIS 选举以及 CSNP 的泛洪，即可以节省网络带宽，又可以加快网络的收敛速度。

表1-6 配置接口网络类型

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的网络类型为 P2P	**isis circuit-type p2p**	可选 缺省情况下，路由器接口网络类型根据物理接口决定



说明

仅当接口的网络类型为广播网，且只有两台路由器接入该广播网时才需要进行该项配置，并且两台路由器都要进行此项配置。

1.4 配置 IS-IS 路由信息控制

1.4.1 配置准备

在配置 IS-IS 路由信息控制之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 使能 IS-IS 功能

1.4.2 配置 IS-IS 链路度量值

IS-IS 有三种方式来配置接口的链路度量值，按照选择顺序依次为：

- 在接口视图下为指定接口配置的链路度量值。
- 在系统视图下全局配置的链路度量值，该配置将对该 IS-IS 进程关联的接口同时生效。
- 自动计算度量值：将根据带宽参考值自动计算接口的链路度量值，当开销值的类型为 **wide** 或 **wide-compatible** 时，可以根据公式“开销=（参考值÷带宽）×10”计算接口的链路度量值。当开销值类型为其他类型时，具体情况如下：接口带宽 10Mbps 时，值为 60；接口带宽 100Mbps 时，值为 50；接口带宽 155Mbps 时，值为 40；接口带宽 622Mbps 时，值为 30；接口带宽 2500Mbps 时，值为 20；接口带宽>2500Mbps 时，值为 10。

如果没有采用上述三种方式中的任一种进行度量值的配置，接口的链路度量值将取系统设置的缺省值 10。

1. 接口配置 IS-IS 链路度量值

表1-7 接口配置 IS-IS 链路度量值

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 开销值的类型	**cost-style** { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }	可选 缺省情况下，IS-IS 开销值的类型为 narrow
退回至系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 IS-IS 接口的链路度量值	**isis cost** *value* [**level-1** | **level-2**]	必选 缺省情况下，没有配置 IS-IS 接口的链路度量值

2. 全局配置 IS-IS 链路度量值

表1-8 全局配置 IS-IS 链路度量值

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-

操作	命令	说明
配置 IS-IS 开销值的类型	**cost-style** { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }	可选 缺省情况下，IS-IS 开销值的类型为 narrow
全局配置 IS-IS 的链路度量值	**circuit-cost** *value* [**level-1** | **level-2**]	必选 缺省情况下，没有全局配置 IS-IS 的链路度量值

3. 配置 IS-IS 自动计算链路度量值

表1-9 配置 IS-IS 自动计算链路度量值

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 开销值的类型	**cost-style** { **wide** | **wide-compatible** }	必选 缺省情况下，IS-IS 开销值的类型为 narrow
使能自动计算接口链路度量值功能	**auto-cost enable**	必选 缺省情况下，自动计算接口链路度量值功能处于关闭状态
配置IS-IS 自动计算链路度量值时依据的带宽参考值	**bandwidth-reference** *value*	可选 缺省情况下，带宽参考值为 100Mbps

1.4.3 配置 IS-IS 路由优先级

一台路由器可同时运行多个路由协议，当多个路由协议都发现到同一目的地的路由时，将选用高优先级路由协议所发现的路由。以下配置用来为 IS-IS 路由设置优先级，使用路由策略可以为特定的路由设置特定的优先级，路由策略的相关知识请参考“IP 路由分册”中的“路由策略配置”。

表1-10 配置 IS-IS 路由优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 路由优先级	**preference** { **route-policy** *route-policy-name* | *preference* } *	必选 缺省情况下，IS-IS 路由优先级为 15

1.4.4 配置 IS-IS 最大等价路由条数

如果到一个目的地有几条开销相同的路径，可以通过等价路由负载分担来提高链路利用率。该配置用以设置 IS-IS 协议的最大等价路由条数。

表1-11 配置 IS-IS 最大等价路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置在负载分担方式下 IS-IS 最大等价路由条数	**maximum load-balancing** *number*	必选 缺省情况下，负载分担方式下 IS-IS 最大等价路由条数为 8

1.4.5 配置 IS-IS 路由聚合

通过配置路由聚合，可以减小路由表规模，还可以减少本路由器生成的 LSP 报文大小和 LSDB 的规模。其中，被聚合的路由可以是 IS-IS 协议发现的路由，也可以是引入的外部路由。

表1-12 配置 IS-IS 路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置聚合路由	**summary** *ip-address* { *mask* | *mask-length* } [**avoid-feedback** | **generate_null0_route** | **tag** *tag* | [**level-1** | **level-1-2** | **level-2**]] *	必选 缺省情况下，没有对路由进行聚合

说明

- 聚合后路由的开销值取所有被聚合路由中最小的开销值。
- 路由器只对本地生成的 LSP 中的路由进行聚合。

1.4.6 配置 IS-IS 发布缺省路由

对于运行 IS-IS 的路由器来说，无法引入缺省路由，因此也无法通过将目的地为 0.0.0.0/0 的路径信息（即缺省路由）通过 LSP 发布给其它路由器，可以通过配置发布一条缺省路由，将目的地为 0.0.0.0/0 的路径信息通过 LSP 发布出去，其它同级别的路由器中将在自己的路由表中新增一条缺省路由。

表1-13 配置 IS-IS 发布缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 发布缺省路由	**default-route-advertise** [**route-policy** *route-policy-name* | [**level-1** | **level-2** | **level-1-2**]] *	必选 缺省情况下，此功能关闭

说明

该配置产生的缺省路由只被发布到同级别的路由器。通过使用路由策略，可以强制 IS-IS 只在路由表中有匹配的路由项时才生成缺省路由。

1.4.7 配置 IS-IS 引入外部路由

IS-IS 将其它路由协议发现的路由当作外部路由处理。在引入其它协议路由时，可指定引入路由的缺省开销。还可以通过配置对引入路由进行过滤。

在实际组网环境中，每台路由器的性能即处理能力不同，如果在处理能力强的高端设备上引入大量外部路由，那么可能会对网络上其它低端设备的性能造成较大的冲击，网络管理员可以通过配置支持的最大引入路由条数，限制引入外部路由的条数，从而最终限制发布路由的数量。

表1-14 配置 IS-IS 引入外部路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
从其它路由协议或其它 IS-IS 进程引入路由信息	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	必选 缺省情况下，IS-IS 不引入其它协议的路由信息 如果 **import-route** 命令中不指定引入的级别，则默认为引入路由到 Level-2 路由表中
配置引入 Level1/Level2 的 IPv4 路由最大条数	**import-route limit** *number*	可选 本命令的缺省值与设备的型号有关，请以设备的实际情况为准

说明

只能引入路由表中状态为 active 的路由，是否为 active 状态可以通过 **display ip routing-table protocol** 命令来查看。

1.4.8 配置 IS-IS 路由过滤

路由过滤就是通过对 ACL、IP 地址前缀列表或路由策略等规则的引用对路由信息的生成进行更加严格的控制，包括过滤通过接收到的 LSP 计算出来的路由信息和过滤从其它路由协议引入的路由信息。

1. 过滤通过接收到的 LSP 计算出来的路由信息

运行 IS-IS 的路由器会把从邻居收到的 LSP 保存到自己维护的链路状态数据库中，并使用 SPF 算法计算出以自己为根的最短路径树，并把计算好的路由信息加入到 IS-IS 路由表中。

通过 ACL、IP 地址前缀列表或路由策略可以对计算出来的将要加入到 IS-IS 路由表中的路由进行过滤，满足条件则加入到 IS-IS 路由表中，否则将不能加入到 IS-IS 路由表中。

表1-15 过滤通过接收到的 LSP 计算出来的路由信息

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
过滤通过接收到的 LSP 计算出来的路由信息	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **import**	必选 缺省情况下，IS-IS 不对通过接收到的 LSP 计算出来的路由信息进行过滤

2. 过滤引入路由信息

IS-IS 可以从其它路由协议或其它 IS-IS 进程引入路由信息，把它直接加入到 IS-IS 的路由表中并通过 LSP 发布出去。

通过 ACL、IP 地址前缀列表或路由策略可以对引入的路由信息进行过滤，满足条件加入到 IS-IS 路由表中，否则将不能加入到 IS-IS 路由表中，没有加入 IS-IS 路由表的路由将不会通过 LSP 发布出去。

表1-16 过滤引入路由信息

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
过滤引入路由信息	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]	必选 缺省情况下，IS-IS 不对引入的路由信息进行过滤

1.4.9 配置 IS-IS 路由渗透

通过 IS-IS 路由渗透功能，Level-1-2 路由器可以将它所知道的其它 Level-1 区域路由信息和 Level-2 区域路由信息发布到 Level-1 路由器。

表1-17 配置 IS-IS 路由渗透

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能 IS-IS 路由渗透	**import-route isis level-2 into level-1** [**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *	必选 缺省情况下，Level-2 区域的路由信息不向 Level-1 区域发布

说明

将 Level-2 区域的路由信息引入到 Level-1 区域时，需要注意：

- 如果指定了过滤策略，则只有通过过滤的路由才能够被发布到 Level-1 区域中。
- 如果要通过路由策略对从 Level-2 区域引入到 Level-1 区域的路由信息进行过滤，必须在 **import-route isis level-2 into level-1** 命令中同时指定要应用的路由策略，否则路由过滤将不会生效；其它路由策略，如在接收或引入路由时指定的路由策略对路由渗透无效。

1.5 配置 IS-IS 调整和优化

1.5.1 配置准备

在配置 IS-IS 调整和优化之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 使能 IS-IS 功能

1.5.2 配置 Hello/CSNP 报文发送时间间隔

表1-18 配置 Hello/CSNP 报文发送时间间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 Hello 报文的发送时间间隔	**isis timer hello** *seconds* [**level-1** | **level-2**]	可选 缺省情况下，Hello 报文的发送时间间隔为 10 秒
配置 DIS 在广播网络上发送 CSNP 报文的时间间隔	**isis timer csnp** *seconds* [**level-1** | **level-2**]	可选 缺省情况下，CSNP 报文的发送时间间隔为 10 秒

说明

DIS 发送 Hello 报文的时间间隔是 **isis timer hello** 设置的时间的 1/3。

1.5.3 配置 Hello 报文失效数目

当路由器在抑制时间内没有收到来自邻居的 Hello 报文时将宣告邻居失效，抑制时间等于 Hello 报文失效数目与 Hello 报文发送时间间隔的乘积，通过设置 Hello 报文失效数目和 Hello 报文的发送时间间隔，可以调整抑制时间，即路由器要花多长时间能够监测到链路已经失效并重新进行路由计算。

表1-19 配置 Hello 报文失效数目

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置 Hello 报文失效数目	**isis timer holding-multiplier** *value* [**level-1** | **level-2**]	可选 缺省情况下，Hello 报文失效数目为 3



说明

在广播链路上，Level-1 和 Level-2 Hello 报文会分别发送，Hello 报文失效数目需要分别设置；在点到点链路中，Level-1 和 Level-2 的 Hello 报文是在同一个点到点 Hello 报文中发送，因此不需要指定 Level-1 或 Level-2。

1.5.4 配置 DIS 优先级

在广播网络中，IS-IS 需要在所有的路由器中选举一个路由器作为 DIS。

对于 IS-IS，Level-1 和 Level-2 的 DIS 是分别选举的，可以为不同级别的 DIS 选举设置不同的优先级。优先级数值越高，被选中的可能性就越大。如果所有路由器的 DIS 优先级相同，将会选择 MAC 地址最大的路由器作为 DIS。

如果所有路由器的 DIS 优先级相同，将会选择 MAC 地址最大的路由器作为 DIS。

表1-20 配置 DIS 优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 DIS 优先级	**isis dis-priority** *value* [**level-1** | **level-2**]	可选 缺省情况下，接口的 DIS 优先级为 64

1.5.5 禁止接口发送和接收 IS-IS 报文

通过禁止接口发送和接收 IS-IS 报文，禁止了该接口与相邻路由器建立邻居关系，但仍然可以把该接口直连网络的路由信息放在 LSP 中从其它接口宣告出去。由于不用建立邻居关系，可以节省带宽和路由器处理时间，同时，其它路由器也可以知道到达该接口直连网络的路由信息。

表1-21 禁止接口发送和接收 IS-IS 报文

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
禁止接口发送和接收 IS-IS 报文	**isis silent**	必选 缺省情况下，接口既发送也接收 IS-IS 报文

1.5.6 配置在 PPP 接口上取消建立邻接关系必须在同一网段的限制

当网络类型为点到点链路时，IS-IS 将检查 Hello 报文的源地址是否与接收 Hello 报文的接口在同一网段，如果不在同一网段，将丢弃接收到的 Hello 报文，无法与对端路由器建立邻接关系。

当接口封装 PPP 协议时，对端的 IP 地址与当前接口可以不在同一网段，在这种情况下，可以配置取消与对端路由器建立邻接关系必须在同一网段的限制，即在 PPP 协议接口上接收 Hello 报文时，如果对端的 IP 地址与当前接口不在同一网段也可以建立邻居关系。

表1-22 配置在 PPP 接口上取消建立邻接关系必须在同一网段的限制

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置在 PPP 接口上取消建立邻接关系必须在同一网段的限制	**isis peer-ip-ignore**	必选 该命令只能在协议类型为 PPP 的接口上配置 缺省情况下，在协议类型为 PPP 的接口接收 ISIS Hello 报文时检查对端的 IP 地址，即要求双方必须在同一网段

1.5.7 配置接口发送小型 Hello 报文

IS-IS 协议报文直接封装在链路层报文头后面，无法实现协议报文在 IP 层的自动分片。因此，运行 IS-IS 的路由器与对端路由器建立邻居关系时，会发送达到链路 MTU 大小的 Hello 报文，双方进行 MTU 大小的通信协商，来保证建立邻居双方接口 MTU 的一致性，从而避免双方 MTU 大小不一致导致较小的 PDU 可以通过，但是较大的 PDU 无法通过。

当邻居路由器双方 MTU 大小一样的时候，为了避免发送过大的 Hello 报文浪费带宽，可以配置接口发送不加入填充 CLV 的小型 Hello 报文。

表1-23 配置接口发送小型 Hello 报文

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口发送不加入填充 CLV 的小型 Hello 报文	**isis small-hello**	必选 缺省情况下，接口发送标准 Hello 报文

1.5.8 配置 LSP 参数

1. 配置 LSP 时间参数

(1) 配置 LSP 最大生存时间

每一个 LSP 都包含一个最大生存时间，当 LSP 驻留在 LSDB 中时，随着时间的推移，每一个 LSP 的最大生存时间将逐渐递减至 0，当 LSP 的最大生存时间为 0 时，IS-IS 将删掉这个过期的 LSP。用户可根据网络规模对 LSP 的最大生存时间进行调整。

表1-24 配置 LSP 最大生存时间

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-

操作	命令	说明
配置 LSP 最大生存时间	**timer lsp-max-age** *seconds*	可选 缺省情况下，LSP 最大生存时间为 1200 秒

(2) 配置 LSP 生成时间间隔

每一个 LSP 都有一个最大生存时间，每个 LSP 都会随着时间的推移而被逐渐老化，因此每个路由器必须定时刷新自己生成的 LSP，以防止 LSP 的最大生存时间减小至 0；并通过定时刷新 LSP 以使整个区域中的 LSP 保持同步。用户可对 LSP 的刷新周期进行配置，提高 LSP 的刷新频率可以加快网络收敛速度，但是将占用更多的带宽。

除了定时刷新可以重新生成 LSP 外，当网络拓扑发生变化，如邻居路由器 up 或 down，接口 Metric 值、System ID 或区域地址发生变化等，将触发路由器重新生成 LSP。为了防止网络拓扑频繁变化而导致 LSP 频繁重新生成，用户可配置 LSP 生成时间间隔，以抑制网络变化频繁导致占用过多的带宽资源和路由器资源。

表1-25 配置 LSP 生成时间间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 LSP 刷新周期	**timer lsp-refresh** *seconds*	可选 缺省情况下，LSP 刷新周期为 900 秒
配置 LSP 生成时间间隔	**timer lsp-generation** *maximum-interval* [*initial-interval* [*second-wait-interval*]] [**level-1** | **level-2**]	可选 缺省情况下，LSP 生成时间间隔为 2 秒

(3) 配置 LSP 发送时间间隔

当 LSDB 的内容发生变化时，IS-IS 将把发生变化的 LSP 扩散出去，用户可以对 LSP 的最小发送时间间隔进行调节。

在点到点链路上，发送的 LSP 需要得到对端的应答，否则将在指定的时间间隔内重新发送该 LSP，重传时间间隔决定了当一个 LSP 在 P2P 链路上丢失时它被重传需要等待的时间。

表1-26 配置 LSP 发送时间间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置发送 LSP 的最小时间间隔以及一次可以最多发送的 LSP 的数目	**isis timer lsp** *time* [**count** *count*]	可选 缺省情况下，LSP 的发送最小时间间隔为 33 毫秒，一次可以最多发送的 LSP 的数目为 5
配置 LSP 在在点到点链路上的重传时间间隔	**isis timer retransmit** *seconds*	可选 缺省情况下，LSP 报文在在点到点链路上的重传间隔为 5 秒

说明

请合理配置 LSP 发送时间间隔，当存在大量 IS-IS 接口或大量路由时，会发送大量的 LSP 报文，导致 LSP 风暴的出现。

2. 配置 LSP 报文长度

IS-IS 协议报文直接封装在链路层报文头后面，无法实现协议报文在 IP 层的自动分片。

为了不影响 LSP 的正常扩散，要求同一区域内所有 IS-IS 路由器生成 LSP 报文的最大长度不能超过该区域内所有路由器 IS-IS 接口 MTU 的最小值。

如果 IS-IS 运行的区域中各 IS-IS 接口的 MTU 值不一致，建议用户对 IS-IS 生成 LSP 报文的最大长度进行配置，将同一区域内所有 IS-IS 路由器生成 LSP 报文的最大长度配置为该区域内所有路由器 IS-IS 接口 MTU 的最小值。如果不进行配置，系统将根据当前设备 IS-IS 接口最小 MTU 值的变化而自动重启 IS-IS 进程动态调整生成 LSP 报文的最大长度，会在一定程度上影响业务的正常运行。

表1-27 配置 LSP 报文长度

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置生成的 Level-1 LSP 和 Level-2 LSP 的最大长度	**lsp-length originate** *size* [**level-1** | **level-2**]	可选 缺省情况下，生成的 Level-1 LSP 和 Level-2 LSP 的最大长度为 1497 字节
配置可以接收 LSP 的最大长度	**lsp-length receive** *size*	可选 缺省情况下，接收的 LSP 报文的最大长度为 1497 字节

3. 配置 LSP 快速扩散功能

通过使能 LSP 快速扩散功能，当 LSP 发生变化而导致 SPF 重新计算时，在 SPF 重新计算前，把导致 SPF 重新计算的 LSP 快速扩散出去，将大大缩短路由器之间由于进行 LSP 同步而导致 LSDB 不一致的时间，提高全网的快速收敛性能。

表1-28 配置 LSP 快速扩散功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 LSP 快速扩散功能	**flash-flood** [**flood-count** *flooding-count* | **max-timer-interval** *flooding-interval* | [**level-1** | **level-2**]] *	必选 缺省情况下，禁止 LSP 快速扩散功能

4. 配置 LSP 分片扩展功能

表1-29 配置 LSP 分片扩展功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能 IS-IS 进程的 LSP 分片扩展功能，并配置 LSP 分片扩展特性的工作模式	**lsp-fragments-extend** [[**level-1** | **level-2** | **level-1-2**] | [**mode-1** | **mode-2**]] *	必选 缺省情况下，LSP 分片扩展功能处于关闭状态
配置 IS-IS 进程的虚拟系统 ID	**virtual-system** *virtual-system-id*	必选 缺省情况下，没有配置 IS-IS 进程的虚拟系统 ID

说明

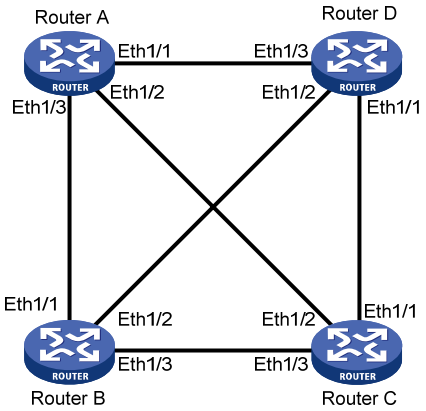
使能 LSP 分片扩展功能时需要注意：

- IS-IS 进程使能分片扩展功能后，使能该 IS-IS 进程的所有接口的 MTU 不能小于 512，否则 LSP 分片扩展功能将不会生效。
- 为了使路由器生成扩展 LSP 分片，应至少配置一个虚拟 System ID。一个 IS-IS 进程最多可配置 50 个虚拟 System ID。

5. 限制 LSP 泛洪

在ATM、FR等NBMA网络中，如果网络联通程度较高、网络中存在多条点到点链路时，如 图 1-14 所示，Router A、Router B、Router C和Router D均使能了IS-IS，Router A新生成一个LSP时，将把该LSP分别从GigabitEthernet1/1、GigabitEthernet1/2 和GigabitEthernet1/3 泛洪出去，Router D 从 GigabitEthernet1/3 收到Router A发送的LSP后，也会把该LSP从GigabitEthernet1/1 和GigabitEthernet1/2 发送给Router B和Router C，而Router B和Router C已经从GigabitEthernet1/1、GigabitEthernet1/2 收到了Router A发送的LSP。LSP的重复扩散会导致带宽的浪费。

图1-14 联通程度较高网络示意图



为了避免这种情况的发生，可以将一些接口配置属于一个 Mesh-Group，也可以配置接口阻塞。

- 将设备的几个接口配置属于一个 Mesh-Group 后，如果从其中的一个接口接收到一个新的 LSP，IS-IS 只把该 LSP 扩散到没有配置 Mesh-Group 的接口以及与当前接口不属于同一个 Mesh-Group 的接口，而不会扩散到到同 Mesh-Group 中的其它接口。
- 配置接口阻塞后，只有该接口从邻居路由器收到要求发送 LSP 的请求时才会发送 LSP，否则不会主动向外发送 LSP。

设置接口加入 Mesh-Group 或对接口进行阻塞时应注意保留一定的冗余度，以免由于链路故障影响 LSP 报文的正常扩散。

表1-30 限制 LSP 泛洪

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口属于 Mesh-Group	**isis mesh-group** *mesh-group-number*	二者必选其一 缺省情况下，接口不属于任何 Mesh-Group 且接口不阻塞
配置接口阻塞	**isis mesh-group mesh-blocked**	

说明

Mesh-Group 只对点到点类型链路的接口起作用。

1.5.9 配置 SPF 参数

IS-IS 协议中，当 LSDB 发生变化时需要进行路由计算。频繁的路由计算会占用大量的系统资源，导致系统性能下降，而周期性地计算 SPF 可以在一定程度上提高效率，计算 SPF 的时间间隔可以由用户根据需要进行配置。

表1-31 配置 SPF 参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 路由计算时间间隔	**timer spf** *maximum-interval* [*initial-interval* [*incremental-interval*]]	可选 缺省情况下，IS-IS 路由计算的时间间隔为 10 秒

1.5.10 配置 LSDB 过载标志位

通过配置 LSDB 过载标志位，IS-IS 将在其发送的 LSP 报文中把 OL 位置位，以通知其它路由器当前路由器发生了问题，无法正确的执行路由选择和报文转发。

当运行 IS-IS 的路由器因为内存不足或其它原因无法记录完整的 LSDB 时，将会导致区域路由的计算错误，在故障排除过程中，通过给怀疑有问题的路由器设置过载标志位，可以将其从 IS-IS 网络中暂时隔离，便于进行故障定位。

表1-32 配置 LSDB 过载标志位

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
设置过载标志位	**set-overload** [**on-startup** [[**start-from-nbr** *system-id* [*timeout1* [*nbr-timeout*]]] | *timeout2*] [**allow** { **interlevel** | **external** } *]	必选 缺省情况下，不设置过载标志位

1.6 配置 IS-IS 验证功能

1.6.1 配置准备

在配置 IS-IS 验证功能之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 使能 IS-IS 功能

1.6.2 配置邻居关系验证

配置邻居关系验证后，验证密码将会按照设定的方式封装到 Hello 报文中，并对接收到的 Hello 报文进行验证密码的检查，通过检查才会形成邻居关系，否则将不会形成邻居关系，用以确认邻居的正确性和有效性，防止与无法信任的路由器形成邻居。

两台路由器要形成邻居关系必须配置相同的验证方式和验证密码。

表1-33 配置邻居关系验证

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置邻居关系验证方式和验证密码	**isis authentication-mode** { **simple** | **md5** } *password* [**level-1** | **level-2**] [**ip** | **osi**]	必选 缺省情况下，接口没有配置邻居关系验证，既不会验证收到的 Hello 报文，也不会把验证密码插入到 Hello 报文中



说明

isis authentication-mode 命令的参数 **level-1** 和 **level-2** 仅在路由器的 GigabitEthernet 接口上是可见的，而且必须先使用 **isis enable** 命令使能该接口。

1.6.3 配置区域验证

通过配置区域验证，可以防止将从不可信任的路由器学习到的路由信息加入到本地 Level-1 的 LSDB 中。

配置区域验证后，验证密码将会按照设定的方式封装到 Level-1 报文（LSP、CSNP、PSNP）中，并对收到的 Level-1 报文进行验证密码的检查。

同一区域内的路由器必须配置相同的验证方式和验证密码。

表1-34 配置区域验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置区域验证方式和验证密码	**area-authentication-mode** { **simple** | **md5** } *password* [**ip** | **osi**]	必选 缺省情况下，系统没有配置区域验证，既不会验证收到的 Level-1 报文，也不会把验证密码插入到 Level-1 报文中

1.6.4 配置路由域验证

通过配置路由域验证，可以防止将不可信的路由信息注入当前路由域。

配置路由域验证后，验证密码将会按照设定的方式封装到 Level-2 报文（LSP、CSNP、PSNP）中，并对收到的 Level-2 报文进行验证密码的检查。

所有骨干层（Level-2）路由器必须配置相同的验证方式和验证密码。

表1-35 配置路由域验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置路由域验证方式和验证密码	**domain-authentication-mode** { **simple** | **md5** } *password* [**ip** | **osi**]	必选 缺省情况下，系统没有配置路由域验证，既不会验证收到的 Level-2 报文，也不会把验证密码插入到 Level-2 报文中

1.7 配置 IS-IS 主机名映射

IS-IS 用 System ID 来在区域内唯一标识主机或路由器，System ID 长度固定为 6 字节。当网络管理员检查 IS-IS 邻居关系的状态、IS-IS 路由表以及 LSDB 中的内容时，点分十进制表示的 System ID 以及 LSP 标识符不够直观，查看也不方便。

主机名映射提供了一种将 System ID 映射到主机名的服务，运行 IS-IS 的路由器维护一个主机名到 System ID 的映射关系表，在维护和管理以及网络故障诊断时，使用主机名比使用 System ID 会更直观，也更容易记忆。

可以通过静态配置和动态生成两种方式生成和维护此关系映射表，需要注意的是：

- 只有使能动态主机名映射功能后，在路由器上使用 **display isis lsdb** 命令显示 IS-IS 链路状态数据库时，才可以看到路由器的主机名而不是 System ID。
- 倘若网络中的一台路由器使能了动态主机名映射功能且在当前路由器也通过静态方式为那台路由器配置了主机名，动态配置的主机名将覆盖当前路由器为其静态配置的主机名称。

1.7.1 配置 IS-IS 静态主机名映射

网络管理员为远端 IS 手工配置 System ID 与主机名称的映射关系。

表1-36 配置 IS-IS 静态主机名映射

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
为远端 IS 配置 System ID 与主机名称的映射关系	**is-name map** *sys-id map-sys-name*	必选 每个 System ID 只能对应一个主机名称

1.7.2 配置 IS-IS 动态主机名映射

静态配置关系映射表要求网络中的每一台路由器为其它路由器配置 System ID 和主机名的映射关系，当网络中路由器数目增多时，网络中每新增一台路由器或修改某台路由器的主机名映射关系，其它路由器都要做相应配置，增加了维护工作量。

使能动态主机名映射功能后，IS-IS 网络中的每台路由器只需要在本机上配置自己的主机名称即可，配置的主机名称将通过动态主机名 CLV 发布出去，最后 IS-IS 网络中使能动态主机名映射功能的路由器都将收集到其它路由器 System ID 与主机名称的映射关系并生成映射表。

同时还可以为广播网中的 DIS 配置局域网名称来代表这个广播网中的伪节点，便于网络管理员查看 LSDB 内容时判断 LSP 是由哪个 DIS 产生的。

表1-37 配置 IS-IS 动态主机名映射

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能动态主机名映射功能并为当前路由器配置主机名称	**is-name** *sys-name*	必选 缺省情况下，动态主机名映射功能处于关闭状态且没有为当前路由器配置主机名称
退回至系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置本地局域网名称	**isis dis-name** *symbolic-name*	可选 缺省情况下，没有配置本地局域网名称 该命令只有在使能了动态主机名进程的路由器上有效。该命令在点到点链路的接口上无效

1.8 配置 IS-IS GR

在普通的路由协议重启的情况下，路由器与其它邻居的邻接关系被拆除。此时会引起网络暂时的中断。

基于 IS-IS 的 GR 可以解决这个问题：重启路由器将重启状态通知给邻居，允许邻居重新建立邻接关系而不终止连接。

基于 IS-IS 的 GR 特性有以下几个特点：

- 发生了协议重启时，应用了 GR 特性的设备可重新向邻居发送连接请求，而不会终止邻接关系。

- 在路由协议重启完毕后，邻居协助当前设备完成路由信息同步，在尽量短的时间内使该设备的各种路由信息恢复到重启前的状态。通过 GR，在整个协议重启过程中，网络路由和转发保持高度稳定。
- 对于第一次启动的路由器，在 LSP 报文中设置过载标记位直到数据库同步，通知其它路由器当前路由器发生了问题，无法正确的执行路由选择和报文转发。

重启间隔时间用来指定路由器的重启时间。重启间隔时间在 IS-IS 的 Hello PDU 中设置为保持时间，这样在该路由器重启的时间内邻居不会断掉与其的邻接关系。

配置重启时抑制 SA（Suppress-Advertisement）位，即在重启路由器的 Hello PDU 中设置抑制发布 SA 位，重启路由器的邻居将继续发布该邻接关系。

分别在作为 GR Restarter 和 GR Helper 的设备上进行以下配置。

表1-38 配置 IS-IS GR

操作	命令	说明
进入系统视图	**system-view**	-
使能 IS-IS 路由进程，进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，系统没有运行 IS-IS
使能 IS-IS 协议的 GR 能力	**graceful-restart**	必选 缺省情况下，IS-IS 协议的 GR 能力处于关闭状态
配置 IS-IS 协议的 GR 重启间隔时间	**graceful-restart interval** *timer*	可选 缺省情况下，IS-IS 协议的 GR 重启间隔时间为 300 秒
配置重启时抑制 SA 位	**graceful-restart suppress-sa**	可选 缺省情况下，重启时不抑制 SA 位

1.9 配置邻接状态输出

表1-39 配置邻接状态输出

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
打开邻接状态输出开关	**log-peer-change**	必选 缺省情况下，邻接状态输出开关处于打开状态


说明

当打开邻接状态输出开关后，IS-IS 邻接状态的变化会输出到配置终端上，直至邻接状态输出开关被关闭。

1.10 使能 IS-IS 的 TRAP 功能

表1-40 使能 IS-IS 的 TRAP 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能 IS-IS 的 TRAP 功能	**is-snmp-traps enable**	必选 缺省情况下，IS-IS 的 TRAP 功能处于使能状态

1.11 配置 IS-IS 和 MIB 绑定

表1-41 配置 IS-IS 和 MIB 绑定

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 IS-IS 和 MIB 绑定	**isis mib-binding** *process-id*	必选 缺省情况下，MIB 操作绑定在 IS-IS 进程 1 上

1.12 IS-IS 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IS-IS 的运行情况，用户可以通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IS-IS 的数据库信息，或者复位特定邻居的数据信息。

表1-42 IS-IS 显示和维护

操作	命令
显示 IS-IS 的摘要信息	**display isis brief** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 调试开关的状态	**display isis debug-switches** { *process-id* | **vpn-instance** *vpn-instance-name* }
显示 IS-IS 协议的 GR 状态	**display isis graceful-restart status** [**level-1** | **level-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示使能 IS-IS 功能接口的信息	**display isis interface** [**statistics** | [*interface-type interface-number*] [**verbose**]] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 的链路状态数据库	**display isis lsdb** [[**l1** | **l2** | **level-1** | **level-2**] | [**lsp-id** *lspid* | **lsp-name** *lspname*] | **local** | **verbose**] * [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS Mesh-Group 的配置信息	**display isis mesh-group** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示系统 ID 到主机名称的映射关系表	**display isis name-table** [*process-id* | **vpn-instance** *vpn-instance-name*]

操作	命令
显示 IS-IS 的邻居信息	**display isis peer** [**verbose | statistics**] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 的 IPv4 路由信息	**display isis route** [**ipv4**] [[**level-1** | **level-2**] | **verbose**] * [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 进行 SPF 计算的日志信息	**display isis spf-log** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 统计信息	**display isis statistics** [**level-1** | **level-2** | **level-1-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]
清除所有 IS-IS 的数据结构信息	**reset isis all** [*process-id* | **vpn-instance** *vpn-instance-name*]
清除 IS-IS 特定邻居的数据信息	**reset isis peer** *system-id* [*process-id* | **vpn-instance** *vpn-instance-name*]

1.13 IS-IS 典型配置举例

1.13.1 IS-IS 基本配置举例

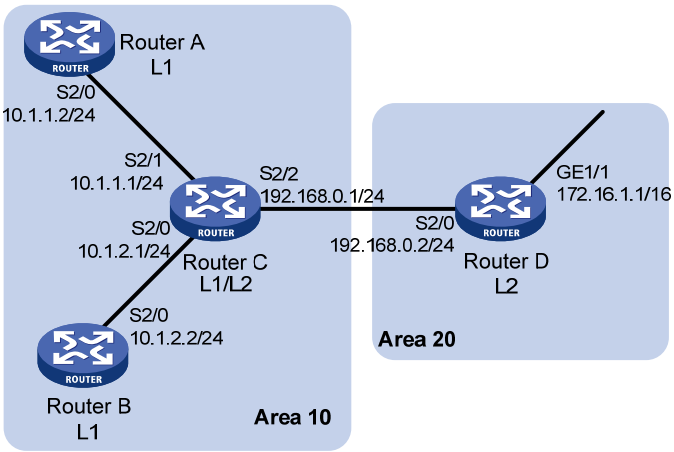
1. 组网需求

如 图 1-15所示，Router A、Router B、Router C和Router D属于同一自治系统，要求他们之间通过IS-IS协议达到IP网络互连的目的。

Router A 和 Router B 为 Level-1 路由器，Router D 为 Level-2 路由器，Router C 作为 Level-1-2 路由器将两个区域相连。Router A、Router B 和 Router C 的区域号为 10，Router D 的区域号为 20。

2. 组网图

图1-15 IS-IS 基本配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IS-IS

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] is-level level-1
```

```
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] isis enable 1
[RouterA-Serial2/0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] is-level level-1
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] isis enable 1
[RouterB-Serial2/0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 10.0000.0000.0003.00
[RouterC-isis-1] quit
[RouterC] interface serial 2/0
[RouterC-Serial2/0] isis enable 1
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] isis enable 1
[RouterC-Serial2/1] quit
[RouterC] interface serial 2/2
[RouterC-Serial2/2] isis enable 1
[RouterC-Serial2/2] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] is-level level-2
[RouterD-isis-1] network-entity 20.0000.0000.0004.00
[RouterD-isis-1] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] isis enable 1
[RouterD-GigabitEthernet1/1] quit
[RouterD] interface serial 2/0
[RouterD-Serial2/0] isis enable 1
[RouterD-Serial2/0] quit
```

(3) 验证配置结果

显示各路由器的 IS-IS LSDB 信息。

```
[RouterA] display isis lsdb

                        Database information for ISIS(1)
```

```
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0001.00-00* 0x0000000d   0xb184     879         68      0/0/0
0000.0000.0002.00-00  0x0000000c   0xcf65     493         68      0/0/0
0000.0000.0003.00-00  0x00000013   0x2f38     594         111     1/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[RouterB] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0001.00-00  0x0000000d   0xb184     707         68      0/0/0
0000.0000.0002.00-00* 0x0000000c   0xcd66     1167        68      0/0/0
0000.0000.0003.00-00  0x00000013   0x2d39     1136        111     1/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[RouterC] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum  Holdtime   Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0001.00-00  0x0000000d   0xc57a    991        68      0/0/0
0000.0000.0002.00-00  0x0000000c   0xef4d    1025       68      0/0/0
0000.0000.0003.00-00* 0x00000013   0x93dd    1026       111     1/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload


                          Level-2 Link State Database

LSPID                 Seq Num      Checksum  Holdtime   Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0003.00-00* 0x00000013   0xbb56    1026       100     0/0/0
```

```
0000.0000.0004.00-00  0x00000005   0xd086    904        84      0/0/0

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[RouterD] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------

                          Level-2 Link State Database

LSPID                 Seq Num      Checksum  Holdtime   Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0003.00-00  0x00000013   0xbb56    910        100     0/0/0
0000.0000.0004.00-00* 0x00000005   0xd086    791        84      0/0/0

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
```

显示各路由器的 IS-IS 路由信息。Level-1 路由器的路由表中应该有一条缺省路由，且下一跳为 Level-1-2 路由器，Level-2 路由器的路由表中应该有所有 Level-1 和 Level-2 的路由。

```
[RouterA] display isis route

                        Route information for ISIS(1)
                        ------------------------------

                     ISIS(1) IPv4 Level-1 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost  ExtCost ExitInterface  NextHop       Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10       NULL    S2/0           Direct        D/L/-
 10.1.2.0/24          20       NULL    S2/0           10.1.1.1      R/-/-
 192.168.0.0/24       20       NULL    S2/0           10.1.1.1      R/-/-
 0.0.0.0/0            10       NULL    S2/0           10.1.1.1      R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[RouterC] display isis route

                        Route information for ISIS(1)
                        ------------------------------

                     ISIS(1) IPv4 Level-1 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop     Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1           Direct      D/L/-
```

```
 10.1.2.0/24          10        NULL    S2/0            Direct      D/L/-
 192.168.0.0/24       10        NULL    S2/2            Direct      D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set


                     ISIS(1) IPv4 Level-2 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1            Direct       D/L/-
 10.1.2.0/24          10        NULL    S2/0            Direct       D/L/-
 192.168.0.0/24       10        NULL    S2/2            Direct       D/L/-
 172.16.0.0/16        20        NULL    S2/2            192.168.0.2  R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
[RouterD] display isis route

                        Route information for ISIS(1)
                        -----------------------------

                     ISIS(1) IPv4 Level-2 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
------------------------------------------------------------------------------
 192.168.0.0/24       10        NULL    S2/0            Direct       D/L/-
 10.1.1.0/24          20        NULL    S2/0            192.168.0.1  R/-/-
 10.1.2.0/24          20        NULL    S2/0            192.168.0.1  R/-/-
 172.16.0.0/16        10        NULL    GE1/0          Direct       D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

1.13.2 配置 IS-IS 的 DIS 选择

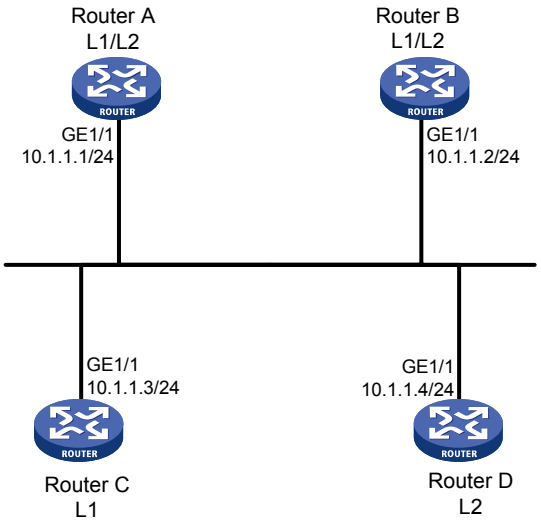
1. 组网需求

如 图 1-16所示，Router A、Router B、Router C和Router D都运行IS-IS路由协议以实现互连，他们属于同一区域 10，网络类型为广播网（以太网）。其中Router A和Router B是Level-1-2 路由器，Router C为Level-1 路由器，Router D为Level-2 路由器。

要求通过改变接口的 DIS 优先级，将 Router A 配置为 Level-1-2 的 DIS 路由器。

2. 组网图

图1-16 配置 IS-IS 的 DIS 选择组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IS-IS

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] isis enable 1
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] isis enable 1
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 10.0000.0000.0003.00
[RouterC-isis-1] is-level level-1
[RouterC-isis-1] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] isis enable 1
[RouterC-GigabitEthernet1/1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] network-entity 10.0000.0000.0004.00
[RouterD-isis-1] is-level level-2
[RouterD-isis-1] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] isis enable 1
[RouterD-GigabitEthernet1/1] quit
```

查看 Router A 的 IS-IS 邻居信息。

```
[RouterA] display isis peer
                         Peer information for ISIS(1)
                         ---------------------------

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0003.01
  State: Up     HoldTime: 21s        Type: L1(L1L2)     PRI: 64

  System Id: 0000.0000.0003
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0003.01
  State: Up     HoldTime: 6s         Type: L1           PRI: 64

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0004.01
  State: Up     HoldTime: 23s        Type: L2(L1L2)     PRI: 64

  System Id: 0000.0000.0004
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0004.01
  State: Up     HoldTime: 23s        Type: L2           PRI: 64
```

显示 Router A 的 IS-IS 接口信息。

```
[RouterA] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State      IPV6.State    MTU    Type   DIS
001         Up             Down       1497   L1/L2  No/No
```

显示 Router C 的 IS-IS 接口信息。

```
[RouterC] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State      IPV6.State    MTU   Type   DIS
001         Up            Down        1497  L1/L2  Yes/No
```

显示 Router D 的 IS-IS 接口信息。

```
[RouterD] display isis interface

                    Interface information for ISIS(1)
                    ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State       IPV6.State   MTU   Type    DIS
001         Up             Down       1497  L1/L2   No/Yes
```

从接口信息中可以看到，在使用缺省 DIS 优先级的情况下，Router C 为 Level-1 的 DIS，Router D 为 Level-2 的 DIS。Level-1 和 Level-2 的伪节点分别是 0000.0000.0003.01 和 0000.0000.0004.01。

配置 Router A 的 DIS 优先级

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] isis dis-priority 100
```

查看 Router A 的 IS-IS 邻居信息。

```
[RouterA] display isis peer

                      Peer information for ISIS(1)
                      ----------------------------

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 29s        Type: L1(L1L2)     PRI: 64

  System Id: 0000.0000.0003
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 22s        Type: L1           PRI: 64

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 22s        Type: L2(L1L2)     PRI: 64

  System Id: 0000.0000.0004
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 22s        Type: L2           PRI: 64
```

查看 Router A 的 IS-IS 接口信息。

```
[RouterA] display isis interface

                    Interface information for ISIS(1)
                    ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State      IPV6.State      MTU   Type   DIS
001         Up              Down        1497  L1/L2  Yes/Yes
```

从上述信息中可以看到，在改变 IS-IS 接口的 DIS 优先级后，Router A 立即成为 Level-1-2 的 DIS，且伪节点是 0000.0000.0001.01。

显示 Router C 的 IS-IS 邻居和接口信息。

```
[RouterC] display isis peer

                        Peer information for ISIS(1)
                        ----------------------------

  System Id: 0000.0000.0001
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 7s         Type: L1           PRI: 100

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 23s        Type: L1           PRI: 64
[RouterC] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State          IPV6.State      MTU   Type   DIS
001         Up                  Down        1497  L1/L2  No/No
```

显示 Router D 的 IS-IS 邻居和接口信息。

```
[RouterD] display isis peer

                        Peer information for ISIS(1)
                        ----------------------------

  System Id: 0000.0000.0001
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 7s         Type: L2           PRI: 100

  System Id: 0000.0000.0002
  Interface: GigabitEthernet1/1             Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 26s        Type: L2           PRI: 64

[RouterD] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: GigabitEthernet1/1
Id      IPV4.State          IPV6.State    MTU    Type    DIS
001         Up                  Down      1497   L1/L2   No/No
```

1.13.3 配置 IS-IS 引入外部路由

1. 组网需求

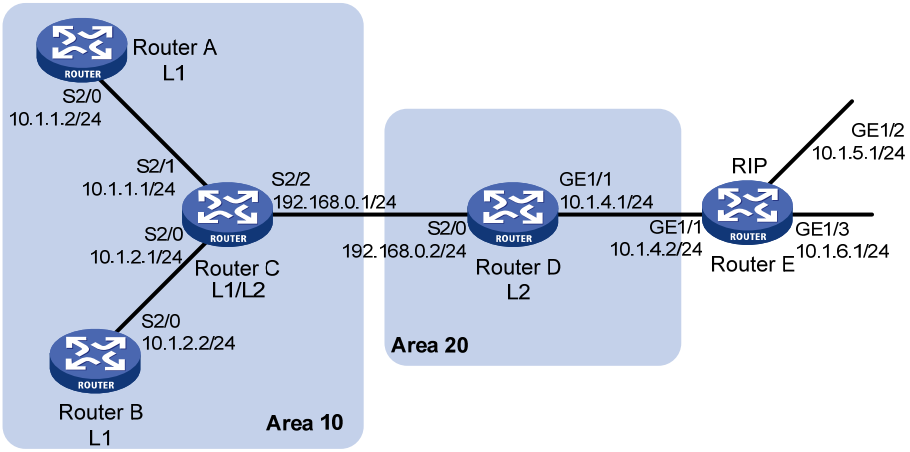
如 图 1-17所示：

- Router A、Router B、Router C 和 Router D 属于同一自治系统，要求他们之间通过 IS-IS 协议达到 IP 网络互连的目的。

- Router A 和 Router B 为 Level-1 路由器，Router D 为 Level-2 路由器，Router C 作为 Level-1-2 路由器将两个区域相连。Router A、Router B 和 Router C 的区域号为 10，Router D 的区域号为 20。
- 在 Router D 的 IS-IS 进程中引入 RIP 路由。

2. 组网图

图1-17 配置 IS-IS 引入外部路由组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IS-IS 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] is-level level-1
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] isis enable 1
[RouterA-Serial2/0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] is-level level-1
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] isis enable 1
[RouterB-Serial2/0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
```

```
[RouterC-isis-1] network-entity 10.0000.0000.0003.00
[RouterC-isis-1] quit
[RouterC] interface serial 2/0
[RouterC-Serial2/0] isis enable 1
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] isis enable 1
[RouterC-Serial2/1] quit
[RouterC] interface serial 2/2
[RouterC-Serial2/2] isis enable 1
[RouterC-Serial2/2] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] is-level level-2
[RouterD-isis-1] network-entity 20.0000.0000.0004.00
[RouterD-isis-1] quit
[RouterD] interface serial 2/0
[RouterD-Serial2/0] isis enable 1
[RouterD-Serial2/0] quit
```

显示各路由器的 IS-IS 路由信息。

```
[RouterA] display isis route

                        Route information for ISIS(1)
                        ----------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost  ExtCost ExitInterface  NextHop       Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10       NULL    S2/0           Direct        D/L/-
 10.1.2.0/24          20       NULL    S2/0           10.1.1.1      R/-/-
 192.168.0.0/24       20       NULL    S2/0           10.1.1.1      R/-/-
 0.0.0.0/0            10       NULL    S2/0           10.1.1.1      R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[RouterC] display isis route

                        Route information for ISIS(1)
                        ----------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop     Flags
```

```
------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1            Direct      D/L/-
 10.1.2.0/24          10        NULL    S2/0            Direct      D/L/-
 192.168.0.0/24       10        NULL    S2/2            Direct      D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set


                     ISIS(1) IPv4 Level-2 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1           Direct       D/L/-
 10.1.2.0/24          10        NULL    S2/0           Direct       D/L/-
 192.168.0.0/24       10        NULL    S2/2           Direct       D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[RouterD] display isis route

                         Route information for ISIS(1)
                         -----------------------------

                     ISIS(1) IPv4 Level-2 Forwarding Table
                     -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
------------------------------------------------------------------------------
 192.168.0.0/24       10        NULL    S2/0           Direct       D/L/-
 10.1.1.0/24          20        NULL    S2/0           192.168.0.1  R/-/-
 10.1.2.0/24          20        NULL    S2/0           192.168.0.1  R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

(3) 在 Router D 和 Router E 之间运行 RIPv2，在 Router D 上配置 IS-IS 进程引入 RIP 路由。

在 Router D 上配置 RIPv2。

```
[RouterD] rip 1
[RouterD-rip-1] network 10.0.0.0
[RouterD-rip-1] version 2
[RouterD-rip-1] undo summary
```

在 Router E 上配置 RIPv2。

```
[RouterE] rip 1
[RouterE-rip-1] network 10.0.0.0
[RouterE-rip-1] version 2
[RouterE-rip-1] undo summary
```

在 Router D 上配置 IS-IS 进程引入 RIP 进程的路由。

```
[RouterD-rip-1] quit
[RouterD] isis 1
[RouterD–isis]import-route rip level-2
```

显示 Router C 的 IS-IS 路由信息。

```
[RouterC] display isis route

                        Route information for ISIS(1)
                        ------------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop     Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1            Direct      D/L/-
 10.1.2.0/24          10        NULL    S2/0            Direct      D/L/-
 192.168.0.0/24       10        NULL    S2/2            Direct      D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set


                    ISIS(1) IPv4 Level-2 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    S2/1           Direct       D/L/-
 10.1.2.0/24          10        NULL    S2/0           Direct       D/L/-
 192.168.0.0/24       10        NULL    S2/2           Direct       D/L/-
 10.1.4.0/24          10        NULL    S2/2           192.168.0.2  R/L/-
 10.1.5.0/24          20        NULL    S2/2           192.168.0.2  R/L/-
 10.1.6.0/24          20        NULL    S2/2           192.168.0.2  R/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

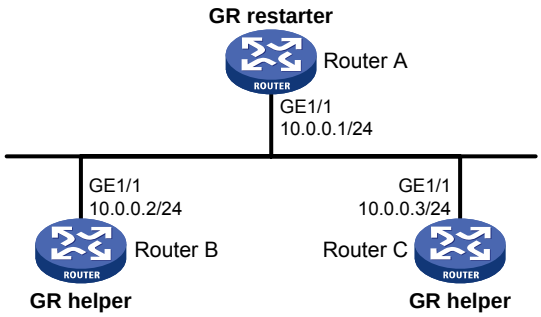
1.13.4 IS-IS GR 配置举例

1. 组网需求

如 图 1-18所示，Router A、Router B和Router C属于同一域。这三台路由器都运行IS-IS协议以实现路由互连。

2. 组网图

图1-18 IS-IS GR 配置组网图



3. 配置步骤

(1) 配置各路由器接口的 IP 地址和 IS-IS 协议

请按照 图 1-18配置各接口的IP地址和子网掩码，具体配置过程略。

配置各路由器之间采用 IS-IS 协议进行互连，确保 Router A、Router B 和 Router C 之间能够在网络层互通，并且各路由器之间能够借助 IS-IS 协议实现动态路由更新，具体配置过程略。

(2) 配置 IS-IS GR

使能 Router A 的 IS-IS 协议的 GR 能力，并配置重启间隔时间。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] graceful-restart
[RouterA-isis-1] graceful-restart interval 150
[RouterA-isis-1] return
```

Router B 和 Router C 的配置与 Router A 相似，配置过程略。

(3) 检验配置效果

Router A 分别与 Router B 和 Router C 建立邻接关系后，三台路由器开始交换路由信息。Router A 的 IS-IS 协议重启，进入重启模式后，通过 GR 机制向邻居重新发送连接请求，同步数据库。使用 **display isis graceful-restart status** 命令，可查看 Router A 上 IS-IS 协议的 GR 状态。

重启 Router A 的 IS-IS 进程。

```
<RouterA> reset isis all 1
Warning : Reset ISIS process? [Y/N]:y
```

查看 Router A 上 IS-IS 协议的 GR 状态。

```
<RouterA> display isis graceful-restart status
                Restart information for IS-IS(1)
--------------------------------------------------------------------
IS-IS(1) Level-1 Restart Status
Restart Interval: 150
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTARTING
  Number of LSPs Awaited: 3
```

```
  T3 Timer Status:
    Remaining Time: 140
  T2 Timer Status:
    Remaining Time: 59

IS-IS(1) Level-2 Restart Status
Restart Interval: 150
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTARTING
  Number of LSPs Awaited: 3
  T3 Timer Status:
    Remaining Time: 140
  T2 Timer Status:
Remaining Time: 59
```

1.13.5 IS-IS 验证配置举例

1. 组网需求

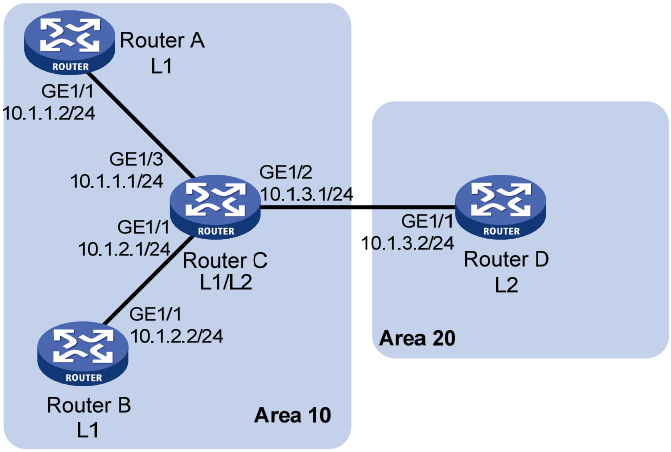
如 图 1-19所示，Router A、Router B、Router C和Router D属于同一路由域，要求它们之间通过IS-IS协议达到IP网络互连的目的。

其中，Router A、Router B 和 Router C 属于同一个区域，区域号为 10，Router D 属于另外一个区域，区域号为 20。

在区域 10 内配置区域验证，防止不可信任的路由信息加入到区域 10 的 LSDB 中；在 Router C 和 Router D 上配置路由域验证，防止将不可信的路由信息注入当前路由域；分别在 Router A、Router B、Router C 和 Router D 上配置邻居关系验证。

2. 组网图

图1-19 IS-IS 验证配置举例图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IS-IS 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] is-level level-1
[RouterA-isis-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] isis enable 1
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] is-level level-1
[RouterB-isis-1] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] isis enable 1
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 10.0000.0000.0003.00
[RouterC-isis-1] quit
 [RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] isis enable 1
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] isis enable 1
[RouterC-GigabitEthernet1/2] quit
[RouterC] interface GigabitEthernet 1/3
[RouterC-GigabitEthernet1/3] isis enable 1
[RouterC-GigabitEthernet1/3] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] network-entity 20.0000.0000.0001.00
[RouterD-isis-1] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] isis enable 1
[RouterD-GigabitEthernet1/1] quit
```

(3) 在 Router A、Router B、Router C 和 Router D 之间建立邻居关系验证

分别在 Router A 的 GigabitEthernet1/1、Router C 的 GigabitEthernet1/3 配置邻居关系验证，验证方式为 MD5 密文，验证密码为“eRg”。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] isis authentication-mode md5 eRg
[RouterA-GigabitEthernet1/1] quit
```

```
[RouterC] interface GigabitEthernet 1/3
[RouterC-GigabitEthernet1/3] isis authentication-mode md5 eRg
[RouterC-GigabitEthernet1/3] quit
```

分别在 Router B 的 GigabitEthernet1/1、Router C 的 GigabitEthernet1/1 配置邻居关系验证，验证方式为 MD5 密文，验证密码为“t5Hr”。

```
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] isis authentication-mode md5 t5Hr
[RouterB-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] isis authentication-mode md5 t5Hr
[RouterC-GigabitEthernet1/1] quit
```

分别在 Router C 的 GigabitEthernet1/2、Router D 的 GigabitEthernet1/1 配置邻居关系验证，验证方式为 MD5 密文，验证密码为“hSec”。

```
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] isis authentication-mode md5 hSec
[RouterC-GigabitEthernet1/2] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] isis authentication-mode md5 hSec
[RouterD-GigabitEthernet1/1] quit
```

(4) 在 Router A、Router B 和 Router C 上配置区域验证，验证方式为 MD5 密文验证，验证密码为“10Sec”。

```
[RouterA] isis 1
[RouterA-isis-1] area-authentication-mode md5 10Sec
[RouterA-isis-1] quit
[RouterB] isis 1
[RouterB-isis-1] area-authentication-mode md5 10Sec
[RouterB-isis-1] quit
[RouterC] isis 1
[RouterC-isis-1] area-authentication-mode md5 10Sec
[RouterC-isis-1] quit
```

(5) 在 Router C 和 Router D 上配置路由域验证，验证方式为 MD5 密文验证，验证密码为“1020Sec”。

```
[RouterC] isis 1
[RouterC-isis-1] domain-authentication-mode md5 1020Sec
[RouterC-isis-1] quit
[RouterD] isis 1
[RouterD-isis-1] domain-authentication-mode md5 1020Sec
```

目 录

1 OSPF配置 1-1
1.1 OSPF简介 1-1
1.1.1 OSPF的基本概念 1-1
1.1.2 OSPF区域 1-3
1.1.3 路由器的类型 1-6
1.1.4 OSPF的网络类型 1-7
1.1.5 DR/BDR 1-8
1.1.6 OSPF的协议报文 1-9
1.1.7 系统支持的OSPF特性 1-17
1.1.8 协议规范 1-20
1.2 OSPF配置任务简介 1-20
1.3 使能OSPF功能 1-22
1.3.1 配置准备 1-22
1.3.2 使能OSPF功能 1-22
1.4 配置OSPF区域 1-23
1.4.1 配置准备 1-23
1.4.2 配置Stub区域特性 1-23
1.4.3 配置NSSA区域 1-24
1.4.4 配置虚连接 1-24
1.5 配置OSPF的网络类型 1-25
1.5.1 配置准备 1-25
1.5.2 配置OSPF接口网络类型为广播 1-26
1.5.3 配置OSPF接口网络类型为NBMA 1-26
1.5.4 配置OSPF接口网络类型为P2MP 1-27
1.5.5 配置OSPF接口网络类型为P2P 1-27
1.6 配置OSPF的路由信息控制 1-27
1.6.1 配置准备 1-27
1.6.2 配置OSPF路由聚合 1-27
1.6.3 配置OSPF对通过接收到的LSA计算出来的路由信息进行过滤 1-28
1.6.4 配置对Type-3 LSA过滤 1-29
1.6.5 配置OSPF接口的开销值 1-29
1.6.6 配置OSPF支持的路由最大数目 1-30
1.6.7 配置OSPF最大等价路由条数 1-30
1.6.8 配置OSPF协议的优先级 1-31
1.6.9 配置OSPF引入外部路由 1-31
1.6.10 配置发布一条主机路由 1-32
1.7 配置OSPF网络调整优化 1-33
1.7.1 配置准备 1-33
1.7.2 配置OSPF报文定时器 1-33
1.7.3 配置接口传送LSA的延迟时间 1-34
1.7.4 配置SPF计算时间间隔 1-34
1.7.5 配置LSA重复到达的最小时间间隔 1-35

1.7.6 配置LSA重新生成的时间间隔 1-35
1.7.7 禁止接口发送OSPF报文 1-36
1.7.8 配置Stub路由器 1-36
1.7.9 配置OSPF验证 1-37
1.7.10 配置DD报文中的MTU 1-37
1.7.11 配置LSDB中External LSA的最大数量 1-38
1.7.12 配置兼容RFC 1583 的外部路由选择规则 1-38
1.7.13 配置邻接状态输出 1-38
1.7.14 配置OSPF网管功能 1-39
1.7.15 使能日志功能 1-39
1.7.16 使能Opaque LSA发布接收能力 1-39
1.7.17 配置OSPF优先接收并处理Hello报文 1-40
1.7.18 配置LSU报文的发送速率 1-40
1.8 配置OSPF GR 1-40
1.8.1 配置GR Restarter 1-41
1.8.2 配置GR Helper 1-41
1.8.3 重启OSPF GR进程 1-42
1.9 OSPF显示和维护 1-43
1.10 典型配置举例 1-44
1.10.1 配置OSPF基本功能 1-44
1.10.2 配置OSPF引入自治系统外部路由 1-47
1.10.3 配置OSPF发布聚合路由 1-49
1.10.4 配置OSPF的Stub区域 1-51
1.10.5 配置OSPF的NSSA区域 1-54
1.10.6 配置OSPF的DR选择 1-56
1.10.7 配置OSPF虚连接 1-60
1.10.8 OSPF GR配置举例 1-62
1.10.9 配置路由过滤 1-65
1.11 常见配置错误举例 1-67
1.11.1 OSPF邻居无法建立 1-67
1.11.2 OSPF路由信息不正确 1-68

1 OSPF 配置

1.1 OSPF 简介

OSPF（Open Shortest Path First，开放最短路径优先）是 IETF 组织开发的一个基于链路状态的内部网关协议。目前针对 IPv4 协议使用的是 OSPF Version 2（RFC 2328）。


说明

本章若没有特别说明，下文中所提到的 OSPF 均指 OSPFv2。

OSPF 具有如下特点：

- 适应范围广——支持各种规模的网络，最多可支持几百台路由器。
- 快速收敛——在网络的拓扑结构发生变化后立即发送更新报文，使这一变化在自治系统中同步。
- 无自环——由于 OSPF 根据收集到的链路状态用最短路径树算法计算路由，从算法本身保证了不会生成自环路由。
- 区域划分——允许自治系统的网络被划分成区域来管理，区域间传送的路由信息被进一步抽象，从而减少了占用的网络带宽。
- 等价路由——支持到同一目的地址的多条等价路由。
- 路由分级——使用 4 类不同的路由，按优先顺序来说分别是：区域内路由、区域间路由、第一类外部路由、第二类外部路由。
- 支持验证——支持基于接口的报文验证，以保证报文交互和路由计算的安全性。
- 组播发送——在某些类型的链路上以组播地址发送协议报文，减少对其他设备的干扰。

1.1.1 OSPF 的基本概念

1. 自治系统（Autonomous System）

一组使用相同路由协议交换路由信息的路由器，缩写为 AS。

2. OSPF 路由的计算过程

同一个区域内，OSPF 协议路由的计算过程可简单描述如下：

- 每台 OSPF 路由器根据自己周围的网络拓扑结构生成 LSA（Link State Advertisement，链路状态通告），并通过更新报文将 LSA 发送给网络中的其它 OSPF 路由器。
- 每台 OSPF 路由器都会收集其它路由器通告的 LSA，所有的 LSA 放在一起便组成了 LSDB（Link State Database，链路状态数据库）。LSA 是对路由器周围网络拓扑结构的描述，LSDB 则是对整个自治系统的网络拓扑结构的描述。
- OSPF 路由器将 LSDB 转换成一张带权的有向图，这张图便是对整个网络拓扑结构的真实反映。各个路由器得到的有向图是完全相同的。

- 每台路由器根据有向图，使用 SPF 算法计算出一棵以自己为根的最短路径树，这棵树给出了到自治系统中各节点的路由。

3. 路由器 ID 号

一台运行 OSPF 协议路由器，每一个 OSPF 进程必须存在自己的 Router ID（路由器 ID）。Router ID 是一个 32 比特无符号整数，可以在一个自治系统中唯一的标识一台路由器。

4. OSPF 的协议报文

OSPF 有五种类型的协议报文：

- Hello 报文：周期性发送，用来发现和维持 OSPF 邻居关系。内容包括一些定时器的数值、DR（Designated Router，指定路由器）、BDR（Backup Designated Router，备份指定路由器）以及自己已知的邻居。
- DD（Database Description，数据库描述）报文：描述了本地 LSDB 中每一条 LSA 的摘要信息，用于两台路由器进行数据库同步。
- LSR（Link State Request，链路状态请求）报文：向对方请求所需的 LSA。两台路由器互相交换 DD 报文之后，得知对端的路由器有哪些 LSA 是本地的 LSDB 所缺少的，这时需要发送 LSR 报文向对方请求所需的 LSA。内容包括所需要的 LSA 的摘要。
- LSU（Link State Update，链路状态更新）报文：向对方发送其所需要的 LSA。
- LSAck（Link State Acknowledgment，链路状态确认）报文：用来对收到的 LSA 进行确认。内容是需要确认的 LSA 的 Header（一个报文可对多个 LSA 进行确认）。

5. LSA 的类型

OSPF 中对链路状态信息的描述都是封装在 LSA 中发布出去，常用的 LSA 有以下几种类型：

- Router LSA（Type1）：由每个路由器产生，描述路由器的链路状态和开销，在其始发的区域内传播。
- Network LSA（Type2）：由 DR 产生，描述本网段所有路由器的链路状态，在其始发的区域内传播。
- Network Summary LSA（Type3）：由 ABR（Area Border Router，区域边界路由器）产生，描述区域内某个网段的路由，并通告给其他区域。
- ASBR Summary LSA（Type4）：由 ABR 产生，描述到 ASBR（Autonomous System Boundary Router，自治系统边界路由器）的路由，通告给相关区域。
- AS External LSA（Type5）：由 ASBR 产生，描述到 AS（Autonomous System，自治系统）外部的路由，通告到所有的区域（除了 Stub 区域和 NSSA 区域）。
- NSSA External LSA（Type7）：由 NSSA（Not-So-Stubby Area）区域内的 ASBR 产生，描述到 AS 外部的路由，仅在 NSSA 区域内传播。
- Opaque LSA：是一个被提议的 LSA 类别，由标准的 LSA 头部后面跟随特殊应用的信息组成，可以直接由 OSPF 协议使用，或者由其它应用分发信息到整个 OSPF 域间接使用。Opaque LSA 分为 Type 9、Type10、Type11 三种类型，泛洪区域不同；其中，Type 9 的 Opaque LSA 仅在本地链路范围进行泛洪，Type 10 的 Opaque LSA 仅在本地区域范围进行泛洪，Type 11 的 LSA 可以在一个自治系统范围进行泛洪。

6. 邻居和邻接

在 OSPF 中，邻居（Neighbor）和邻接（Adjacency）是两个不同的概念。

OSPF 路由器启动后，便会通过 OSPF 接口向外发送 Hello 报文。收到 Hello 报文的 OSPF 路由器会检查报文中所定义的参数，如果双方一致就会形成邻居关系。

形成邻居关系的双方不一定都能形成邻接关系，这要根据网络类型而定。只有当双方成功交换 DD 报文，交换 LSA 并达到 LSDB 的同步之后，才形成真正意义上的邻接关系。

1.1.2 OSPF 区域

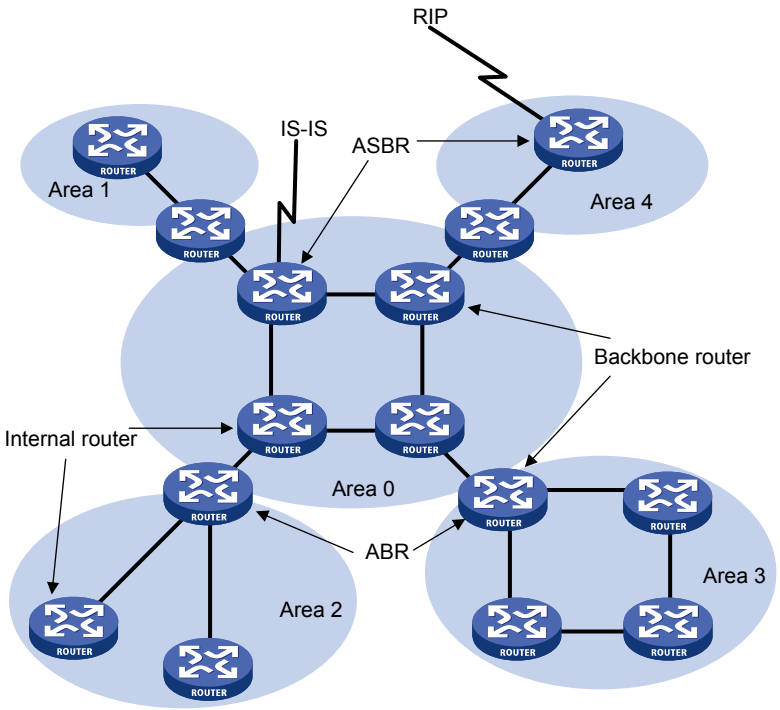
1. 区域划分

随着网络规模日益扩大，当一个大型网络中的路由器都运行 OSPF 路由协议时，路由器数量的增多会导致 LSDB 非常庞大，占用大量的存储空间，并使得运行 SPF 算法的复杂度增加，导致 CPU 负担很重。

在网络规模增大之后，拓扑结构发生变化的概率也增大，网络会经常处于“振荡”之中，造成网络中会有大量的 OSPF 协议报文在传递，降低了网络的带宽利用率。更为严重的是，每一次变化都会导致网络中所有的路由器重新进行路由计算。

OSPF协议通过将自治系统划分成不同的区域（Area）来解决上述问题。区域是从逻辑上将路由器划分为不同的组，每个组用区域号（Area ID）来标识。如 图 1-1所示。

图1-1 OSPF 区域划分



区域的边界是路由器，而不是链路。一个路由器可以属于不同的区域，但是一个网段（链路）只能属于一个区域，或者说每个运行 OSPF 的接口必须指明属于哪一个区域。划分区域后，可以在区域边界路由器上进行路由聚合，以减少通告到其他区域的 LSA 数量，还可以将网络拓扑变化带来的影响最小化。

2. 骨干区域与虚连接

(1) 骨干区域（Backbone Area）

OSPF 划分区域之后，并非所有的区域都是平等的关系。其中有一个区域是与众不同的，它的区域号（Area ID）是 0，通常被称为骨干区域。骨干区域负责区域之间的路由，非骨干区域之间的路由信息必须通过骨干区域来转发。对此，OSPF 有两个规定：

- 所有非骨干区域必须与骨干区域保持连通；
- 骨干区域自身也必须保持连通。

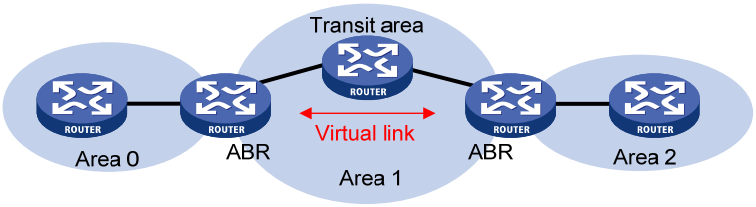
但在实际应用中，可能会因为各方面条件的限制，无法满足这个要求。这时可以通过配置 OSPF 虚连接（Virtual Link）予以解决。

(2) 虚连接（Virtual Link）

虚连接是指在两台 ABR 之间通过一个非骨干区域而建立的一条逻辑上的连接通道。它的两端必须是 ABR，而且必须在两端同时配置方可生效。为虚连接两端提供一条非骨干区域内部路由的区域称为传输区（Transit Area）。

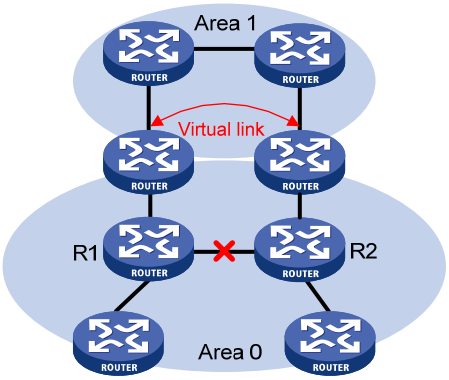
在 图 1-2中，Area2 与骨干区域之间没有直接相连的物理链路，但可以在ABR上配置虚连接，使Area2通过一条逻辑链路与骨干区域保持连通。

图1-2 虚连接示意图之一



虚连接的另外一个应用是提供冗余的备份链路，当骨干区域因链路故障不能保持连通时，通过虚连接仍然可以保证骨干区域在逻辑上的连通性。如 图 1-3所示。

图1-3 虚连接示意图之二



虚连接相当于在两个 ABR 之间形成了一个点到点的连接，因此，在这个连接上，和物理接口一样可以配置接口的各参数，如发送 Hello 报文间隔等。

两台 ABR 之间直接传递 OSPF 报文信息，它们之间的 OSPF 路由器只是起到一个转发报文的作用。由于协议报文的目的地址不是中间这些路由器，所以这些报文对于它们而言是透明的，只是当作普通的 IP 报文来转发。

3. Stub 区域

Stub 区域是一些特定的区域，Stub 区域的 ABR 不允许注入 Type5 LSA，在这些区域中路由器的路由表规模以及路由信息传递的数量都会大大减少。

为了进一步减少 Stub 区域中路由器的路由表规模以及路由信息传递的数量，可以将该区域配置为 Totally Stub（完全 Stub）区域，该区域的 ABR 不会将区域间的路由信息和外部路由信息传递到本区域。

（Totally） Stub 区域是一种可选的配置属性，但并不是每个区域都符合配置的条件。通常来说，（Totally） Stub 区域位于自治系统的边界。

为保证到本自治系统的其他区域或者自治系统外的路由依旧可达，该区域的 ABR 将生成一条缺省路由，并发布给本区域中的其他非 ABR 路由器。

配置（Totally） Stub 区域时需要注意下列几点：

- 骨干区域不能配置成（Totally） Stub 区域。
- 如果要将一个区域配置成 Stub 区域，则该区域中的所有路由器必须都要配置 **stub** 命令。
- 如果要将一个区域配置成 Totally Stub 区域，该区域中的所有路由器必须配置 **stub** 命令，该区域的 ABR 路由器需要配置 **stub [no-summary]**命令。
- （Totally） Stub 区域内不能存在 ASBR，即自治系统外部的路由不能在本区域内传播。
- 虚连接不能穿过（Totally） Stub 区域。

4. NSSA 区域

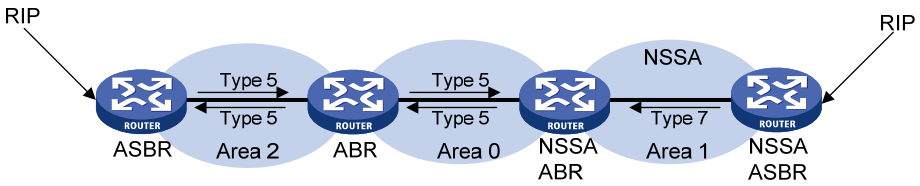
NSSA（Not-So-Stubby Area）区域是 Stub 区域的变形，与 Stub 区域有许多相似的地方。NSSA 区域也不允许 Type5 LSA 注入，但可以允许 Type7 LSA 注入。Type7 LSA 由 NSSA 区域的 ASBR 产生，在 NSSA 区域内传播。当 Type7 LSA 到达 NSSA 的 ABR 时，由 ABR 将 Type7 LSA 转换成 Type5 LSA，传播到其他区域。

如 图 1-4所示，运行OSPF协议的自治系统包括 3 个区域：区域 1、区域 2 和区域 0，另外两个自治系统运行RIP协议。区域 1 被定义为NSSA区域，区域 1 接收的RIP路由传播到NSSA ASBR后，由NSSA ASBR产生Type7 LSA在区域 1 内传播，当Type7 LSA到达NSSA ABR后，转换成Type5 LSA传播到区域 0 和区域 2。

另一方面，运行 RIP 的自治系统的 RIP 路由通过区域 2 的 ASBR 产生 Type5 LSA 在 OSPF 自治系统中传播。但由于区域 1 是 NSSA 区域，所以 Type5 LSA 不会到达区域 1。

与 Stub 区域一样，虚连接也不能穿过 NSSA 区域。

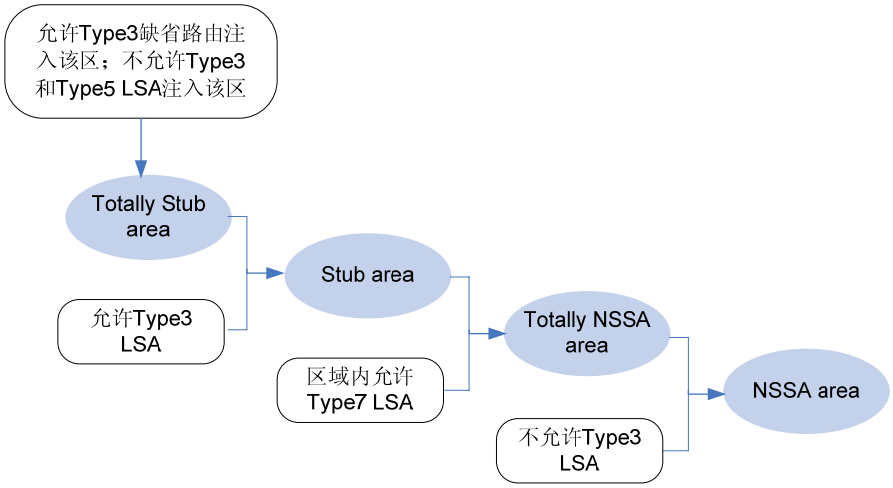
图1-4 NSSA 区域



5. 各区域特性

如 图 1-5所示，列出了各个区域之间的区别。

图1-5 区域特性示意图



各区域特性：

- Totally Stub 区域：允许 ABR 发布的 Type3 缺省路由，不允许自治系统外部路由和区域间的路由；
- Stub 区域：和 Totally Stub 区域不同在于该区域允许区域间路由；
- Nssa 区域：和 Stub 区域不同在于该区域允许自治系统外部路由的引入，由 ASBR 发布 Type7 LSA 通告给本区域；
- Totally Nssa 区域：和 Nssa 区域不同在于该区域不允许区域间路由。

1.1.3 路由器的类型

1. 路由器的分类

OSPF 路由器根据在 AS 中的不同位置，可以分为以下四类：

(1) 区域内路由器（Internal Router）

该类路由器的所有接口都属于同一个 OSPF 区域。

(2) 区域边界路由器 ABR（Area Border Router）

该类路由器可以同时属于两个以上的区域，但其中一个必须是骨干区域（骨干区域的介绍请参见下一小节）。ABR 用来连接骨干区域和非骨干区域，它与骨干区域之间既可以是物理连接，也可以是逻辑上的连接。

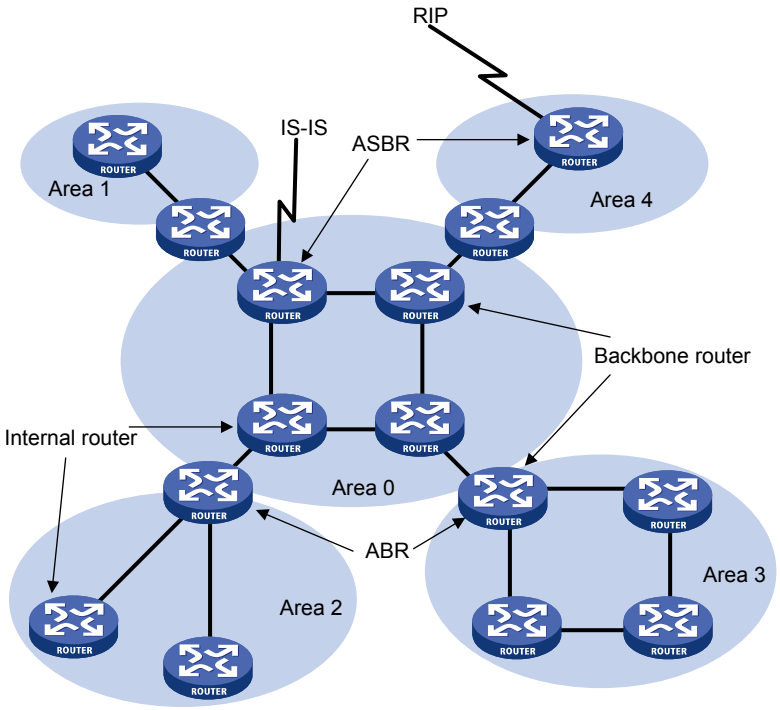
(3) 骨干路由器（Backbone Router）

该类路由器至少有一个接口属于骨干区域。因此，所有的 ABR 和位于 Area0 的内部路由器都是骨干路由器。

(4) 自治系统边界路由器 ASBR

与其他 AS 交换路由信息的路由器称为 ASBR。ASBR 并不一定位于 AS 的边界，它有可能是区域内路由器，也有可能是 ABR。只要一台 OSPF 路由器引入了外部路由的信息，它就成为 ASBR。

图1-6 OSPF 路由器的类型



2. 路由类型

OSPF 将路由分为四类，按照优先级从高到低的顺序依次为：

- 区域内路由（Intra Area）
- 区域间路由（Inter Area）
- 第一类外部路由（Type1 External）
- 第二类外部路由（Type2 External）

区域内和区域间路由描述的是 AS 内部的网络结构，外部路由则描述了应该如何选择到 AS 以外目的地址的路由。OSPF 将引入的 AS 外部路由分为两类：Type1 和 Type2。

第一类外部路由是指接收的是 IGP（Interior Gateway Protocol，内部网关协议）路由（例如静态路由和 RIP 路由）。由于这类路由的可信程度较高，并且和 OSPF 自身路由的开销具有可比性，所以到第一类外部路由的开销等于本路由器到相应的 ASBR 的开销与 ASBR 到该路由目的地址的开销之和。

第二类外部路由是指接收的是 EGP（Exterior Gateway Protocol，外部网关协议）路由。由于这类路由的可信度比较低，所以 OSPF 协议认为从 ASBR 到自治系统之外的开销远远大于在自治系统之内到达 ASBR 的开销。所以计算路由开销时将主要考虑前者，即到第二类外部路由的开销等于 ASBR 到该路由目的地址的开销。如果计算出开销值相等的两条路由，再考虑本路由器到相应的 ASBR 的开销。

1.1.4 OSPF 的网络类型

1. OSPF 的 4 种网络类型

OSPF 根据链路层协议类型将网络分为下列四种类型：

- 广播（Broadcast）类型：当链路层协议是 Ethernet、FDDI 时，OSPF 缺省认为网络类型是 Broadcast。在该类型的网络中，通常以组播形式（224.0.0.5：含义是 OSPF 路由器的预留 IP

组播地址；和 224.0.0.6：含义是 OSPF DR 的预留 IP 组播地址）发送 Hello 报文、LSU 报文和 LSAck 报文；以单播形式发送 DD 报文和 LSR 报文。

- NBMA（Non-Broadcast Multi-Access，非广播多点可达网络）类型：当链路层协议是帧中继、ATM 或 X.25 时，OSPF 缺省认为网络类型是 NBMA。在该类型的网络中，以单播形式发送协议报文。
- P2MP（Point-to-MultiPoint，点到多点）类型：没有一种链路层协议会被缺省的认为是 P2MP 类型。点到多点必须是由其他的网络类型强制更改的。常用做法是将 NBMA 改为点到多点的网络。在该类型的网络中，缺省情况下，以组播形式（224.0.0.5）发送协议报文。可以根据用户需要，以单播形式发送协议报文。
- P2P（Point-to-Point，点到点）类型：当链路层协议是 PPP、HDLC 时，OSPF 缺省认为网络类型是 P2P。在该类型的网络中，以组播形式（224.0.0.5）发送协议报文。

2. NBMA 网络的配置原则

NBMA 网络是指非广播、多点可达的网络，比较典型的有 ATM 和帧中继网络。

对于接口的网络类型为 NBMA 的网络需要进行一些特殊的配置。由于无法通过广播 Hello 报文的形式发现相邻路由器，必须手工为该接口指定相邻路由器的 IP 地址，以及该相邻路由器是否有 DR 选举权等。

NBMA 网络必须是全连通的，即网络中任意两台路由器之间都必须有一条虚电路直接可达。如果部分路由器之间没有直接可达的链路时，应将接口配置成 P2MP 类型。如果路由器在 NBMA 网络中只有一个对端，也可将接口类型配置为 P2P 类型。

NBMA 与 P2MP 网络之间的区别如下：

- NBMA 网络是指那些全连通的、非广播、多点可达网络。而 P2MP 网络，则并不需要一定是全连通的。
- 在 NBMA 网络中需要选举 DR 与 BDR，而在 P2MP 网络中没有 DR 与 BDR。
- NBMA 是一种缺省的网络类型，而 P2MP 网络必须是由其它的网络强制更改的。最常见的做法是将 NBMA 网络改为 P2MP 网络。
- NBMA 网络采用单播发送报文，需要手工配置邻居。P2MP 网络采用组播方式发送报文。

1.1.5 DR/BDR

1. DR/BDR 简介

在广播网和 NBMA 网络中，任意两台路由器之间都要交换路由信息。如果网络中有 n 台路由器，则需要建立 n（n-1）/2 个邻接关系。这使得任何一台路由器的路由变化都会导致多次传递，浪费了带宽资源。为解决这一问题，OSPF 协议定义了指定路由器 DR（Designated Router），所有路由器都只将信息发送给 DR，由 DR 将网络链路状态发送出去。

如果 DR 由于某种故障而失效，则网络中的路由器必须重新选举 DR，再与新的 DR 同步。这需要较长的时间，在这段时间内，路由的计算是不正确的。为了能够缩短这个过程，OSPF 提出了 BDR（Backup Designated Router，备份指定路由器）的概念。

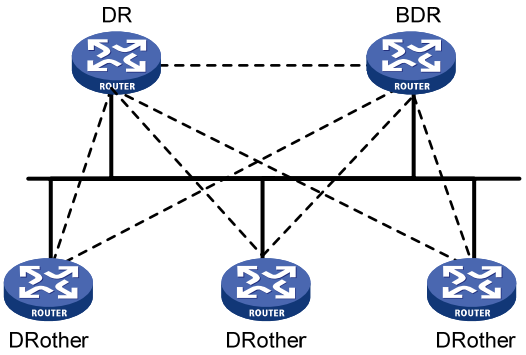
BDR 实际上是对 DR 的一个备份，在选举 DR 的同时也选举出 BDR，BDR 也和本网段内的所有路由器建立邻接关系并交换路由信息。当 DR 失效后，BDR 会立即成为 DR。由于不需要重新选举，

并且邻接关系事先已建立，所以这个过程是非常短暂的。当然这时还需要再重新选举出一个新的BDR，虽然一样需要较长的时间，但并不会影响路由的计算。

运行 OSPF 进程的网络中，既不是 DR 也不是 BDR 的路由器为 DR Other。DR Other 仅与 DR 和 BDR 之间建立邻接关系，DR Other 之间不交换任何路由信息。这样就减少了广播网和 NBMA 网络上各路由器之间邻接关系的数量，同时减少网络流量，节约了带宽资源。

如 图 1-7所示，用实线代表以太网物理连接，虚线代表建立的邻接关系。可以看到，采用DR/BDR机制后，5 台路由器之间只需要建立 7 个邻接关系就可以了。

图1-7 DR 和 BDR 示意图



2. DR/BDR 选举过程

DR 和 BDR 是由同一网段中所有的路由器根据路由器优先级、Router ID 通过 Hello 报文选举出来的，只有优先级大于 0 的路由器才具有选举资格。

进行 DR/BDR 选举时每台路由器将自己选出的 DR 写入 Hello 报文中，发给网段上的每台运行 OSPF 协议的路由器。当处于同一网段的两台路由器同时宣布自己是 DR 时，路由器优先级高者胜出。如果优先级相等，则 Router ID 大者胜出。如果一台路由器的优先级为 0，则它不会被选举为 DR 或 BDR。

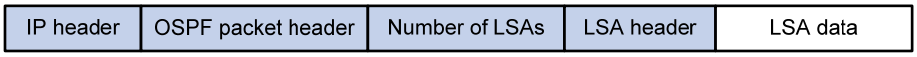
需要注意的是：

- 只有在广播或 NBMA 类型接口才会选举 DR，在点到点或点到多点类型的接口上不需要选举 DR。
- DR 是某个网段中的概念，是针对路由器的接口而言的。某台路由器在一个接口上可能是 DR，在另一个接口上有可能是 BDR，或者是 DR Other。
- 路由器的优先级可以影响 DR/BDR 的选举过程，但是当 DR/BDR 已经选举完毕，就算一台具有更高优先级的路由器变为有效，也不会替换该网段中已经存在的 DR/BDR 成为新的 DR/BDR。
- DR 并不一定就是路由器优先级最高的路由器接口；同理，BDR 也并不一定就是路由器优先级次高的路由器接口。

1.1.6 OSPF 的协议报文

OSPF报文直接封装为IP报文协议报文，协议号为 89。一个比较完整的OSPF报文（以LSU报文为例）结构如 图 1-8所示。

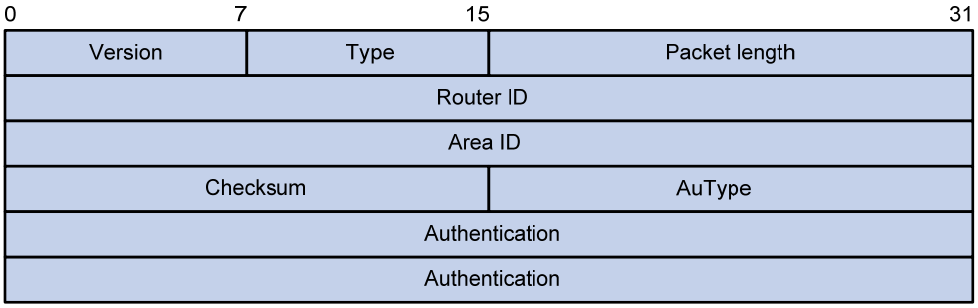
图1-8 OSPF 报文结构



1. OSPF 报文头

OSPF有五种报文类型，它们有相同的报文头。如 图 1-9所示。

图1-9 OSPF 报文头格式



主要字段的解释如下：

- Version：OSPF 的版本号。对于 OSPFv2 来说，其值为 2。
- Type：OSPF 报文的类型。数值从 1 到 5，分别对应 Hello 报文、DD 报文、LSR 报文、LSU 报文和 LSAck 报文。
- Packet length：OSPF 报文的总长度，包括报文头在内，单位为字节。
- Router ID：始发该 LSA 的路由器的 ID。
- Area ID：始发 LSA 的路由器所在的区域 ID。
- Checksum：对整个报文的校验和。
- AuType：验证类型。可分为不验证、简单（明文）口令验证和 MD5 验证，其值分别为 0、1、2。
- Authentication：其数值根据验证类型而定。当验证类型为 0 时未作定义，为 1 时此字段为密码信息，类型为 2 时此字段包括 Key ID、MD5 验证数据长度和序列号的信息。

说明

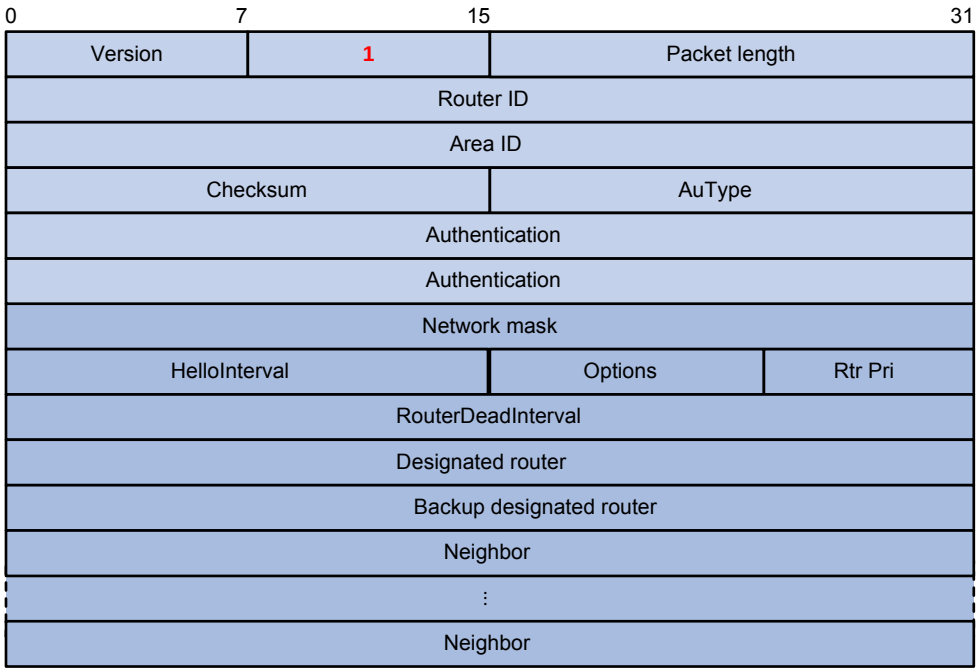
MD5 验证数据添加在 OSPF 报文后面，不包含在 Authenticaiton 字段中。

---

2. Hello 报文（Hello Packet）

最常用的一种报文，周期性的发送给邻居路由器用来维持邻居关系以及DR/BDR的选举，内容包括一些定时器的数值、DR、BDR以及自己已知的邻居。Hello报文格式如 图 1-10所示。

图1-10 Hello 报文格式



主要字段解释如下：

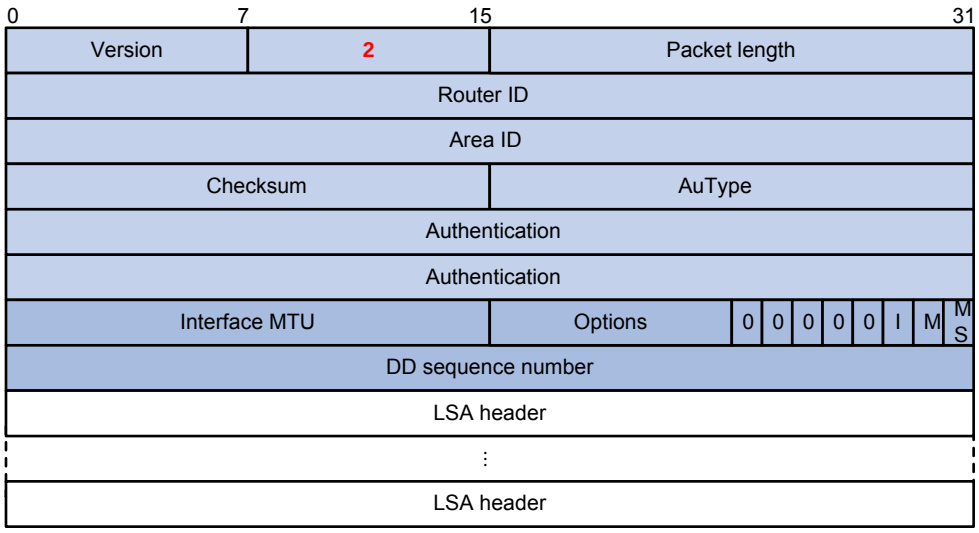
- Network mask：发送 Hello 报文的接口所在网络的掩码，如果相邻两台路由器的网络掩码不同，则不能建立邻居关系。
- HelloInterval：发送 Hello 报文的时间间隔。如果相邻两台路由器的 Hello 间隔时间不同，则不能建立邻居关系。
- Rtr Pri：路由器优先级。如果设置为 0，则该路由器接口不能成为 DR/BDR。
- RouterDeadInterval：失效时间。如果在此时间内未收到邻居发来的 Hello 报文，则认为邻居失效。如果相邻两台路由器的失效时间不同，则不能建立邻居关系。
- Designated router：指定路由器的接口的 IP 地址。
- Backup designated router：备份指定路由器的接口的 IP 地址。
- Neighbor：邻居路由器的 Router ID。

3. DD 报文（Database Description Packet）

两台路由器进行数据库同步时，用 DD 报文来描述自己的 LSDB，内容包括 LSDB 中每一条 LSA 的摘要（摘要是指 LSA 的 Header，通过该 Header 可以唯一标识一条 LSA）。这样做是为了减少路由器之间传递信息的量，因为 LSA 的 Header 只占一条 LSA 的整个数据量的一小部分，根据 Header，对端路由器就可以判断出是否已有这条 LSA。

DD报文格式如 图 1-11所示。

图1-11 DD 报文格式



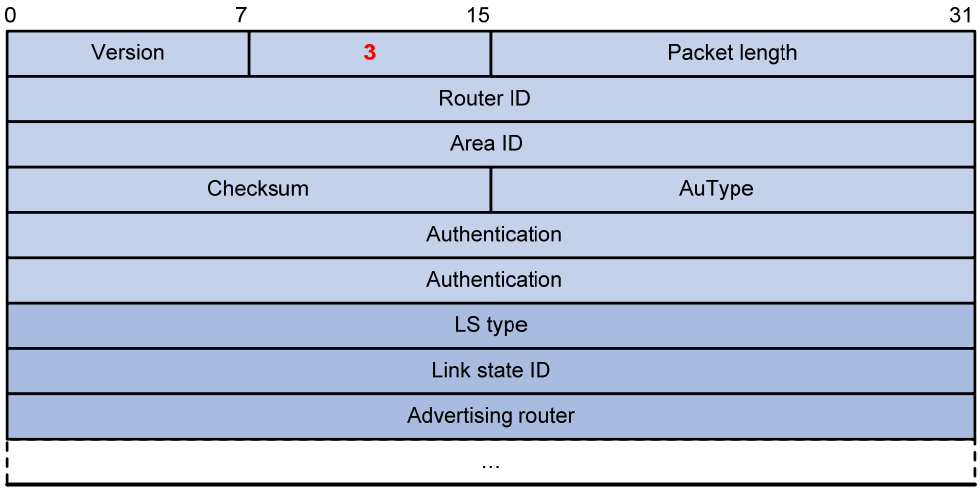
主要字段的解释如下：

- Interface MTU：在不分片的情况下，此接口最大可发出的 IP 报文长度。
- I（Initial）：当发送连续多个 DD 报文时，如果这是第一个 DD 报文，则置为 1，否则置为 0。
- M（More）：当连续发送多个 DD 报文时，如果这是最后一个 DD 报文，则置为 0。否则置为 1，表示后面还有其他的 DD 报文。
- MS（Master/Slave）：当两台 OSPF 路由器交换 DD 报文时，首先需要确定双方的主（Master）从（Slave）关系，Router ID 大的一方会成为 Master。当值为 1 时表示发送方为 Master。
- DD Sequence Number：DD 报文序列号，由 Master 方规定起始序列号，每发送一个 DD 报文序列号加 1，Slave 方使用 Master 的序列号作为确认。主从双方利用序列号来保证 DD 报文传输的可靠性和完整性。

4. LSR 报文（Link State Request Packet）

两台路由器互相交换过DD报文之后，知道对端的路由器有哪些LSA是本地的LSDB所缺少的，这时需要发送LSR报文向对方请求所需的LSA。内容包括所需要的LSA的摘要。LSR报文格式如 图 1-12 所示。

图1-12 LSR 报文格式



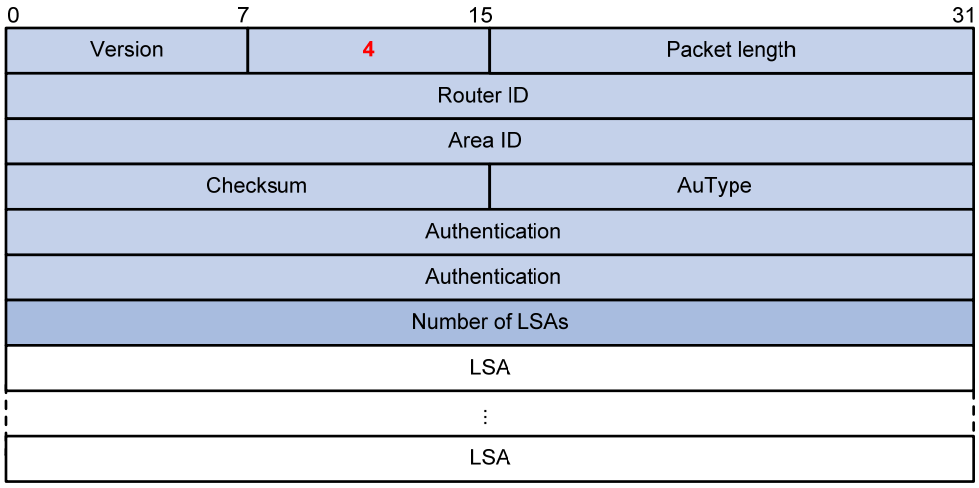
主要字段解释如下：

- LS type：LSA 的类型号。例如 Type1 表示 Router LSA。
- Link State ID：链路状态标识，根据 LSA 的类型而定。
- Advertising Router：产生此 LSA 的路由器的 Router ID。

5. LSU 报文（Link State Update Packet）

LSU报文用来向对端路由器发送它所需要的LSA，内容是多条LSA（全部内容）的集合。LSU报文格式如 图 1-13所示。

图1-13 LSU 报文格式



主要字段解释如下：

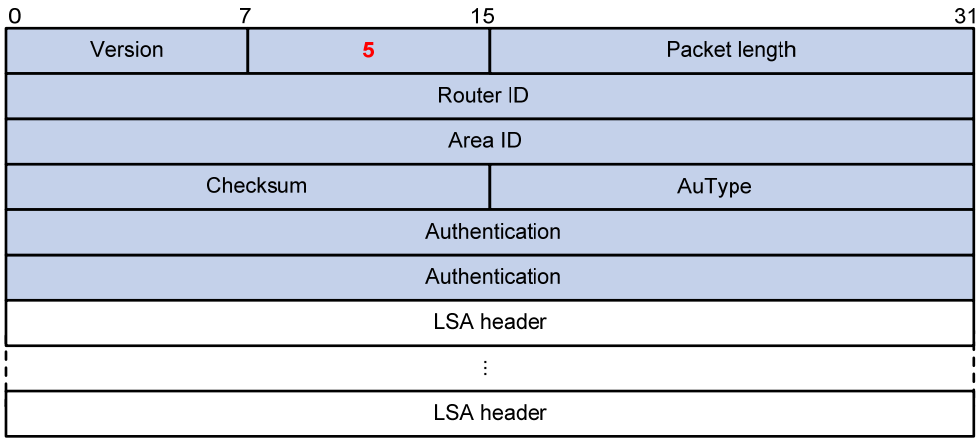
Number of LSAs：该报文包含的 LSA 的数量。

LSA：该报文包含的 LSA。

6. LSAck 报文（Link State Acknowledgment Packet）

LSAck报文用来对接收到的LSU报文进行确认，内容是需要确认的LSA的Header。一个LSAck报文可对多个LSA进行确认。报文格式如 图 1-14所示。

图1-14 LSAck 报文格式



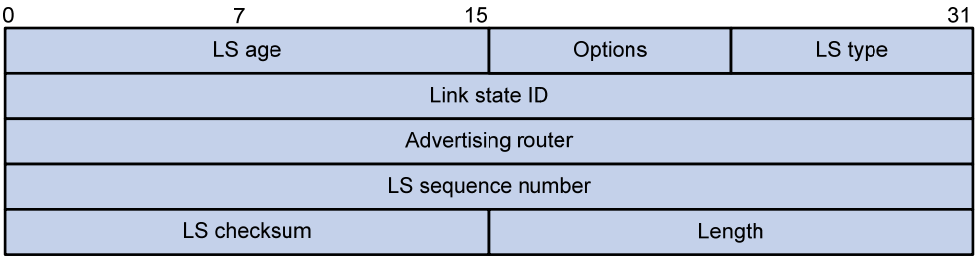
主要字段解释如下：

LSA Headers：该报文包含的 LSA 头部。

7. LSA 头格式

所有的LSA都有相同的报文头，其格式如 图 1-15所示。

图1-15 LSA 的头格式



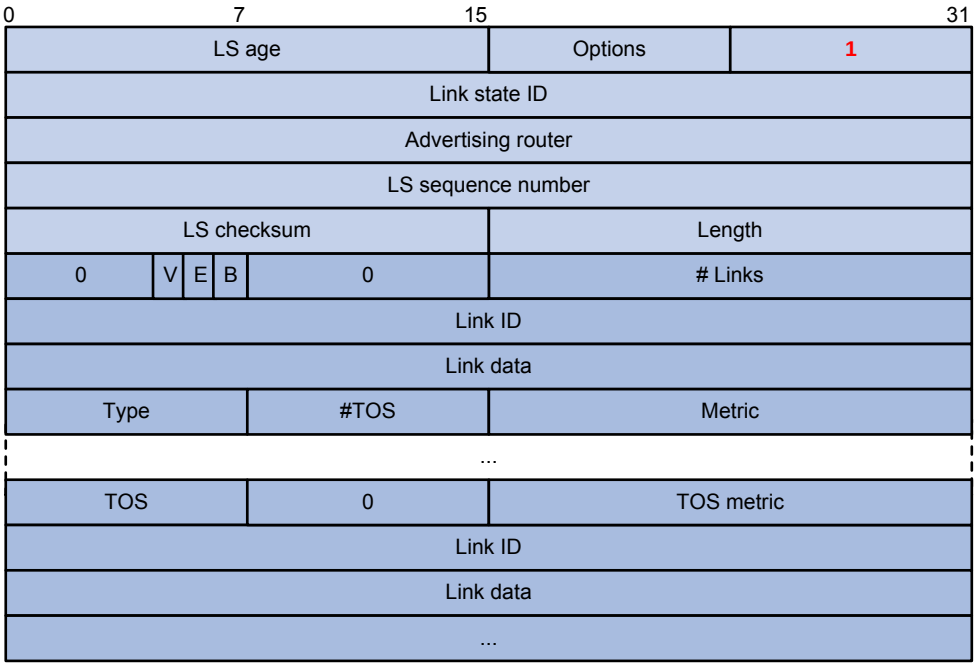
主要字段的解释如下：

- LS age：LSA 产生后所经过的时间，以秒为单位。LSA 在本路由器的链路状态数据库（LSDB）中会随时间老化（每秒钟加 1），但在网络的传输过程中却不会。
- LS type：LSA 的类型。
- Link State ID：具体数值根据 LSA 的类型而定。
- Advertising Router：始发 LSA 的路由器的 ID。
- LS sequence number：LSA 的序列号，其他路由器根据这个值可以判断哪个 LSA 是最新的。
- LS checksum：除了 LS age 字段外，关于 LSA 的全部信息的校验和。
- length：LSA 的总长度，包括 LSA Header，以字节为单位。

8. LSA 类型

(1) Router LSA

图1-16 Router LSA 格式



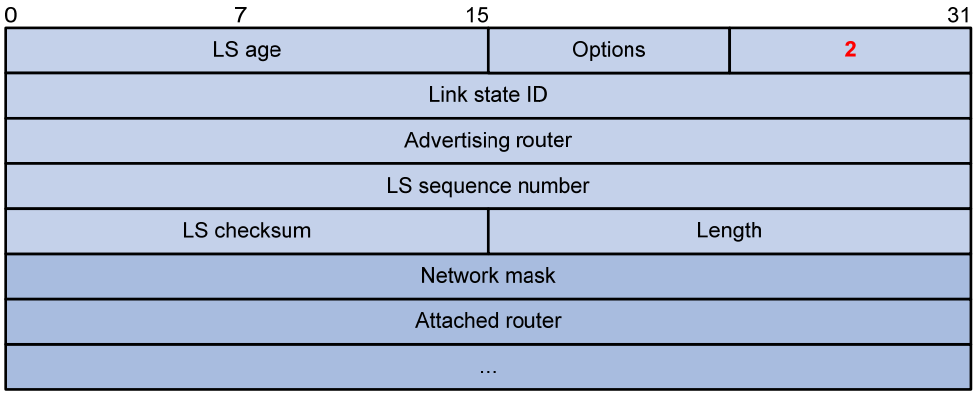
主要字段的解释如下：

- Link State ID：产生此 LSA 的路由器的 Router ID。
- V（Virtual Link）：如果产生此 LSA 的路由器是虚连接的端点，则置为 1。
- E（External）：如果产生此 LSA 的路由器是 ASBR，则置为 1。
- B（Border）：如果产生此 LSA 的路由器是 ABR，则置为 1。
- # Links：LSA 中所描述的链路信息的数量，包括路由器上处于某区域中的所有链路和接口。
- Link ID：链路标识，具体的数值根据链路类型而定。
- Link data：链路数据，具体的数值根据链路类型而定。
- Type：链路类型，取值为 1 表示通过点对点链路与另一路由器相连，取值为 2 表示连接到传送网络，取值为 3 表示连接到 Stub 网络，取值为 4 表示虚连接。
- #TOS：描述链路的不同方式的数量。
- Metric：链路的开销。
- TOS：服务类型。
- TOS metric：指定服务类型的链路的开销。

(2) Network LSA

Network LSA 由广播网或 NBMA 网络中的 DR 发出，LSA 中记录了这一网段上所有路由器的 Router ID。

图1-17 Network LSA 格式



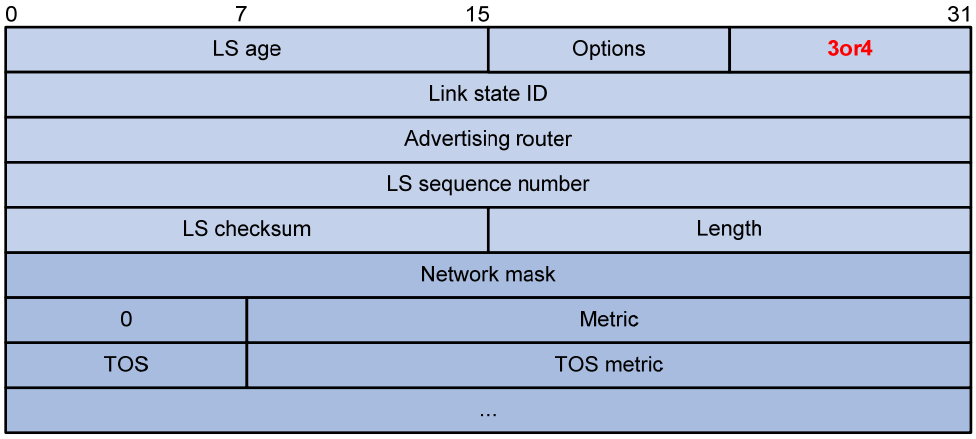
主要字段的解释如下：

- Link State ID：DR 的 IP 地址。
- Network mask：广播网或 NBMA 网络地址的掩码。
- Attached router：连接在同一个网段上的所有与 DR 形成了完全邻接关系的路由器的 Router ID，也包括 DR 自身的 Router ID。

(3) Summary LSA

Network Summary LSA（Type3 LSA）和 ASBR Summary LSA（Type4 LSA）除 Link State ID 字段有所不同外，有着相同的格式，它们都是由 ABR 产生。

图1-18 Summary LSA 格式



主要字段的解释如下：

- Link State ID：对于 Type3 LSA 来说，它是所通告的区域外的网络地址；对于 Type4 来说，它是所通告区域外的 ASBR 的 Router ID。
- Network mask：Type3 LSA 的网络地址掩码。对于 Type4 LSA 来说没有意义，设置为 0.0.0.0。
- Metric：到目的地址的路由开销。

说明

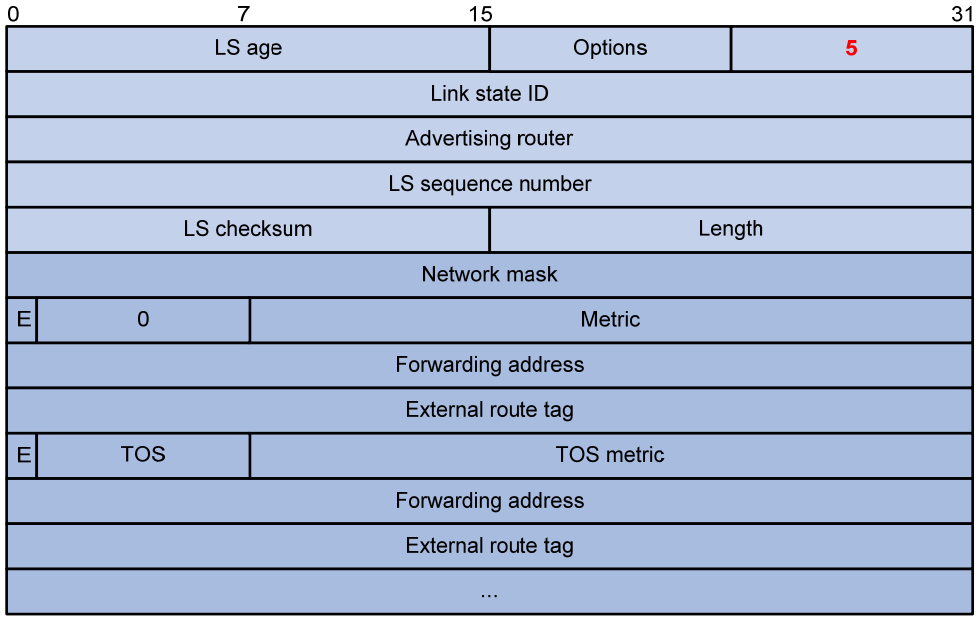
Type3 的 LSA 可以用来通告缺省路由，此时 Link State ID 和 Network Mask 都设置为 0.0.0.0。

---

(4) AS External LSA

由 ASBR 产生，描述到 AS 外部的路由信息。

图1-19 AS External LSA 格式



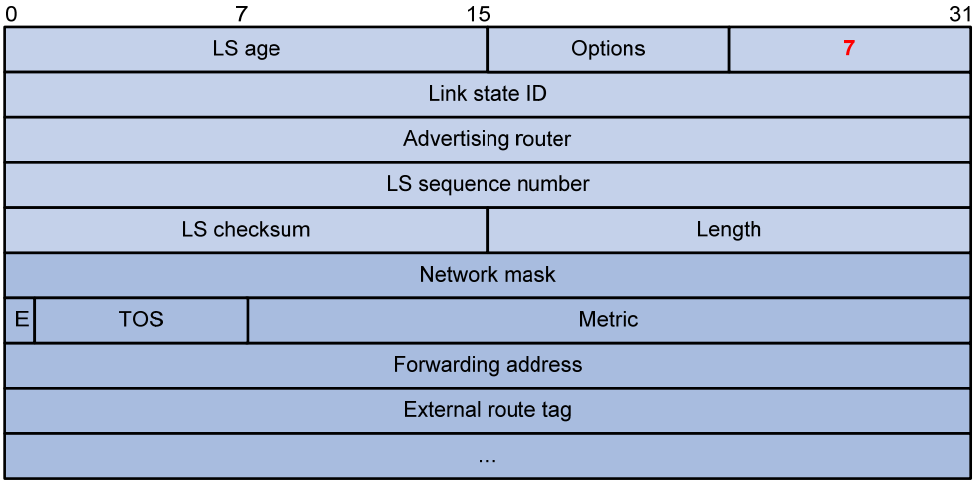
主要字段的解释如下：

- Link State ID：所要通告的其他外部 AS 的目的地址，如果通告的是一条缺省路由，那么链路状态 ID（Link State ID）和网络掩码（Network Mask）字段都将设置为 0.0.0.0。
- Network mask：所通告的目的地址的掩码。
- E（External Metric）：外部度量值的类型。如果是第 2 类外部路由就设置为 1，如果是第 1 类外部路由则设置为 0。关于外部路由类型的详细描述请参见 1.1.2 5. 各区域特性部分。
- Metirc：路由开销。
- Forwarding Address：到所通告的目的地址的报文将被转发到的地址。
- External Route Tag：添加到外部路由上的标记。OSPF 本身并不使用这个字段，它可以用来对外部路由进行管理。

(5) NSSA External LSA

由NSSA区域内的ASBR产生，且只能在NSSA区域内传播。其格式与AS External LSA相同，如 图1-20所示。

图1-20 NSSA External LSA 格式



1.1.7 系统支持的 OSPF 特性

1. 多进程

OSPF 支持多进程，在同一台路由器上可以运行多个不同的 OSPF 进程，它们之间互不影响，彼此独立。不同 OSPF 进程之间的路由交互相当于不同路由协议之间的路由交互。支持多个 OSPF 进程公用一个 RID。

路由器的一个接口只能属于某一个 OSPF 进程。

2. 验证

OSPF 支持报文验证功能，只有通过验证的 OSPF 报文才能接收，否则将不能正常建立邻居关系。

一个区域中所有的接口的验证类型（不支持验证、支持明文验证或者支持 MD5 密文验证）必须一致。一个网段中所有接口的验证字口令必须一致。

3. 热备份

分布式结构的路由器支持 OSPF 热备份（Hot Standby，HSB）特性。OSPF 将会把主用主控板（Active Main Board，AMB）上必要的信息备份到备用主控板（Standby Main Board，SMB）上。当 AMB 发生故障时，SMB 将会取代它的功能，使 OSPF 可以不受任何影响地正常运行。

OSPF 支持两种热备份方式：

- 备份所有 OSPF 数据，一旦发生 AMB 和 SMB 切换，OSPF 能够马上恢复正常运行。
- 只备份 OSPF 配置信息，发生 AMB 和 SMB 切换时，OSPF 进行 Graceful Restart（GR），从邻居那里获得邻接关系，并对 LSDB 进行同步。

GR 是平缓重启路由器的一种功能，主要用于 HA（High Availability，高可靠性），不会对其他路由器造成影响。

一台路由器上的 OSPF 进程重启后，它的邻居路由器就会把它从邻居列表中删除，并通知给其他路由器，这样就要重新计算 SPF。如果路由器关闭的时间很短，可能只是几秒钟，就没有必要影响整个网络的拓扑结构。几秒钟之后路由器恢复工作，又需要重新建立邻接关系并计算 SPF。

为了避免不必要的 SPF 计算，当一台路由器重启时，会通知与它相邻的路由器它只是暂时断开，马上就会恢复正常。这样，邻居路由器就不会将进行 GR 操作的路由器从邻居列表中删除，其他路由器也不会知道有路由器重启。

重启的路由器恢复工作之后，将会通过 GR 相关的同步机制重新从邻居路由器处同步 LSDB。

4. OSPF GR



说明

有关 GR（Graceful Restart，平滑重启）的原理介绍请参见“系统分册”中的“GR 概述”。

基于 OSPF 的 GR Restarter 进行协议重启后，为了与其 OSPF 邻居重新同步链路状态数据库，它必须完成下列两项任务：

- 在不改变邻接关系的前提下，重新获取网络中的有效 OSPF 邻居信息；
- 重新获取网络链路状态数据库的内容。

在 OSPF 协议重启前，GR Restarter 与 GR Helper 协商 GR 能力。，如果 GR Restarter 发生重启，在重起过程中，GR Helper 继续宣告与 GR Restarter 的邻接状态不变。

OSPF 协议重启完毕后，GR Restarter 会立即向其邻接的 GR Helper 发送一个 OSPF GR 信号。这样，OSPF 邻居就不会复位与其的邻居关系。在收到其 OSPF 邻居的响应后，GR Restarter 会重新恢复与其的邻居关系列表。

邻居关系重新建立后，GR Restarter 与其所有具备 GR 感知能力的 OSPF 邻居之间同步数据库，并交换路由信息。交换完成后，GR Restarter 根据新的路由转发信息更新路由表和转发表，删除失效的路由，完成 OSPF 协议收敛。

5. TE 与 DS-TE

OSPF TE（Traffic Engineering，流量工程）支持建立和维护 TE 的 LSP（Label Switch Path，标签交换路径）。

MPLS（MultiProtocol Label Switching，多协议标签交换）在构建 CR LSP（Constraint-based Routed LSP，基于约束路由的 LSP）时，需要了解本区域中所有链路的流量属性信息。它通过 OSPF 来获取链路的流量工程信息。

OSPF 支持 Opaque LSA，可以用来承载流量工程信息。

DS-TE（DiffServ Aware TE）主要用于优化和分配网络传输资源、对流进行分类、指定每条流在链路带宽中所占的比例。流量工程是基于划分后的类（细粒度聚合类）实现的，而不是聚合的类（粗粒度聚合类）。这一点改善了性能和带宽的利用率。

为了支持 MPLS 中 DS-TE 的应用，OSPF 支持子 TLV（Local Overbooking Multiplier TLV）和 BC（Bandwidth Constraint，带宽约束）TLV。


说明

OSPF TE 的详细配置请参见“MPLS 分册”中的“MPLS TE 配置”。

---

6. IGP Shortcut 和 Forwarding Adjacency

OSPF 支持 IGP Shortcut 和 Forwarding Adjacency 特性，这两个特性允许 OSPF 使用 LSP 作为到达某个目的地址的出接口。否则，即使存在到达某个目的地址的 LSP，OSPF 也不能使用它作为出接口。

IGP Shortcut 和 Forwarding Adjacency 的区别在于：

- 如果仅使能了 Forwarding Adjacency 特性，OSPF 也可以使用 LSP 到达目的地址；
- 如果仅使能了 IGP Shortcut，则只有使能此特性的路由器才可以在路由中使用 LSP。


说明

此特性的详细配置请参见“MPLS 分册”中的“MPLS TE 配置”。

---

7. OSPF 支持多实例

OSPF 支持多实例，可以运行在 VPN 网络中的 PE 上。

在 BGP MPLS VPN 网络中，属于同一个 VPN 的多个 Site 可以使用 OSPF 作为内部路由协议。然而，它们会被看作属于不同的自治系统来处理。这样，在一个节点学到的 OSPF 路由，将被作为外部路由传送给另一节点。这种处理方式导致了比较高的 OSPF 路由协议流量，并带来了一些原本可以避免的网络管理问题。

目前，可以在 PE 上通过配置域 ID 来区分 Site 所在的 VPN。属于同一个 VPN 的不同 Site 之间彼此看作是直接相连的。这样，PE 路由器之间交换 OSPF 路由信息时就好像是通过一条专线相连，改善了网络管理并使 OSPF 的应用更为有效。


说明

此特性的相关配置请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

---

8. OSPF 伪连接

OSPF 伪连接（sham link）是 MPLS VPN 骨干网上两个 PE 路由器之间的点到点链路。

通常情况下，BGP 对等体之间通过 BGP 扩展团体属性在 MPLS VPN 骨干网上承载路由信息。另一端 PE 上运行的 OSPF 可利用这些信息来生成 PE 到 CE 的 Type3 LSA，这些路由是区域间路由。

但是，如果路由器和它同一区域内的 PE 路由器相连，且建立到达特定目的地址的内部区域路由（后门路由），那么 VPN 流量就将总是穿越这条后门路由，而不是骨干路由。这是因为在路由表中建立的 OSPF 内部区域路由的优先级较高。为了避免这一异常现象，可以在 PE 路由器之间配置一条 unnumbered 的点到点伪连接。这样，就可以通过一条低开销的内部区域路由到达 PE 路由器。



说明

OSPF 伪连接的相关配置请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

1.1.8 协议规范

与 OSPF 相关的协议规范有：

- RFC 1765：OSPF Database Overflow
- RFC 2328：OSPF Version 2
- RFC 3101：OSPF Not-So-Stubby Area (NSSA) Option
- RFC 3137：OSPF Stub Router Advertisement
- RFC 3630：Traffic Engineering Extensions to OSPF Version 2
- RFC 4811：OSPF Out-of-Band LSDB Resynchronization
- RFC 4812：OSPF Restart Signaling
- RFC 4813：OSPF Link-Local Signaling

1.2 OSPF 配置任务简介

在一个运行 OSPF 的路由域里，有各种类型的路由器，区域内路由器、骨干路由器、ABR、ASBR 等。

无论是哪种类型的路由器，都必须 先使能 OSPF，否则 OSPF 协议将无法正常运行；使能 OSPF 基本功能后，路由器可以使用一些缺省配置，如 OSPF 各种协议报文的发送时间间隔、LSA 延迟时间、SPF 计算时间间隔等，也可以根据需要修改缺省配置。

在 OSPF 路由器上进行各项配置的时候都应该先做好网络规划，在配置同一区域内的路由器时，大多数的配置数据都应该以区域为基础来统一考虑，错误的配置可能会导致相邻路由器之间无法相互传递信息，甚至导致路由信息的阻塞或者产生路由环。

表1-1 OSPF 配置任务简介

配置任务		说明	详细配置
使能 OSPF 基本功能		必选	1.3
配置 OSPF 的区域	配置 Stub 区域	可选	1.4.2
	配置 NSSA 区域	可选	1.4.3
	配置虚连接	可选	1.4.4

配置任务		说明	详细配置
配置 OSPF 的网络类型	配置 OSPF 接口网络类型为广播	可选	1.5.2
	配置 OSPF 接口网络类型为 NBMA	可选	1.5.3
	配置 OSPF 接口网络类型为 P2MP	可选	1.5.4
	配置 OSPF 接口网络类型为 P2P	可选	1.5.5
配置 OSPF 的路由信息控制	配置 OSPF 路由聚合	可选	1.6.2
	配置 OSPF 通过接收到的 LSA 计算出来的路由信息进行过滤	可选	1.6.3
	配置对 Type-3 LSA 进行过滤	可选	1.6.4
	配置 OSPF 的接口的开销值	可选	1.6.5
	配置 OSPF 支持的路由最大数目	可选	1.6.6
	配置 OSPF 最大等价路由条数	可选	1.6.7
	配置 OSPF 的协议优先级	可选	1.6.8
	配置 OSPF 引入外部路由	可选	1.6.9
	配置发布一条主机路由	可选	1.6.10
配置 OSPF 网络调整优化	配置 OSPF 报文定时器	可选	1.7.2
	配置接口传送 LSA 的延迟时间	可选	1.7.3
	配置 SPF 计算时间间隔	可选	1.7.4
	配置 LSA 重复到达的最小时间间隔	可选	1.7.5
	配置 LSA 重新生成的时间间隔	可选	1.7.6
	禁止接口发送 OSPF 报文	可选	1.7.7
	配置 Stub 路由器	可选	1.7.8
	配置 OSPF 验证	可选	1.7.9
	配置 DD 报文中的 MTU	可选	1.7.10
	配置 LSDB 中 External LSA 的最大数量	可选	1.7.11
	配置兼容 RFC1583 的外部路由选择规则	可选	1.7.12
	配置邻接状态输出	可选	1.7.13
	配置 OSPF 网管功能	可选	1.7.13
	使能日志功能	可选	1.7.15
	使能 Opaque LSA 发布接收能力	可选	1.7.15
	配置 OSPF 优先处理 Hello 报文功能	可选	1.7.17
	配置 OSPF 发送更新报文的速度	可选	1.7.18
配置 OSPF GR	配置 GR Restarter	可选	1.8.1
	配置 GR Helper	可选	1.8.2
	重启 OSPF GR 进程	可选	1.8.3

1.3 使能 OSPF 功能

在 OSPF 的各项配置任务中，必须先使能 OSPF 功能，才能配置其它的功能特性启动 OSPF、。

1.3.1 配置准备

在配置 OSPF 之前，需完成以下任务：

- 配置链路层协议，保证链路层通信正常。
- 配置接口的网络层地址，使各相邻节点网络层可达。

1.3.2 使能 OSPF 功能

OSPF 基本功能配置包括：

要在路由器上使能 OSPF 功能，OSPF 基本功能配置包括：必须先创建 OSPF 进程、指定该进程关联的区域以及区域包括的网段；对于当前路由器来说，如果某个路由器的接口 IP 地址落在某个区域的网段内，则该接口属于这个区域，这个接口使能了 OSPF 功能，OSPF 将把这个接口的直连路由宣告出去。

Router ID 用来在一个自治系统中唯一的标识一台路由器，一台路由器如果要运行 OSPF 协议，则必须存在 Router ID。

- 用户可以在创建 OSPF 进程的时候指定 Router ID，配置时，必须保证自治系统中任意两台路由器的 ID 都不相同。通常的做法是将路由器的 ID 配置为与该路由器某个接口的 IP 地址一致。
- 如果在创建 OSPF 进程的时候没有指定 Router ID，则缺省使用全局 Router ID。关于全局 Router ID 的详细说明，请参考“IP 路由分册”中的“IP 路由概述”。

目前，系统支持 OSPF 多进程和 OSPF 多实例：

- 当在一台路由器上启动多个 OSPF 进程时，需要指定不同的进程号。OSPF 进程号是本地概念，不影响与其它路由器之间的报文交换。因此，不同的路由器之间，即使进程号不同也可以进行报文交换。
- 可以配置 OSPF 在 VPN 实例中运行。如果指定了 VPN 实例，那么此 OSPF 进程属于指定的 VPN 实例。

表1-2 配置 OSPF 基本功能

操作	命令	说明
进入系统视图	**system-view**	-
创建 OSPF，进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	必选 缺省情况下，系统不运行 OSPF
配置 OSPF 进程描述	**description** *description*	可选 缺省情况下，没有配置进程描述
配置 OSPF 区域，进入 OSPF 区域视图	**area** *area-id*	必选 缺省情况下，没有配置 OSPF 区域
配置区域描述	**description** *description*	可选 缺省情况下，没有配置区域描述

操作	命令	说明
配置区域所包含的网段并在指定网段的接口上使能 OSPF	**network** *ip-address wildcard-mask*	必选 缺省情况下，接口不属于任何区域且 OSPF 功能处于关闭状态

说明

- **ospf** 命令中参数 **vpn-instance** *instance-name* 的支持情况与设备的型号有关，请以设备的实际情况为准。为提高可读性，在手册的描述中将不另行说明。
- 一个网段只能属于一个区域。
- 建议用户为每个 OSPF 进程配置进程描述信息，帮助识别进程的用途，以便于记忆和管理。
- 建议用户为每个区域配置区域描述信息，帮助识别区域的用途，以便于记忆和管理。

1.4 配置 OSPF 区域

网络管理员对整个网络划分区域完毕后，可以根据组网需要进一步将区域配置成 Stub 区域或 NSSA 区域。

当非骨干区域不能与骨干区域保持连通，或者骨干区域因为各方面条件的限制无法保持连通时，可以通过配置 OSPF 虚连接予以解决。

1.4.1 配置准备

在配置 OSPF 的区域特性之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 使能 OSPF 功能

1.4.2 配置 Stub 区域特性

对于位于 AS 边缘的一些非骨干区域，我们可以在该区域的所有路由器上配置 **stub** 命令，把该区域配置为 Stub 区域。这样，描述自治系统外部路由的 Type5 LSA 不会在 Sutb 区域里泛洪，减小了路由表的规模。ABR 生成一条缺省路由，所有到达自治系统外部的报文都交给 ABR 进行转发。

如果想进一步减少 Stub 区域路由表规模以及路由信息传递的数量，那么在 ABR 上配置 **stub** 命令时指定 **no-summary** 参数，可以将该区域配置为 Totally Stub（完全 Stub）区域。这样，自治系统外部路由和区域间的路由信息都不会传递到本区域，所有目的地是自治系统外和区域外的报文都交给 ABR 进行转发。

表1-3 配置 Stub 区域

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置当前区域为 Stub 区域	**stub** [**no-summary**]	可选 缺省情况下，没有区域被设置为 Stub 区域

操作	命令	说明
配置 ABR 发送到 Stub 区域缺省路由的开销	**default-cost** *cost*	可选 缺省情况下，ABR 发送到 Stub 区域的缺省路由的开销为 1

说明

- 所有连接到 Stub 区域的路由器必须使用 **stub** 命令将该区域配置成 Stub 属性。
- **default-cost** 命令只有在 Stub 区域的 ABR 上配置才能生效。
- 骨干区域不能配置成（Totally） Stub 区域。
- （Totally） Stub 区域内不能存在 ASBR，即自治系统外部的路由不能在本区域内传播。
- 虚连接不能穿过（Totally） Stub 区域。

1.4.3 配置 NSSA 区域

Stub 区域不能引入外部路由，为此又产生了 NSSA 区域的概念。NSSA 区域允许外部路由通告到 OSPF 路由域内部，同时也保持其余部分的 Stub 区域的特征。

在 NSSA 区域中，ASBR 通过发布 Type7 LSA 来通告到达外部的网络，Type7 LSA 将在整个 NSSA 区域内泛洪，到达 NSSA 区域的 ABR 时，就会转换成 Type5 LSA（AS External LSA），并通告到其他区域。

表1-4 配置 NSSA 区域

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置当前区域为 NSSA 区域	**nssa** [**default-route-advertise** | **no-import-route** | **no-summary** | **translate-always** | **translator-stability-interval** *value*] *	可选 缺省情况下，没有区域被设置为 NSSA 区域
配置发送到 NSSA 区域缺省路由的开销	**default-cost** *cost*	可选 缺省情况下，发送到 NSSA 区域的缺省路由的开销为 1

说明

- NSSA 区域内的所有路由器必须使用 **nssa** 命令将该区域配置成 NSSA 属性。
- **default-cost** 命令只有在 NSSA 区域的 ABR/ASBR 上配置才能生效。

1.4.4 配置虚连接

在划分区域之后，非骨干区域之间的 OSPF 路由更新是通过骨干区域来完成交换的。对此，OSPF 要求所有非骨干区域必须与骨干区域保持连通，并且骨干区域自身也要保持连通。

但在实际应用中，可能会因为各方面条件的限制，无法满足这个要求。这时可以通过在 ABR 上配置 OSPF 虚连接予以解决。

表1-5 配置虚连接

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
创建并配置虚连接	**vlink-peer** *router-id* [**hello** *seconds* | **retransmit** *seconds* | **trans-delay** *seconds* | **dead** *seconds* | **simple** [**plain** | **cipher**] *password* | { **md5** | **hmac-md5** } *key-id* [**plain** | **cipher**] *password*] *	可选 为使虚连接生效，在虚连接的两端都需配置此命令，并且两端配置的 **hello**、**dead** 等参数必须一致

1.5 配置 OSPF 的网络类型

OSPF 的网络类型有四种：广播、NBMA、P2MP 和 P2P。

当接口封装的链路层协议不同时，OSPF 接口网络类型的缺省值也不同：

- 广播：当接口封装的链路层协议是 Ethernet、FDDI 时，接口网络类型缺省值为广播；
- NBMA：当接口封装的链路层协议是 ATM、帧中继、HDLC 或 X.25 时，接口网络类型缺省值为 NBMA；
- 点对点：当接口封装的链路层协议是 PPP、LAPB 或 POS 时，接口网络类型缺省值为点对点。

用户可以根据需要更改接口的网络类型，例如：

- 当 NBMA 网络通过配置地址映射成为全联通网络时（即网络中任意两台路由器之间都存在一条虚电路而直接可达），可以将网络类型更改为广播，不需要手工配置邻居，简化配置。
- 当广播网络中有部分路由器不支持组播时，那么可以将网络类型更改为 NBMA。
- NBMA 网络要求必须是全连通的，即网络中任意两台路由器之间都必须有一条虚电路直接可达；如果 NBMA 网络不是全连通而是部分连通时，可以将网络类型更改为点对多点，达到简化配置、节省开销的目的。
- 如果路由器在 NBMA 网络中只有一个对端，也可将接口类型配置为 P2P，节省网络开销。

1.5.1 配置准备

在配置 OSPF 的网络类型之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点之间网络层可达
- 使能 OSPF 功能

1.5.2 配置 OSPF 接口网络类型为广播

表1-6 配置 OSPF 接口的网络类型为广播

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPF 接口的网络类型为广播	**ospf network-type broadcast**	可选 缺省情况下，接口的网络类型根据接口封装的链路层协议而定
配置 OSPF 接口的路由器优先级	**ospf dr-priority** *priority*	可选 缺省情况下，接口的路由器优先级为 1

1.5.3 配置 OSPF 接口网络类型为 NBMA

把接口类型配置为 NBMA 后，需要进行一些特殊的配置。

由于无法通过广播 Hello 报文的形式动态发现相邻路由器，必须手工为接口指定相邻接口的 IP 地址，该相邻接口是否有选举权等（*dr-priority* 参数的值仅表示路由器是否具有 DR 选举权，为 0 表示不具有 DR 选举权，大于 0 时表示具有 DR 选举权）。

表1-7 配置 OSPF 接口网络类型为 NBMA

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPF 接口的网络类型为 NBMA	**ospf network-type nbma**	可选 缺省情况下，接口的网络类型根据接口封装的链路层协议而定
配置 OSPF 接口的路由器优先级	**ospf dr-priority** *priority*	可选 缺省情况下，接口的路由器优先级为 1
退回系统视图	**quit**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 NBMA 网络的邻居	**peer** *ip-address* [**cost** *value* | **dr-priority** *dr-priority*]	必选

说明

使用 **peer** 命令和使用 **ospf dr-priority** 命令设置的优先级具有不同的用途：

- peer 命令设置的优先级用于表示邻居是否具有选举权。如果在配置邻居时将优先级指定为 0，则本地路由器认为该邻居不具备选举权，不向该邻居发送 Hello 报文，这种配置可以减少在 DR 和 BDR 选举过程中网络上的 Hello 报文数量。但如果本地路由器是 DR 或 BDR，它也会向优先级为 0 的邻居发送 Hello 报文，以建立邻接关系。
- **ospf dr-priority** 命令设置的优先级用于实际的 DR 选举。

1.5.4 配置 OSPF 接口网络类型为 P2MP

表1-8 配置 OSPF 接口的路由器优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPF 接口的网络类型为 P2MP	**ospf network-type p2mp[unicast]**	必选 缺省情况下，接口的网络类型根据物理接口而定 当把接口类型配置为 P2MP 单播后，OSPF 协议在该接口上发送的报文均为单播报文。由于无法通过广播 Hello 报文的形式动态发现相邻路由器，必须手工为接口指定相邻接口的 IP 地址

1.5.5 配置 OSPF 接口网络类型为 P2P

表1-9 配置 OSPF 接口网络类型为 P2P

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPF 接口的网络类型为 P2P	**ospf network-type p2p**	必选 缺省情况下，接口的网络类型根据物理接口而定

1.6 配置 OSPF 的路由信息控制

通过本节的配置，可以控制 OSPF 的路由信息的发布与接收，并引入其他协议的路由。

1.6.1 配置准备

在配置 OSPF 路由信息控制之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点之间网络层可达
- 使能 OSPF 功能
- 如果对路由信息进行过滤，则需要配置对应的过滤列表

1.6.2 配置 OSPF 路由聚合

路由聚合是指 ABR 或 ASBR 将具有相同前缀的路由信息聚合，只发布一条路由到其它区域。

AS 被划分成不同的区域后，每一个区域通过 OSPF 边界路由器（ABR）相连，区域间可以通过路由聚合来减少路由信息，减小路由表的规模，提高路由器的运算速度。

ABR 在计算出一个区域的区域内路由之后，根据聚合相关设置，将其中多条 OSPF 路由聚合成一条发送到区域之外。例如某个区域内有三条区域内路由 19.1.1.0/24，19.1.2.0/24，19.1.3.0/24，如果在 ABR 上配置了路由聚合，将三条路由聚合成一条 19.1.0.0/16，则 ABR 就只生成一条聚合后的 LSA，并发布给其它区域的路由器。

1. 配置区域边界路由器（ABR）路由聚合

如果区域里存在一些连续的网段，则可以在 ABR 上配置路由聚合，将这些连续的网段聚合成一个网段，ABR 向其它区域发送路由信息时，以网段为单位生成 Type3 LSA。

这样 ABR 只发送一条聚合后的 LSA，所有属于聚合网段范围的 LSA 将不再会被单独发送出去，既可以减少其它区域中 LSDB 的规模，也减小了因为网络拓扑变化带来的影响。

表1-10 配置 ABR 路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置 OSPF 的 ABR 路由聚合	**abr-summary** *ip-address* { *mask* | *mask-length* } [**advertise** | **not-advertise**] [**cost** *cost*]	必选 此命令只有在 ABR 上配置才会有效 缺省情况下，ABR 不会对路由进行聚合

2. 配置自治系统边界路由器（ASBR）对引入的路由进行聚合

ASBR 引入外部路由后，每一条路由都会放在单独的一条 ASE LSA 中向外宣告；通过配置路由聚合，路由器只把聚合后的路由放在 ASE LSA 中向外宣告，减少了 LSDB 中 LSA 的数量。

在 ASBR 上配置路由聚合后，将对聚合地址范围内的 Type5 LSA 进行聚合，如果 ASBR 在 NSSA 区域里面，将对聚合地址范围内的 Type7 LSA 进行聚合。

如果本地路由器同时是 ASBR 和 ABR，将对由 Type7 LSA 转化成的 Type5 LSA 进行聚合处理。

表1-11 配置 ASBR 路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 的 ASBR 路由聚合	**asbr-summary** *ip-address* { *mask* | *mask-length* } [**tag** *tag* | **not-advertise** | **cost** *cost*] *	必选 此命令只有在 ASBR 上配置才会有效 缺省情况下，ASBR 不会对引入的路由进行聚合

1.6.3 配置 OSPF 对通过接收到的 LSA 计算出来的路由信息进行过滤

说明

- 关于地址前缀列表的详细介绍，请参见“IP 路由分册”中的“路由策略配置”。
- 关于路由策略的详细介绍，请参见“IP 路由分册”中的“路由策略配置”。

OSPF 是基于链路状态的动态路由协议，路由信息是根据接收到的 LSA 计算出来的，可以对通过接收到的 LSA 计算出来的 OSPF 路由信息进行过滤。

一共有四种过滤方式：

- 基于要加入到路由表的路由信息的目的地址进行过滤，可以通过配置访问控制列表或 IP 地址前缀列表来指定过滤条件；
- 基于要加入到路由表的路由信息的下一跳进行过滤，可以通过在命令中配置 **gateway** 参数来指定过滤条件。
- 基于要加入到路由表的路由信息的目的地址和下一跳进行过滤，可以通过配置访问控制列表或 IP 地址前缀列表指定过滤目的地址的条件，同时配置 gateway 参数来指定过滤下一跳的条件；
- 基于路由策略对要加入到路由表的路由信息进行过滤，可以通过在命令中配置 route-policy 参数来指定过滤条件。

表1-12 配置 OSPF 对通过接收到的 LSA 计算出来的路由信息进行过滤

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置对接收的路由进行过滤	**filter-policy** { *acl-number* [**gateway** *ip-prefix-name*] | **gateway** *ip-prefix-name* | **ip-prefix** *ip-prefix-name* [**gateway** *ip-prefix-name*] | **route-policy** *route-policy-name* } **import**	必选 缺省情况下，不对接收到的路由信息进行过滤

1.6.4 配置对 Type-3 LSA 过滤

通过在 ABR 上配置 Type-3 LSA 过滤，可以对进入 ABR 所在区域或 ABR 向其它区域发布的 Type-3 LSA 进行过滤。

表1-13 配置过滤 Type-3 LSA

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置对 Type-3 LSA 进行过滤	**filter** { *acl-number* | **ip-prefix** *ip-prefix-name* } { **import** | **export** }	必选 缺省情况下，没有对 Type-3 LSA 进行过滤

1.6.5 配置 OSPF 接口的开销值

OSPF 有两种方式来配置接口的开销值：

- 在接口视图下直接配置开销值；
- 配置接口的带宽参考值，OSPF 根据带宽参考值自动计算接口的开销值，计算公式为：接口开销＝带宽参考值÷接口带宽，当计算出来的开销值大于 65535，开销取最大值 65535。

如果没有在接口视图下显式的配置此接口的开销值，OSPF 会根据该接口的带宽自动计算其开销值。

1. 配置接口的开销值

表1-14 配置 OSPF 接口的开销值

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
设置 OSPF 接口的开销值	**ospf cost** *value*	可选 缺省情况下，接口按照当前的带宽自动计算开销，对于 Loopback 接口，缺省值为 0

2. 配置带宽参考值

表1-15 配置带宽参考值

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置带宽参考值	**bandwidth-reference** *value*	可选 缺省情况下，带宽参考值为 100Mbps

1.6.6 配置 OSPF 支持的路由最大数目

表1-16 配置 OSPF 支持的路由最大数目

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 支持的路由最大数目	**maximum-routes** { **external** | **inter** | **intra** } *number*	可选 缺省情况下，OSPF 支持的路由最大数目，**external** 为 500000，**inter** 为 20000，**intra** 为 2000

1.6.7 配置 OSPF 最大等价路由条数

如果到一个目的地有几条开销相同的路径，可以实现等价路由负载分担，IP 报文在这几个链路上负载分担，以提高链路利用率。该配置用以设置 OSPF 协议的最大等价路由条数。

表1-17 配置 OSPF 最大等价路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 最大等价路由条数	**maximum load-balancing** *maximum*	可选 缺省情况下，OSPF 最大等价路由条数为 8

1.6.8 配置 OSPF 协议的优先级

由于路由器上可能同时运行多个动态路由协议，就存在各个路由协议之间路由信息共享和选择的问题。系统为每一种路由协议设置一个优先级，在不同协议发现同一条路由时，优先级高的路由将被优先选择。

表1-18 配置 OSPF 协议的优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 协议的路由优先级	**preference** [**ase**] [**route-policy** *route-policy-name*] *value*	可选 缺省情况下，OSPF 内部路由的优先级为 10，OSPF 外部路由的优先级为 150

1.6.9 配置 OSPF 引入外部路由

1. 配置 OSPF 引入其它协议的路由

如果在路由器上不仅运行 OSPF，还运行着其它路由协议，可以配置 OSPF 引入其它协议生成的路由，如 RIP、ISIS、BGP、静态路由或者直连路由，把这些路由信息通过 Type5 LSA 或 Type7 LSA 向外宣告。

OSPF 还可以对引入的路由进行过滤，只将满足过滤条件的外部路由转换为 Type5 LSA 或 Type7 LSA 发布出去。

表1-19 配置 OSPF 引入其它协议的路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 引入其它协议的路由	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] *	必选 缺省情况下，没有引入其他协议的路由信息
配置对引入的路由进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*]]	可选 缺省情况下，没有对引入的路由信息进行过滤



说明

只能引入路由表中状态为 active 的路由，是否为 active 状态可以通过 **display ip routing-table protocol** 命令来查看。

2. 配置 OSPF 引入缺省路由

OSPF 不能通过 **import-route** 命令从其它协议引入缺省路由，如果想把缺省路由引入到 OSPF 路由区域，必须要使用下面命令配置 OSPF 引入缺省路由。

表1-20 配置 OSPF 引入缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 OSPF 引入缺省路由	**default-route-advertise** [**always** | **cost** *cost* | **type** *type* | **route-policy** *route-policy-name*] *	可选 缺省情况下，没有引入缺省路由
	default-route-advertise summary cost *cost*	

说明

default-route-advertise summary cost 命令仅在 VPN 中应用，以 Type-3 LSA 引入缺省路由，PE 路由器会将引入的缺省路由发布给 CE 路由器。

3. 配置引入路由的相关参数

当 OSPF 引入外部路由时，还可以配置一些开销、路由数量、标记和类型等参数的缺省值。路由标记可以用来标识协议相关的信息，如 OSPF 从 BGP 引入路由时，可以用来标记自治系统的编号。

表1-21 配置引入路由时的相关参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置引入外部路由时的参数缺省值（开销、路由数量、标记、类型）	**default** { **cost** *cost* | **limit** *limit* | **tag** *tag* | **type** *type* } *	可选 缺省情况下，OSPF 引入外部路由的缺省值如下： • 路由度量值为 1 • 单位时间内引入外部路由数量的上限为 1000 • 外部路由标记值为 1 • 引入的外部路由类型为 Type2

1.6.10 配置发布一条主机路由

表1-22 配置发布一条主机路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置并发布一条主机路由	**host-advertise** *ip-address cost*	可选 缺省情况下，OSPF 不发布所包含网段之外的主机路由

1.7 配置 OSPF 网络调整优化

用户可以从以下几个方面来调整和优化 OSPF 网络：

- 通过改变 OSPF 的报文定时器，可以调整 OSPF 网络的收敛速度以及协议报文带来的网络负荷。在一些低速链路上，需要考虑接口传送 LSA 的延迟时间。
- 通过调整 SPF 计算间隔时间，可以抑制由于网络频繁变化带来的资源消耗问题。
- 在安全性较高的网络中，可以通过配置 OSPF 验证特性，来提高 OSPF 网络的安全性。
- OSPF 同时支持网管功能，可以配置 OSPF MIB 与某一进程绑定，以及发送 Trap 消息和日志功能。

1.7.1 配置准备

在调整和优化 OSPF 网络之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点之间网络层可达
- 使能 OSPF 功能

1.7.2 配置 OSPF 报文定时器

用户可以在接口上配置下列 OSPF 报文定时器：

- Hello 定时器：接口向邻居发送 Hello 报文的时间间隔，OSPF 邻居之间的 Hello 定时器的值要保持一致，且应与路由收敛速度、网络负荷大小成反比。
- Poll 定时器：在 NBMA 网络中，路由器向状态为 down 的邻居路由器发送轮询 Hello 报文的时间间隔。
- 邻居失效时间：在邻居失效时间内，如果接口还没有收到邻居发送的 Hello 报文，路由器就会宣告该邻居无效。
- 接口重传 LSA 的时间间隔：路由器向它的邻居通告一条 LSA 后，需要对方进行确认。若在重传间隔时间内没有收到对方的确认报文，就会向邻居重传这条 LSA。

表1-23 配置 OSPF 报文定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 Hello 定时器	**ospf timer hello** *seconds*	可选 缺省情况下，P2P、Broadcast 类型接口发送 Hello 报文的时间间隔为 10 秒，P2MP、NBMA 类型接口发送 Hello 报文的时间间隔为 30 秒
配置 Poll 定时器	**ospf timer poll** *seconds*	可选 缺省情况下，发送轮询 Hello 报文的时间间隔为 120 秒
配置邻居失效时间	**ospf timer dead** *seconds*	可选 缺省情况下，P2P、Broadcast 类型接口的 OSPF 邻居失效时间为 40 秒，P2MP、NBMA 类型接口的 OSPF 邻居失效时间为 120 秒

操作	命令	说明
配置接口重传 LSA 的时间间隔	**ospf timer retransmit** *interval*	可选 缺省情况下，时间间隔为 5 秒

说明

- 修改了网络类型后，Hello 定时器与邻居失效时间都将恢复缺省值。
- 在同一接口上邻居失效时间应至少为 Hello 时间间隔的 4 倍。
- 轮询 Hello 报文的时间间隔至少应为 Hello 时间间隔的 4 倍。
- 相邻路由器重传 LSA 时间间隔的值不要设置得太小，否则将会引起不必要的重传。通常应该大于一个报文在两台路由器之间传送一个来回的时间。

1.7.3 配置接口传送 LSA 的延迟时间

考虑到 OSPF 报文在链路上传送时也需要花费时间，所以 LSA 的老化时间（age）在传送之前要增加一定的延迟时间，在低速链路上需要对该项配置进行重点考虑。

表1-24 配置接口传送 LSA 的延迟时间

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口传送 LSA 的延迟时间	**ospf trans-delay** *seconds*	可选 缺省情况下，接口传送 LSA 的延迟时间为 1 秒

1.7.4 配置 SPF 计算时间间隔

当 OSPF 的链路状态数据库（LSDB）发生改变时，需要重新计算最短路径。如果网络频繁变化，且每次变化都立即计算最短路径，将会占用大量系统资源，并影响路由器的效率。通过调节 SPF 计算时间间隔，可以抑制由于网络频繁变化带来的影响。

表1-25 配置 SPF 计算时间间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 SPF 计算时间间隔	**spf-schedule-interval** *maximum-interval* [*minimum-interval* [*incremental-interval*]]	可选 缺省情况下，SPF 计算的时间间隔为 5 秒

说明

本命令在网络变化不频繁的情况下将连续路由计算的时间间隔缩小到 *minimum-interval*，而在网络变化频繁的情况下可以进行相应惩罚，增加 *incremental-interval**2^{n-2}（n 为连续触发路由计算的次数），将等待时间按照配置的惩罚增量延长，最大不超过 *maximum-interval*。

1.7.5 配置 LSA 重复到达的最小时间间隔

如果在重复到达的最小时间间隔收到一条 LSA 类型、LS ID、生成路由器 ID 均相同的 LSA 则直接丢弃，这样就可以抑制网络频繁变化可能导致的占用过多带宽资源和路由器资源。

表1-26 配置 LSA 的重复接收最小间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 LSA 重复到达的最小时间间隔	**lsa-arrival-interval** *interval*	可选 缺省情况下，LSA 重复到达的最小时间间隔为 1000 毫秒

说明

建议 **lsa-arrival-interval** 命令配置的 *interval* 小于或等于 **lsa-generation-interval** 命令所配置的 *minimum-interval*。

1.7.6 配置 LSA 重新生成的时间间隔

通过调节 LSA 重新生成的时间间隔，可以抑制网络频繁变化可能导致的占用过多带宽资源和路由器资源。

表1-27 配置 LSA 发送间隔

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 LSA 重新生成的时间间隔	**lsa-generation-interval** *maximum-interval* [*initial-interval* [*incremental-interval*]]	可选 缺省情况下，最大时间间隔为 5 秒，最小时间间隔为 0 毫秒，惩罚增量为 5000 毫秒

说明

本命令在网络变化不频繁的情况下将 LSA 重新生成时间间隔缩小到 *minimum-interval*，而在网络变化频繁的情况下可以进行相应惩罚，增加 *incremental-interval* × 2^{n-2}（n 为连续触发路由计算的次数），将等待时间按照配置的惩罚增量延长，最大不超过 *maximum-interval*。

1.7.7 禁止接口发送 OSPF 报文

如果要使 OSPF 路由信息不被某一网络中的路由器获得，可以禁止接口发送 OSPF 报文。

表1-28 禁止接口发送 OSPF 报文

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
禁止接口发送 OSPF 报文	**silent-interface** { *interface-type interface-number* | **all** }	可选 缺省情况下，允许接口发送 OSPF 报文

说明

- 不同的进程可以对同一接口禁止发送 OSPF 报文，但 **silent-interface** 命令只对本进程已经使能的 OSPF 接口起作用，对其它进程的接口不起作用。
- 将运行 OSPF 协议的接口指定为 Silent 状态后，该接口的直连路由仍可以由同一路由器的其它接口通过 Router-LSA 发布出去，但 OSPF 报文将被阻塞，接口上无法建立邻居关系。这样可以增强 OSPF 的组网适应能力，减少系统资源的消耗。

1.7.8 配置 Stub 路由器

Stub 路由器用来控制流量，它告知其他 OSPF 路由器不要使用这个 Stub 路由器来转发数据，但可以拥有一个到 Stub 路由器的路由。

通过将当前路由器配置为 Stub 路由器，在该路由器发布的 Router-LSA 中，当链路类型取值为 3 表示连接到 Stub 网络时，链路度量值不变；当链路类型为 1、2、4 分别表示通过点对点链路与另一路由器相连、连接到传送网络、虚连接时，链路度量值将设置为最大值 65535。

这样其邻居计算出这条路由的开销就会很大，如果邻居上有到这个目的地址开销更小的路由，则数据不会通过这个 Stub 路由器转发。

表1-29 配置 Stub 路由器

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 Stub 路由器	**stub-router**	必选 缺省情况下，没有路由器被配置为 Stub 路由器

说明

Stub 路由器与 Stub 区域无关。

1.7.9 配置 OSPF 验证

从安全性角度来考虑，为了避免路由信息外泄或者对 OSPF 路由器进行恶意攻击，OSPF 提供报文验证功能。

OSPF 路由器建立邻居关系时，在发送的报文中会携带配置好的口令，接收报文时进行密码验证，只有通过验证的报文才能接收，否则将不会接收报文，不能正常建立邻居。

要配置 OSPF 报文验证，同一个区域的所有路由器上都需要配置区域验证模式，且配置的验证模式必须相同，同一个网段内的路由器需要配置相同的接口验证模式和口令。

表1-30 配置 OSPF 验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
进入 OSPF 区域视图	**area** *area-id*	-
配置 OSPF 区域的验证模式	**authentication-mode** { **simple** | **md5** }	必选 缺省情况下，没有配置区域验证模式
退回 OSPF 视图	**quit**	-
退回系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPF 接口的验证模式（简单验证）	**ospf authentication-mode simple** [**plain** | **cipher**] *password*	二者必选其一 缺省情况下，接口不对 OSPF 报文进行验证
配置 OSPF 接口的验证模式（MD5 验证）	**ospf authentication-mode** { **md5** | **hmac-md5** } *key-id* [**plain** | **cipher**] *password*	

1.7.10 配置 DD 报文中的 MTU

一般情况下，接口发送 DD 报文时不使用接口的实际 MTU 值，而是用 0 代替。进行此配置后，将使用接口的实际 MTU 值填写 DD 报文 Interface MTU 字段。

表1-31 配置 DD 报文中的 MTU

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 DD 报文中 MTU 域的值为发送该报文接口的 MTU 值	**ospf mtu-enable**	可选 缺省情况下，接口发送的 DD 报文中 MTU 域的值为 0

1.7.11 配置 LSDB 中 External LSA 的最大数量

表1-32 配置 LSDB 中 External LSA 的最大数量

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 LSDB 中 External LSA 的最大数量	**lsdb-overflow-limit** *number*	可选 缺省情况下，不对 LSDB 中 External LSA 的最大条目数进行限制

1.7.12 配置兼容 RFC 1583 的外部路由选择规则

当通过多条 LSA 计算出同一条外部路由时，在 RFC 2328 中定义的选路规则与 RFC 1583 的有所不同，进行此配置可以兼容 RFC 1583 中定义的方式。当 RFC 2328 兼容 RFC 1583 时，优选骨干区的区域内路由；当 RFC 2328 不兼容 RFC 1583 时，优选非骨干区的区域内路由，这样做的目的是尽量减少骨干区的负担。

表1-33 配置兼容 RFC 1583 的外部路由选择规则

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置兼容 RFC 1583 的外部路由选择规则	**rfc1583 compatible**	可选 缺省情况下，使能兼容 RFC 1583 的选路规则



说明

为了避免路由环路，对于是否兼容 RFC 1583 的外部路由选择规则，同一路由域内的路由器建议配置相同，即要么配置所有路由器都兼容 RFC 1583 的外部路由选择规则，要么配置所有路由器都不兼容 RFC 1583 的外部路由选择规则

1.7.13 配置邻接状态输出

表1-34 配置邻接状态输出

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
使能 OSPF 日志记录功能，记录邻居关系连接以及断开事件的日志信息	**log-peer-change**	可选 缺省情况下，使能 OSPF 日志记录功能

1.7.14 配置 OSPF 网管功能

开启 OSPF 模块的 Trap 功能后，该模块会生成 Trap 报文，用于报告该模块的重要事件。Trap 报文的级别如下：

- 错误级别 Trap 报文，级别为 level3；
- 警告级别 Trap 报文，级别为 level4；
- 正常出现但是重要信息级别 Trap 报文，级别为 level5；
- 通知信息级别 Trap 报文，级别为 level6。

生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-35 配置 OSPF 网管功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 OSPF MIB 绑定	**ospf mib-binding** *process-id*	可选 缺省情况下，MIB 绑定在第一个启动的 OSPF 进程上。
使能 OSPF 的 TRAP 功能	**snmp-agent trap enable ospf** [*process-id*] [**ifauthfail** | **ifcfgerror** | **ifrxbadpkt** | **ifstatechange** | **iftxretransmit** | **lsdbapproachoverflow** | **lsdboverflow** | **maxagelsa** | **nbrstatechange** | **originatelsa** | **vifcfgerror** | **virifauthfail** | **virifrxbadpkt** | **virifstatechange** | **viriftxretransmit** | **virnbrstatechange**] *	可选 缺省情况下，OSPF 的 TRAP 功能处于使能状态
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
使能日志信息	**enable log** [**config** | **error** | **state**]	可选 缺省情况下，关闭日志信息

1.7.15 使能日志功能

表1-36 使能日志功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
使能日志功能	**enable log** [**config** | **error** | **state**]	必选 缺省情况下，日志功能处于关闭状态

1.7.16 使能 Opaque LSA 发布接收能力

通过使能 Opaque LSA 发布接收能力，OSPF 可以接收和发布 Type 9、Type 10 和 Type 11 的 Opaque LSA。

表1-37 使能 Opaque LSA 发布接收能力

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
使能 Opaque LSA 发布接收能力	**opaque-capability enable**	可选 缺省情况下，Opaque LSA 发布接收能力处于关闭状态

1.7.17 配置 OSPF 优先接收并处理 Hello 报文

为了保证协议的正常运行，路由器要同时接收和处理 Hello 报文和其它类型的协议报文，当路由器与多个邻居路由器同时建立邻居关系并且路由表中的路由条数比较多时，需要路由器接收和处理的报文数量会很大，可以通过配置 OSPF 优先接收和处理 Hello 报文，来确保邻居关系的稳定性。

表1-38 配置 OSPF 优先接收并处理 Hello 报文

操作	命令	说明
进入系统视图	**system-view**	-
配置 OSPF 优先接收和处理 Hello 报文	**ospf packet-process prioritized-treatment**	必选 缺省情况下，OSPF 不优先接收和处理 Hello 报文

1.7.18 配置 LSU 报文的发送速率

如果路由器路由表里的路由条目很多，在与邻居进行 LSDB 同步时，可能需要发送大量 LSU，有可能会对当前设备和网络带宽带来影响；因此，路由器将 LSU 报文分为多个批次进行发送，并且对 OSPF 接口每次允许发送的 LSU 报文的最大个数做出限制。

用户可根据需要配置 OSPF 接口发送 LSU 报文的时间间隔以及接口一次发送 LSU 报文的最大个数。

表1-39 配置 LSU 报文的发送速率

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	-
配置 LSU 报文的发送速率	**transmit-pacing interval** *interval* **count** *count*	可选 缺省情况下，OSPF 接口发送 LSU 报文的时间间隔为 20 毫秒，一次最多发送 3 个 LSU 报文

1.8 配置 OSPF GR

说明

一台设备可以同时充当 GR Restarter 和 GR Helper。

1.8.1 配置 GR Restarter

可以在 GR Restarter 上配置基于 OSPF 的 IETF 标准或非 IETF 标准的 GR 能力。在作为 GR Restarter 的设备上进行如下配置：

1. 配置 IETF 标准 GR Restarter

表1-40 配置 IETF 标准 GR Restarter

操作	命令	说明
进入系统视图	**system-view**	-
启动 OSPF，进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	必选 缺省情况下，系统没有运行 OSPF
使能 Opaque LSA 发布接收能力	**opaque-capability enable**	必选 缺省情况下，Opaque LSA 发布接收能力处于关闭状态
使能 OSPF 协议的 IETF 标准 GR 能力	**graceful-restart ietf**	必选 缺省情况下，OSPF 协议的 IETF 标准 GR 能力处于关闭状态
配置 OSPF 协议的 GR 重启间隔时间	**graceful-restart interval** *timer*	可选 缺省情况下，OSPF 协议的 GR 重启间隔时间为 120 秒

2. 配置非 IETF 标准 GR Restarter

表1-41 配置非 IETF 标准 GR Restarter

操作	命令	说明
进入系统视图	**system-view**	-
启动 OSPF，进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	必选 缺省情况下，系统没有运行 OSPF
使能 OSPF 本地链路信令能力	**enable link-local-signaling**	必选 缺省情况下，OSPF 本地链路信令能力处于关闭状态
使能 OSPF 带外同步能力	**enable out-of-band-resynchronization**	必选 缺省情况下，OSPF 带外同步能力处于关闭状态
使能 OSPF 协议的非 IETF 标准 GR 能力	**graceful-restart** [**nonstandard**]	必选 缺省情况下，OSPF 协议的非 IETF 标准 GR 能力处于关闭状态
配置 OSPF 协议的 GR 重启间隔时间	**graceful-restart interval** *timer*	可选 缺省情况下，OSPF 协议的 GR 重启间隔时间为 120 秒

1.8.2 配置 GR Helper

可以在作为 GR Helper 的设备上配置基于 OSPF 的 IETF 标准或非 IETF 标准的 GR Helper 能力。在作为 GR Helper 的设备上进行如下配置：

1. 配置 IETF 标准 GR Helper

表1-42 配置 IETF 标准 GR Helper

操作	命令	说明
进入系统视图	**system-view**	-
启动 OSPF，进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	必选 缺省情况下，系统没有运行 OSPF
使能 Opaque LSA 发布接收能力	**opaque-capability enable**	必选 缺省情况下，Opaque LSA 发布接收能力处于关闭状态
配置允许设备可以作哪些 OSPF 邻居的 GR Helper	**graceful-restart help** { *acl-number* | **prefix** *prefix-list* }	可选 缺省情况下，设备可以作任一 OSPF 邻居的 GR Helprer

2. 配置非 IETF 标准 GR Helper

表1-43 配置非 IETF 标准 GR Helper

操作	命令	说明
进入系统视图	**system-view**	-
启动 OSPF，进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	必选 缺省情况下，系统没有运行 OSPF
使能 OSPF 本地链路信令能力	**enable link-local-signaling**	必选 缺省情况下，OSPF 本地链路信令能力处于关闭状态
使能 OSPF 带外同步能力	**enable out-of-band-resynchronization**	必选 缺省情况下，OSPF 带外同步能力处于关闭状态
配置允许设备可以作哪些 OSPF 邻居的 GR Helper	**graceful-restart help** { *acl-number* | **prefix** *prefix-list* }	可选 缺省情况下，设备可以作任一 OSPF 邻居的 GR Helper

1.8.3 重启 OSPF GR 进程

在分布式双主控设备进行主备倒换或者路由器上进行如下操作均可以重新启动 OSPF 协议的 GR 进程。

以 IETF 标准方式实现 GR 时，要求这些 GR 设备上必须预先使能以下两种能力：

- Opaque LSA 发布能力；
- IETF 标准 GR 能力。

以 IETF 标准方式实现 GR 时，要求这些 GR 设备上必须预先使能以下三种能力：

- 本地链路信令能力；
- 带外同步能力；
- 非 IETF 标准 GR 能力。

表1-44 重启 OSPF GR 进程

操作	命令	说明
重启 OSPF 协议的 GR 进程	**reset ospf** [*process-id*] **process graceful-restart**	必选 请在用户视图下执行该命令

1.9 OSPF 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 OSPF 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以复位 OSPF 计数器或连接。

表1-45 OSPF 显示和维护

操作	命令
显示 OSPF 的概要信息	**display ospf** [*process-id*] **brief**
显示 OSPF 的统计信息	**display ospf** [*process-id*] **cumulative**
显示 OSPF 的 LSDB 信息	**display ospf** [*process-id*] **lsdb** [**brief** | [{ **ase** | **router** | **network** | **summary** | **asbr** | **nssa** | **opaque-link** | **opaque-area** | **opaque-as** } [*link-state-id*]] [**originate-router** *advertising-router-id* | **self-originate**]]
显示 OSPF 邻居的信息	**display ospf** [*process-id*] **peer** [**verbose |** [*interface-type interface-number*] [*neighbor-id*]]
显示 OSPF 各区域邻居的统计信息	**display ospf** [*process-id*] **peer statistics**
显示 OSPF 下一跳信息	**display ospf** [*process-id*] **nexthop**
显示 OSPF 路由表的信息	**display ospf** [*process-id*] **routing** [**interface** *interface-type interface-number*] [**nexthop** *nexthop-address*]
显示 OSPF 虚连接信息	**display ospf** [*process-id*] **vlink**
显示 OSPF 请求列表	**display ospf** [*process-id*] **request-queue** [*interface-type interface-number*] [*neighbor-id*]
显示 OSPF 重传列表	**display ospf** [*process-id*] **retrans-queue** [*interface-type interface-number*] [*neighbor-id*]
显示 OSPF ABR 及 ASBR 信息	**display ospf** [*process-id*] **abr-asbr**
显示 OSPF 接口信息	**display ospf** [*process-id*] **interface** [**all** | *interface-type interface-number*]
显示 OSPF 错误信息	**display ospf** [*process-id*] **error**
显示 OSPF ASBR 聚合信息	**display ospf** [*process-id*] **asbr-summary** [*ip-address* { *mask* | *mask-length* }]
清除 OSPF 计数器	**reset ospf** [*process-id*] **counters** [**neighbor** [*interface-type interface-number*] [*router-id*]]
重启 OSPF 进程	**reset ospf** [*process-id*] **process** [**graceful-restart**]
重新向 OSPF 引入外部路由	**reset ospf** [*process-id*] **redistribution**

1.10 典型配置举例

> 注意
> 在配置举例中，只列出了与 OSPF 配置相关的命令。

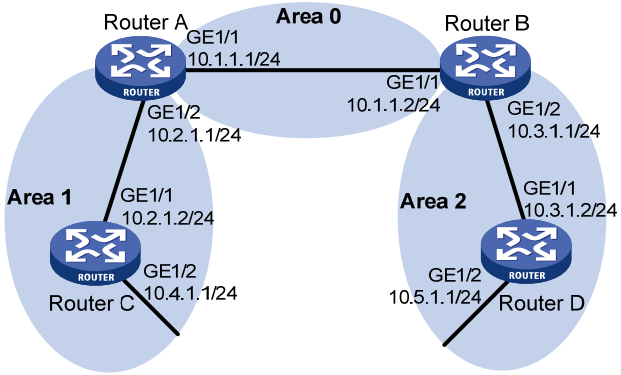
1.10.1 配置 OSPF 基本功能

1. 组网需求

- 所有的路由器都运行 OSPF，并将整个自治系统划分为 3 个区域。
- 其中 Router A 和 Router B 作为 ABR 来转发区域之间的路由。
- 配置完成后，每台路由器都应学到 AS 内的到所有网段的路由。

2. 组网图

图1-21 OSPF 基本配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] area 1
[RouterA-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.1] quit
[RouterA-ospf-1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
```

```
[RouterB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] area 2
[RouterB-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.2] quit
[RouterB-ospf-1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf
[RouterC-ospf-1] area 1
[RouterC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.1] network 10.4.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.1] quit
[RouterC-ospf-1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ospf
[RouterD-ospf-1] area 2
[RouterD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.2] network 10.5.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.2] quit
[RouterD-ospf-1] quit
```

(3) 检验配置结果

查看 Router A 的 OSPF 邻居。

```
[RouterA] display ospf peer verbose
          OSPF Process 1 with Router ID 10.2.1.1
                  Neighbors

 Area 0.0.0.0 interface 10.1.1.1(GigabitEthernet1/1)'s neighbors
 Router ID: 10.3.1.1         Address: 10.1.1.2         GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 10.1.1.1  BDR: 10.1.1.2  MTU: 0
   Dead timer due in 37  sec
   Neighbor is up for 06:03:59
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 5

                  Neighbors

 Area 0.0.0.1 interface 10.2.1.1(GigabitEthernet1/2)'s neighbors
 Router ID: 10.4.1.1         Address: 10.2.1.2         GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 10.2.1.1  BDR: 10.2.1.2  MTU: 0
   Dead timer due in 32  sec
   Neighbor is up for 06:03:12
   Authentication Sequence: [ 0 ]
```

```
   Neighbor state change count: 5
```

显示 Router A 的 OSPF 路由信息。

```
[RouterA] display ospf routing

          OSPF Process 1 with Router ID 10.2.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
10.2.1.0/24         1       Transit 10.2.1.1        10.2.1.1        0.0.0.1
 10.3.1.0/24         2        Inter   10.1.1.2        10.3.1.1        0.0.0.0
 10.4.1.0/24         2       Stub    10.2.1.2        10.4.1.1        0.0.0.1
 10.5.1.0/24         3       Inter   10.1.1.2        10.3.1.1        0.0.0.0
 10.1.1.0/24         1        Transit 10.1.1.1        10.2.1.1        0.0.0.0

 Total Nets: 5
 Intra Area: 3  Inter Area: 2  ASE: 0  NSSA: 0
```

显示 Router A 的 LSDB。

```
[RouterA] display ospf lsdb

          OSPF Process 1 with Router ID 10.2.1.1
                  Link State Database

                         Area: 0.0.0.0
Type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    10.2.1.1        10.2.1.1          1069  36    80000012       0
 Router    10.3.1.1        10.3.1.1           780  36    80000011       0
 Network   10.1.1.1        10.2.1.1          1069  32    80000010       0
 Sum-Net   10.5.1.0        10.3.1.1           780  28    80000003       1
 Sum-Net   10.2.1.0        10.2.1.1          1069  28    8000000F       2
 Sum-Net   10.3.1.0        10.3.1.1           780  28    80000014       2
 Sum-Net   10.4.1.0        10.2.1.1           769  28    8000000F       1
                         Area: 0.0.0.1
 Type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    10.2.1.1        10.2.1.1           769  36    80000012       0
 Router    10.4.1.1        10.4.1.1          1663  48    80000012       0
 Network   10.2.1.1        10.2.1.1           769  32    80000010       0
 Sum-Net   10.5.1.0        10.2.1.1           769  28    80000003       3
 Sum-Net   10.3.1.0        10.2.1.1          1069  28    8000000F       2
 Sum-Net   10.1.1.0        10.2.1.1          1069  28    8000000F       1
 Sum-Asbr  10.3.1.1        10.2.1.1          1069  28    8000000F       1
```

查看 Router D 的路由表。

```
[RouterD] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
```

```
                Routing Tables

 Routing for Network
Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24         3       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24         1       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24         4       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24         1       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24         2       Inter   10.3.1.1        10.3.1.1        0.0.0.2

 Total Nets: 5
 Intra Area: 2  Inter Area: 3  ASE: 0  NSSA: 0
```

在 Router D 上使用 Ping 测试连通性。

```
[RouterD] ping 10.4.1.1
  PING 10.4.1.1: 56  data bytes, press CTRL_C to break
    Request time out
    Reply from 10.4.1.1: bytes=56 Sequence=2 ttl=253 time=15 ms
    Reply from 10.4.1.1: bytes=56 Sequence=3 ttl=253 time=1 ms
    Reply from 10.4.1.1: bytes=56 Sequence=4 ttl=253 time=16 ms
    Reply from 10.4.1.1: bytes=56 Sequence=5 ttl=253 time=1 ms

  --- 10.4.1.1 ping statistics ---
    5 packet(s) transmitted
    4 packet(s) received
    20.00% packet loss
    round-trip min/avg/max = 1/8/16 ms
```

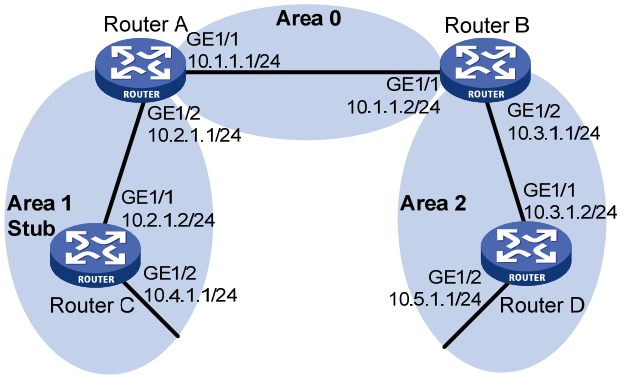
1.10.2 配置 OSPF 引入自治系统外部路由

1. 组网需求

- 所有的路由器都运行 OSPF，整个自治系统划分为 3 个区域。
- 其中 Router A 和 Router B 作为 ABR 来转发区域之间的路由。
- 在 Router C 上配置为 ASBR 引入外部路由（静态路由），且路由信息可正确的在 AS 内传播。

2. 组网图

图1-22 配置 OSPF 引入自治系统外部路由



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置OSPF（同前例 1.10.1 ）

(3) 配置引入自治系统外部路由

在 Router C 上配置一条到目的网段 3.1.2.0/24 的静态路由。

```
<RouterC> system-view
[RouterC] ip route-static 3.1.2.1 24 10.4.1.2
```

在 Router C 上配置 OSPF 引入静态路由。

```
[RouterC] ospf 1
[RouterC-ospf-1] import-route static
```

(4) 检验配置效果

查看 Router D 的 ABR/ASBR 信息。

```
<RouterD> display ospf abr-asbr

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Table to ABR and ASBR

 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.3.1.1        0.0.0.2         10    10.3.1.1        ABR
 Inter       10.4.1.1        0.0.0.2         22    10.3.1.1        ASBR
```

查看 Router D 的 OSPF 路由表。

```
<RouterD> display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                   Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
 3.1.2.0/24         1        Type2   1           10.3.1.1        10.4.1.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

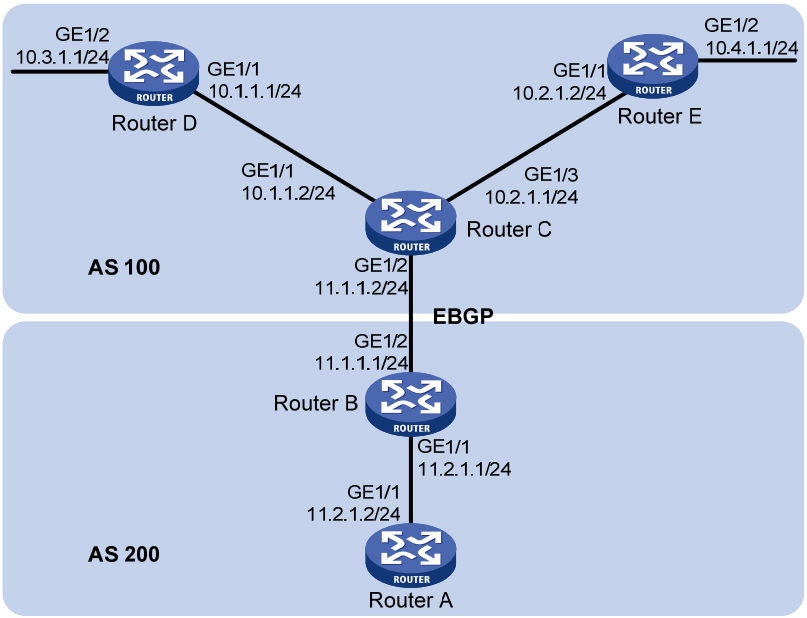
1.10.3 配置 OSPF 发布聚合路由

1. 组网需求

- Router A 和 Router B 位于 AS200 内，AS200 内使用 OSPF 做为 IGP 协议；
- Router C、Router D 和 Router E 位于 AS100 内，AS100 内使用 OSPF 做为 IGP 协议；
- Router B 和 Router C 之间建立 EBGP 连接，在 Router C 上配置 BGP 引入 OSPF 路由；
- 在 Router B 上配置 OSPF 进程引入 BGP 路由，为了减小 Router A 的路由表规模，在 Router B 上配置路由聚合，只发布聚合后的路由 10.0.0.0/8。

2. 组网图

图1-23 配置 OSPF 发布聚合路由



3. 配置步骤

(1) 配置接口的 IP 地址（略）

(2) 配置 OSPF

配置 Router A。

```
<RouterA> system-view
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf
[RouterC-ospf-1] area 0
[RouterC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] quit
[RouterC-ospf-1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ospf
[RouterD-ospf-1] area 0
[RouterD-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.0] quit
```

配置 Router E。

```
<RouterE> system-view
[RouterE] ospf
[RouterE-ospf-1] area 0
[RouterE-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[RouterE-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[RouterE-ospf-1-area-0.0.0.0] quit
[RouterE-ospf-1] quit
```

(3) 配置 BGP

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 200
[RouterB-bgp] peer 11.1.1.2 as 100
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 100
[RouterC-bgp] peer 11.1.1.1 as 200
[RouterC-bgp] import-route ospf
```

(4) 在 Router B 上配置路由引入

在 Router B 上配置 OSPF 引入 BGP 路由

```
[RouterB] ospf
[RouterB-ospf-1] import-route bgp
```

查看 Router A 的路由表信息。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8
```

```
Destination/Mask    Proto   Pre  Cost        NextHop         Interface

10.1.1.0/24         O_ASE   150  1           11.2.1.1        GE1/1
10.2.1.0/24         O_ASE   150  1           11.2.1.1        GE1/1
10.3.1.0/24         O_ASE   150  1           11.2.1.1        GE1/1
10.4.1.0/24         O_ASE   150  1           11.2.1.1        GE1/1
11.2.1.0/24         Direct  0    0           11.2.1.2        GE1/1
11.2.1.2/32         Direct  0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct  0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0    0           127.0.0.1       InLoop0
```

(5) 在 Router B 上配置路由聚合，只发布聚合路由 10.0.0.0/8。

```
[RouterB-ospf-1] asbr-summary 10.0.0.0 8
```

查看 Router A 的路由表信息。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto   Pre  Cost        NextHop         Interface

10.0.0.0/8          O_ASE   150  2           11.2.1.1        GE1/1
11.2.1.0/24         Direct  0    0           11.2.1.2        GE1/1
11.2.1.2/32         Direct  0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct  0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0    0           127.0.0.1       InLoop0
```

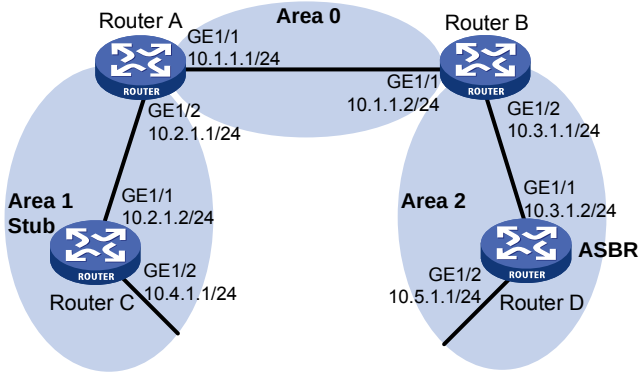
1.10.4 配置 OSPF 的 Stub 区域

1. 组网需求

- 所有的路由器都运行 OSPF，整个自治系统划分为 3 个区域。
- 其中 Router A 和 Router B 作为 ABR 来转发区域之间的路由，Router D 作为 ASBR 引入了外部路由（静态路由）。
- 要求将 Area1 配置为 Stub 区域，减少通告到此区域内的 LSA 数量，但不影响路由的可达性。

2. 组网图

图1-24 配置 OSPF Stub 区域组网图



3. 配置步骤

(1) 配置接口的 IP 地址（略）

(2) 配置OSPF（同前例 1.10.1 ）

(3) 配置 Router D 引入静态路由

```
[RouterD] ip route-static 3.1.2.1 24 10.5.1.2
[RouterD] ospf
[RouterD-ospf-1] import-route static
[RouterD-ospf-1] quit
```

查看 Router C 的 ABR/ASBR 信息。

```
[RouterC] display ospf abr-asbr

          OSPF Process 1 with Router ID 10.4.1.1
                 Routing Table to ABR and ASBR

 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.2.1.1        0.0.0.1         3     10.2.1.1        ABR
 Inter       10.3.1.1        0.0.0.1         5     10.2.1.1        ABR
 Inter       10.5.1.1        0.0.0.1         7     10.2.1.1        ASBR
```

查看 Router C 的 OSPF 路由表。

```
[RouterC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Transit 10.2.1.2        10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
 3.1.2.0/24         1        Type2   1           10.2.1.1        10.5.1.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```



说明

当 Router C 所在区域为普通区域时，可以看到路由表中存在 AS 外部的路由。

(4) 配置 Area1 为 Stub 区域

配置 Router A。

```
[RouterA] ospf
[RouterA-ospf-1] area 1
[RouterA-ospf-1-area-0.0.0.1] stub
[RouterA-ospf-1-area-0.0.0.1] quit
[RouterA-ospf-1] quit
```

配置 Router C。

```
[RouterC] ospf
[RouterC-ospf-1] area 1
[RouterC-ospf-1-area-0.0.0.1] stub
[RouterC-ospf-1-area-0.0.0.1] quit
[RouterC-ospf-1] quit
```

显示 Router C 的路由表。

```
[RouterC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.2.1.0/24        3        Transit 10.2.1.2        10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 4  ASE: 0  NSSA: 0
```



说明

当把 Router C 所在区域配置为 Stub 区域时，已经看不到 AS 外部的路由，取而代之的是一条缺省路由。

在 ABR 上配置禁止向区域通告 Type3 LSA。

```
[RouterA] ospf
[RouterA-ospf-1] area 1
[RouterA-ospf-1-area-0.0.0.1] stub no-summary
[RouterA-ospf-1-area-0.0.0.1] quit
```

查看 Router C 的 OSPF 路由表。

```
[RouterC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables
```

```
Routing for Network
Destination        Cost     Type    NextHop         AdvRouter       Area
0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
10.2.1.0/24        3        Transit 10.2.1.2        10.4.1.1        0.0.0.1
10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1

Total Nets: 3
Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

说明

禁止向 Stub 区域通告 Summary LSA 后，Stub 路由器的路由表项进一步减少，只保留了一条通往区域外部的缺省路由。

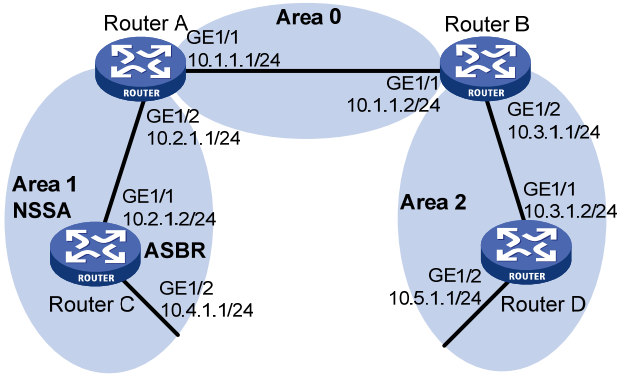
1.10.5 配置 OSPF 的 NSSA 区域

1. 组网需求

- 所有的路由器都运行 OSPF，整个自治系统划分为 3 个区域。
- 其中 Router A 和 Router B 作为 ABR 来转发区域之间的路由。
- 要求将 Area1 配置为 NSSA 区域，同时将 Router C 配置为 ASBR 引入外部路由（静态路由），且路由信息可正确的在 AS 内传播。

2. 组网图

图1-25 配置 OSPF NSSA 区域组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置OSPF（同前例 1.10.1 ）

(3) 配置 Area1 区域为 NSSA 区域

配置 Router A。

```
[RouterA] ospf
[RouterA-ospf-1] area 1
[RouterA-ospf-1-area-0.0.0.1] nssa
[RouterA-ospf-1-area-0.0.0.1] quit
```

配置 Router C。

```
[RouterC] ospf
[RouterC-ospf-1] area 1
[RouterC-ospf-1-area-0.0.0.1] nssa
[RouterC-ospf-1-area-0.0.0.1] quit
[RouterC-ospf-1] quit
```



- 建议在 ABR（这里的 Router A）上配置 **default-route-advertise no-summary** 参数，相当于把该区域配置为 NSSA Totally Stub 区域，这样可以进一步减少 NSSA 路由器的路由表数量。
- 其他 NSSA 路由器只需配置 **nssa** 命令就可以。

查看 Router C 的 OSPF 路由表。

```
[RouterC] display ospf routing
          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Transit 10.2.1.2        10.4.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Total Nets: 5
 Intra Area: 2  Inter Area: 3  ASE: 0  NSSA: 0
```

(4) 配置 Router C 引入静态路由

```
[RouterC] ip route-static 3.1.2.1 24 10.4.1.2
[RouterC] ospf
[RouterC-ospf-1] import-route static
[RouterC-ospf-1] quit
```

查看 Router D 的 OSPF 路由表。

```
[RouterD] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2
```

```
Routing for ASEs
Destination        Cost     Type    Tag         NextHop         AdvRouter
3.1.2.0/24         1        Type2   1           10.3.1.1        10.2.1.1

Total Nets: 6
Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

说明

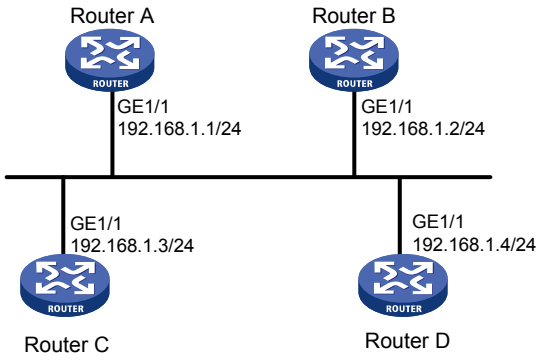
在 Router D 上可以看到 NSSA 区域引入的一条 AS 外部的路由。

1.10.6 配置 OSPF 的 DR 选择

1. 组网需求

- Router A、Router B、Router C、Router D 在同一网段，运行 OSPF 协议；
- 配置 Router A 为 DR，Router C 为 BDR。

图1-26 配置 OSPF 的 DR 选择组网图



2. 配置思路

(1) 配置各接口的 IP 地址；

(2) 配置 OSPF 基本功能；

(3) 改变路由器接口的路由器优先级使 Router A 成为 DR，Router C 成为 BDR。

3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] router id 1.1.1.1
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] router id 2.2.2.2
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] router id 3.3.3.3
[RouterC] ospf
[RouterC-ospf-1] area 0
[RouterC-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] quit
[RouterC-ospf-1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] router id 4.4.4.4
[RouterD] ospf
[RouterD-ospf-1] area 0
[RouterD-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.0] quit
[RouterD-ospf-1] quit
```

查看 Router A 的邻居信息。

```
<RouterA> display ospf peer verbose

          OSPF Process 1 with Router ID 1.1.1.1
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.1(GigabitEthernet1/1)'s neighbors
 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 38  sec
   Neighbor is up for 00:01:31
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:01:28
   Authentication Sequence: [ 0 ]

 Router ID: 4.4.4.4          Address: 192.168.1.4      GR State: Normal
```

```
  State: Full  Mode: Nbr is Master  Priority: 1
  DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
  Dead timer due in 31  sec
  Neighbor is up for 00:01:28
  Authentication Sequence: [ 0 ]
```

可以看到 Router D 为 DR，Router C 为 BDR。

(3) 配置接口上的路由器优先级

配置 Router A。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospf dr-priority 100
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ospf dr-priority 0
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ospf dr-priority 2
[RouterC-GigabitEthernet1/1] quit
```

查看 Router D 的邻居信息。

```
<RouterD> display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(GigabitEthernet1/1)'s neighbors
 Router ID: 1.1.1.1      Address: 192.168.1.1      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 100
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:11:17
   Authentication Sequence: [ 0 ]

 Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 0
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 35  sec
   Neighbor is up for 00:11:19
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 2
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 33  sec
   Neighbor is up for 00:11:15
```

```
  Authentication Sequence: [ 0 ]
```

可以看到，网络中 DR/BDR 并没有改变。



说明

网络中 DR/BDR 已经存在的情况下，接口上的路由器优先级的配置并不会立即生效。

(4) 重启 OSPF 进程

重启 Router D 的进程。

```
<RouterD> reset ospf 1 process
Warning : Reset OSPF process? [Y/N]:y
```

查看 Router D 的邻居信息。

```
<RouterD> display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(GigabitEthernet1/1)'s neighbors
 Router ID: 1.1.1.1          Address: 192.168.1.1      GR State: Normal
   State: Full  Mode: Nbr is Slave  Priority: 100
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 39  sec
   Neighbor is up for 00:01:40
   Authentication Sequence: [ 0 ]

 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 0
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 35  sec
   Neighbor is up for 00:01:44
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is Slave  Priority: 2
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 39  sec
   Neighbor is up for 00:01:41
   Authentication Sequence: [ 0 ]
```

可以看到 Router A 成为 DR，Router C 为 BDR。



说明

当路由器的邻居关系稳定后：

- 如果邻居的状态是 Full，这说明它和邻居之间形成了邻接关系；
- 如果邻居的状态是 2-Way，则说明它们都不是 DR 或 BDR，两者之间不需要交换 LSA。

查看 OSPF 接口的状态。

```
<RouterA> display ospf interface

          OSPF Process 1 with Router ID 1.1.1.1
                  Interfaces

 Area: 0.0.0.0
 IP Address      Type      State   Cost  Pri   DR             BDR
 192.168.1.1     Broadcast DR      1     100   192.168.1.1    192.168.1.3

[RouterB] display ospf interface

          OSPF Process 1 with Router ID 2.2.2.2
                  Interfaces

 Area: 0.0.0.0
 IP Address      Type      State    Cost  Pri   DR             BDR
 192.168.1.2     Broadcast DROther  1     0     192.168.1.1    192.168.1.3
```

说明

如果 OSPF 接口的状态是 DROther，则说明它既不是 DR，也不是 BDR。

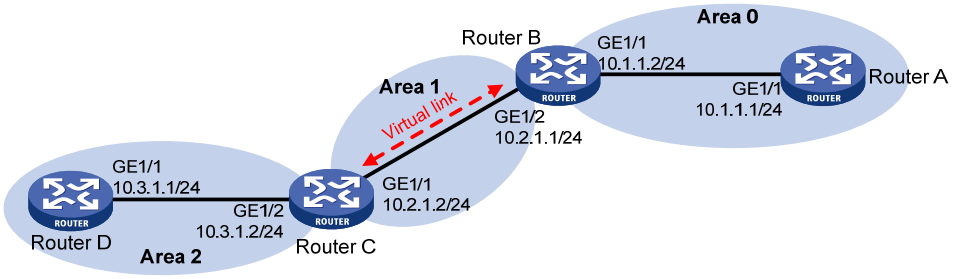
1.10.7 配置 OSPF 虚连接

1. 组网需求

- Area2 没有与 Area0 直接相连。Area1 被用作传输区域（Transit Area）来连接 Area2 和 Area0。Router B 和 Router C 之间配置一条虚连接。
- 配置完成后，Router B 能够学到 Area2 中的路由。

2. 组网图

图1-27 配置 OSPF 虚连接的组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ospf 1 router-id 1.1.1.1
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf 1 router-id 2.2.2.2
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] area 1
[RouterB–ospf-1-area-0.0.0.1] network 10.2.0.0 0.0.0.255
[RouterB–ospf-1-area-0.0.0.1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf 1 router-id 3.3.3.3
[RouterC-ospf-1] area 1
[RouterC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.1] quit
[RouterC-ospf-1] area 2
[RouterC–ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[RouterC–ospf-1-area-0.0.0.2] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ospf 1 router-id 4.4.4.4
[RouterD-ospf-1] area 2
[RouterD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[RouterD-ospf-1-area-0.0.0.2] quit
```

查看 Router B 的 OSPF 路由表。

```
[RouterB] display ospf routing

          OSPF Process 1 with Router ID 2.2.2.2
                  Routing Tables
 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0
 Total Nets: 2
 Intra Area: 2  Inter Area: 0  ASE: 0  NSSA: 0
```



说明

由于 Area0 没有与 Area2 直接相连，所以 Router B 的路由表中没有 Area2 中的路由。

(3) 配置虚连接

配置 Router B。

```
[RouterB] ospf
[RouterB-ospf-1] area 1
[RouterB-ospf-1-area-0.0.0.1] vlink-peer 3.3.3.3
[RouterB-ospf-1-area-0.0.0.1] quit
[RouterB-ospf-1] quit
```

配置 Router C。

```
[RouterC] ospf 1
[RouterC-ospf-1] area 1
[RouterC-ospf-1-area-0.0.0.1] vlink-peer 2.2.2.2
[RouterC-ospf-1-area-0.0.0.1] quit
```

查看 Router B 的 OSPF 路由表。

```
[RouterB] display ospf routing
          OSPF Process 1 with Router ID 2.2.2.2
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.3.1.0/24        5        Inter   10.2.1.2        3.3.3.3         0.0.0.0
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

可以看到，Router B 已经学到了 Area2 的路由 10.3.1.0/24。

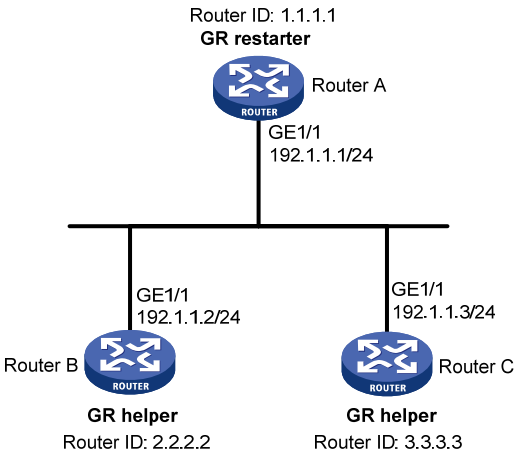
1.10.8 OSPF GR 配置举例

1. 组网需求

- Router A、Router B 和 Router C 既属于同一自治系统，也属于同一 OSPF 域，通过 OSPF 协议实现网络互连，并提供 GR 机制。
- Router A 作为非 IETF 标准 GR Restarter，Router B 和 Router C 作为 GR Helper 并且通过 GR 机制与 Router A 保持带外同步。

2. 组网图

图1-28 OSPF GR 配置组网图（路由应用）



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface gigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.1.1.1 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
[RouterA] router id 1.1.1.1
[RouterA] ospf 100
[RouterA-ospf-100] enable link-local-signaling
[RouterA-ospf-100] enable out-of-band-resynchronization
[RouterA-ospf-100] graceful-restart
[RouterA-ospf-100] area 0
[RouterA-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[RouterA-ospf-100-area-0.0.0.0] return
```

配置 Router B

```
<RouterB> system-view
[RouterB] interface gigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 10.1.1.0 255.255.255.0
[RouterB-GigabitEthernet1/1] quit
[RouterB] router id 2.2.2.2
[RouterB] ospf 100
[RouterB-ospf-100] enable link-local-signaling
[RouterB-ospf-100] enable out-of-band-resynchronization
[RouterB-ospf-100] graceful-restart help 2000
[RouterB-ospf-100] area 0
[RouterB-ospf-100-area-0.0.0.0] network 10.1.1.0 0.0.0.255
```

配置 Router C

```
<RouterC> system-view
 [RouterC] interface gigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ip address 192.1.1.3 255.255.255.0
```

```
[RouterC-GigabitEthernet1/1] quit
[RouterC] router id 3.3.3.3
[RouterC] ospf 100
[RouterC-ospf-100] enable link-local-signaling
[RouterC-ospf-100] enable out-of-band-resynchronization
[RouterC-ospf-100] graceful-restart help 2000
[RouterC-ospf-100] area 0
[RouterC-ospf-100-area-0.0.0.0] network 10.1.1.0 0.0.0.255
```

(2) 检验配置效果

运行稳定后，打开 Router A 的 OSPF 平滑启动事件调试信息开关。在 Router A 上重启 OSPF 协议的 GR 进程。

```
<RouterA> debugging ospf event graceful-restart
<RouterA> terminal monitor
<RouterA> terminal debugging
<RouterA> reset ospf 100 process graceful-restart
Warning : Reset OSPF process? [Y/N]:y
%Dec 12 09:36:12:500 2006 RouterA RM/3/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
192.1.1.1(Ethernet1/1) from Full to Down
OSPF 1: Intf 192.1.1.1 Rcv InterfaceDown State BackupDR -> Down.
OSPF 1 nonstandard GR Started for OSPF Router
OSPF 1 notify RM that OSPF process will enter GR.
OSPF 1 created GR wait timer, timeout interval is 40(s).
OSPF 1 created GR Interval timer,timeout interval is 120(s).
OSPF 1: Intf 192.1.1.1 Rcv InterfaceUp State Down -> Waiting.
OSPF 1: Intf 192.1.1.1 Rcv BackupSeen State Waiting -> BackupDR.
OSPF 1 created OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
%Oct 22 09:36:12:566 2008 RouterA RM/3/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
192.1.1.2(Ethernet1/1) from Loading to Full
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 deleted OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 Gr Wait Timeout timer fired.
OSPF 1 deleted GR wait timer.
OSPF 1 deleted GR Interval timer.
OSPF 1 GR Completed for OSPF Router
OSPF 1 notified RM that OSPF process left GR.
RM notified that all protocol left GR.
OSPF 1 started flushing STALE LSA after all protocol left GR.
OSPF 1: Flush Stale Area LSAs
OSPF 1: Start Flush Stale ASE + NSSA LSAs
OSPF 1: End Flush Stale ASE + NSSA LSAs
```

从上面的信息可以看出 Router A 在 Router B 的协助下完成了 GR。

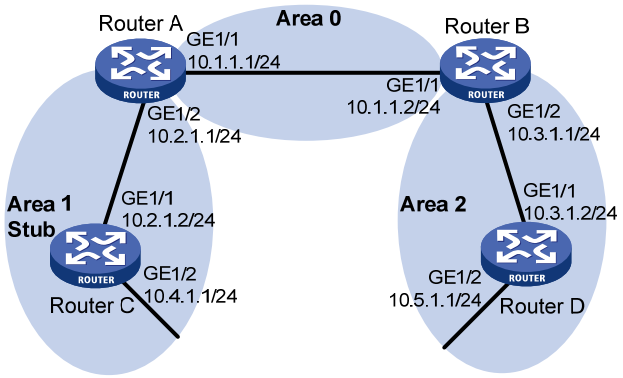
1.10.9 配置路由过滤

1. 组网需求

- 所有的路由器都运行 OSPF，整个自治系统划分为 3 个区域。
- 其中 Router A 和 Router B 作为 ABR 来转发区域之间的路由。
- 在 Router C 上配置为 ASBR 引入外部路由（静态路由），并在 Router C 上配置过滤策略，对引入的一条路由（3.1.3.0/24）进行过滤。
- 在 Router A 上配置路由策略，对路由（10.5.1.0/24）进行过滤。

2. 组网图

图1-29 配置路由过滤



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置OSPF（同前例 1.10.1 ）

(3) 配置引入自治系统外部路由

在 Router C 上配置一条到目的网段 3.1.1.0/24 的静态路由。

```
<RouterC> system-view
[RouterC] ip route-static 3.1.1.0 24 10.4.1.2
```

在 Router C 上配置一条到目的网段 3.1.2.0/24 的静态路由。

```
[RouterC] ip route-static 3.1.2.0 24 10.4.1.2
```

在 Router C 上配置一条到目的网段 3.1.3.0/24 的静态路由。

```
[RouterC] ip route-static 3.1.3.0 24 10.4.1.2
```

在 Router C 上配置 OSPF 引入静态路由。

```
[RouterC] ospf 1
[RouterC-ospf-1] import-route static
[RouterC-ospf-1] quit
```

在 Router A 上查看路由信息。

```
<RouterA> display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12
```

```
Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE  150  1            10.2.1.2        GE1/2
3.1.2.0/24          O_ASE  150  1            10.2.1.2        GE1/2
3.1.3.0/24          O_ASE  150  1            10.2.1.2        GE1/2
10.1.1.0/24         Direct 0    0            10.1.1.1        GE1/1
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.1        GE1/2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         OSPF   10   4            10.1.1.2        GE1/1
10.4.1.0/24         OSPF   10   13           10.2.1.2        GE1/2
10.5.1.0/24         OSPF   10   14           10.1.1.2        GE1/1
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

(4) 在 Router C 配置对路由 3.1.3.0/24 进行过滤

配置 IPv4 地址前缀列表。

```
[RouterC] ip ip-prefix prefix1 index 1 deny 3.1.3.0 24
[RouterC] ip ip-prefix prefix1 index 2 permit 3.1.1.0 24
[RouterC] ip ip-prefix prefix1 index 3 permit 3.1.2.0 24
```

配置对引入的静态路由信息进行过滤，过滤掉路由 3.1.3.0/24。

```
[RouterC] ospf 1
[RouterC-ospf-1] filter-policy ip-prefix prefix1 export static
```

在 Router A 上查看路由信息。

```
<RouterA> display ip routing-table
Routing Tables: Public
         Destinations : 11       Routes : 11

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE  150  1            10.2.1.2        GE1/2
3.1.2.0/24          O_ASE  150  1            10.2.1.2        GE1/2
10.1.1.0/24         Direct 0    0            10.1.1.1        GE1/1
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.1        GE1/2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         OSPF   10   4            10.1.1.2        GE1/1
10.4.1.0/24         OSPF   10   13           10.2.1.2        GE1/2
10.5.1.0/24         OSPF   10   14           10.1.1.2        GE1/1
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

可以看到，到目的网段 3.1.3.0/24 的路由被过滤掉了。

(5) 在 Router A 上配置对路由 10.5.1.1/24 进行过滤

在 Router A 上配置访问控制列表。

```
<RouterA> system-veiw
[RouterA] acl number 2000
```

```
[RouterA-acl-basic-2000] rule 0 deny source 10.5.1.0 0.0.0.255
[RouterA-acl-basic-2000] rule 1 permit source any
[RouterA-acl-basic-2000] quit
```

配置对通过 LSA 计算出来的路由信息 10.5.1.0/24 进行过滤。

```
[RouterA] ospf 1
[RouterA-ospf-1] filter-policy 2000 import
[RouterA-ospf-1] quit
```

在 Router A 上查看路由信息。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto   Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE   150  1            10.2.1.2        GE1/2
3.1.2.0/24          O_ASE   150  1            10.2.1.2        GE1/2
10.1.1.0/24         Direct  0    0            10.1.1.1        GE1/1
10.1.1.1/32         Direct  0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct  0    0            10.2.1.1        GE1/2
10.2.1.1/32         Direct  0    0            127.0.0.1       InLoop0
10.3.1.0/24         OSPF    10   4            10.1.1.2        GE1/1
10.4.1.0/24         OSPF    10   13           10.2.1.2        GE1/2
127.0.0.0/8         Direct  0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct  0    0            127.0.0.1       InLoop0
```

可以看到，到 10.5.1.1/24 的路由被过滤掉了。

1.11 常见配置错误举例

1.11.1 OSPF 邻居无法建立

1. 故障现象

OSPF 邻居无法建立

2. 分析

如果物理连接和下层协议正常，则检查接口上配置的 OSPF 参数，必须保证与相邻路由器的参数一致，区域号相同，网段与掩码也必须一致（点到点与虚连接的网段与掩码可以不同）。

相邻的两台路由器接口的网络类型必须一致。若网络类型为广播网或 NBMA，则至少有一个接口的路由器优先级应大于零。

3. 处理过程

(1) 使用 **display ospf peer** 命令查看 OSPF 邻居状态。

(2) 使用 **display ospf interface** 命令查看 OSPF 接口的信息。

(3) 检查物理连接及下层协议是否正常运行，可通过 **Ping** 命令测试。若从本地路由器 Ping 对端路由器不通，则表明物理连接和下层协议有问题。

(4) 检查 OSPF 定时器，在同一接口上邻居失效时间应至少为 Hello 报文发送时间间隔的 4 倍。

(5) 如果是 NBMA 网络，则应该使用 **peer ip-address** 命令手工指定邻居。

(6) 如果网络类型为广播网或 NBMA，则至少有一个接口的路由器优先级大于零。

1.11.2 OSPF 路由信息不正确

1. 故障现象

OSPF 不能发现其他区域的路由

2. 分析

应保证骨干区域与所有的区域相连接。若一台路由器配置了两个以上的区域，则至少有一个区域应与骨干区域相连。骨干区域也不能配置成 Stub 区域。

在 Stub 区域内的路由器不能接收外部 AS 的路由。如果一个区域配置成 Stub 区域，则与这个区域相连的所有路由器都应将此区域配置成 Stub 区域。

3. 处理过程

(1) 使用 **display ospf peer** 命令查看 OSPF 邻居状态。

(2) 使用 **display ospf interface** 命令查看 OSPF 接口的信息。

(3) 使用 **display ospf lsdb** 查看 LSDB 的信息是否完整。

(4) 使用 **display current-configuration configuration ospf** 命令查看区域是否配置正确。若配置了两个以上的区域，则至少有一个区域与骨干区域相连。

(5) 如果某区域是 Stub 区域，则该区域中的所有路由器都要配置 **stub** 命令；如果某区域是 NSSA 区域，则该区域中的所有路由器都要配置 **nssa** 命令。

(6) 如果配置了虚连接，使用 **display ospf vlink** 命令查看 OSPF 虚连接是否正常。

目 录

1 RIP配置 1-1
1.1 简介 1-1
1.1.1 RIP工作机制 1-1
1.1.2 RIP的启动和运行过程 1-2
1.1.3 RIP的版本 1-2
1.1.4 RIP的报文格式 1-3
1.1.5 支持的RIP特性 1-5
1.1.6 协议规范 1-5
1.2 配置RIP的基本功能 1-5
1.2.1 配置准备 1-5
1.2.2 配置RIP的基本功能 1-5
1.3 配置RIP路由特性 1-7
1.3.1 配置接口附加度量值 1-7
1.3.2 配置RIP-2 路由聚合 1-8
1.3.3 禁止RIP接收主机路由 1-8
1.3.4 配置RIP发布缺省路由 1-9
1.3.5 配置RIP对接收/发布的路由进行过滤 1-9
1.3.6 配置RIP协议优先级 1-10
1.3.7 配置RIP引入外部路由 1-10
1.4 调整和优化RIP网络 1-11
1.4.1 配置RIP定时器 1-11
1.4.2 配置水平分割和毒性逆转 1-11
1.4.3 配置最大等价路由条数 1-12
1.4.4 配置RIP-1 报文的零域检查 1-13
1.4.5 配置源地址检查 1-13
1.4.6 配置RIP-2 报文的认证方式 1-13
1.4.7 配置RIP邻居 1-14
1.4.8 配置RIP和MIB绑定 1-14
1.4.9 配置RIP报文的发送速率 1-15
1.5 RIP显示和维护 1-15
1.6 RIP典型配置举例 1-15
1.6.1 配置RIP的版本 1-15
1.6.2 配置RIP引入外部路由 1-17
1.6.3 配置RIP接口附加度量值 1-19
1.6.4 配置RIP发布聚合路由 1-21
1.7 常见配置错误举例 1-24
1.7.1 收不到邻居的RIP更新报文 1-24
1.7.2 RIP网络发生路由振荡 1-24

1 RIP 配置

1.1 简介

RIP（Routing Information Protocol，路由信息协议）是一种较为简单的内部网关协议（Interior Gateway Protocol，IGP），主要用于规模较小的网络中，比如校园网以及结构较简单的地区性网络。对于更为复杂的环境和大型网络，一般不使用 RIP。

由于 RIP 的实现较为简单，在配置和维护管理方面也远比 OSPF 和 IS-IS 容易，因此在实际组网中仍有广泛的应用。

1.1.1 RIP 工作机制

1. RIP 的基本概念

RIP 是一种基于距离矢量（Distance-Vector，D-V）算法的协议，它通过 UDP 报文进行路由信息的交换，使用的端口号为 520。

RIP 使用跳数来衡量到达目的地址的距离，跳数称为度量值。在 RIP 中，路由器到与它直接相连网络的跳数为 0，通过一个路由器可达的网络的跳数为 1，其余依此类推。为限制收敛时间，RIP 规定度量值取 0～15 之间的整数，大于或等于 16 的跳数被定义为无穷大，即目的网络或主机不可达。由于这个限制，使得 RIP 不适合应用于大型网络。

为提高性能，防止产生路由环路，RIP 支持水平分割（Split Horizon）和毒性逆转（Poison Reverse）功能。

2. RIP 的路由数据库

每个运行 RIP 的路由器管理一个路由数据库，该路由数据库包含了到所有可达目的地的路由项，这些路由项包含下列信息：

- 目的地址：主机或网络的地址。
- 下一跳地址：为到达目的地，需要经过的相邻路由器的接口 IP 地址。
- 出接口：本路由器转发报文的出接口。
- 度量值：本路由器到达目的地的开销。
- 路由时间：从路由项最后一次被更新到现在所经过的时间，路由项每次被更新时，路由时间重置为 0。
- 路由标记（Route Tag）：用于标识外部路由，在路由策略中可根据路由标记对路由信息进行灵活的控制。关于路由策略的详细信息，请参见“IP 路由分册”中的“路由策略配置”。

3. RIP 定时器

RIP 受四个定时器的控制，分别是 Update、Timeout、Suppress 和 Garbage-Collect。

- Update 定时器，定义了发送路由更新的时间间隔。
- Timeout 定时器，定义了路由老化时间。如果在老化时间内没有收到关于某条路由的更新报文，则该条路由在路由表中的度量值将会被设置为 16。

- Suppress 定时器，定义了 RIP 路由处于抑制状态的时长。当一条路由的度量值变为 16 时，该路由将进入抑制状态。在被抑制状态，只有来自同一邻居且度量值小于 16 的路由更新才会被路由器接收，取代不可达路由。
- Garbage-Collect 定时器，定义了一条路由从度量值变为 16 开始，直到它从路由表里被删除所经过的时间。在 Garbage-Collect 时间内，RIP 以 16 作为度量值向外发送这条路由的更新，如果 Garbage-Collect 超时，该路由仍没有得到更新，则该路由将从路由表中被彻底删除。

4. 防止路由环路

RIP 是一种基于 D-V 算法的路由协议，由于它向邻居通告的是自己的路由表，存在发生路由环路的可能性。

RIP 通过以下机制来避免路由环路的产生：

- 计数到无穷（Counting to infinity）：将度量值等于 16 的路由定义为不可达（infinity）。在路由环路发生时，某条路由的度量值将会增加到 16，该路由被认为不可达。
- 水平分割（Split Horizon）：RIP 从某个接口学到的路由，不会从该接口再发回给邻居路由器。这样不但减少了带宽消耗，还可以防止路由环路。
- 毒性逆转（Poison Reverse）：RIP 从某个接口学到路由后，将该路由的度量值设置为 16（不可达），并从原接口发回邻居路由器。利用这种方式，可以清除对方路由表中的无用信息。
- 触发更新（Triggered Updates）：RIP 通过触发更新来避免在多个路由器之间形成路由环路的可能，而且可以加速网络的收敛速度。一旦某条路由的度量值发生了变化，就立刻向邻居路由器发布更新报文，而不是等到更新周期的到来。

1.1.2 RIP 的启动和运行过程

RIP 启动和运行的整个过程可描述如下：

- 路由器启动 RIP 后，便会向相邻的路由器发送请求报文（Request message），相邻的 RIP 路由器收到请求报文后，响应该请求，回送包含本地路由表信息的响应报文（Response message）。
- 路由器收到响应报文后，更新本地路由表，同时向相邻路由器发送触发更新报文，通告路由更新信息。相邻路由器收到触发更新报文后，又向其各自的相邻路由器发送触发更新报文。在一连串的触发更新广播后，各路由器都能得到并保持最新的路由信息。
- RIP 在缺省情况下每隔 30 秒向相邻路由器发送本地路由表，运行 RIP 协议的相邻路由器在收到报文后，对本地路由进行维护，选择一条最佳路由，再向其各自相邻网络发送更新信息，使更新的路由最终能达到全局有效。同时，RIP 采用老化机制对超时的路由进行老化处理，以保证路由的实时性和有效性。

1.1.3 RIP 的版本

RIP 有两个版本：RIP-1 和 RIP-2。

RIP-1 是有类别路由协议（Classful Routing Protocol），它只支持以广播方式发布协议报文。RIP-1 的协议报文无法携带掩码信息，它只能识别 A、B、C 类这样的自然网段的路由，因此 RIP-1 不支持不连续子网（Discontiguous Subnet）。

RIP-2 是一种无类别路由协议（Classless Routing Protocol），与 RIP-1 相比，它有以下优势：

- 支持路由标记，在路由策略中可根据路由标记对路由进行灵活的控制。
- 报文中携带掩码信息，支持路由聚合和 CIDR（Classless Inter-Domain Routing，无类域间路由）。
- 支持指定下一跳，在广播网上可以选择到最优下一跳地址。
- 支持组播路由发送更新报文，减少资源消耗。
- 支持对协议报文进行验证，并提供明文验证和 MD5 验证两种方式，增强安全性。

说明

RIP-2 有两种报文传送方式：广播方式和组播方式，缺省将采用组播方式发送报文，使用的组播地址为 224.0.0.9。当接口运行 RIP-2 广播方式时，也可接收 RIP-1 的报文。

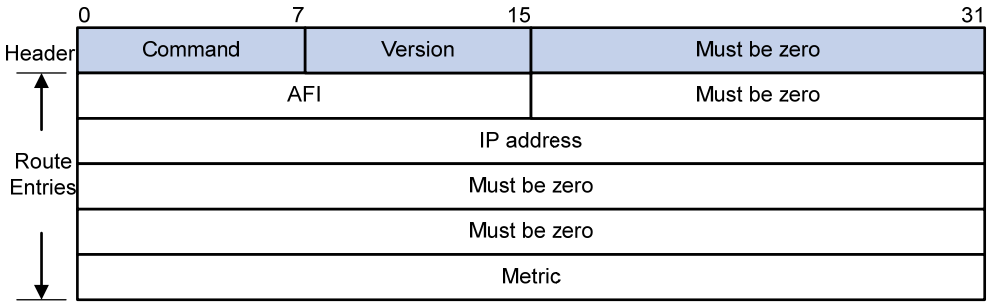
1.1.4 RIP 的报文格式

1. RIP-1 的报文格式

RIP 报文由头部（Header）和多个路由表项（Route Entries）部分组成。在一个 RIP 报文中，最多可以有 25 个路由表项（如果是 RIP-2 的验证报文，由于第一个路由表项作为验证项，所以最多可以有 24 个路由表项）。

RIP-1 的报文格式如 图 1-1所示。

图1-1 RIP-1 的报文格式



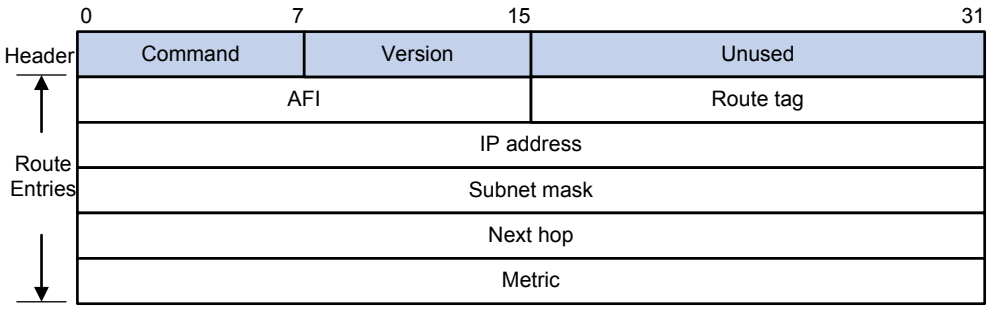
各字段的解释如下：

- Command：标识报文的类型。值为 1 时表示 Request 报文，向邻居请求全部或部分路由信息；值为 2 表示 Response 报文，发送全部或部分路由信息，1 个 Response 报文中最多包含 25 个路由表项。
- Version：RIP 的版本号。对于 RIP-1 来说其值为 0x01。
- Must be zero：必须为 0 字段。
- AFI（Address Family Identifier）：地址族标识，其值为 2 时表示 IP 协议。对于 Request 报文，此字段为 0。
- IP Address：该路由的目的 IP 地址，可以是自然网段地址、子网地址或主机地址。
- Metric：路由的度量值。对于 Request 报文，此字段为 16。

2. RIP-2 的报文格式

RIP-2 的报文格式与RIP-1 类似，如 图 1-2所示。

图1-2 RIP-2 的报文格式



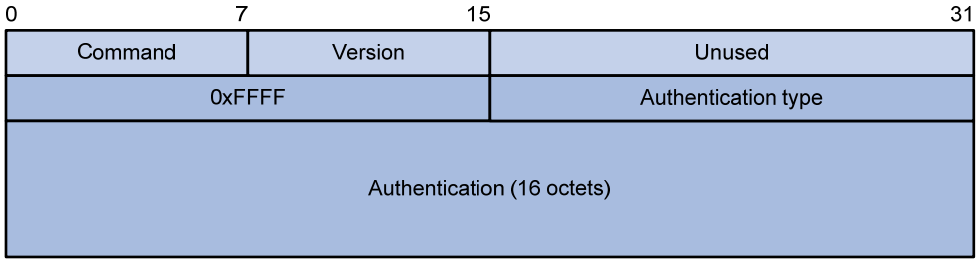
其中，与 RIP-1 不同的字段有：

- Version：RIP 的版本号。对于 RIP-2 来说其值为 0x02。
- Route Tag：路由标记。
- IP Address：该路由的目的 IP 地址，可以是自然网段地址、子网地址或主机地址。
- Subnet Mask：目的地址的掩码。
- Next Hop：如果为 0.0.0.0，则表示发布此条路由信息的路由器地址就是最优下一跳地址，否则表示提供了一个比发布此条路由信息的路由器更优的下一条地址。

3. RIP-2 的验证报文格式

RIP-2 为了支持报文验证，使用第一个路由表项（Route Entry）作为验证项，并将AFI字段的值设为 0xFFFF标识报文携带认证信息，如 图 1-3所示。

图1-3 RIP-2 的验证报文格式



各字段的解释如下：

- Authentication Type：验证类型。值为 2 时表示明文验证，值为 3 时表示 MD5 验证。
- Authentication：验证字，当使用明文验证时包含了密码信息；当使用 MD5 验证时包含了 Key ID、MD5 验证数据长度和序列号的信息。

说明

- RFC1723 中只定义了明文验证方式，关于 MD5 验证的详细信息，请参考 RFC2453 “RIP Version 2” 。
- 当 RIP 的版本为 RIP-1 时，虽然在接口视图下仍然可以配置验证方式，但由于 RIP-1 不支持认证，因此该配置不会生效。

1.1.5 支持的 RIP 特性

目前设备支持以下 RIP 特性：

- 支持 RIP-1 和 RIP-2。
- 支持 RIP 多实例。它可以作为 VPN 内部路由协议，在 BGP/MPLS VPN 网络中的 CE 和 PE 之间运行，相关概念及应用请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

1.1.6 协议规范

与 RIP 相关的协议规范有：

- RFC1058：Routing Information Protocol
- RFC1723：RIP Version 2 - Carrying Additional Information
- RFC1721：RIP Version 2 Protocol Analysis
- RFC1722：RIP Version 2 Protocol Applicability Statement
- RFC1724：RIP Version 2 MIB Extension
- RFC2082：RIP-2 MD5 Authentication
- RFC2091：Triggered Extensions to RIP to Support Demand Circuits
- RFC2453：RIP Version 2

1.2 配置 RIP 的基本功能

1.2.1 配置准备

在配置 RIP 的基本功能之前，需完成以下任务：

- 配置链路层协议
- 配置接口的网络层地址，使相邻节点的网络层可达

1.2.2 配置 RIP 的基本功能

1. 启动 RIP，配置指定网段范围内的接口运行 RIP

表1-1 启动 RIP，配置指定的接口运行 RIP

操作	命令	说明
进入系统视图	**system-view**	-
创建RIP进程并进入RIP视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，RIP 进程处于关闭状态 **vpn-instance** *vpn-instance-name* 参数的支持情况与设备的型号有关，请以设备的实际情况为准
在指定网段接口上使能RIP	**network** *network-address*	必选 缺省情况下，接口上的 RIP 功能处于关闭状态

说明

- 如果在启动 RIP 前在接口视图下配置了 RIP 相关命令，这些配置只有在 RIP 启动后才会生效。
- RIP 只在指定网段的接口上运行；对于不在指定网段上的接口，RIP 既不在它上面接收和发送路由，也不将它的接口路由转发出去。因此，RIP 启动后必须指定其工作网段。
- **network** 0.0.0.0 命令用来在所有接口上使能 RIP。
- RIP 不支持将同一物理接口下的不同网段使能到不同的 RIP 进程中。

2. 配置接口的工作状态

用户可对接口的工作状态进行配置：

- 配置接口工作在抑制状态，即接口只接收路由更新报文而不发送路由更新报文；
- 配置接口接收 RIP 报文；
- 配置接口发送 RIP 报文。

表1-2 配置接口的工作状态

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置接口工作在抑制状态	**silent-interface** { *interface-type interface-number* |**all** }	可选 缺省情况下，所有使能 RIP 的接口发送路由更新报文
退回系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
允许接口接收 RIP 报文	**rip input**	可选 缺省情况下，所有使能 RIP 的接口接收 RIP 报文
允许接口发送 RIP 报文	**rip output**	可选 缺省情况下，所有使能 RIP 的接口发送 RIP 报文

3. 配置 RIP 版本

用户可以在 RIP 视图下配置 RIP 版本，也可在接口上配置 RIP 版本：

- 当全局和接口都没有进行 RIP 版本配置时，接口发送 RIP-1 广播报文，可以接收 RIP-1 广播报文、RIP-1 单播报文、RIP-2 广播报文、RIP-2 组播报文、RIP-2 单播报文。
- 只有在接口上没有进行 RIP 版本配置时，全局配置才会生效，否则将以接口配置的为准。
- 当接口运行的 RIP 版本为 RIP-1 时，发送 RIP-1 广播报文，可以接收 RIP-1 广播报文、RIP-1 单播报文。
- 当接口运行的 RIP 版本为 RIP-2 且工作在组播方式时，发送 RIP-2 组播报文，可以接收 RIP-2 单播报文、RIP-2 广播报文、RIP-2 组播报文。
- 当接口运行的 RIP 版本为 RIP-2 且工作在广播方式时，发送 RIP-2 广播报文，可以接收 RIP-1 广播报文、RIP-1 单播报文、RIP-2 广播报文、RIP-2 组播报文、RIP-2 单播报文。

表1-3 配置 RIP 版本号

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置全局 RIP 版本	**version** { **1** | **2** }	可选 缺省情况下，如果接口配置了 RIP 版本，以接口配置的为准，如果接口也没有配置，接口只能发送 RIP-1 广播报文，可以接收 RIP-1 广播报文、RIP-1 单播报文、RIP-2 广播报文、RIP-2 组播报文、RIP-2 单播报文
退回系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口运行的 RIP 版本	**rip version** { **1** | **2** [**broadcast** | **multicast**] }	可选

1.3 配置 RIP 路由特性

在实际应用中，有时候需要对 RIP 路由信息进行更为精确的控制以满足复杂网络环境的需要。在配置 RIP 路由特性之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 配置 RIP 基本功能

1.3.1 配置接口附加度量值

附加度量值是在 RIP 路由原来度量值的基础上所增加的度量值（跳数），包括发送附加度量值和接收附加度量值。发送附加度量值不会改变路由表中的路由度量值，仅当接口发送 RIP 路由信息时才会添加到发送路由上；接收附加度量值会影响接收到的路由度量值，接口接收到一条合法的 RIP 路由时，在将其加入路由表前会把度量值附加到该路由上，当附加度量值与原路由度量值之和大于 16，该条路由的度量值取 16。

表1-4 配置接口附加度量值

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口接收 RIP 路由时的附加度量值	**rip metricin** [**route-policy** *route-policy-name*] *value*	可选 缺省情况下，接口接收 RIP 路由时的附加度量值为 0
配置接口发送 RIP 路由时的附加度量值	**rip metricout** [**route-policy** *route-policy-name*] *value*	可选 缺省情况下，接口发送 RIP 路由时的附加路由度量值为 1

1.3.2 配置 RIP-2 路由聚合

路由聚合是指将同一自然网段内的不同子网的路由聚合成一条自然掩码的路由向外（其它网段）发送，目的是为了减小路由表的规模，从而减少网络上的流量。

1. 配置 RIP-2 自动路由聚合功能

RIP-2 支持路由聚合。

需要注意的是，当需要将所有子网路由广播出去时，需要关闭 RIP-2 的自动路由聚合功能。

表1-5 配置 RIP-2 自动路由聚合功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能 RIP-2 自动路由聚合功能	**summary**	可选 缺省情况下，RIP-2 自动路由聚合功能处于使能状态

2. 配置发布一条聚合路由

用户可在指定接口配置 RIP-2 发布一条聚合路由。

表1-6 配置发布一条聚合路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
关闭 RIP-2 自动路由聚合功能	**undo summary**	必选 缺省情况下，RIP-2 自动路由聚合功能处于使能状态
退至系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置发布一条聚合路由	**rip summary-address** *ip-address* { *mask* | *mask-length* }	必选



说明

在接口配置发布一条聚合路由时必须关闭 RIP-2 的自动路由聚合功能。

1.3.3 禁止 RIP 接收主机路由

在某些特殊情况下，路由器会收到大量来自同一网段的主机路由。这些路由对于路由寻址没有多少作用，却占用了大量的资源；此时可配置 RIP 禁止接收主机路由，以节省网络资源。

表1-7 禁止 RIP 接收主机路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
禁止 RIP 接收主机路由	**undo host-route**	必选 缺省情况下，允许 RIP 接收主机路由

说明

禁止接收主机路由仅对 RIPv2 有效，对 RIPv1 无效。

1.3.4 配置 RIP 发布缺省路由

用户可以配置 RIP 以指定度量值向邻居发布一条缺省路由。

- 用户可以在 RIP 视图下配置 RIP 进程的所有接口向邻居发布缺省路由，也可以在接口下配置指定 RIP 接口向邻居发布缺省路由。
- 如果接口没有进行发布缺省路由的相关配置，则以 RIP 进程下的配置为准，否则将以接口配置为准。
- 如果 RIP 进程配置了发布缺省路由，但希望该进程下的某个接口不发送缺省路由（只发布普通路由），可以通过在接口下配置 **rip default-route no-originate** 命令实现。

表1-8 配置 RIP 发布缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 发布缺省路由	**default-route** { **only** | **originate** } [**cost** *cost*]	可选 缺省情况下 RIP 不向邻居发送缺省路由
退回系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 RIP 接口发布缺省路由	**rip default-route** { { **only** | **originate** } [**cost** *cost*] | **no-originate** }	可选 缺省情况下，RIP 接口是否发布缺省路由以 RIP 进程配置的为准

说明

配置发布缺省路由的 RIP 路由器不接收来自 RIP 邻居的缺省路由。

1.3.5 配置 RIP 对接收/发布的路由进行过滤

路由器提供路由信息过滤功能，通过指定访问控制列表和地址前缀列表，可以配置入口或出口过滤策略，对接收和发布的路由进行过滤。在接收路由时，还可以指定只接收来自某个邻居的 RIP 报文。

表1-9 配置 RIP 对接收/发布的路由进行过滤

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **gateway** *ip-prefix-name* | **ip-prefix** *ip-prefix-name* [**gateway** *ip-prefix-name*] } **import** [*interface-type interface-number*]	必选 缺省情况下，RIP 不对接收的路由信息进行过滤
对发布的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*] | *interface-type interface-number*]	必选 缺省情况下，RIP 不对发布的路由信息进行过滤

说明

- **filter-policy import** 命令对从邻居收到的 RIP 路由进行过滤，没有通过过滤的路由将不被加入路由表，也不向邻居发布该路由。
- **filter-policy export** 命令对本机所有路由的发布进行过滤，包括使用 **import-route** 引入的路由和从邻居学到的 RIP 路由。

1.3.6 配置 RIP 协议优先级

在路由器中可能会运行多个 IGP 路由协议，如果想让 RIP 路由具有比从其它路由协议学来的路由更高的优先级，需要配置小的优先级值。优先级的高低将最后决定 IP 路由表中的路由是通过哪种路由算法获取的最佳路由。

表1-10 配置 RIP 协议优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 路由的优先级	**preference** [**route-policy** *route-policy-name*] *value*	可选 缺省情况下，RIP 路由的优先级为 100

1.3.7 配置 RIP 引入外部路由

如果在路由器上不仅运行 RIP，还运行着其它路由协议，可以配置 RIP 引入其它协议生成的路由，如 OSPF、ISIS、BGP、静态路由或者直连路由。

表1-11 配置 RIP 引入外部路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置引入路由的缺省度量值	**default cost** *value*	可选 缺省情况下，引入路由的缺省度量值为 0

操作	命令	说明
引入外部路由	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	必选 缺省情况下，RIP 不引入其它路由

说明

只能引入路由表中状态为 active 的路由，是否为 active 状态可以通过 **display ip routing-table protocol** 命令来查看。

1.4 调整和优化 RIP 网络

在某些特殊的网络环境中，需要对 RIP 网络的性能进行调整和优化，在调整 RIP 之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点网络层可达
- 配置 RIP 基本功能

1.4.1 配置 RIP 定时器

表1-12 配置 RIP 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 定时器的值	**timers** { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *	可选 缺省情况下： • Update 定时器的值为 30 秒 • Timeout 定时器的值为 180 秒 • Suppress 定时器的值为 120 秒 • Garbage-collect 定时器的值为 120 秒

说明

在配置 RIP 定时器时需要注意，定时器值的调整应考虑网络的性能，并在所有运行 RIP 的路由器上进行统一配置，以免增加不必要的网络流量或引起网络路由震荡。

1.4.2 配置水平分割和毒性逆转

说明

如果同时配置了水平分割和毒性逆转，则只有毒性逆转功能生效。

1. 配置水平分割

配置水平分割可以使得从一个接口学到的路由不能通过此接口向外发布，用于避免相邻路由器间的路由环路。

表1-13 配置水平分割

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能水平分割功能	**rip split-horizon**	可选 缺省情况下，水平分割功能处于使能状态



说明

- 在帧中继和 X.25 等 NBMA（Non-Broadcast Multi-Access，非广播多点可达）网络中，当主接口和点到多点子接口配置了多条虚电路时，为了保证路由信息的正确传播，需要关闭水平分割功能。关于帧中继和 X.25 的详细信息，请参见“接入分册”中的“帧中继配置”和“LAPB 和 X.25 配置”。
- 在点到点链路上关闭水平分割功能是无效的。

2. 配置毒性逆转

配置毒性逆转可以使得从一个接口学到的路由还可以从这个接口向外发布，但这些路由的度量值已设置为 16，即不可达。

表1-14 配置毒性逆转

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能毒性逆转功能	**rip poison-reverse**	必选 缺省情况下，毒性逆转功能处于关闭状态

1.4.3 配置最大等价路由条数

表1-15 配置最大等价路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 最大等价路由条数	**maximum load-balancing** *number*	可选 本命令的缺省值与设备的型号有关，请以设备的实际情况为准

1.4.4 配置 RIP-1 报文的零域检查

RIP-1 报文中的有些字段必须为零，称之为零域。用户可配置 RIP-1 在接收报文时对零域进行检查，零域值不为零的 RIP-1 报文将不被处理。如果用户能确保所有报文都是可信任的，则可以不进行该项检查，以节省 CPU 处理时间。

由于 RIP-2 的报文没有零域，此项配置对 RIP-2 无效。

表1-16 配置 RIP-1 报文的零域检查

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能对 RIP-1 报文的零域检查功能	**checkzero**	可选 缺省情况下，对 RIP-1 报文的零域检查功能处于使能状态

1.4.5 配置源地址检查

通过配置对接收到的 RIP 路由更新报文进行源 IP 地址检查：

- 对于在接口上接收的报文，RIP 将检查该报文源地址和接口的 IP 地址是否处于同一网段，如果不在同一网段则丢弃该报文。
- 对于串口上接收的报文，RIP 检查该报文的源地址是否是对端接口的 IP 地址，如果不是则丢弃该报文。

表1-17 配置源地址检查

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
使能对接收到的 RIP 路由更新报文进行源 IP 地址检查功能	**validate-source-address**	可选 缺省情况下，对接收到的 RIP 路由更新报文进行源 IP 地址检查功能处于使能状态



说明

当存在 RIP 非直连的邻居时，应该取消源地址检查。

1.4.6 配置 RIP-2 报文的认证方式

RIP-2 支持两种认证方式：明文认证和 MD5 密文认证。

明文认证不能提供安全保障，未加密的认证字随报文一同传送，所以明文认证不能用于安全性要求较高的情况。

表1-18 配置 RIP-2 报文的认证方式

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 RIP-2 报文的认证方式	**rip authentication-mode** { **md5** { **rfc2082** *key-string key-id* | **rfc2453** *key-string* } | **simple** *password* }	必选



说明

当 RIP 的版本为 RIP-1 时，虽然在接口视图下仍然可以配置验证方式，但由于 RIP-1 不支持认证，因此该配置不会生效。

1.4.7 配置 RIP 邻居

通常情况下，RIP 使用广播或组播地址发送报文，如果在不支持广播或组播报文的链路上运行 RIP，则必须手工指定 RIP 的邻居。

表1-19 配置 RIP 邻居

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 邻居	**peer** *ip-address*	必选
取消对接收到的 RIP 路由更新报文进行源 IP 地址检查操作	**undo validate-source-address**	必选 缺省情况下，对接收到的 RIP 路由更新报文进行源 IP 地址检查



说明

- 当 RIP 邻居与当前设备直连时不推荐使用 **peer** *ip-address* 命令，因为这样可能会造成对端同时收到同一路由信息的组播（或广播）和单播两种形式的报文。
- 当指定的邻居和本地路由器非直接连接，则必须取消对更新报文的源地址进行检查。

1.4.8 配置 RIP 和 MIB 绑定

通过将 MIB 操作绑定在指定的 RIP 进程上，可以让指定的 RIP 进程接收 SNMP 请求。

表1-20 配置 RIP 和 MIB 绑定

操作	命令	说明
进入系统视图	**system-view**	-
配置 RIP 和 MIB 绑定	**rip mib-binding** *process-id*	可选 缺省情况下，MIB 操作绑定在 RIP 进程 1 上

1.4.9 配置 RIP 报文的发送速率

RIP 周期性地将路由信息放在 RIP 报文中向邻居发送。

如果路由表里的路由条目数量很多，同时发送大量 RIP 协议报文有可能会对当前设备和网络带宽带来冲击；因此，路由器将 RIP 协议报文分为多个批次进行发送，并且对 RIP 接口每次允许发送的 RIP 协议报文最大个数做出限制。

用户可根据需要配置接口发送 RIP 报文的时间间隔以及接口一次发送 RIP 报文的最大个数。

表1-21 配置 RIP 报文的发送速率

操作	命令	说明
进入系统视图	**system-view**	-
启动 RIP 并进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	-
配置 RIP 报文的发送速率	**output-delay** *time* **count** *count*	可选 缺省情况下，接口发送 RIP 报文的时间间隔为 20 毫秒，一次最多发送 3 个 RIP 报文

1.5 RIP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 RIP 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以重启 RIP 进程或清除指定 RIP 进程的统计信息。

表1-22 RIP 显示和维护

操作	命令
显示 RIP 的当前运行状态及配置信息	**display rip** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 RIP 发布数据库的所有激活路由	**display rip** *process-id* **database**
显示 RIP 的接口信息	**display rip** *process-id* **interface** [*interface-type interface-number*]
显示指定 RIP 进程的路由信息	**display rip** *process-id* **route** [*ip-address* { *mask* | *mask-length* } | **peer** *ip-address* |**statistics**]
重启 RIP 进程	reset rip process-id process
清除指定 RIP 进程维护的计数器的统计信息	**reset rip** *process-id* **statistics**

1.6 RIP 典型配置举例

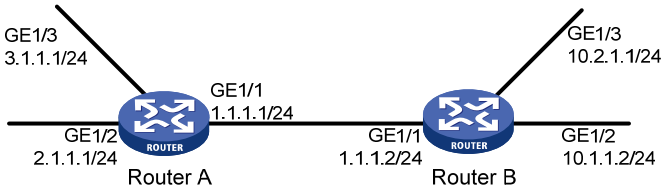
1.6.1 配置 RIP 的版本

1. 组网需求

如 图 1-4所示，要求在Router A和Router B的所有接口上使能RIP，并使用RIP-2 进行网络互连。

2. 组网图

图1-4 配置 RIP 版本组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 使能 RIP 功能

配置 Router A。

```
<RouterA> system-view
[RouterA] rip
[RouterA-rip-1] network 1.0.0.0
[RouterA-rip-1] network 2.0.0.0
[RouterA-rip-1] network 3.0.0.0
```

配置 Router B。

```
<RouterB> system-view
[RouterB] rip
[RouterB-rip-1] network 1.0.0.0
[RouterB-rip-1] network 10.0.0.0
```

查看 Router A 的 RIP 路由表。

```
[RouterA] display rip 1 route
Route Flags: R - RIP, T - TRIP
             P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
 Peer 1.1.1.2  on GigabitEthernet 1/1
      Destination/Mask        Nexthop     Cost    Tag    Flags    Sec
         10.0.0.0/8            1.1.1.2      1       0     RA       9
```

从路由表中可以看出，RIP-1 发布的路由信息使用的是自然掩码。

(3) 配置 RIP 的版本

在 Router A 上配置 RIP-2。

```
[RouterA] rip
[RouterA-rip-1] version 2
[RouterA-rip-1] undo summary
```

在 Router B 上配置 RIP-2。

```
[RouterB] rip
[RouterB-rip-1] version 2
[RouterB-rip-1] undo summary
```

查看 Router A 的 RIP 路由表。

```
[RouterA] display rip 1 route
Route Flags: R - RIP, T - TRIP
             P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
 Peer 1.1.1.2  on GigabitEthernet 1/1
      Destination/Mask        Nexthop      Cost    Tag    Flags    Sec
         10.0.0.0/8           1.1.1.2        1       0     RA       87
         10.1.1.0/24          1.1.1.2        1       0     RA       19
         10.2.1.0/24          1.1.1.2        1       0     RA       19
```

从路由表中可以看出，RIP-2 发布的路由中带有更为精确的子网掩码信息。

说明

由于 RIP 路由信息的老化时间较长，所以在配置 RIP-2 版本后的一段时间里，路由表中还会存在 RIP-1 的路由信息。

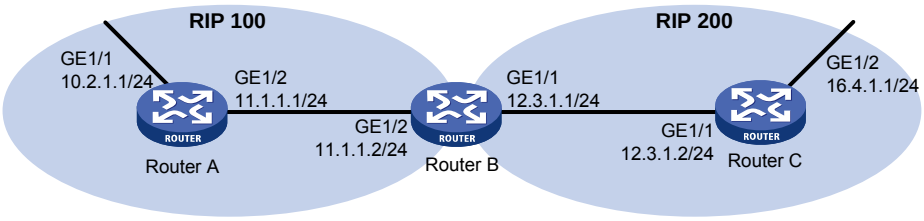
1.6.2 配置 RIP 引入外部路由

1. 组网需求

- Router B 上运行两个 RIP 进程：RIP100 和 RIP200。Router B 通过 RIP100 和 Router A 交换路由信息，通过 RIP200 和 Router C 交换路由信息。
- 在 Router B 上配置 RIP 进程 200 引入外部路由，引入直连路由和 RIP 进程 100 的路由，使得 Router C 能够学习到达 10.1.1.0/24 和 11.1.1.0/24 的路由，但 Router A 不能学习到达 12.3.1.0/24 和 16.4.1.0/24 的路由。
- 在 Router B 配置过滤策略，对引入的 RIP100 的一条路由（10.2.1.1/24）进行过滤，使其不发布给 Router C。

2. 组网图

图1-5 配置 RIP 引入外部路由组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 RIP 基本功能

在 Router A 上启动 RIP 进程 100，并配置 RIP 版本号为 2。

```
<RouterA> system-view
[RouterA] rip 100
[RouterA-rip-100] network 10.0.0.0
[RouterA-rip-100] network 11.0.0.0
```

```
[RouterA-rip-100] version 2
[RouterA-rip-100] undo summary
[RouterA-rip-100] quit
```

在 Router B 上启动两个 RIP 进程，进程号分别为 100 和 200，并配置 RIP 版本号为 2。

```
<RouterB> system-view
[RouterB] rip 100
[RouterB-rip-100] network 11.0.0.0
[RouterB-rip-100] version 2
[RouterB-rip-100] undo summary
[RouterB-rip-100] quit
[RouterB] rip 200
[RouterB-rip-200] network 12.0.0.0
[RouterB-rip-200] version 2
[RouterB-rip-200] undo summary
[RouterB-rip-200] quit
```

在 Router C 上启动 RIP 进程 200，并配置 RIP 版本号为 2。

```
<RouterC> system-view
[RouterC] rip 200
[RouterC-rip-200] network 12.0.0.0
[RouterC-rip-200] network 16.0.0.0
 [RouterC-rip-200] version 2
[RouterC-rip-200] undo summary
```

查看 Router C 的路由表信息。

```
[RouterC] display ip routing-table
Routing Tables: Public
         Destinations : 6        Routes : 6
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
12.3.1.0/24         Direct 0    0            12.3.1.2        GE1/1
12.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0            16.4.1.1        GE1/2
16.4.1.1/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

(3) 配置 RIP 引入外部路由

在 Router B 配置 RIP 进程 200 引入外部路由，引入直连路由和 RIP 进程 100 的路由。

```
[RouterB] rip 200
[RouterB-rip-200] import-route rip 100
[RouterB-rip-200] import-route direct
[RouterB-rip-200] quit
```

查看路由引入后 Router C 的路由表信息。

```
[RouterC] display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         RIP    100  1            12.3.1.1        GE1/1
```

```
11.1.1.0/24         RIP    100  1           12.3.1.1        GE1/1
12.3.1.0/24         Direct 0    0           12.3.1.2        GE1/1
12.3.1.2/32         Direct 0    0           127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0           16.4.1.1        GE1/2
16.4.1.1/32         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
```

(4) 配置 RIP 对引入的路由进行过滤

在 Router B 上配置 ACL，并对引入的 RIP 进程 100 的路由进行过滤，使其不发布给 Router C。

```
[RouterB] acl number 2000
[RouterB-acl-basic-2000] rule deny source 10.2.1.1 0.0.0.255
[RouterB-acl-basic-2000] rule permit
[RouterB-acl-basic-2000] quit
[RouterB] rip 200
[RouterB-rip-200] filter-policy 2000 export rip 100
```

查看过滤后 Router C 的路由表。

```
[RouterC] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

11.1.1.0/24         RIP    100  1           12.3.1.1        GE1/1
12.3.1.0/24         Direct 0    0           12.3.1.2        GE1/1
12.3.1.2/32         Direct 0    0           127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0           16.4.1.1        GE1/2
16.4.1.1/32         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
```

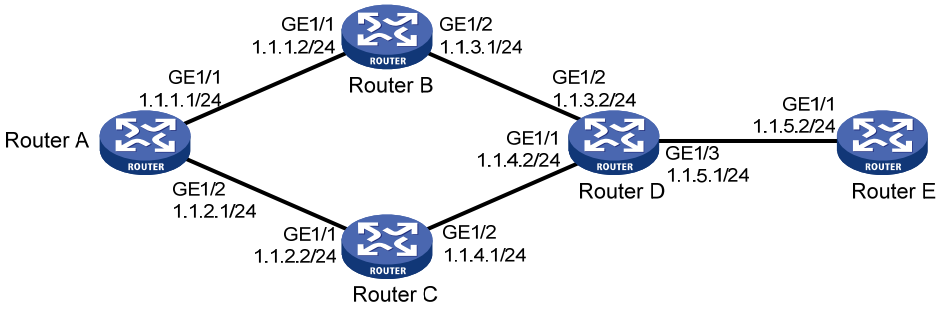
1.6.3 配置 RIP 接口附加度量值

1. 组网需求

- 在 Router A、Router B、Router C、Router D 和 Router E 的所有接口上使能 RIP，并使用 RIP-2 进行网络互连。
- Router A 有两条链路可以到达 Router D，其中，通过 Router B 到达 Router D 的链路比通过 Router C 到达 Router D 的链路更加稳定。通过在 Router A 的 GigabitEthernet1/2 上配置接口接收 RIP 路由的附加度量值，使得 Router A 优选从 Router B 学到的 1.1.5.0/24 网段的路由。

2. 组网图

图1-6 配置 RIP 接口附加度量值



3. 配置步骤

(1) 配置各接口的地址（略）

(2) 配置 RIP 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] rip
[RouterA-rip-1] network 1.0.0.0
[RouterA-rip-1] version 2
[RouterA-rip-1] undo summary
[RouterA-rip-1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] rip
[RouterB-rip-1] network 1.0.0.0
[RouterB-rip-1] version 2
[RouterB-rip-1] undo summary
```

配置 Router C。

```
<RouterC> system-view
[RouterB] rip
[RouterC-rip-1] network 1.0.0.0
[RouterC-rip-1] version 2
[RouterC-rip-1] undo summary
```

配置 Router D。

```
<RouterD> system-view
[RouterD] rip
[RouterD-rip-1] network 1.0.0.0
[RouterD-rip-1] version 2
[RouterD-rip-1] undo summary
```

配置 Router E。

```
<RouterE> system-view
[RouterE] rip
[RouterE-rip-1] network 1.0.0.0
```

```
[RouterE-rip-1] version 2
[RouterE-rip-1] undo summary
```

查看 Router A 的 IP 路由表。

```
[RouterA] display rip 1 database
   1.0.0.0/8, cost 0, ClassfulSumm
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, Rip-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, Rip-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 1, nexthop 1.1.2.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.2.2
```

可以看到，到达网段 1.1.5.0/24 有两条 RIP 路由，下一跳分别是 Router B（IP 地址为 1.1.1.2）和 Router C（IP 地址为 1.1.2.2），cost 值都是 2；到达网段 1.1.4.0/24 的下一跳是 Router B，cost 为 1。

(3) 配置 RIP 接口附加度量值

在 Router A 上配置接口 GigabitEthernet1/2 的接口附加度量值为 3。

```
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] rip metricin 3
[RouterA-GigabitEthernet1/2] display rip 1 database
   1.0.0.0/8, cost 0, ClassfulSumm
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, Rip-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, Rip-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
```

可以看到，到达网段 1.1.5.0/24 的 RIP 路由仅有一条，下一跳是 Router B（IP 地址为 1.1.1.2），cost 值为 2。

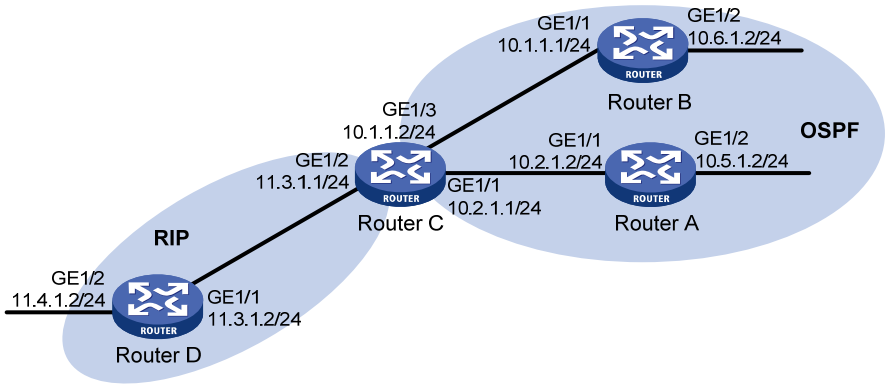
1.6.4 配置 RIP 发布聚合路由

1. 组网需求

- Router A、Router B 运行 OSPF，Router D 运行 RIP，Router C 同时运行 OSPF 和 RIP。
- 在 Router C 上配置 RIP 进程引入 OSPF 路由，使 Router D 有到达 10.1.1.0/24、10.2.1.0/24、10.5.1.0/24 和 10.6.1.0/24 网段的路由。
- 为了减小 Router D 的路由表规模，在 Router C 上配置路由聚合，只发布聚合后的路由 10.0.0.0/8。

2. 组网图

图1-7 配置 RIP 发布聚合路由



3. 配置步骤

(1) 配置各接口的地址（略）

(2) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] network 10.6.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf
[RouterC-ospf-1] area 0
[RouterC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] quit
```

(3) 配置 RIP 基本功能

配置 Router C。

```
<RouterC> system-view
[RouterC] rip 1
[RouterC-rip-1] network 11.3.1.0
[RouterC-rip-1] version 2
```

```
[RouterC-rip-1] undo summary
```

配置 Router D。

```
<RouterD> system-view
[RouterD] rip 1
[RouterD-rip-1] network 11.0.0.0
[RouterD-rip-1] version 2
[RouterD-rip-1] undo summary
[RouterD-rip-1] quit
```

在 Router C 上配置 RIP 引入外部路由，引入 OSPF 进程 1 的路由和直连路由。

```
[RouterC-rip-1] import-route direct
[RouterC-rip-1] import-route ospf 1
```

查看 Router D 的路由表信息。

```
[RouterD] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         RIP    100  1            11.3.1.1        GE1/1
10.2.1.0/24         RIP    100  1            11.3.1.1        GE1/1
10.5.1.0/24         RIP    100  1            11.3.1.1        GE1/1
10.6.1.0/24         RIP    100  1            11.3.1.1        GE1/1
11.3.1.0/24         Direct 0    0            11.3.1.2        GE1/1
11.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0            11.4.1.2        GE1/2
11.4.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

(4) 在 Router C 上配置路由聚合，只发布聚合路由 10.0.0.0/8。

```
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] rip summary-address 10.0.0.0 8
```

查看 Router D 的路由表信息。

```
[RouterD] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.0.0.0/8          RIP    100  1            11.3.1.1        GE1/1
11.3.1.0/24         Direct 0    0            11.3.1.2        GE1/1
11.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0            11.4.1.2        GE1/2
11.4.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

1.7 常见配置错误举例

1.7.1 收不到邻居的 RIP 更新报文

1. 故障现象

在链路正常的情况下收不到邻居的 RIP 更新报文。

2. 故障分析

RIP 启动后，必须使用 **network** 命令使能相应的接口。如果对接口的工作状态单独进行了配置，应确认没有抑制相关接口或禁止其收发 RIP 报文。

如果在对端路由器上配置的是组播方式发送 RIP 报文，在本地路由器上也应该配置为组播方式。

3. 故障排除

(1) 执行 **display current-configuration** 命令，检查 RIP 的配置。

(2) 执行 **display rip** 命令，检查是否抑制了相关 RIP 接口。

1.7.2 RIP 网络发生路由振荡

1. 故障现象

在链路正常的情况下，运行 RIP 的网络发生路由振荡，查看路由表时发现部分路由时有时无。

2. 故障分析

在 RIP 网络中，应确保全网定时器的配置一致，且各定时器之间的关系合理，如 Timeout 定时器的值应大于 Update 定时器的值。

3. 故障排除

(1) 使用 **display rip** 命令查看 RIP 定时器的配置

(2) 使用 **timers** 将全网的定时器配置一致

目 录

1 路由策略配置 1-1
1.1 路由策略简介 1-1
1.1.1 路由策略与策略路由 1-1
1.1.2 过滤器 1-1
1.1.3 路由策略的应用 1-2
1.2 配置过滤列表 1-2
1.2.1 配置准备 1-2
1.2.2 配置地址前缀列表 1-2
1.2.3 配置AS路径过滤列表 1-4
1.2.4 配置团体属性列表 1-4
1.2.5 配置扩展团体属性列表 1-4
1.3 配置路由策略 1-4
1.3.1 配置准备 1-5
1.3.2 创建一个路由策略 1-5
1.3.3 配置if-match子句 1-5
1.3.4 配置apply子句 1-7
1.4 路由策略的显示和维护 1-8
1.5 路由策略典型配置举例 1-9
1.5.1 在IPv4 路由引入中应用路由策略 1-9
1.5.2 在IPv6 路由引入中应用路由策略 1-11
1.5.3 配置应用路由策略过滤接收的BGP路由 1-13
1.6 常见错误配置举例 1-15
1.6.1 无法实现IPv4 路由信息过滤 1-15
1.6.2 无法实现IPv6 路由信息过滤 1-16

1 路由策略配置

1.1 路由策略简介

1.1.1 路由策略与策略路由

路由策略（Routing Policy）是为了改变网络流量所经过的途径而修改路由信息的技术，主要通过改变路由属性（包括可达性）来实现。

策略路由（Policy Routing）是一种依据用户制定的策略进行路由选择的机制。有关策略路由的详细介绍请参见“IP 业务分册”中的“IP 单播策略路由配置”。

路由器在发布与接收路由信息时，可能需要实施一些策略，以便对路由信息进行过滤，例如只接收或发布满足一定条件的路由信息。一种路由协议可能需要引入其它的路由协议发现的路由信息，路由器在引入其它路由协议的路由信息时，可能只需要引入一部分满足条件的路由信息，并控制所引入的路由信息的某些属性，以使其满足本协议的要求。

为实现路由策略，首先要定义将要实施路由策略的路由信息的特征，即定义一组匹配规则。可以以路由信息中的不同属性作为匹配依据进行设置，如目的地址、发布路由信息的路由器地址等。匹配规则可以预先设置好，然后再将它们应用于路由的发布、接收和引入等过程的路由策略中。

1.1.2 过滤器

路由协议可以引用访问控制列表、地址前缀列表、AS 路径访问列表、团体属性列表、扩展团体属性列表和 Route-policy 几种过滤器。下面对各种过滤器逐一进行介绍。

1. 访问控制列表

访问控制列表包括针对 IPv4 报文的 ACL 和针对 IPv6 报文的 ACL。用户在定义 ACL 时可以指定 IP(v6) 地址和子网范围，用于匹配路由信息的目的网段地址或下一跳地址。ACL 的有关配置请参见“安全分册”中的“ACL 配置”。

2. 地址前缀列表

地址前缀列表包括 IPv4 地址前缀列表和 IPv6 地址前缀列表。

地址前缀列表的作用类似于 ACL，但比它更为灵活，且更易于用户理解。使用地址前缀列表过滤路由信息时，其匹配对象为路由信息的目的地址信息域；另外，用户可以指定 **gateway** 选项，指明只接收某些路由器发布的路由信息。关于 **gateway** 选项的设置请参见“IP 路由分册”中的“RIP 命令”和“OSPF 命令”。

一个地址前缀列表由前缀列表名标识。每个前缀列表可以包含多个表项，每个表项可以独立指定一个网络前缀形式的匹配范围，并用一个索引号来标识，索引号指明了在地址前缀列表中进行匹配检查的顺序。

每个表项之间是“或”的关系，在匹配的过程中，路由器按升序依次检查由索引号标识的各个表项，只要有某一表项满足条件，就意味着通过该地址前缀列表的过滤（不再进入下一个表项的测试）。

3. AS 路径访问列表（as-path）

as-path 仅用于 BGP。BGP 的路由信息中，包含有自治系统路径域。as-path 就是针对自治系统路径域指定匹配条件。

4. 团体属性列表（community-list）

community-list 仅用于 BGP。BGP 的路由信息包中，包含一个 community 属性域，用来标识一个团体。community-list 就是针对团体属性域指定匹配条件。

5. 扩展团体属性列表（extcommunity-list）

extcommunity-list 仅用于 BGP。BGP 扩展团体属性有两种，一种是用于 VPN 的 Route-Target（路由目标）扩展团体，另一种则是 Source of Origin（源节点）扩展团体。扩展团体属性列表就是针对这两种属性指定匹配条件。

6. 路由策略（Route-policy）

路由策略是一种比较复杂的过滤器，它不仅可以匹配路由信息的某些属性，还可以在条件满足时改变路由信息的属性。路由策略可以使用前面几种过滤器定义自己的匹配规则。

一个路由策略可以由多个节点（node）构成，每个节点是匹配检查的一个单元，在匹配过程中，系统按节点序号升序依次检查各个节点。

每个节点可以由一组 **if-match** 和 **apply** 子句组成。**if-match** 子句定义匹配规则，匹配对象是路由信息的一些属性。同一节点中的不同 **if-match** 子句是“与”的关系，只有满足节点内所有 **if-match** 子句指定的匹配条件，才能通过该节点的匹配测试。**apply** 子句指定动作，也就是在通过节点的匹配后，对路由信息的一些属性进行设置。

一个路由策略的不同节点间是“或”的关系，如果通过了其中一节点，就意味着通过该路由策略，不再对其他节点进行匹配测试。

1.1.3 路由策略的应用

路由策略主要有两种应用方式：

- 路由协议在引入其它路由协议发现的路由时，通过路由策略只引入满足条件的路由信息。
- 路由协议在发布或接收路由信息时，通过路由策略对路由信息进行过滤，只接收或发布满足给定条件的路由信息。

1.2 配置过滤列表

1.2.1 配置准备

在配置过滤列表之前，需要准备以下数据：

- 前缀列表名称
- 匹配的地址范围
- 扩展团体属性列表序号

1.2.2 配置地址前缀列表

1. 配置 IPv4 地址前缀列表

IPv4 地址前缀列表由列表名标识，每个前缀列表可以包含多个表项。各表项以网络前缀的形式，独立指定一个匹配范围，并使用索引号标识。

在匹配过程中，系统按索引号升序依次检查各个表项，只要路由信息满足一个表项，就认为通过该过滤列表，不再去匹配其他表项。

表1-1 配置 IPv4 地址前缀列表

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv4 地址前缀列表	**ip ip-prefix** *ip-prefix-name* [**index** *index-number*] { **permit** | **deny** } *ip-address mask-length* [**greater-equal** *min-mask-length*] [**less-equal** *max-mask-length*]	必选 缺省情况下，没有配置 IPv4 地址前缀列表



说明

如果所有表项都是 **deny** 模式，则任何路由都不能通过该过滤列表。这种情况下，建议在多条 **deny** 模式的表项后定义一条 **permit** 0.0.0.0 0 **less-equal** 32 表项，允许其它所有 IPv4 路由信息通过。

例如，按如下配置可以保证仅过滤掉 10.1.0.0/16、10.2.0.0/16、10.3.0.0/16 三个网段的路由，而其它网段的路由信息可以通过。

```
<Sysname> system-view
[Sysname] ip ip-prefix abc index 10 deny 10.1.0.0 16
[Sysname] ip ip-prefix abc index 20 deny 10.2.0.0 16
[Sysname] ip ip-prefix abc index 30 deny 10.3.0.0 16
[Sysname] ip ip-prefix abc index 40 permit 0.0.0.0 0 less-equal 32
```

2. 配置 IPv6 地址前缀列表

IPv6 地址前缀列表由列表名标识，每个前缀列表可以包含多个表项。各表项可以独立指定一个网络前缀形式的匹配范围，并使用索引号标识。

在匹配的过程中，系统按索引号升序依次检查各个表项，只要路由信息满足一个表项，就认为通过该过滤列表，不再去匹配其他表项。

表1-2 配置 IPv6 地址前缀列表

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 地址前缀列表	**ip ipv6-prefix** *ipv6-prefix-name* [**index** *index-number*] { **deny** | **permit** } *ipv6-address prefix-length* [**greater-equal** *min-prefix-length*] [**less-equal** *max-prefix-length*]	必选 缺省情况下，没有配置 IPv6 地址前缀列表



说明

如果所有表项都是 **deny** 模式，则任何路由都不能通过该过滤列表。这种情况下，需要在多条 **deny** 模式的表项后定义一条 **permit** :: 0 **less-equal** 128 的表项，以允许其它所有 IPv6 路由信息通过。

例如，按如下配置可以保证仅过滤掉 2000:1::/48、2000:2::/48、2000:3::/48 三个网段的路由，而其它网段的路由信息可以通过。

```
<Sysname> system-view
[Sysname] ip ipv6-prefix abc index 10 deny 2000:1:: 48
[Sysname] ip ipv6-prefix abc index 20 deny 2000:2:: 48
```

```
[Sysname] ip ipv6-prefix abc index 30 deny 2000:3:: 48
[Sysname] ip ipv6-prefix abc index 40 permit :: 0 less-equal 128
```

1.2.3 配置 AS 路径过滤列表

一个 AS 过滤列表可以包含多个表项。在匹配过程中，各表项之间是“或”的关系，即只要路由信息通过该列表中的一条表项，就认为通过该 AS 路径过滤列表。

表1-3 配置 AS 路径过滤列表

操作	命令	说明
进入系统视图	**system-view**	-
配置 AS 路径过滤列表	**ip as-path** *as-path-number* { **deny** | **permit** } *regular-expression*	必选 缺省情况下，没有配置 AS 路径过滤列表

1.2.4 配置团体属性列表

一个团体属性列表可以定义多个表项。在匹配过程中，各表项之间是“或”的关系，即只要路由信息通过该列表中的一条表项，就认为通过该团体属性列表。

表1-4 配置团体属性列表

操作		命令	说明
进入系统视图		**system-view**	-
配置团体属性列表	配置基本团体属性列表	**ip community-list** *basic-comm-list-num* { **deny** | **permit** } [*community-number-list*] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] *	二者必选其一 缺省情况下，没有配置团体属性列表
	配置高级团体属性列表	**ip community-list** *adv-comm-list-num* { **deny** | **permit** } *regular-expression*	

1.2.5 配置扩展团体属性列表

一个扩展团体属性列表可以定义多个表项。在匹配过程中，各表项之间是“或”的关系，即只要路由信息通过该列表中的一条表项，就认为通过该扩展团体属性列表。

表1-5 配置扩展团体属性列表

操作	命令	说明
进入系统视图	**system-view**	-
配置扩展团体属性列表	**ip extcommunity-list** *ext-comm-list-number* { **deny** | **permit** } { **rt** *route-target* }&<1-16>	必选 缺省情况下，没有配置扩展团体属性列表

1.3 配置路由策略

路由策略用来根据路由信息的某些属性过滤路由信息，并改变与路由策略规则匹配的路由信息的属性。匹配条件可以使用前面几种过滤列表。

一个路由策略可由多个节点构成，每个节点又分为：

- if-match 子句：定义匹配规则，即路由信息通过当前 Route-policy 所需满足的条件，匹配对象是路由信息的某些属性。
- apply 子句：指定动作，也就是在满足由 if-match 子句指定的过滤条件后所执行的一些配置命令，对路由的某些属性进行修改。

1.3.1 配置准备

在配置路由策略之前，需完成以下任务：

- 配置过滤列表
- 配置路由协议

在配置之前，需要准备以下数据：

- 路由策略的名称、节点序号
- 匹配条件
- 要修改的路由属性值

1.3.2 创建一个路由策略

表1-6 创建一个路由策略

操作	命令	说明
进入系统视图	**system-view**	-
创建路由策略并进入该路由策略视图	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	必选 缺省情况下，没有创建路由策略

说明

- **permit** 指定节点的匹配模式为允许模式。当路由信息通过该节点的过滤后，将执行该节点的 **apply** 子句，不进入下一个节点的测试；如果路由信息没有通过该节点过滤，将进入下一个节点继续测试。
- **deny** 指定节点的匹配模式为拒绝模式（此模式下 **apply** 子句不会被执行）。当路由项满足该节点的所有 **if-match** 子句时，将被拒绝通过该节点，不进入下一个节点的测试；如果路由项不满足该节点的 **if-match** 子句，将进入下一个节点继续测试。
- 如果路由策略中定义了一个以上的节点，则各节点中至少应该有一个节点的匹配模式是 **permit**。当路由策略用于路由信息过滤时，如果某路由信息没有通过任一节点，则认为该路由信息没有通过该路由策略。如果路由策略的所有节点都是 **deny** 模式，则没有路由信息能通过该路由策略。

1.3.3 配置 if-match 子句

表1-7 配置 if-match 子句

操作	命令	说明
进入系统视图	**system-view**	-
进入路由策略视图	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	必选

操作		命令	说明
配置 IPv4 的路由信息的匹配条件	配置 IPv4 路由信息的目的地址范围的匹配条件	**if-match acl** *acl-number*	可选 缺省情况下，没有配置 IPv4 路由信息的目的地址范围的匹配条件
		if-match ip-prefix *ip-prefix-name*	
	配置 IPv4 的路由信息的下一跳地址或源地址的匹配条件	**if-match ip** { **next-hop** | **route-source** } { **acl** *acl-number* | **ip-prefix** *ip-prefix-name* }	可选 缺省情况下，没有配置 IPv4 的路由信息的下一跳地址或源地址的匹配条件
配置 IPv6 的路由信息的匹配条件		**if-match ipv6** { **address** | **next-hop** | **route-source** } { **acl** *acl-number* | **prefix-list** *ipv6-prefix-name* }	可选 缺省情况下，没有配置 IPv6 的路由信息的匹配条件
配置 BGP 路由信息的 AS 路径域的匹配条件		**if-match as-path** *as-path-number*&<1-16>	可选 缺省情况下，没有配置 BGP 路由信息的 AS 路径域的匹配条件
匹配 BGP 路由信息的团体属性		**if-match community** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	可选 缺省情况下，不匹配 BGP 路由信息的团体属性。
配置路由信息的路由开销的匹配条件		**if-match cost** *value*	可选 缺省情况下，没有配置路由信息的路由开销的匹配条件
配置 BGP 扩展团体属性的匹配条件		**if-match extcommunity** *ext-comm-list-number*&<1-16>	可选 缺省情况下，没有配置 BGP 扩展团体属性的匹配条件
配置路由信息的出接口的匹配条件		**if-match interface** { *interface-type interface-number* }&<1-16>	可选 缺省情况下，没有配置路由信息的出接口的匹配条件
配置路由信息的 MPLS 标签的匹配条件		**if-match mpls-label**	可选 缺省情况下，没有配置路由信息的 MPLS 标签的匹配条件
配置路由信息的类型的匹配条件		**if-match route-type** { **internal** | **external-type1** | **external-type2** | **external-type1or2** | **is-is-level-1** | **is-is-level-2** | **nssa-external-type1** | **nssa-external-type2** | **nssa-external-type1or2** } *	可选 缺省情况下，没有配置路由信息的类型的匹配条件
配置 RIP、OSPF、IS-IS 路由信息的标记域的匹配条件		**if-match tag** *value*	可选 缺省情况下，没有配置 RIP、OSPF、IS-IS 路由信息的标记域的匹配条件

说明

- 对于同一个 Route-policy 节点，在匹配的过程中，各个 **if-match** 子句间是“与”的关系，即路由信息必须同时满足所有匹配条件，才可以执行 **apply** 子句的动作。
- 在一个节点中，可以没有 **if-math** 子句，也可以有多个 **if-match** 子句。当不指定 **if-match** 子句时，如果该节点的匹配模式为允许模式，则所有路由信息都会通过该节点的过滤；如果该节点的匹配模式为拒绝模式，则所有路由信息都会被拒绝。
- 路由策略应使用非 VPN 的 ACL 进行路由过滤。
- IPv4 路由策略和 IPv6 路由策略在配置 **if-match** 子句时，不同之处在于匹配路由信息的目的地址、下一跳和源地址的命令不同。

1.3.4 配置 apply 子句

表1-8 配置 apply 子句

操作		命令	说明
进入系统视图		**system-view**	-
创建路由策略并进入该路由策略视图		**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	必选 缺省情况下，不创建路由策略
配置 BGP 路由信息的 AS_PATH 属性		**apply as-path** *as-number*&<1-10> [**replace**]	可选 缺省情况下，没有配置 BGP 路由信息的 AS_PATH 属性
删除指定的 BGP 团体属性		**apply comm-list** *comm-list-number* **delete**	可选 缺省情况下，不删除路由团体属性
配置 BGP 路由信息的团体属性		**apply community** { **none** | **additive** | { *community-number*&<1-16> | *aa:nn*&<1-16> | **internet** | **no-export-subconfed** | **no-export** | **no-advertise** }* [**additive**] }	可选 缺省情况下，没有配置 BGP 路由信息的团体属性
配置路由信息的路由开销		**apply cost** [**+** | **-**] *value*	可选 缺省情况下，没有配置路由信息的路由开销
配置路由信息的开销类型		**apply cost-type** { **external** | **internal** | **type-1** | **type-2** }	可选 缺省情况下，没有配置路由开销类型
配置 BGP 扩展团体属性		**apply extcommunity** { **rt** { *as-number*:*nn* | *ip-address*:*nn* } }&<1-16> [**additive**]	可选 缺省情况下，没有配置 BGP 扩展团体属性
配置路由信息的下一跳地址	配置 IPv4 路由信息的下一跳地址	**apply ip-address next-hop** *ip-address*	可选 缺省情况下，没有配置 IPv4 路由信息的下一跳地址
	配置 IPv6 路由信息的下一跳地址	**apply ipv6 next-hop** *ipv6-address*	可选 缺省情况下，没有配置 IPv6 路由信息的下一跳地址
配置引入路由到 IS-IS 某个级别的区域		**apply isis** { **level-1** | **level-1-2** | **level-2** }	可选 缺省情况下，没有配置引入路由到 IS-IS 某个级别的区域

操作	命令	说明
配置 BGP 路由信息的本地优先级	**apply local-preference** *preference*	可选 缺省情况下，没有配置 BGP 路由信息的本地优先级
配置 MPLS 标签	**apply mpls-label**	可选 缺省情况下，没有配置 MPLS 标签
配置BGP路由信息的ORIGIN属性	**apply origin** { **igp** | **egp** *as-number* | **incomplete** }	可选 缺省情况下，没有配置 BGP 路由信息的 ORIGIN 属性
配置路由协议的优先级	**apply preference** *preference*	可选 缺省情况下，没有配置路由协议的优先级
配置 BGP 路由信息的首选值	**apply preferred-value** *preferred-value*	可选 缺省情况下，没有配置 BGP 路由信息的首选值
配置 RIP、OSPF、IS-IS 路由信息的标记域	**apply tag** *value*	可选 缺省情况下，没有配置 RIP、OSPF、IS-IS 路由信息的标记域

说明

- IPv4 路由策略和 IPv6 路由策略在配置 **apply** 子句时，不同之处在于设置路由信息的下一跳地址的命令不同。
- 对于引入的 IPv4 路由，使用 **apply ip-address next-hop** 命令设置下一跳地址无效；对于引入的 IPv6 路由，使用 **apply ipv6 next-hop** 命令设置下一跳地址无效。

1.4 路由策略的显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后路由策略的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除路由策略的统计信息。

表1-9 路由策略的显示和维护

操作	命令
显示 BGP AS 路径过滤列表信息	**display ip as-path** [*as-path-number*]
显示 BGP 团体属性列表信息	**display ip community-list** [*basic-community-list-number* | *adv-community-list-number*]
显示 BGP 扩展团体属性列表信息	**display ip extcommunity-list** [*ext-comm-list-number*]
显示 IPv4 地址前缀列表的统计信息	**display ip ip-prefix** [*ip-prefix-name*]
显示 IPv6 地址前缀列表的统计信息	**display ip ipv6-prefix** [*ipv6-prefix-name*]
显示路由策略信息	**display route-policy** [*route-policy-name*]
清除 IPv4 地址前缀列表的统计信息	**reset ip ip-prefix** [*ip-prefix-name*]
清除 IPv6 地址前缀列表的统计信息	**reset ip ipv6-prefix** [*ipv6-prefix-name*]

1.5 路由策略典型配置举例

1.5.1 在 IPv4 路由引入中应用路由策略

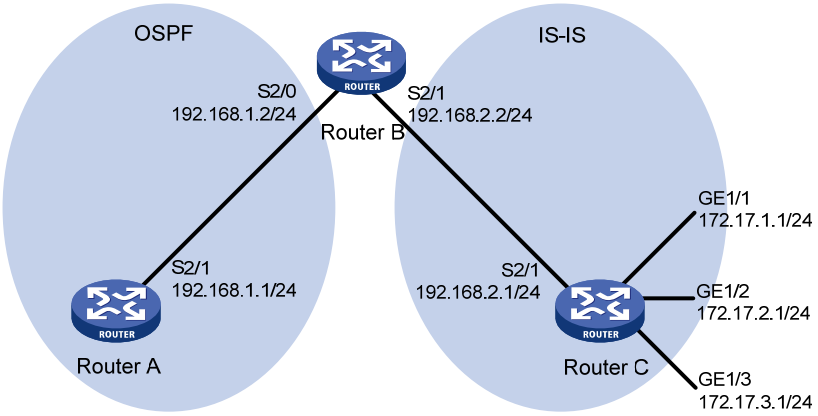
1. 组网需求

如下图所示，Router B 与 Router A 之间通过 OSPF 协议交换路由信息，与 Router C 之间通过 IS-IS 协议交换路由信息。

要求在 Router B 上配置路由引入，将 IS-IS 路由引入到 OSPF 中去，并同时使用路由策略设置路由的属性。其中，设置 172.17.1.0/24 的路由的开销为 100，设置 172.17.2.0/24 的路由的 Tag 属性为 20。

2. 组网图

图1-1 在 IPv4 路由引入中应用路由策略组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 IS-IS 路由协议

配置 Router C。

```
<RouterC> system-view
[RouterC] isis
[RouterC-isis-1] is-level level-2
[RouterC-isis-1] network-entity 10.0000.0000.0001.00
[RouterC-isis-1] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] isis enable
[RouterC-Serial2/1] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] isis enable
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] isis enable
[RouterC-GigabitEthernet1/2] quit
[RouterC] interface GigabitEthernet 1/3
[RouterC-GigabitEthernet1/3] isis enable
```

```
[RouterC-GigabitEthernet1/3] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis
[RouterB-isis-1] is-level level-2
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] isis enable
[RouterB-Serial2/1] quit
```

(3) 配置 OSPF 路由协议及路由引入

配置 Router A，启动 OSPF。

```
<RouterA> system-view
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
```

配置 RouterB，启动 OSPF，并引入 IS-IS 路由。

```
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] import-route isis 1
[RouterB-ospf-1] quit
```

查看 Router A 的 OSPF 路由表，可以看到引入的路由。

```
[RouterA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop        AdvRouter     Area
 192.168.1.0/24     1        Transit 192.168.1.1    192.168.1.1   0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag        NextHop        AdvRouter
 172.17.1.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.2.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.3.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 192.168.2.0/24     1        Type2   1          192.168.1.2    192.168.2.2

 Total Nets: 5
 Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0
```

(4) 配置过滤列表

配置编号为 2002 的 ACL，允许 172.17.2.0/24 的路由通过。

```
[RouterB] acl number 2002
[RouterB-acl-basic-2002] rule permit source 172.17.2.0 0.0.0.255
[RouterB-acl-basic-2002] quit
```

配置名为 prefix-a 的地址前缀列表，允许 172.17.1.0/24 的路由通过。

```
[RouterB] ip ip-prefix prefix-a index 10 permit 172.17.1.0 24
```

(5) 配置路由策略

```
[RouterB] route-policy isis2ospf permit node 10
[RouterB-route-policy] if-match ip-prefix prefix-a
[RouterB-route-policy] apply cost 100
[RouterB-route-policy] quit
[RouterB] route-policy isis2ospf permit node 20
[RouterB-route-policy] if-match acl 2002
[RouterB-route-policy] apply tag 20
[RouterB-route-policy] quit
[RouterB] route-policy isis2ospf permit node 30
[RouterB-route-policy] quit
```

(6) 在路由引入时应用路由策略

配置 Router B，设置在路由引入时应用路由策略。

```
[RouterB] ospf
[RouterB-ospf-1] import-route isis 1 route-policy isis2ospf
[RouterB-ospf-1] quit
```

查看 Router A 的 OSPF 路由表，可以看到目的地址为 172.17.1.0/24 的路由的开销为 100，目的地址为 172.17.2.0/24 的路由的标记域（Tag）为 20，而其他外部路由没有变化。

```
[RouterA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables
 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter     Area
 192.168.1.0/24     1        Transit 192.168.1.1     192.168.1.1   0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop       AdvRouter
 172.17.1.0/24      100      Type2   1           192.168.1.2   192.168.2.2
 172.17.2.0/24      1        Type2   20          192.168.1.2   192.168.2.2
 172.17.3.0/24      1        Type2   1           192.168.1.2   192.168.2.2
 192.168.2.0/24     1        Type2   1           192.168.1.2   192.168.2.2

 Total Nets: 5
 Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0
```

1.5.2 在 IPv6 路由引入中应用路由策略

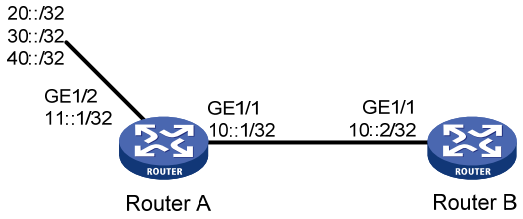
1. 组网需求

- Router A 与 Router B 通信，都运行 RIPng 协议。

- 使能 Router A 上的 RIPng 协议，配置三条静态路由。
- 设置在引入静态路由时应用路由策略，使三条静态路由部分引入、部分被屏蔽掉——20::/32 和 40::/32 网段的路由是可见的，30::/32 网段的路由则被屏蔽。
- 通过在 Router B 上查看 RIPng 路由表，验证路由策略是否生效。

2. 组网图

图1-2 在 IPv6 路由引入中应用路由策略组网图



3. 配置步骤

(1) 配置 Router A

配置接口 GigabitEthernet1/1 和 GigabitEthernet1/2 的 IPv6 地址，封装 PPP 协议。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ipv6 address 10::1 32
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ipv6 address 11::1 32
[RouterA-GigabitEthernet1/2] quit
```

在接口下使能 RIPng。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ripng 1 enable
[RouterA-GigabitEthernet1/1] quit
```

配置三条静态路由。

```
[RouterA] ipv6 route-static 20:: 32 GigabitEthernet 1/2
[RouterA] ipv6 route-static 30:: 32 GigabitEthernet 1/2
[RouterA] ipv6 route-static 40:: 32 GigabitEthernet 1/2
```

配置路由策略。

```
[RouterA] ip ipv6-prefix a index 10 permit 30:: 32
[RouterA] route-policy static2ripng deny node 0
[RouterA-route-policy] if-match ipv6 address prefix-list a
[RouterA-route-policy] quit
[RouterA] route-policy static2ripng permit node 10
[RouterA-route-policy] quit
```

启动 RIPNG 协议，同时应用路由策略 static2ripng 对引入的静态路由进行过滤。

```
[RouterA] ripng
[RouterA-ripng-1] import-route static route-policy static2ripng
```

(2) 配置 Router B

配置接口 GigabitEthernet1/1 的 IPv6 地址。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ipv6 address 10::2 32
```

在接口下使能 RIPng。

```
[RouterB-GigabitEthernet1/1] ripng 1 enable
[RouterB-GigabitEthernet1/1] quit
```

启动 RIPng 协议。

```
[RouterB] ripng
```

查看 RIPng 路由表。

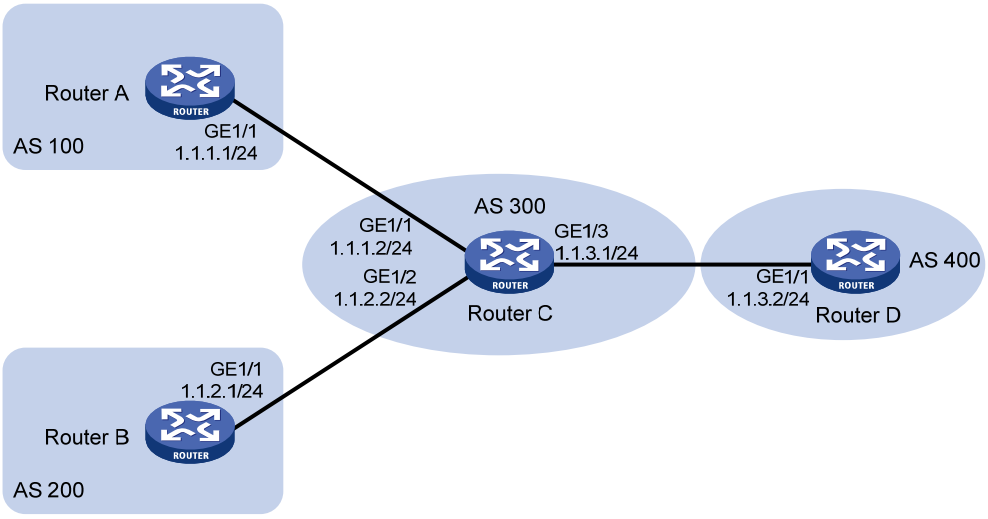
```
[RouterB-ripng-1] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------
 Peer FE80::7D58:0:CA03:1  on GigabitEthernet1/1
 Dest 10::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 18 Sec
 Dest 20::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 8 Sec
 Dest 40::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 3 Sec
```

1.5.3 配置应用路由策略过滤接收的 BGP 路由

1. 组网需求

- 所有路由器均运行 BGP 协议，Router A、Router B 和 Router C 之间建立 EBGP 连接，Router C 和 Router D 之间建立 EBGP 连接。
- 通过配置路由策略，使得 Router D 拒绝接收来自 AS200 的路由。

2. 组网图



3. 配置步骤

(1) 配置接口 IP 地址（略）

(2) 配置 BGP 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] bgp 100
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] peer 1.1.1.2 as-number 300
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 200
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 1.1.2.2 as-number 300
```

配置 Router C。

```
<RouterC> system-view
[RouterC] bgp 300
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] peer 1.1.1.1 as-number 100
[RouterC-bgp] peer 1.1.2.1 as-number 200
[RouterC-bgp] peer 1.1.3.2 as-number 400
```

配置 Router D。

```
<RouterD> system-view
[RouterD] bgp 400
[RouterD-bgp] router-id 4.4.4.4
[RouterD-bgp] peer 1.1.3.1 as-number 300
[RouterD-bgp] quit
```

配置 Router A，将网段路由 4.4.4.4/24、5.5.5.5/24、6.6.6.6/24 发布到 BGP 路由表中。

```
[RouterA-bgp] network 4.4.4.4 24
[RouterA-bgp] network 5.5.5.5 24
[RouterA-bgp] network 6.6.6.6 24
```

配置 Router B，将网段路由 7.7.7.7/24、8.8.8.8/24、9.9.9.9/24 发布到 BGP 路由表中。

```
[RouterB-bgp] network 7.7.7.7 24
[RouterB-bgp] network 8.8.8.8 24
[RouterB-bgp] network 9.9.9.9 24
```

查看 Router D 的 BGP 路由表。

```
[RouterD-bgp] display bgp routing-table

 Total Number of Routes: 6

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
```

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

 *>  4.4.4.0/24         1.1.3.1                                0       300 100i
 *>  5.5.5.0/24         1.1.3.1                                0       300 100i
 *>  6.6.6.0/24         1.1.3.1                                0       300 100i
 *>  7.7.7.0/24         1.1.3.1                                0       300 200i
 *>  8.8.8.0/24         1.1.3.1                                0       300 200i
 *>  9.9.9.0/24         1.1.3.1                                0       300 200i
```

从路由表可以看出，Router D 同时学到了来自 AS 100 的路由 4.4.4.0/24、5.5.5.0/24、6.6.6.0/24 和来自 AS 200 的路由 7.7.7.0/24、8.8.8.0/24、9.9.9.0/24。

(3) 配置 Router D 拒绝接收来自 AS200 的路由。

在 Router D 上配置编号为 1 的 AS 路径过滤列表。

```
[RouterD] ip as-path 1 permit .*200.*
```

在 Router D 上配置名称为“rt1”的路由策略。

```
[RouterD] route-policy rt1 deny node 1
[RouterD] if-match as-path 1
[RouterD] route-policy rt2 permit node 2
```

在 Router D 上，配置对从对等体 1.1.3.1 接收的路由指定名称为“rt1”的路由策略。

```
[RouterD] bgp 400
[RouterD] peer 1.1.3.1 route-policy rt1 import
```

查看 Router D 的 BGP 路由表。

```
[RouterD] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

 *>  4.4.4.0/24         1.1.3.1                                0       300 100i
 *>  5.5.5.0/24         1.1.3.1                                0       300 100i
 *>  6.6.6.0/24         1.1.3.1                                0       300 100i
```

从路由表可以看出，Router D 只学到了来自 AS 100 的路由 4.4.4.0/24、5.5.5.0/24、6.6.6.0/24。

1.6 常见错误配置举例

1.6.1 无法实现 IPv4 路由信息过滤

1. 错误现象

路由协议运行正常的情况下无法实现路由信息过滤。

2. 分析

地址前缀列表的各个表项中至少应该有一个表项的匹配模式是 **permit** 模式。Route-policy 的各个节点中至少应该有一个节点的匹配模式是 **permit** 模式。

3. 处理过程

(1) 使用 **display ip ip-prefix** 命令查看前缀列表的配置。

(2) 使用 **display route-policy** 命令查看路由策略的配置。

1.6.2 无法实现 IPv6 路由信息过滤

1. 错误现象

路由协议运行正常的情况下无法实现路由信息过滤。

2. 分析

地址前缀列表的各个表项中至少应该有一个表项的匹配模式是 **permit** 模式。Route-policy 的各个节点中至少应该有一个节点的匹配模式是 **permit** 模式。

3. 处理过程

(1) 使用 **display ip ipv6-prefix** 命令查看前缀列表的配置。

(2) 使用 **display route-policy** 命令查看路由策略的配置。

目 录

1 静态路由配置 1-1
1.1 简介 1-1
1.1.1 静态路由 1-1
1.1.2 缺省路由 1-1
1.1.3 静态路由应用 1-1
1.2 配置静态路由 1-2
1.2.1 配置准备 1-2
1.2.2 配置静态路由 1-2
1.3 检测静态路由下一跳是否可达 1-3
1.3.1 配置BFD检测静态路由下一跳是否可达 1-3
1.3.2 配置静态路由与Track联动检测下一跳是否可达 1-4
1.4 静态路由显示和维护 1-5
1.5 静态路由典型配置举例 1-5

1 静态路由配置

1.1 简介

1.1.1 静态路由

静态路由是一种特殊的路由，由管理员手工配置；配置静态路由后，去往指定目的地的数据报文将按照管理员指定的路径进行转发。

在组网结构比较简单的网络中，只需配置静态路由就可以实现网络互通。恰当地设置和使用静态路由可以改善网络的性能，并可为重要的网络应用保证带宽。

静态路由的缺点在于：不能自动适应网络拓扑结构的变化，当网络发生故障或者拓扑发生变化后，可能会出现路由不可达，导致网络中断，此时必须由网络管理员手工修改静态路由的配置。

1.1.2 缺省路由

如果到达某个指定网络的数据报文在路由器的路由表里找不到对应的表项，那么该报文将被路由器丢弃。

通过给当前路由器配置一条缺省路由，那些在路由表里找不到匹配路由表入口项的数据报文将会转发给另外一台路由器（如果这台路由器的路由能力比较强，包括到达大部分所有网络的路由信息），由另外一台路由器进行报文的转发。

缺省路由是在路由器没有找到匹配的路由表入口项时才使用的路由：

- 如果报文的目的地址不能与路由表的任何入口项相匹配，那么该报文将选取缺省路由；
- 如果没有缺省路由且报文的目的地不在路由表中，那么该报文将被丢弃，将向源端返回一个 ICMP 报文报告该目的地址或网络不可达。

缺省路由有两种生成方式：

- 第一种是通过网络管理员在路由器上配置到网络 0.0.0.0（掩码也为 0.0.0.0）的静态路由，对于一个到来的数据报文，如果在当前路由器里找不到匹配的路由表项，将会把报文发给在配置的静态路由里指定的下一跳路由器；
- 第二种是通过动态路由协议生成（如 OSPF、IS-IS 和 RIP），由路由能力比较强的路由器将缺省路由发布给其它路由器，其它路由器在自己的路由表里生成指向那台路由器的缺省路由。

1.1.3 静态路由应用

配置静态路由时，需要了解以下内容：

(1) 目的地址与掩码

在 **ip route-static** 命令中，IPv4 地址为点分十进制格式，掩码可以用点分十进制表示，也可用掩码长度（即掩码中连续‘1’的位数）表示。

(2) 出接口和下一跳地址

在配置静态路由时，可指定出接口，也可指定下一跳地址。指定出接口还是指定下一跳地址要视具体情况而定，下一跳地址不能为本地接口 IP 地址，否则路由不会生效。

实际上，所有的路由项都必须明确下一跳地址。在发送报文时，首先根据报文的目的地址寻找路由表中与之匹配的路由。只有指定了下一跳地址，链路层才能找到对应的链路层地址，并转发报文。

指定出接口时需要注意：

- 对于 Null0 和 Loopback 接口，配置了出接口就不再配置下一跳地址。
- 对于点到点接口，即使不知道对端地址，也可以在路由器配置时指定出接口。这样，即使对端地址发生了改变也无须改变该路由器的配置。如封装 PPP 协议的接口，通过 PPP 协商获取对端的 IP 地址，这时可以不指定下一跳地址，只需指定出接口即可。
- 对于 NBMA、P2MP 等接口，它们支持点到多点网络，这时除了配置 IP 路由外，还需在链路层建立二次路由，即 IP 地址到链路层地址的映射。通常情况下，建议在配置出接口时，同时配置下一跳 IP 地址。
- 在配置静态路由时，建议不要直接指定广播类型接口作出接口（如以太网接口、Virtual-Template、VLAN 接口等）。因为广播类型的接口，会导致出现多个下一跳，无法唯一确定下一跳。在某些特殊应用中，如果必须配置广播接口（如以太网接口、Virtual-Template、VLAN 接口等）为出接口，则必须同时指定其对应的下一跳地址。

(3) 其它属性

对于不同的静态路由，可以为它们配置不同的优先级，从而更灵活地应用路由管理策略。例如：配置到达相同目的地的多条路由，如果指定相同优先级，则可实现负载分担，如果指定不同优先级，则可实现路由备份。

1.2 配置静态路由

1.2.1 配置准备

在配置静态路由之前，需完成以下任务：

- 配置相关接口的物理参数
- 配置相关接口的链路层属性
- 配置相关接口的 IP 地址

1.2.2 配置静态路由

表1-1 配置静态路由

操作	命令	说明
进入系统视图	**system-view**	-
配置静态路由	**ip route-static** *dest-address* { *mask* | *mask-length* } { *next-hop-address* | *interface-type interface-number next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } [**preference** *preference-value*] [**tag** *tag-value*] [*description description-text*]	必选 缺省情况下，静态路由的优先级 **preference** 为 60，静态路由 **tag** 值为 0，未配置描述信息
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* | *mask-length* } { *next-hop-address* [**public**] | *interface-type interface-number next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

操作	命令	说明
配置静态路由的缺省优先级	**ip route-static default-preference** *default-preference-value*	可选 缺省情况下，静态路由的缺省优先级为 60

说明

- 在配置静态路由时，如果先指定下一跳地址，然后将该地址配置为本地接口（如以太网接口、VLAN 接口等）的 IP 地址，静态路由不会生效。
- 如果在配置静态路由时没有指定优先级，就会使用缺省优先级。重新设置缺省优先级后，新设置的缺省优先级仅对新增的静态路由有效。
- 设置静态路由的 Tag 值，可以在路由策略中根据 Tag 值对路由进行灵活的控制。
- 配置静态路由命令对 VPN 实例的支持情况与设备的型号有关，请以设备的实际情况为准。
- 在使用 **ip route-static** 配置静态路由时，如果将目的地址与掩码配置为全零（0.0.0.0 0.0.0.0），则表示配置的是缺省路由。

1.3 检测静态路由下一跳是否可达

当网络发生故障或者拓扑发生变化后，可能会出现路由不可达，导致网络中断，因此为了提升现有网络的稳定性，需要对静态路由下一跳的可达性进行快速检测，当下一跳不可达时可以快速切换到备份路由。

可以通过以下两种方式检测静态路由下一跳是否可达，需要注意的是，两种方式不能同时使用，只能使用其中的一种。

1.3.1 配置 BFD 检测静态路由下一跳是否可达

BFD（Bidirectional Forwarding Detection，双向转发检测）提供了一个通用的、标准化的、介质无关、协议无关的快速故障检测机制，可以为各上层协议如路由协议、MPLS 等统一地快速检测两台路由器间双向转发路径的故障。关于 BFD 的详细介绍，请参考“IP 路由分册”中的“BFD 配置”。

配置静态路由后，可以使能 BFD 功能，对静态路由下一跳的可达性进行快速检测。

1. 配置准备

在配置通过 BFD 检测静态路由下一跳是否可达之前，需要使能 BFD 功能，具体配置过程请参考“IP 路由分册”中的“BFD 配置”。

2. 配置过程

表1-2 配置 BFD 检测静态路由下一跳是否可达

操作	命令	说明
进入系统视图	**system-view**	-
配置 BFD 检测静态路由下一跳是否可达	**ip route-static** *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* [**bfd** { **control-packet** | **echo-packet** }] [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	必选 缺省情况下，没有使用 BFD 功能检查静态路由下一跳的可达性
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* [**bfd** { **control-packet** | **echo-packet** }] [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

说明

- 通过控制报文方式实现 BFD 功能时，对端必须创建 BFD 会话，否则 BFD 功能将无法正常运行。通过 echo 报文方式实现 BFD 功能时，对端不需要创建 BFD 会话 BFD 功能也可以正常运行。关于 BFD 的详细介绍，请参考“IP 路由分册”中的“BFD 配置”。
- 路由振荡时，使能 BFD 检测功能可能会加剧振荡，需谨慎使用。

1.3.2 配置静态路由与 Track 联动检测下一跳是否可达

如果在配置静态路由时只指定了下一跳而没有指定出接口，可以通过配置静态路由与 Track 建立联动来检测静态路由的有效性：

- 当 Track 项状态为 positive 时，静态路由的下一跳可达，配置的静态路由将生效。
- 当 Track 项状态为 negative 时，静态路由的下一跳可不达，配置的静态路由无效。

关于 Track 的详细介绍，请参考“系统分册”中的“Track 配置”。

1. 配置准备

配置静态路由与 Track 建立联动来检测静态路由下一跳是否可达，需要先创建 Track 项，具体配置过程请参考“系统分册”中的“Track 配置”。

2. 配置过程

表1-3 配置静态路由与 Track 建立联动检测下一跳是否可达

操作	命令	说明
进入系统视图	**system-view**	-
配置静态路由与 Track 联动检测静态路由下一跳是否可达	**ip route-static** *dest-address* { *mask* | *mask-length* } { *next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } **track** *track-entry-number* [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	必选 缺省情况下，没有配置静态路由与 Track 联动
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* | *mask-length* } { *next-hop-address* **track** *track-entry-number* [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address* **track** *track-entry-number* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

说明

- 配置静态路由支持 Track 监测功能，该条静态路由可以是已经创建的，也可以是未创建的。对于已经创建的静态路由，只是将静态路由与 Track 项关联，并根据 Track 项的状态来判断静态路由的有效性；对于未创建的静态路由，首先要生成该静态路由，然后将其与 Track 项关联。
- 如果 Track 模块通过 NQA 探测私网静态路由中下一跳的可达性，静态路由下一跳的 VPN 实例号与 NQA 测试组配置的实例号必须相同，才能进行正常的探测。
- 需要注意在静态路由进行迭代时，Track 项监测的应该是静态路由真正的下一跳，而不是配置的下一跳。否则，可能导致错误地将有效路由判断为无效路由。

1.4 静态路由显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令查看静态路由配置的运行情况并检验配置结果。

在系统视图下执行 **delete** 命令可以删除配置的所有静态路由。

表1-4 静态路由显示和维护

操作	命令
查看当前的配置文件信息	**display current-configuration**
查看 IP 路由表摘要信息	**display ip routing-table**
查看 IP 路由表详细信息	**display ip routing-table verbose**
查看静态路由表信息	**display ip routing-table protocol static** [**inactive** | **verbose**]
删除所有静态路由	**delete** [**vpn-instance** *vpn-instance-name*] **static-routes all**

说明

删除所有静态路由命令对 VPN 实例的支持情况与设备的型号有关，请以设备的实际情况为准。

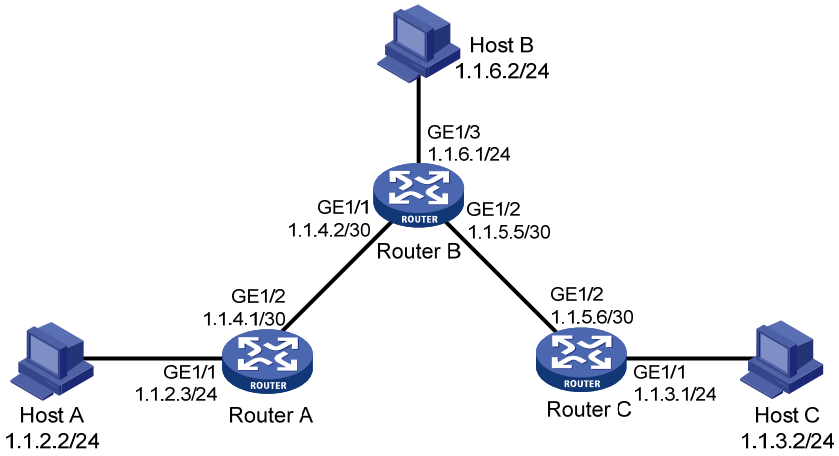
1.5 静态路由典型配置举例

1. 组网需求

路由器各接口及主机的 IP 地址和掩码如下图所示。要求采用静态路由，使图中任意两台主机之间都能互通。

2. 组网图

图1-1 静态路由配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置静态路由

在 Router A 上配置缺省路由。

```
<RouterA> system-view
[RouterA] ip route-static 0.0.0.0 0.0.0.0 1.1.4.2
```

在 Router B 上配置两条静态路由。

```
<RouterB> system-view
[RouterB] ip route-static 1.1.2.0 255.255.255.0 1.1.4.1
[RouterB] ip route-static 1.1.3.0 255.255.255.0 1.1.5.6
```

在 Router C 上配置缺省路由。

```
<RouterC> system-view
[RouterC] ip route-static 0.0.0.0 0.0.0.0 1.1.5.5
```

(3) 配置主机

配置 Host A 的缺省网关为 1.1.2.3，Host B 的缺省网关为 1.1.6.1， Host C 的缺省网关为 1.1.3.1，具体配置过程略。

(4) 查看配置结果

显示 Router A 的 IP 路由表。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7

Destination/Mask    Proto   Pre  Cost        NextHop         Interface

0.0.0.0/0           Static 60   0           1.1.4.2         GE1/2
1.1.2.0/24          Direct 0    0           1.1.2.3         GE1/1
1.1.2.3/32          Direct 0    0           127.0.0.1       InLoop0
1.1.4.0/30          Direct 0    0           1.1.4.1         GE1/2
```

```
1.1.4.1/32          Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
```

显示 Router B 的 IP 路由表。

```
[RouterB] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

1.1.2.0/24          Static 60   0           1.1.4.1         GE1/1
1.1.3.0/24          Static 60   0           1.1.5.6         GE1/2
1.1.4.0/30          Direct 0    0           1.1.4.2         GE1/1
1.1.4.2/32          Direct 0    0           127.0.0.1       InLoop0
1.1.5.4/30          Direct 0    0           1.1.5.5         GE1/2
1.1.5.5/32          Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
1.1.6.0/24          Direct 0    0           1.1.6.1         GE1/3
1.1.6.1/32          Direct 0    0           127.0.0.1       InLoop0
```

在 Host B 上使用 **ping** 命令验证 Host A 是否可达（假定主机安装的操作系统为 Windows XP）。

```
C:\Documents and Settings\Administrator> ping 1.1.2.2

Pinging 1.1.2.2 with 32 bytes of data:

Reply from 1.1.2.2: bytes=32 time=1ms TTL=128
Reply from 1.1.2.2: bytes=32 time=1ms TTL=128
Reply from 1.1.2.2: bytes=32 time=1ms TTL=128
Reply from 1.1.2.2: bytes=32 time=1ms TTL=128

Ping statistics for 1.1.2.2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

在 Host B 上使用 **tracert** 命令验证 Host A 是否可达。

```
C:\Documents and Settings\Administrator>tracert 1.1.2.2

Tracing route to 1.1.2.2 over a maximum of 30 hops

  1    <1 ms    <1 ms    <1 ms  1.1.6.1
  2    <1 ms    <1 ms    <1 ms  1.1.4.1
  3     1 ms    <1 ms    <1 ms  1.1.2.2

Trace complete.
```

目 录

1 IPv6 BGP配置 1-1
1.1 IPv6 BGP简介 1-1
1.2 配置任务简介 1-2
1.3 配置IPv6 BGP的基本功能 1-2
1.3.1 配置准备 1-2
1.3.2 配置IPv6 对等体 1-3
1.3.3 配置IPv6 BGP发布本地IPv6 路由 1-3
1.3.4 配置路由首选值 1-3
1.3.5 配置IPv6 BGP连接所使用的本地接口 1-4
1.3.6 配置非直接相连的邻居建立EBGP连接 1-4
1.3.7 配置IPv6 对等体/IPv6 对等体组的描述信息 1-5
1.3.8 禁止与IPv6 对等体/IPv6 对等体组建立会话 1-5
1.3.9 记录指定IPv6 对等体/IPv6 对等体组的会话状态和事件信息 1-6
1.4 控制路由信息的发布与接收 1-6
1.4.1 配置准备 1-6
1.4.2 配置IPv6 BGP引入其他路由 1-6
1.4.3 配置IPv6 BGP路由聚合 1-7
1.4.4 配置向IPv6 对等体/IPv6 对等体组发送缺省路由 1-7
1.4.5 配置路由信息的发布策略 1-7
1.4.6 配置路由信息的接收策略 1-8
1.4.7 配置IPv6 BGP与IGP路由同步 1-9
1.4.8 配置路由衰减 1-9
1.5 配置IPv6 BGP的路由属性 1-10
1.5.1 配置准备 1-10
1.5.2 配置IPv6 BGP路由管理的优先级、缺省LOCAL_PREF及NEXT_HOP属性 1-10
1.5.3 配置MED属性 1-11
1.5.4 配置AS_PATH属性 1-11
1.6 调整和优化IPv6 BGP网络 1-12
1.6.1 配置准备 1-12
1.6.2 配置IPv6 BGP的时钟 1-12
1.6.3 配置IPv6 BGP软复位 1-13
1.6.4 配置最大等价路由的条数 1-14
1.7 组建大型IPv6 BGP网络 1-14
1.7.1 配置准备 1-14
1.7.2 配置IPv6 BGP对等体组 1-15
1.7.3 配置IPv6 BGP团体 1-16
1.7.4 配置IPv6 BGP路由反射器 1-17
1.8 IPv6 BGP显示和维护 1-17
1.8.1 IPv6 BGP显示 1-17
1.8.2 复位IPv6 BGP连接 1-18
1.8.3 清除IPv6 BGP信息 1-18

1.9 IPv6 BGP典型配置举例 1-19
1.9.1 IPv6 BGP基本配置 1-19
1.9.2 配置IPv6 BGP路由反射 1-21
1.10 IPv6 BGP常见错误配置举例 1-22
1.10.1 IPv6 BGP对等体关系不能建立 1-22

1 IPv6 BGP 配置

说明

本章只列出了 IPv6 BGP 专有的配置与操作，其他相关内容请参见“IP 路由分册”中的“BGP 配置”。

1.1 IPv6 BGP 简介

传统的 BGP-4 只能管理 IPv4 的路由信息，对于使用其它网络层协议（如 IPv6 等）的应用，在跨自治系统传播时就受到一定限制。

为了提供对多种网络层协议的支持，IETF 对 BGP-4 进行了扩展，形成 IPv6 BGP，目前的 IPv6 BGP 标准是 RFC 2858（Multiprotocol Extensions for BGP-4，BGP-4 多协议扩展）。

为了实现对 IPv6 协议的支持，IPv6 BGP 需要将 IPv6 网络层协议的信息反映到 NLRI（Network Layer Reachable Information，网络层可达信息）及 NEXT_HOP 属性中。

IPv6 BGP 中引入的两个 NLRI 属性分别是：

- MP_REACH_NLRI：Multiprotocol Reachable NLRI，多协议可达 NLRI。用于发布可达路由及下一跳信息。
- MP_UNREACH_NLRI：Multiprotocol Unreachable NLRI，多协议不可达 NLRI。用于撤销不可达路由。

IPv6 BGP 中的 NEXT_HOP 属性用 IPv6 地址来表示，可以是 IPv6 全球单播地址或者链路本地地址。

IPv6 BGP 是利用 BGP 的多协议扩展属性，来达到在 IPv6 网络中应用的目的，BGP 协议原有的消息机制和路由机制并没有改变。

1.2 配置任务简介

表1-1 配置任务简介

配置任务		说明	详细配置
配置 IPv6 BGP 基本功能	配置 IPv6 对等体	必选	1.3.2
	配置 IPv6 BGP 发布本地 IPv6 路由	可选	1.3.3
	配置路由首选值	可选	1.3.4
	配置 IPv6 BGP 连接所使用的本地接口	可选	1.3.5
	配置 EBGP 连接的最大跳数	可选	1.3.6
	配置 IPv6 对等体/IPv6 对等体组的描述信息	可选	1.3.7
	禁止与 IPv6 对等体/IPv6 对等体组建立会话	可选	1.3.8
	记录指定 IPv6 对等体/IPv6 对等体组的会话状态和事件信息	可选	1.3.9
控制路由信息的发布与接收	配置 IPv6 BGP 引入其他路由	可选	1.4.2
	配置 IPv6 BGP 路由聚合	可选	1.4.3
	配置向 IPv6 对等体/IPv6 对等体组发送缺省路由	可选	1.4.4
	配置路由信息的发布策略	可选	1.4.5
	配置路由信息的接收策略	可选	1.4.6
	配置 IPv6 BGP 与 IGP 路由同步	可选	1.4.7
	配置路由衰减	可选	1.4.8
配置 IPv6 BGP 的路由属性	配置 IPv6 BGP 路由管理的优先级、缺省 LOCAL_PREF 及 NEXT_HOP 属性	可选	1.5.2
	配置 MED 属性	可选	1.5.3
	配置 AS_PATH 属性	可选	1.5.4
调整和优化 IPv6 BGP 网络	配置 IPv6 BGP 的时钟	可选	1.6.2
	配置 IPv6 BGP 软复位	可选	1.6.3
	配置最大等价路由的条数	可选	1.6.4
组建大型 IPv6 BGP 网络	配置 IPv6 BGP 对等体组	可选	1.7.2
	配置 IPv6 BGP 团体	可选	1.7.3
	配置 IPv6 BGP 路由反射器	可选	1.7.4

1.3 配置 IPv6 BGP 的基本功能

1.3.1 配置准备

在配置 IPv6 BGP 基本功能之前，需完成以下任务：

- 配置接口的网络层地址
- 使能 IPv6 功能

说明

必须首先创建对等体组，才能配置对等体组的基本功能。关于对等体组的创建请参见“1.7.2 配置IPv6 BGP对等体组”。

1.3.2 配置 IPv6 对等体

表1-2 配置 IPv6 对等体

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
为路由器指定 ID	**router-id** *router-id*	可选 如果 Loopback 和其它接口没有配置 IP 地址，则该任务为必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置对等体	**peer** *ipv6-address* **as-number** *as-number*	必选

1.3.3 配置 IPv6 BGP 发布本地 IPv6 路由

表1-3 配置 IPv6 BGP 发布本地 IPv6 路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
将路由通告到 IPv6 BGP 路由表中	**network** *ipv6-address prefix-length* [**short-cut** | **route-policy** *route-policy-name*]	必选 缺省情况下，IPv6 BGP 不通告任何路由

1.3.4 配置路由首选值

表1-4 配置路由首选值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
为从 IPv6 对等体/IPv6 对等体组接收的路由分配首选值	**peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value*	可选 缺省情况下，从对 IPv6 对等体/IPv6 对等体组接收的路由的首选值为 0

注意

如果同时通过路由策略和 **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* 命令为从对等体/对等体组接收的 IPv6 路由设置首选值，通过路由策略过滤的 IPv6 路由的首选值将取路由策略中设置的首选值而不是 **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* 命令里设置的首选值，只有当路由策略里设置的首选值为 0 时，IPv6 路由的首选值才会取 **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* 命令里设置的值；没有通过路由策略过滤的 IPv6 路由的首选值仍取 **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* 命令里设置的首选值。通过路由策略配置 BGP IPv6 路由信息首选值的相关配置可参考命令 **peer** { *group-name* | *ipv4-address* | *ipv6-address* } **route-policy** *route-policy-name* { **import** | **export** } 和“IP 路由分册/路由策略命令”中的 **apply preferred-value** *preferred-value*。

1.3.5 配置 IPv6 BGP 连接所使用的本地接口

表1-5 配置 IPv6 BGP 连接所使用的本地接口

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
指定与 IPv6 对等体/IPv6 对等体组创建 BGP 会话时建立 TCP 连接使用的源接口	**peer** { *ipv6-group-name* | *ipv6-address* } **connect-interface** *interface-type interface-number*	必选 缺省情况下，IPv6 BGP 使用到达 IPv6 对等体的最佳路由的出接口作为与 IPv6 对等体/IPv6 对等体组创建 BGP 会话时建立 TCP 连接的源接口

说明

- 为了提高 IPv6 BGP 连接的可靠性和稳定性，可将 IPv6 BGP 连接所使用的本地接口配置成 Loopback 接口，这样当网络中存在冗余链路时，不会因为其中某个接口或链路的故障而使 IPv6 BGP 连接中断。
- 当 BGP 对等体之间建立 BGP 连接时，如果没有明确指定建立 TCP 连接的源接口，可能会由于无法根据到达 BGP 对等体的最优路由确定 TCP 连接源接口从而导致无法建立 TCP 连接，因此建议用户在此情况下配置 BGP 对等体时明确配置 BGP 会话建立 TCP 连接的源接口为指定接口。

1.3.6 配置非直接相连的邻居建立 EBGP 连接

表1-6 配置非直接相连的邻居建立 EBGP 连接

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-

操作	命令	说明
配置允许同非直接相连网络上的邻居建立 EBGP 连接	**peer** { *ipv6-group-name* | *ipv6-address* } **ebgp-max-hop** [*hop-count*]	必选 缺省情况下，不允许同非直接相连网络上的邻居建立 EBGP 连接

注意

通常情况下，EBGP 对等体之间必须具有直连的物理链路，如果不满足这一要求，则必须使用 **peer ebgp-max-hop** 命令允许它们之间经过多跳建立 TCP 连接。但是，对于直连 EBGP 使用 Loopback 接口建立邻居关系，不需要 **peer ebgp-max-hop** 命令的配置。

1.3.7 配置 IPv6 对等体/IPv6 对等体组的描述信息

表1-7 配置 IPv6 对等体/IPv6 对等体组的描述信息

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置 IPv6 对等体/IPv6 对等体组的描述信息	**peer** { *ipv6-group-name* | *ipv6-address* } **description** *description-text*	可选 缺省情况下，IPv6 对等体/IPv6 对等体组无描述信息

说明

如果配置对等体组的描述信息，需要先创建对等体组。

1.3.8 禁止与 IPv6 对等体/IPv6 对等体组建立会话

表1-8 禁止与 IPv6 对等体/IPv6 对等体组建立会话

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
禁止与 IPv6 对等体/IPv6 对等体组建立会话	**peer** { *ipv6-group-name* | *ipv6-address* } **ignore**	可选 缺省情况下，允许与 IPv6 BGPIPv6 对等体/IPv6 对等体组建立会话

1.3.9 记录指定 IPv6 对等体/IPv6 对等体组的会话状态和事件信息

表1-9 记录指定 IPv6 对等体/IPv6 对等体组的会话状态和事件信息

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
全局使能 BGP 日志记录功能	**log-peer-change**	可选 缺省情况下，使能 BGP 日志记录功能
进入 IPv6 地址族视图	**ipv6-family**	-
记录指定 IPv6 对等体/IPv6 对等体组的会话状态和事件信息	**peer** { *ipv6-group-name* | *ipv6-address* } **log-change**	可选 缺省情况下，记录 IPv6 对等体/IPv6 对等体组的会话状态和事件信息



说明

有关 **log-peer-change** 命令的详细信息请参见“IP 路由分册”中的“BGP 命令”。

1.4 控制路由信息的发布与接收

控制 IPv6 BGP 的路由信息的发布和接收，包括对路由信息进行过滤，应用路由策略，路由衰减。

1.4.1 配置准备

在控制路由信息的发布与接收之前，需完成以下任务：

- 使能 IPv6 功能
- 配置 IPv6 BGP 基本功能

1.4.2 配置 IPv6 BGP 引入其他路由

表1-10 配置 IPv6 BGP 引入和过滤外部路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
允许将缺省路由引入到 IPv6 BGP 路由表中	**default-route imported**	可选 缺省情况下，IPv6 BGP 不允许将缺省路由引入到 IPv6 BGP 路由表中
配置引入其它协议路由信息并通告	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 缺省情况下，IPv6 BGP 不引入且不通告其它协议的路由

说明

如果没有配置 **default-route imported** 命令，则使用 **import-route** 命令引入 IGP 路由时，不能引入 IGP 的缺省路由。

1.4.3 配置 IPv6 BGP 路由聚合

在中型或大型 BGP 网络中，在向对等体发布 IPv6 路由信息时，可以配置路由聚合，减小对等体路由表中的路由数量。BGP 仅支持 IPv6 路由的手动聚合。

表1-11 配置 IPv6 路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 IPv6 地址族视图	**ipv6-family**	-
配置手动路由聚合	**aggregate** *ipv6-address prefix-length* [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	必选 缺省情况下，不进行路由聚合

1.4.4 配置向 IPv6 对等体/IPv6 对等体组发送缺省路由

表1-12 配置向 IPv6 对等体/IPv6 对等体组发送缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置向 IPv6 对等体/IPv6 对等体组发送缺省路由	**peer** { *ipv6-group-name* | *ipv6-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	必选 缺省情况下，不向 IPv6 对等体/IPv6 对等体组发布缺省路由

说明

执行 **peer default-route-advertise** 命令后，不论本地路由表中是否存在缺省路由，都将向指定 IPv6 对等体/IPv6 对等体组发布一条下一跳地址为本地地址的缺省路由。

1.4.5 配置路由信息的发布策略

表1-13 配置路由信息的发布策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-

操作	命令	说明
对发布的路由进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol process-id*]	必选 缺省情况下，没有对发布的路由信息进行过滤
配置出方向路由策略	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，没有指定 IPv6 对等体/IPv6 对等体组的路由策略
配置基于 IPv6 ACL 的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **export**	必选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组配置基于 IPv6 ACL 的过滤策略
配置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **export**	必选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组配置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略
配置基于 IPv6 地址前缀列表的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **export**	必选 缺省情况下，没有指定 IPv6 对等体/IPv6 对等体组基于 IPv6 前缀列表的过滤策略

说明

IPv6 BGP 对引入的路由信息进行过滤后，会将符合条件的路由信息发布给 IPv6 BGP 对等体（Export）。指定 *protocol* 参数可以只对特定路由协议的信息进行过滤，如果没有指定此参数，则对所有要发布的 IPv6 BGP 路由信息进行过滤，包括引入的路由和使用 **network** 命令发布的路由。

1.4.6 配置路由信息的接收策略

表1-14 配置路由信息的接收策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
对接收的路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	必选 缺省情况下，不对接收的路由信息进行过滤
对接收的路由信息应用路由策略	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **import**	必选 缺省情况下，没有指定 IPv6 对等体/IPv6 对等体组的路由策略
配置基于 IPv6 ACL 的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **import**	必选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组设置基于 IPv6 ACL 的路由过滤策略
配置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **import**	必选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组设置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略

操作	命令	说明
配置基于 IPv6 地址前缀列表的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **import**	必选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组设置基于 IPv6 前缀列表的路由过滤策略
配置允许从 IPv6 对等体/IPv6 对等体组接收的最大 IPv6 地址前缀数	**peer** { *group-name* | *ipv4-address* | *ipv6-address* } **route-limit** *prefix-number* [{ **alert-only** | **reconnect** *reconnect-time* } | *percentage*] *	可选 缺省情况下，允许从 IPv6 对等体/IPv6 对等体组接收的 IPv6 地址最大地址前缀数无限制

说明

- 对 IPv6 BGP 接收的路由进行过滤，只有满足某些条件的路由才能被 IPv6 BGP 接收，并加到路由表中。
- 对等体组的成员可以与所在的组使用不同的入方向路由策略，即接收路由时，各对等体可以选择自己的策略。

1.4.7 配置 IPv6 BGP 与 IGP 路由同步

使能同步特性后，如果一个 AS 由一个非 IPv6 BGP 路由器提供转发服务，那么该 AS 中的 IPv6 BGP 发言者不能对外部 AS 发布路由信息，除非该 AS 中的所有路由器都知道更新的路由信息。

IPv6 BGP 路由器收到一条 IBGP 路由，缺省只检查该路由的下一跳是否可达。如果设置了同步特性，该 IBGP 路由只有在 IGP 也发布了这条路由时才会被同步并发布给 EBGP 对等体。

表1-15 配置 IPv6 BGP 与 IGP 路由同步

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置 IPv6 BGP 与 IGP 路由同步	**synchronization**	必选 缺省情况下，IPv6 BGP 和 IGP 路由不同步

1.4.8 配置路由衰减

表1-16 配置路由衰减

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置 IPv6 BGP 路由衰减	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*]*	可选 缺省情况下，没有配置路由衰减

1.5 配置 IPv6 BGP 的路由属性

本节主要介绍使用各类路由属性来改变 IPv6 BGP 的选路策略。包括如下属性：

- IPv6 BGP 协议的优先级
- 缺省 LOCAL_PREF 属性值
- MED 属性
- NEXT_HOP 属性
- AS_PATH 属性

1.5.1 配置准备

在配置 IPv6 BGP 的路由属性之前，需完成以下任务：

- 使能 IPv6 功能
- 配置 IPv6 BGP 基本功能

1.5.2 配置 IPv6 BGP 路由管理的优先级、缺省 LOCAL_PREF 及 NEXT_HOP 属性

表1-17 配置 IPv6 BGP 协议的优先级、缺省 LOCAL_PREF 及 NEXT_HOP 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
设定 IPv6 BGP 路由管理的优先级	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	可选 缺省情况下，EBGP 路由的管理优先级为 255，IBGP 路由的管理优先级为 255，本地产生的 IPv6 BGP 路由的管理优先级为 130
配置本地优先级的缺省值	**default local-preference** *value*	可选 缺省情况下，本地优先级的缺省值为 100
配置发布路由时将自身地址作为下一跳	**peer** { *ipv6-group-name* | *ipv6-address* } **next-hop-local**	可选 缺省情况下，向 EBGPIPv6 对等体/IPv6 对等体组发布路由时，将自身地址作为下一跳；向 IBGPIPv6 对等体/IPv6 对等体组发布路由时，不将自身地址作为下一跳

说明

- 在某些组网环境中，为保证 IBGP 邻居能够找到正确的下一跳，可以配置在向 IBGPIPv6 对等体/IPv6 对等体组发布路由时，改变下一跳地址为自身地址。如果配置了 IPv6 BGP 负载分担，则不论是否配置了 **peer next-hop-local** 命令，本地路由器向 IBGPIPv6 对等体/IPv6 对等体组发布路由时都先将下一跳地址改变为自身地址。
- 在第三方下一跳（即两个 BGP 连接在同一网段的广播网）这种特殊的组网环境中，缺省情况下，向 EBGPIPv6 对等体/IPv6 对等体组发布路由时，不将自身地址作为下一跳；只有配置了 **peer next-hop-local** 命令，才将自身地址作为下一跳。

1.5.3 配置 MED 属性

表1-18 配置 MED 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置系统的缺省 MED 值	**default med** *med-value*	可选 缺省情况下，系统缺省的 MED 值为 0
配置允许比较来自不同 AS 邻居的路由的 MED 属性值	**compare-different-as-med**	可选 缺省情况下，不允许比较来自不同 AS 邻居的路由的 MED 属性值
配置根据路由来自的 AS 进行分组对 MED 排序优选	**bestroute compare-med**	可选 缺省情况下，不根据路由来自的 AS 进行分组对 MED 排序优选
配置允许比较联盟对等体的路由按 MED 值进行优选	**bestroute med-confederation**	可选 缺省情况下，比较联盟对等体的路由时不考虑 MED 值

1.5.4 配置 AS_PATH 属性

表1-19 配置 AS_PATH 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数	**peer** { *ipv6-group-name* | *ipv6-address* } **allow-as-loop** [*number*]	可选 缺省情况下，不允许本地 AS 号重复
为 IPv6 对等体/IPv6 对等体组配置一个虚拟的本地自治系统号	**peer** { *ipv6-group-name* | *ipv6-address* } **fake-as** *as-number*	可选 缺省情况下，没有为 IPv6 对等体/IPv6 对等体组配置虚拟的本地自治系统号
禁止路由器将 AS_PATH 当作选路算法中的一个因素	**bestroute as-path-neglect**	可选 缺省情况下，路由器可以将 AS_PATH 当作选路算法中的一个因素
配置发送 IPv6 BGP 更新报文时不携带私有 AS 编号	**peer** { *ipv6-group-name* | *ipv6-address* } **public-as-only**	可选 缺省情况下，发送 IPv6 BGP 更新报文时携带私有自治系统号
配置用本地 AS 号替换 AS_PATH 属性中指定 IPv6 对等体/IPv6 对等体组的 AS 号	**peer** { *ipv6-group-name* | *ipv6-address* } **substitute-as**	可选 缺省情况下，没有用本地 AS 号替换 AS_PATH 属性中指定 IPv6 对等体/IPv6 对等体组的 AS 号

1.6 调整和优化 IPv6 BGP 网络

本节主要介绍配置 IPv6 BGP 时钟、软复位 IPv6 BGP 连接和配置最大等价路由的条数。

(1) IPv6 BGP 时钟

当对等体间建立了 IPv6 BGP 连接后，它们定时向对端发送 Keepalive 消息，以防止路由器认为 IPv6 BGP 连接已中断。若路由器在设定的连接保持时间（Holdtime）内未收到对端的 Keepalive 消息或任何其它类型的报文，则认为此 IPv6 BGP 连接已中断，从而退出此 IPv6 BGP 连接。

路由器在与对等体建立 IPv6 BGP 连接时，将比较双方保持时间，以数值较小者做为协商后的保持时间。如果协商结果为 0，则不发送 Keepalive 消息，且不检测 Holdtime 是否超时。

(2) 软复位 IPv6 BGP 连接

IPv6 BGP 的选路策略改变后，为了使新的策略生效，必须复位 IPv6 BGP 连接，但这样会造成短暂的 IPv6 BGP 连接中断。在目前的实现中，IPv6 BGP 支持 Route-refresh 功能。当策略改变后，系统可以在不中断 IPv6 BGP 连接的情况下，自动对 IPv6 BGP 路由表进行动态刷新。

在所有 IPv6 BGP 路由器使能 Route-refresh 功能的情况下，如果 IPv6 BGP 的路由策略发生了变化，本地路由器会向对等体发布 Router-refresh 消息，收到此消息的对等体会将其路由信息重新发给本地 IPv6 BGP 路由器。这样，在不中断 IPv6 BGP 连接的情况下，就可以对 IPv6 BGP 路由表进行动态更新，并应用新的策略。

如果网络中存在有不支持 Route-Refresh 的路由器，则需要配置 **peer keep-all-routes** 命令，将其所有路由更新保存在本地；当路由策略发生变化时，系统将对 IPv6 BGP 路由表进行动态更新，并应用新的策略。

1.6.1 配置准备

在调整 IPv6 BGP 的时钟之前，需完成以下任务：

- 使能 IPv6 能力
- 配置 IPv6 BGP 基本功能

1.6.2 配置 IPv6 BGP 的时钟

表1-20 配置 IPv6 BGP 的时钟

步骤		操作	命令
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	必选
进入 IPv6 地址族视图		**ipv6-family**	-
配置 IPv6 BGP 定时器	配置 IPv6 BGP 的存活时间间隔与保持时间	**timer keepalive** *keepalive* **hold** *holdtime*	可选 缺省情况下，存活时间间隔为 60 秒，保持时间为 180 秒
	配置指定 IPv6 对等体/IPv6 对等体组的存活和保持时间	**peer** { *ipv6-group-name* | *ipv6-address* } **timer keepalive** *keepalive* **hold** *holdtime*	

步骤	操作	命令
配置 IPv6 对等体/IPv6 对等体组的发送同一路由更新报文的时间间隔	**peer** { *ipv6-group-name* | *ipv6-address* } **route-update-interval** *interval*	可选 缺省情况下，向 IBGP 对等体发送同一路由更新的时间间隔为 15 秒，向 EBGP 对等体发送同一路由更新的时间间隔为 30 秒



说明

- 使用 **timer** 命令配置的定时器比使用 **peer timer** 命令配置的定时器优先级要低。
- 设置的保持时间应该至少为存活时间间隔的三倍。

1.6.3 配置 IPv6 BGP 软复位

1. 使能 Route-refresh 功能

表1-21 使能 Route-refresh 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
使能 IPv6 BGP 路由刷新功能	**peer** { *ipv6-group-name* | *ipv6-address* } **capability-advertise route-refresh**	可选 缺省情况下，使能 IPv6 BGP 路由刷新功能

2. 手工软复位 IPv6 BGP

表1-22 手工软复位 IPv6 BGP

步骤	操作	命令
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
保存所有来自 IPv6 对等体/IPv6 对等体组的原始路由信息，即使这些路由没有通过已配置的入口策略	**peer** { *ipv6-group-name* | *ipv6-address* } **keep-all-routes**	可选 缺省情况下，不保存 IPv6 对等体/IPv6 对等体组的原始路由信息
返回用户视图	**return**	-
手工对 IPv6 BGP 连接进行软复位	**refresh bgp ipv6** { **all** | *ipv6-address* | **group** *ipv6-group-name* | **external** | **internal** } { **export** | **import** }	必选

说明

配置命令 **peer keep-all-routes** 后，不论是否使用了过滤策略，都将保存指定对等体发来的所有路由更新。当软复位 IPv6 BGP 连接时，这些信息可以用来重新生成 IPv6 BGP 路由。

1.6.4 配置最大等价路由的条数

表1-23 配置最大等价路由的条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置进行 IPv6 BGP 负载分担的路由条数	**balance** *number*	必选 缺省情况下，不进行负载分担

1.7 组建大型 IPv6 BGP 网络

为方便管理，减少重复配置操作，管理员将一些享有相同更新策略的 IPv6 BGP 对等体划分到一个逻辑组织，这些组织就称为对等体组。只须对一个等体组配置策略，该策略对组内所有成员都有效。

当对等体组的配置变化时，每个组员的配置也相应变化。对某些属性可以指定组员的 IPv6 地址进行配置，从而使指定组员在这些属性上不受对等体组配置的影响。

通常将同一自治系统的对等体配置为一个对等体组，当然，也可以将其它自治系统的对等体加入到该组。所有 IBGP 对等体可配置为另一个对等体组，也可以根据业务逻辑创建对等体组。

对等体组可以使一组对等体共享相同的策略，而利用团体可以使多个 AS 中的一组 IPv6 BGP 路由器共享相同的策略。团体是一个路由属性，在 IPv6 BGP 对等体之间传播，它并不受到 AS 范围的限制。

为保证 IBGP 对等体之间的连通性，需要在 IBGP 对等体之间建立全连接关系。当 IBGP 对等体数目很多时，建立全连接网的开销很大。路由反射器和联盟是解决这个问题的两种方法。

IPv6 BGP 联盟的配置方式与 BGP 相同，这里不再赘述。以下主要介绍：

- 配置 IPv6 BGP 对等体组
- 配置 IPv6 BGP 团体
- 配置 IPv6 BGP 路由反射器

1.7.1 配置准备

在配置 IPv6 BGP 对等体组之前，需完成以下任务：

- 相邻节点网络层互通
- 启动 BGP，配置 Router ID

1.7.2 配置 IPv6 BGP 对等体组

1. 创建 IBGP 对等体组

表1-24 创建 IBGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
进入 IPv6 地址族视图	**ipv6-family**	-
创建 IBGP 对等体组	**group** *ipv6-group-name* [**internal**]	必选
将对等体加入已存在的对等体组	**peer** *ipv6-address* **group** *ipv6-group-name* [**as-number** *as-number*]	必选 缺省情况下，对等体不属于任何对等体组

2. 创建纯 EBGP 对等体组

表1-25 创建纯 EBGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
进入 IPv6 地址族视图	**ipv6-family**	-
创建 EBGP 对等体组	**group** *ipv6-group-name* **external**	必选
配置对等体组的 AS 号	**peer** *ipv6-group-name* **as-number** *as-number*	必选 缺省情况下，对等体组无 AS 号
在已创建的对等体组中加入 IPv6 对等体	**peer** *ipv6-address* **group** *ipv6-group-name*	必选 缺省情况下，对等体不属于任何对等体组

说明

- 在创建纯 EBGP 对等体时，需要指定对等体组的自治系统号。
- 如果对等体组中已经加入了对等体，那么不能够为该对等体组指定自治系统号。

3. 创建混合 EBGP 对等体组

表1-26 创建混合 EBGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
进入 IPv6 地址族视图	**ipv6-family**	-
创建 EBGP 对等体组	**group** *ipv6-group-name* **external**	必选

操作	命令	说明
设置 IPv6 对等体的 AS 号	**peer** *ipv6-address* **as-number** *as-number*	必选 缺省情况下，IPv6 对等体无 AS 号
加入已创建的 IPv6 对等体	**peer** *ipv6-address* **group** *ipv6-group-name*	必选 缺省情况下，对等体不属于任何对等体组



在创建混合 EBGP 对等体组时，需要单独创建对等体，并可设置不同的自治系统号，但不能设置对等体组的自治系统号。

1.7.3 配置 IPv6 BGP 团体

1. 配置向 IPv6 对等体/IPv6 对等体组发布团体属性

表1-27 配置向 IPv6 对等体/IPv6 对等体组发布团体属性

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
进入 IPv6 地址族视图	**ipv6-family**	-
配置将团体属性传给 IPv6 对等体/IPv6 对等体组	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-community**	必选 缺省情况下，不将团体属性发布给任何 IPv6 对等体/IPv6 对等体组
配置将扩展团体属性传给 IPv6 对等体/IPv6 对等体组	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-ext-community**	必选 缺省情况下，不将扩展团体属性发布给任何 IPv6 对等体/IPv6 对等体组

2. 对发布的路由信息应用路由策略

表1-28 对发布的路由信息应用路由策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置出方向路由策略	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，不指定 IPv6 对等体/IPv6 对等体组的路由策略

说明

- 配置 IPv6 BGP 团体时，必须使用路由策略来定义具体的团体属性，然后在发布路由信息时应用此路由策略。
- 关于路由策略的配置，请参见“IP 路由分册”中的“路由策略配置”。

1.7.4 配置 IPv6 BGP 路由反射器

表1-29 配置 IPv6 BGP 路由反射器

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 IPv6 地址族视图	**ipv6-family**	-
配置将本机作为路由反射器，并将 IPv6 对等体/IPv6 对等体组作为路由反射器的客户	**peer** { *ipv6-group-name* | *ipv6-address* } **reflect-client**	必选 缺省情况下，没有配置路由反射器及其客户
使能客户机之间的路由反射	**reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
配置路由反射器的集群 ID	**reflector cluster-id** *cluster-id*	可选 缺省情况下，每个路由反射器是使用自己的 Router ID 作为集群 ID

说明

- 通常情况下，路由反射器的客户之间不要求是全连接的，路由缺省通过反射器从一个客户反射到其它客户；如果客户之间是全连接的，可以禁止客户间的反射，以便减少开销。
- 当一个集群里有多个路由反射器时，需要给所有位于同一个集群内的路由反射器配置相同的 Cluster_ID，以避免路由循环。

1.8 IPv6 BGP 显示和维护

1.8.1 IPv6 BGP 显示

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 BGP 的运行情况，通过查看显示信息验证配置的效果。

表1-30 显示 IPv6 BGP 的运行状态

操作	命令
显示对等体组信息	**display bgp ipv6 group** [*ipv6-group-name*]
显示 IPv6 BGP 发布的 IPv6 路由信息	**display bgp ipv6 network**
显示 AS 路径信息	**display bgp ipv6 paths** [*as-regular-expression*]
显示 IPv6 对等体/IPv6 对等体组信息	**display bgp ipv6 peer** [*group-name* **log-info** | *ipv4-address* **verbose** | *ipv6-address* { **log-info** | **verbose** } | **verbose**]

操作	命令
显示 IPv6 BGP 路由信息	**display bgp ipv6 routing-table** [*ipv6-address prefix-length*]
显示匹配指定 AS 路径过滤列表的路由	**display bgp ipv6 routing-table as-path-acl** *as-path-acl-number*
显示指定 IPv6 BGP 团体的路由信息	**display bgp ipv6 routing-table community** [*aa:nn*<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**]* [**whole-match**]
显示匹配指定 IPv6 BGP 团体列表的路由信息	**display bgp ipv6 routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>
显示 IPv6 BGP 衰减的路由	**display bgp ipv6 routing-table dampened**
显示 IPv6 BGP 路由衰减参数	**display bgp ipv6 routing-table dampening parameter**
显示来自不同自治系统的 IPv6 BGP 路由	**display bgp ipv6 routing-table different-origin-as**
显示 IPv6 BGP 路由振荡统计信息	**display bgp ipv6 routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *network-address* [*prefix-length* [**longer-match**]]]
显示 IPv6 BGP 的 IPv6 标签路由信息	**display bgp ipv6 routing-table label**
显示向指定的 IPv4/IPv6 BGP 对等体发送或者从指定的 IPv4/IPv6 BGP 对等体收到的路由信息	**display bgp ipv6 routing-table peer** { *ipv4-address* | *ipv6-address* } { **advertised-routes** | **received-routes** } [*network-address prefix-length* | **statistic**]
显示匹配 AS 正则表达式的 IPv6 路由信息	**display bgp ipv6 routing-table regular-expression** *as-regular-expression*
显示 IPv6 BGP 的路由统计信息	**display bgp ipv6 routing-table statistic**

1.8.2 复位 IPv6 BGP 连接

当 BGP 路由策略或协议发生变化后，如果需要通过复位 IPv6 BGP 连接使新的配置生效，请在用户视图下进行下列配置。

表1-31 复位 IPv6 BGP 连接

操作	命令
手工软复位 IPv6 BGP 连接	**refresh bgp ipv6** { *ipv4-address* | *ipv6-address* | **all** | **external** | **group** *ipv6-group-name* | **internal** } { **export** | **import** }
复位所有 IPv6 BGP 连接	**reset bgp ipv6** { *as-number* | *ipv4-address* | *ipv6-address* [**flap-info**] | **all** | **external** | **group** *group-name* | **internal** }

1.8.3 清除 IPv6 BGP 信息

在用户视图下执行 **reset** 命令可清除 IPv6 BGP 路由的统计信息。

表1-32 清除 IPv6 BGP 信息

操作	命令
清除 IPv6 BGP 路由的衰减信息并释放被抑制的路由	**reset bgp ipv6 dampening** [*ipv6-address prefix-length*]
清除 IPv6 BGP 路由的振荡统计信息	**reset bgp ipv6 flap-info** [*ipv6-address/prefix-length* | **as-path-acl** *as-path-acl-number* | **regexp** *as-path-regexp*]]

1.9 IPv6 BGP 典型配置举例

说明

IPv6 BGP 的配置举例多数与 BGP 的配置举例相似，本节不再一一列出。其他配置举例请参见“IP 路由分册”中的“BGP 配置”。

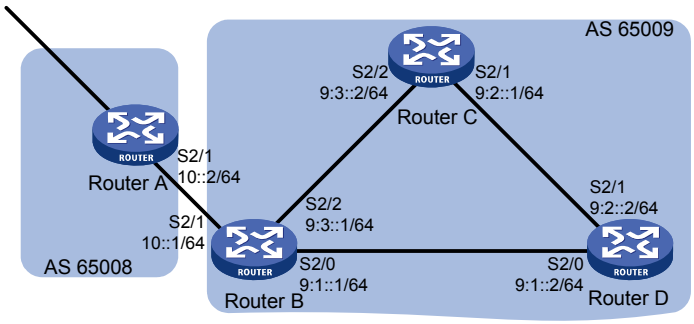
1.9.1 IPv6 BGP 基本配置

1. 组网需求

- 如 图 1-1所示，所有路由器均运行IPv6 BGP协议；
- Router A 和 Router B 之间建立 EBGP 连接；
- Router B、Router C 和 Router D 之间建立 IBGP 全连接。

2. 组网图

图1-1 IPv6 BGP 基本配置组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 IBGP 连接

配置 Router B。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] bgp 65009
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] ipv6-family
[RouterB-bgp-af-ipv6] peer 9:1::2 as-number 65009
[RouterB-bgp-af-ipv6] peer 9:3::2 as-number 65009
[RouterB-bgp-af-ipv6] quit
[RouterB-bgp] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] bgp 65009
[RouterC-bgp] router-id 3.3.3.3
```

```
[RouterC-bgp] ipv6-family
[RouterC-bgp-af-ipv6] peer 9:3::1 as-number 65009
[RouterC-bgp-af-ipv6] peer 9:2::2 as-number 65009
[RouterC-bgp-af-ipv6] quit
[RouterC-bgp] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ipv6
[RouterD] bgp 65009
[RouterD-bgp] router-id 4.4.4.4
[RouterD-bgp] ipv6-family
[RouterD-bgp-af-ipv6] peer 9:1::1 as-number 65009
[RouterD-bgp-af-ipv6] peer 9:2::1 as-number 65009
[RouterD-bgp-af-ipv6] quit
[RouterD-bgp] quit
```

(3) 配置 EBGP 连接

配置 Router A。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] bgp 65008
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] ipv6-family
[RouterA-bgp-af-ipv6] peer 10::1 as-number 65009
[RouterA-bgp-af-ipv6] quit
[RouterA-bgp] quit
```

配置 Router B。

```
[RouterB] bgp 65009
[RouterB-bgp] ipv6-family
[RouterB-bgp-af-ipv6] peer 10::2 as-number 65008
[RouterB-bgp-af-ipv6] quit
[RouterB-bgp] quit
```

查看 Router B 的对等体信息。

```
[RouterB] display bgp ipv6 peer

 BGP local router ID : 2.2.2.2
 Local AS number : 65009
 Total number of peers : 3                 Peers in established state : 3

  Peer       V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  10::2      4 65008        3        3     0       0 00:01:16 Established
  9:3::2     4 65009        2        3     0       0 00:00:40 Established
  9:1::2     4 65009        2        4     0       0 00:00:19 Established
```

查看 Router C 的对等体信息。

```
[RouterC] display bgp ipv6 peer

 BGP local router ID : 3.3.3.3
 Local AS number : 65009
 Total number of peers : 2                 Peers in established state : 2

  Peer       V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  9:3::1     4 65009        4        4     0       0 00:02:18 Established
  9:2::2     4 65009        4        5     0       0 00:01:52 Established
```

可以看出，Router A 和 Router B 之间建立的 EBGP 连接，Router B、Router C 和 Router D 之间建立了 IBGP 连接。

1.9.2 配置 IPv6 BGP 路由反射

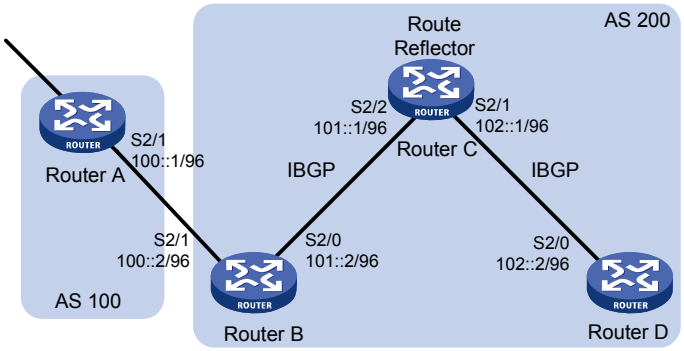
1. 组网需求

Router B 接收了一个经过 EBGP 的更新报文并将之传给 Router C。Router C 被配置为路由反射器，它有两个客户机：Router B 和 Router D。

Router B 和 Router D 间不需要建立 IBGP 连接，当 Router C 接收了来自 Router B 的路由更新时，它将此信息反射给 Router D；反之，接收到来自 Router D 的路由信息也会被反射给 Router B。

2. 组网图

图1-2 配置 IPv6 BGP 路由反射器组网图



错误！不能识别的开关参数。

3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 IPv6 BGP 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] bgp 100
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] ipv6-family
[RouterA-bgp-af-ipv6] peer 100::2 as-number 200
[RouterA-bgp-af-ipv6] network 1:: 64
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] bgp 200
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] ipv6-family
[RouterB-bgp-af-ipv6] peer 100::1 as-number 100
[RouterB-bgp-af-ipv6] peer 101::1 as-number 200
[RouterB-bgp-af-ipv6] peer 101::1 next-hop-local
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] bgp 200
[RouterC-bgp] router-id 3.3.3.3
[RouterC-bgp] ipv6-family
[RouterC-bgp-af-ipv6] peer 101::2 as-number 200
[RouterC-bgp-af-ipv6] peer 102::2 as-number 200
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ipv6
[RouterD] bgp 200
[RouterD-bgp] router-id 4.4.4.4
[RouterD-bgp] ipv6-family
[RouterD-bgp-af-ipv6] peer 102::1 as-number 200
```

(3)　配置路由反射器

配置 Router C 作为路由反射器，Router B 和 Router D 是它的两个客户机。

```
[RouterC-bgp-af-ipv6] peer 101::2 reflect-client
[RouterC-bgp-af-ipv6] peer 102::2 reflect-client
```

在 Router B 上用命令 **display bgp ipv6 routing-table**，可以看到 Router B 已知道网络 1::/64 的存在。

在 Router D 上用命令 **display bgp ipv6 routing-table**，可以看到 Router D 也知道网络 1::/64 的存在。

1.10 IPv6 BGP 常见错误配置举例

1.10.1 IPv6 BGP 对等体关系不能建立

1. 错误现象

使用 **display bgp ipv6 peer** 命令查看 IPv6 BGP 对等体的信息，发现与对端的连接无法进入 Established 状态。

2. 分析

IPv6 BGP 邻居的建立需要能够使用 179 端口建立 TCP 会话，以及能够正确交换 Open 消息。

3. 处理过程

(1) 用 **display current-configuration** 命令检查邻居的 AS 号配置是否正确。

(2) 用 **display bgp ipv6 peer** 命令检查邻居的 IPv6 地址是否正确。

(3) 如果使用 Loopback 接口，检查是否配置了 **peer connect-interface** 命令。

(4) 如果是物理上非直连的 EBGP 邻居，检查是否配置了 **peer ebgp-max-hop** 命令。

(5) 检查路由表中是否存在到邻居的可用路由。

(6) 使用 **ping** 命令检查链路是否畅通。

(7) 使用 **display tcp ipv6 status** 命令检查 TCP 连接是否正常。

(8) 检查是否配置了禁止 TCP 端口 179 的 ACL。

目 录

1 IPv6 IS-IS配置 1-1
1.1 IPv6 IS-IS简介 1-1
1.2 配置IPv6 IS-IS的基本特性 1-1
1.2.1 配置准备 1-1
1.2.2 配置IS-IS的IPv6 基本特性 1-2
1.3 配置IPv6 IS-IS的路由信息控制 1-2
1.3.1 配置准备 1-2
1.3.2 配置IPv6 IS-IS的路由信息控制 1-2
1.4 IPv6 IS-IS显示和维护 1-3
1.5 IPv6 IS-IS典型配置举例 1-4

1 IPv6 IS-IS 配置

说明

IPv6 IS-IS 实现了 IPv4 IS-IS 的所有功能，与 IPv4 IS-IS 的区别在于发布的是 IPv6 路由信息，本章只列出了 IPv6 IS-IS 专有的配置任务，其他相关配置任务请参见“IP 路由分册”中的“IS-IS 配置”。

1.1 IPv6 IS-IS 简介

IS-IS（Intermediate System-to-Intermediate System intra-domain routing information exchange protocol，IS-IS 路由协议）支持多种网络层协议，其中包括 IPv6 协议，支持 IPv6 协议的 IS-IS 路由协议又称为 IPv6 IS-IS 动态路由协议。IETF 的 draft-ietf-isis-ipv6-05 中规定了 IS-IS 为支持 IPv6 所新增的内容，主要是新添加的支持 IPv6 协议的两个 TLVs（Type-Length-Values）和一个新的 NLPID（Network Layer Protocol Identifier，网络层协议标识符）。

TLV 是 LSP（Link State PDU 或 Link State Packet）中的一个可变长字段值。新增的两个 TLV 分别是：

- IPv6 Reachability：类型值为 236（0xEC），通过定义路由信息前缀、度量值等信息来说明网络的可达性。
- IPv6 Interface Address：类型值为 232（0xE8），它对应于 IPv4 中的“IP Interface Address”TLV，只不过把原来的 32 比特的 IPv4 地址改为 128 比特的 IPv6 地址。

NLPID 是标识网络层协议报文的一个 8 比特字段，IPv6 的 NLPID 值为 142（0x8E）。如果 IS-IS 路由器支持 IPv6，那么它必须以这个 NLPID 值向外发布路由信息。

IS-IS 的相关知识请参见“IP 路由分册”中的“IS-IS 配置”。

1.2 配置 IPv6 IS-IS 的基本特性

说明

在 IPv6 网络环境中，可以通过配置 IPv6 IS-IS 路由协议来实现 IPv6 网络的互连。

1.2.1 配置准备

在配置之前，需完成以下任务：

- 全局使能 IPv6 能力
- 配置接口的网络层地址，使各相邻节点网络层可达
- 启动 IS-IS

1.2.2 配置 IS-IS 的 IPv6 基本特性

表1-1 配置 IS-IS 的 IPv6 基本特性

操作	命令	说明
进入系统视图	**system-view**	-
启动 IS-IS 路由进程，进入 IS-IS 视图	**isis** [*process-id*]	必选 缺省情况下，系统没有运行 IS-IS
设置 IS-IS 路由进程的网络实体名称（NET）	**network-entity** *net*	必选 缺省情况下，没有配置 NET
使能 IS-IS 进程的 IPv6 能力	**ipv6 enable**	必选 缺省情况下，没有使能 IS-IS 路由进程的 IPv6 能力
退回到系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能接口 IS-IS 路由进程的 IPv6 能力并指定要关联的 IS-IS 进程号	**isis ipv6 enable** [*process-id*]	必选 缺省情况下，接口上没有使能 IS-IS 路由进程的 IPv6 能力

1.3 配置 IPv6 IS-IS 的路由信息控制

1.3.1 配置准备

在进行 IPv6 IS-IS 的路由特性配置之前，需完成 IPv6 IS-IS 基本配置。

1.3.2 配置 IPv6 IS-IS 的路由信息控制

表1-2 配置 IPv6 IS-IS 的路由信息控制

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*]	-
配置 IPv6 IS-IS 路由的优先级	**ipv6 preference** { **route-policy** *route-policy-name* | *preference* } *	可选 缺省情况下，IPv6 IS-IS 路由的优先级为 15
配置 IPv6 IS-IS 聚合路由	**ipv6 summary** *ipv6-prefix prefix-length* [**avoid-feedback** | **generate_null0_route** | [**level-1** | **level-1-2** | **level-2**] | **tag** *tag*] *	可选 缺省情况下，没有 IPv6 聚合路由
配置生成 IPv6 IS-IS 缺省路由	**ipv6 default-route-advertise** [[**level-1** | **level-2** | **level-1-2**] | **route-policy** *route-policy-name*] *	可选 缺省情况下，不生成 IPv6 IS-IS 缺省路由
配置 IPv6 IS-IS 对接收的路由进行过滤	**ipv6 filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** *route-policy-name* } **import**	可选 缺省情况下，IPv6 IS-IS 不对接收的路由信息进行过滤
配置 IPv6 IS-IS 引入其他协议的路由信息	**ipv6 import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | [**level-1** | **level-2** | **level-1-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	可选 缺省情况下，IPv6 IS-IS 不引入其他协议的路由信息

操作	命令	说明
配置引入 Level1/Level2 的 IPv6 路由最大条数	**ipv6 import-route limit** *number*	可选 本命令的缺省值为 300000
配置 IPv6 IS-IS 对引入的路由进行过滤	**ipv6 filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]	可选 缺省情况下，IPv6 IS-IS 不对引入的路由进行过滤
使能 IPv6 IS-IS 路由渗透功能	**ipv6 import-route isisv6 level-2 into level-1** [**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *	可选 缺省情况下，禁止 IPv6 IS-IS 路由渗透功能
配置在负载分担方式下 IPv6 IS-IS 等价路由的最大数量	**ipv6 maximum load-balancing** *number*	可选 本命令的缺省值为 8

说明

- **ipv6 filter-policy export** 命令一般和 **ipv6 import-route** 命令结合使用，如果没有指定 *protocol* 参数，将对当前路由器从所有协议引入的路由进行过滤；如果指定了 *protocol* 参数，则只对当前路由器从特定协议引入的路由进行过滤。
- ACL 的相关知识请参见“安全分册”中的“ACL 配置”。
- 路由策略、IPv6 地址前缀列表的相关知识请参见“IP 路由分册”中的“路由策略配置”。

1.4 IPv6 IS-IS 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 IS-IS 的运行情况，用户可以通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv6 IS-IS 的数据库信息，或者复位特定邻居的数据信息。

表1-3 IPv6 IS-IS 显示和维护

操作	命令
显示 IS-IS 的摘要信息	**display isis brief**
显示 IS-IS 当前 debug 开关的状态	**display isis debug-switches** { *process-id* | **vpn-instance** *vpn-instance-name* }
显示使能了 IS-IS 的接口信息	**display isis interface** [**statistics** | [*interface-type interface-number*] [**verbose**]] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 的许可证信息	**display isis license**
显示 IS-IS 的链路状态数据库	**display isis lsdb** [[**l1** | **l2** | **level-1** | **level-2**] | [[**lsp-id** *lsp-id* | **lsp-name** *lspname* | **local**] | **verbose**] *] * [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 的 mesh-group	**display isis mesh-group** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示本地路由器名称到系统 ID 的映射关系表	**display isis name-table** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 的邻居信息	**display isis peer** [**verbose | statistics**] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IPv6 IS-IS 路由信息	**display isis route ipv6** [[**level-1** | **level-2**] | **verbose**] * [*process-id*]

操作	命令
显示 IS-IS 的 SPF 计算日志记录	**display isis spf-log** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS 进程的统计信息	**display isis statistics** [**level-1** | **level-2** | **level-1-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]
清除所有 IS-IS 的数据结构信息	**reset isis all** [*process-id* | **vpn-instance** *vpn-instance-name*]
清除 IS-IS 特定邻居的数据信息	**reset isis peer** *system-id* [*process-id* | **vpn** *vpn-instance-name*]

1.5 IPv6 IS-IS 典型配置举例

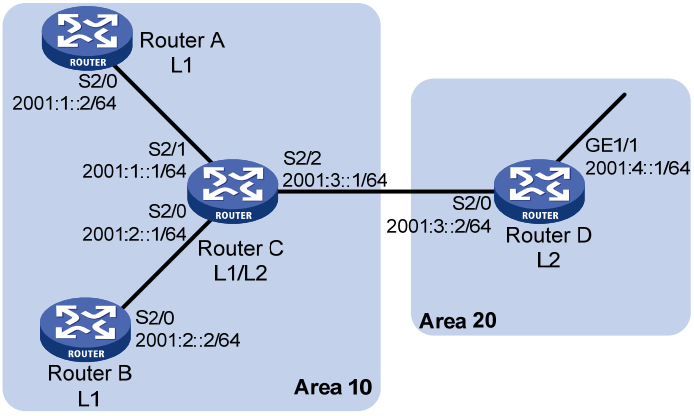
1. 组网需求

如下图所示，Router A、Router B、Router C 和 Router D 属于同一自治系统，所有路由器已使能了 IPv6 能力，要求他们之间通过 IPv6 IS-IS 协议达到 IPv6 网络互连的目的。

其中 Router A 和 Router B 是 Level-1 路由器，Router D 是 Level-2 路由器，Router C 是 Level-1-2 路由器。Router A、Router B 和 Router C 属于区域 10，而 Router D 属于区域 20。

2. 组网图

图1-1 IPv6 IS-IS 基本配置组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 IPv6 IS-IS

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] is-level level-1
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] ipv6 enable
[RouterA-isis-1] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] isis ipv6 enable 1
[RouterA-Serial2/0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] is-level level-1
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] ipv6 enable
[RouterB-isis-1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] isis ipv6 enable 1
[RouterB-Serial2/0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 10.0000.0000.0003.00
[RouterC-isis-1] ipv6 enable
[RouterC-isis-1] quit
[RouterC] interface serial 2/0
[RouterC-Serial2/0] isis ipv6 enable 1
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] isis ipv6 enable 1
[RouterC-Serial2/1] quit
[RouterC] interface serial 2/2
[RouterC-Serial2/2] isis ipv6 enable 1
[RouterC-Serial2/2] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] is-level level-2
[RouterD-isis-1] network-entity 20.0000.0000.0004.00
[RouterD-isis-1] ipv6 enable
[RouterD-isis-1] quit
[RouterD] interface serial 2/0
[RouterD-Serial2/0] isis ipv6 enable 1
[RouterD-Serial2/0] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] isis ipv6 enable 1
[RouterD-GigabitEthernet1/1] quit
```

目 录

1 OSPFv3 配置 1-1
1.1 OSPFv3 简介 1-1
1.1.1 OSPFv3 概述 1-1
1.1.2 OSPFv3 的协议报文 1-1
1.1.3 OSPFv3 的LSA类型 1-2
1.1.4 OSPFv3 的定时器 1-2
1.1.5 系统支持的OSPFv3 特性 1-3
1.1.6 协议规范 1-3
1.2 OSPFv3 配置任务简介 1-4
1.3 使能OSPFv3 功能 1-4
1.3.1 配置准备 1-4
1.3.2 使能OSPFv3 功能 1-4
1.4 配置OSPFv3 的区域属性 1-5
1.4.1 配置准备 1-5
1.4.2 配置OSPFv3 的Stub区域 1-5
1.4.3 配置OSPFv3 的虚连接 1-6
1.5 配置OSPFv3 网络类型 1-6
1.5.1 配置准备 1-6
1.5.2 配置OSPFv3 接口的网络类型 1-7
1.5.3 配置NBMA或者P2MP网络的邻居 1-7
1.6 配置OSPFv3 的路由信息控制 1-7
1.6.1 配置准备 1-7
1.6.2 配置OSPFv3 的路由聚合 1-7
1.6.3 配置OSPFv3 对通过接收到的LSA计算出来的路由信息进行过滤 1-8
1.6.4 配置OSPFv3 接口的开销值 1-8
1.6.5 配置OSPFv3 最大等价路由条数 1-9
1.6.6 配置OSPFv3 协议的优先级 1-9
1.6.7 配置OSPFv3 引入外部路由 1-10
1.7 调整和优化OSPFv3 网络 1-10
1.7.1 配置准备 1-11
1.7.2 配置OSPFv3 定时器 1-11
1.7.3 配置接口的DR优先级 1-12
1.7.4 忽略DD报文中的MTU检查 1-12
1.7.5 禁止接口发送OSPFv3 报文 1-12
1.7.6 配置邻接状态变化输出开关 1-13
1.8 配置OSPFv3 GR 1-13
1.8.1 配置GR Restarter 1-13
1.8.2 配置GR Helper 1-13
1.9 OSPFv3 显示和维护 1-14
1.10 OSPFv3 配置举例 1-15
1.10.1 配置OSPFv3 区域 1-15

1.10.2 配置OSPFv3 的DR选择 1-18
1.10.3 配置OSPFv3 引入外部路由 1-21
1.10.4 配置OSPFv3 GR 1-24
1.11 常见配置错误举例 1-26
1.11.1 OSPFv3 邻居无法建立 1-26
1.11.2 OSPFv3 路由信息不正确 1-26

说明

SR6602 不支持配置 GR Restarter。

1 OSPFv3 配置

1.1 OSPFv3 简介

1.1.1 OSPFv3 概述

OSPFv3 是 OSPF（Open Shortest Path First，开放式最短路径优先）版本 3 的简称，主要提供对 IPv6 的支持，遵循的标准为 RFC 2740（OSPF for IPv6）。

OSPFv3 和 OSPFv2 在很多方面是相同的：

- Router ID，Area ID 仍然是 32 位的。
- 相同类型的报文：Hello 报文，DD（Database Description，数据库描述）报文，LSR（Link State Request，链路状态请求）报文，LSU（Link State Update，链路状态更新）报文和 LSAck（Link State Acknowledgment，链路状态确认）报文。
- 相同的邻居发现机制和邻接形成机制。
- 相同的 LSA 扩散机制和老化机制。

OSPFv3 和 OSPFv2 的不同主要有：

- OSPFv3 是基于链路（Link）运行，OSPFv2 是基于网段（Network）运行。
- OSPFv3 在同一条链路上可以运行多个实例。
- OSPFv3 是通过 Router ID 来标识邻接的邻居。OSPFv2 则是通过 IP 地址来标识邻接的邻居。

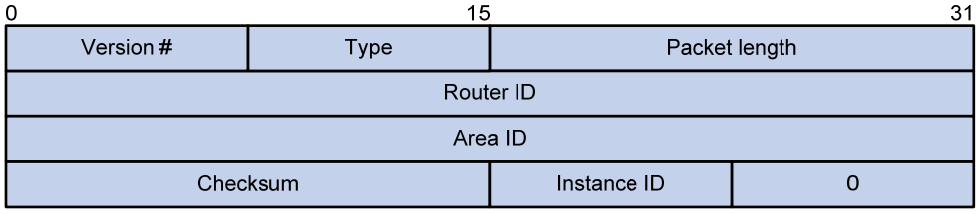
1.1.2 OSPFv3 的协议报文

和 OSPFv2 一样，OSPFv3 也有五种报文类型，分别是 Hello 报文、DD 报文、LSR 报文、LSU 报文和 LSAck 报文。

这五种报文有相同的报文头，但是它和 OSPFv2 的报文头有一些区别，其长度只有 16 字节，且没有认证字段。另外就是多了一个 Instance ID 字段，用来支持在同一条链路上运行多个实例。

OSPFv3 的报文头如 图 1-1所示。

图1-1 OSPFv3 报文头示意图



主要字段的解释如下：

- Version #：OSPF 的版本号。对于 OSPFv3 来说，其值为 3。
- Type：OSPF 报文的类型。数值从 1 到 5，分别对应 Hello 报文、DD 报文、LSR 报文、LSU 报文和 LSAck 报文。
- Packet Length：OSPF 报文的总长度，包括报文头在内，单位为字节。
- Instance ID：同一条链路上的实例标识。
- 0：保留位，必须为 0。

1.1.3 OSPFv3 的 LSA 类型

LSA（Link State Advertisement，链路状态通告）是 OSPFv3 协议计算和维护路由信息的主要来源。在 RFC2740 中定义了七类 LSA，描述如下：

- Router-LSA：由每个路由器生成，描述本路由器的链路状态和开销，只在路由器所处区域内传播。
- Network-LSA：由广播网络和 NBMA（Non-Broadcast Multi-Access）网络的 DR（Designated Router，指定路由器）生成，描述本网段接口的链路状态，只在 DR 所处区域内传播。
- Inter-Area-Prefix-LSA：和 OSPFv2 中的 Type-3 LSA 类似，该 LSA 由 ABR（Area Border Router，区域边界路由器）生成，在与该 LSA 相关的区域内传播。每一条 Inter-Area-Prefix-LSA 描述了一条到达本自治系统内其他区域的 IPv6 地址前缀（IPv6 Address Prefix）的路由。
- Inter-Area-Router-LSA：和 OSPFv2 中的 Type-4 LSA 类似，该 LSA 由 ABR 生成，在与该 LSA 相关的区域内传播。每一条 Inter-Area-Router-LSA 描述了一条到达本自治系统内的 ASBR（Autonomous System Border Router，自治系统边界路由器）的路由。
- AS-external-LSA：由 ASBR 生成，描述到达其它 AS（Autonomous System，自治系统）的路由，传播到整个 AS（Stub 区域除外）。缺省路由也可以用 AS-external-LSA 来描述。
- Link-LSA：路由器为每一条链路生成一个 Link-LSA，在本地链路范围内传播。每一个 Link-LSA 描述了该链路上所连接的 IPv6 地址前缀及路由器的 Link-local 地址。
- Intra-Area-Prefix-LSA：每个 Intra-Area-Prefix-LSA 包含路由器上的 IPv6 前缀信息，Stub 区域信息或穿越区域（Transit Area）的网段信息，该 LSA 在区域内传播。由于 Router-LSA 和 Network-LSA 不再包含地址信息，导致了 Intra-Area-Prefix-LSA 的引入。

在 RFC 5187 中定义了第十一类 LSA，Grace-LSA。Grace-LSA 由 Restarter 在重启时候生成的，在本地链路范围内传播。这个 LSA 描述了重启设备的重启原因和重启时间间隔，目的是通知邻居本设备将进入 GR（Graceful Restart，平滑重启）。

1.1.4 OSPFv3 的定时器

OSPFv3 的定时器包括：

- OSPFv3 的报文定时器。
- LSA 的延迟时间。
- SPF 定时器。
- GR 定时器。

1. OSPFv3 报文定时器

Hello 报文周期性地被发送至邻居路由器，用于发现与维持邻居关系、选举 DR 与 BDR。需要注意的是，网络邻居间的 Hello 时间间隔必须一致，并且 Hello 时钟的值与路由收敛速度、网络负荷大小成反比。

在一定时间间隔内，如果路由器未收到对方的 Hello 报文，则认为对端路由器失效，这个时间间隔被称为相邻路由器间的失效时间。

当一台路由器向它的邻接发送一条 LSA 后，需要等到对方的确认报文。若在设定的重传间隔时间内没有收到对方的确认报文，就会向邻接重传这条 LSA。重传间隔的值必须大于一个报文在两台路由器之间传送一个来回的时间。

2. LSA 的延迟时间

由于 LSA 在本路由器的 LSDB（Link State Database，链路状态数据库）中会随时间老化（每秒加 1），但在网络的传输过程中却不会随时间老化，所以有必要在发送之前就将 LSA 的老化时间增加上传送延迟时间。对于低速网络，该项配置尤为重要。

3. SPF 定时器

当 OSPFv3 的 LSDB 发生改变时，需要重新计算最短路径，如果每次改变都立即计算最短路径，将占用大量资源，并会影响路由器的效率，通过调节 SPF（Shortest Path First，最短路径优先）的计算延迟时间和间隔时间，可以避免在网络频繁变化时过多的占用资源。

4. GR 定时器

在进行 GR 时，可能出现重建邻居失败的情况，这将导致设备长期处于 GR 过程中。GR 定时器即 GR 的最长时间，当它超时时则退出 GR，从而避免设备长期处于 GR 过程中。

1.1.5 系统支持的 OSPFv3 特性

在目前的实现中，支持以下 OSPFv3 特性：

- RFC2740 规定的基本特性；
- OSPFv3 Stub 区域；
- OSPFv3 多进程（Multi-Process），可以在一台路由器上运行多个 OSPFv3 进程。
- OSPFv3 GR

1.1.6 协议规范

与 OSPFv3 相关的协议规范有：

- RFC 2740：OSPF for IPv6
- RFC 2328：OSPF Version 2
- RFC 5187：OSPFv3 Graceful Restart

1.2 OSPFv3 配置任务简介

表1-1 OSPFv3 配置任务简介

配置任务		说明	详细配置
使能 OSPFv3 功能		必选	1.3
配置 OSPFv3 的区域属性	配置 OSPFv3 的 Stub 区域	可选	1.4.2
	配置 OSPFv3 的虚连接	可选	1.4.3
配置 OSPFv3 网络类型	配置 OSPFv3 接口的网络类型	可选	1.5.2
	配置 NBMA 或者 P2MP 网络的邻居	可选	1.5.3
配置 OSPFv3 的路由信息控制	配置 OSPFv3 的路由聚合	可选	1.6.2
	配置 OSPFv3 对通过接收到的 LSA 计算出来的路由信息进行过滤	可选	1.6.3
	配置 OSPFv3 接口的开销值	可选	1.6.4
	配置 OSPFv3 最大等价路由条数	可选	1.6.5
	配置 OSPFv3 协议的优先级	可选	1.6.6
	配置 OSPFv3 引入外部路由	可选	1.6.7
调整和优化 OSPFv3 网络	配置 OSPFv3 定时器	可选	1.7.2
	配置接口的 DR 优先级	可选	1.7.3
	忽略 DD 报文中的 MTU 检查	可选	1.7.4
	禁止接口收发 OSPFv3 报文	可选	1.7.5
	配置邻接状态变化输出开关	可选	1.7.6
配置 GR 能力	配置 GR Restarter	可选	1.8.1
	使能 GR Helper	可选	1.8.2

1.3 使能 OSPFv3 功能

1.3.1 配置准备

在使能 OSPFv3 功能之前，需完成以下任务：

- 各相邻节点网络层可达
- 使能 IPv6 报文转发功能

1.3.2 使能 OSPFv3 功能

要在路由器上使能 OSPFv3 功能，必须先创建 OSPFv3 进程、指定该进程的 Router ID 以及在接口上使能 OSPFv3 功能。

Router ID 用来在一个自治系统中唯一的标识一台路由器。在 OSPFv3 中，用户必须手工配置一个 Router ID，而且必须保证自治系统中任意两台路由器的 Router ID 都不相同。因此，为了保证 OSPFv3 运行的稳定性，在进行网络规划时，应确定路由器 ID 的划分并手工配置。需要注意的是，如果在同一台路由器上运行了多个 OSPFv3 进程，必须为不同的进程指定不同的 Router ID。

在一台路由器上可以创建多个 OSPFv3 进程，OSPFv3 进程号是本地概念，不影响与其它路由器之间的报文交换。因此，不同的路由器之间，即使进程号不同也可以进行报文交换。

表1-2 使能 OSPFv3 功能

操作	命令	说明
进入系统视图	**system-view**	-
创建 OSPFv3 进程，进入 OSPFv3 视图	**ospfv3** [*process-id*]	必选 缺省情况下，系统不运行 OSPFv3 进程
配置路由器的 ID	**router-id** *router-id*	必选
进入接口视图	**interface** *interface-type interface-number*	-
在接口上使能 OSPFv3	**ospfv3** *process-id* **area** *area-id* [**instance** *instance-id*]	必选 缺省情况下，接口上没有使能 OSPFv3 协议

1.4 配置 OSPFv3 的区域属性

OSPFv3 支持 Stub 区域和虚连接的配置，其原理及应用环境与 OSPFv2 相同。

OSPFv3 划分区域后，可以减少网络中 LSA 的数量，OSPFv3 的扩展性也得以增强。对于位于 AS 边缘的一些非骨干区域，为了更多的缩减其路由表规模和降低 LSA 的数量，可以将它们配置为 Stub 区域。

在划分区域之后，非骨干区域之间的 OSPFv3 路由更新是通过骨干区域来交换完成的。对此，OSPFv3 要求所有非骨干区域必须与骨干区域保持连通，并且骨干区域自身也要保持连通。但在实际应用中，可能会因为各方面条件的限制，无法满足这个要求。这时可以通过配置 OSPFv3 虚连接予以解决。

1.4.1 配置准备

在配置 OSPFv3 的区域属性之前，需完成以下任务：

- 使能 IPv6 报文转发功能
- 配置 OSPFv3 基本功能

1.4.2 配置 OSPFv3 的 Stub 区域

表1-3 配置 OSPFv3 的 Stub 区域

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
进入 OSPFv3 区域视图	**area** *area-id*	-
配置一个区域为 Stub 区域	**stub** [**no-summary**]	必选 缺省情况下，没有区域被配置为 Stub 区域
配置发送到 Stub 区域的缺省路由的开销值	**default-cost** *value*	可选 缺省情况下，发送到 Stub 区域的缺省路由的开销值为 1

说明

- OSPFv3 的区域不能直接删除，只有区域视图下的所有配置都删除后，且该区域中相关接口的状态都为 down，此区域才会自动删除。
- 对于位于 Stub 区域中的所有路由器都必须配置 **stub** 命令，参数 **no-summary** 只能在 ABR 上配置。
- 如果 ABR 使用了 **stub** 命令中的参数 **no-summary**，则此 ABR 只向区域内发布一条描述缺省路由 Inter-Area-Prefix-LSA。这种既没有 AS-external-LSA，也没有 Inter-Area-Prefix-LSA、Inter-Area-Router-LSA 的 Stub 区域，又称为 Totally Stub 区域。

1.4.3 配置 OSPFv3 的虚连接

对于没有和骨干区域直接相连的非骨干区域，或者不连续的骨干区域，可以使用该配置建立逻辑上的连通性。

表1-4 配置 OSPFv3 的虚连接

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
进入 OSPFv3 区域视图	**area** *area-id*	-
创建并配置虚连接	**vlink-peer** *router-id* [**hello** *seconds* | **retransmit** *seconds* | **trans-delay** *seconds* | **dead** *seconds* | **instance** *instance-id*] *	必选

说明

- 虚连接的两端必须是 ABR，而且必须在两端同时配置才可生效。
- 虚连接不能在给使能了 GR 能力的进程下的区域进行配置。

1.5 配置 OSPFv3 网络类型

OSPFv3 根据链路层协议类型将网络分为四种不同的类型。

缺省情况下，当接口封装的链路层协议不同时，OSPFv3 接口网络类型的缺省值也不同，例如：

- 当接口封装的链路层协议是 PPP 时，OSPFv3 接口网络类型的缺省值为点对点；
- 当接口封装的链路层协议是 Ethernet 时，OSPFv3 接口网络类型的缺省值为广播类型。

用户可以根据需要配置 OSPFv3 接口的网络类型，例如：

- 由于 NBMA 网络必须是全连通的，即网络中任意两台路由器之间都必须有一条虚电路直接可达。但在很多情况下，这个要求无法满足，这时就需要通过命令强制改变网络的类型。
- 对于 NBMA 网络，如果部分路由器之间没有直接可达的链路时，应将接口的网络类型配置为 P2MP。如果路由器在 NBMA 网络中只有一个对端，也可将接口类型配置为 P2P。

1.5.1 配置准备

在配置 OSPFv3 的网络类型之前，需完成以下任务：

- 使能 IPv6 报文转发功能
- 配置 OSPFv3 基本功能

1.5.2 配置 OSPFv3 接口的网络类型

表1-5 配置 OSPFv3 接口的网络类型

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 OSPFv3 接口的网络类型	**ospfv3 network-type** { **broadcast** | **nbma** | **p2mp** [**non-broadcast**] | **p2p** } [**instance** *instance-id*]	可选 缺省情况下，接口的网络类型根据物理接口而定

1.5.3 配置 NBMA 或者 P2MP 网络的邻居

当路由器的接口类型为如下网络类型时，需要为其指定相邻路由器 IP 地址：

- NBMA 网络
- P2MP 网络（仅当接口选择单播形式发送报文时，需要此配置）

由于无法通过广播 Hello 报文的形式发现相邻路由器，必须手工指定相邻路由器的本地链路地址。对于 NBMA 网络，可以指定该相邻路由器是否有选举权等。

表1-6 配置 NBMA 或者 P2MP(单播)网络的邻居

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 NBMA 或者 P2MP(单播)网络的邻居	**ospfv3 peer** *ipv6-address* [**dr-priority** *dr-priority*] [**instance** *instance-id*]	必选

1.6 配置 OSPFv3 的路由信息控制

通过这部分的配置，可以控制 OSPFv3 的路由信息的发布与接收，并引入外部路由。

1.6.1 配置准备

在配置 OSPFv3 的路由信息控制之前，需完成以下任务：

- 使能 IPv6 报文转发功能
- 配置 OSPFv3 基本功能

1.6.2 配置 OSPFv3 的路由聚合

如果该区域中存在多个连续的网段，则可以在 ABR 上配置 **abr-summary** 命令将它们聚合成一个网段，ABR 只发送一条聚合后的 LSA，所有落入本命令指定的聚合网段范围的 LSA 将不再会被单独发送出去，这样可减小其它区域中 LSDB 的规模。

表1-7 配置 OSPFv3 的路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
进入 OSPFv3 区域视图	**area** *area-id*	-
配置 OSPFv3 区域路由聚合	**abr-summary** *ipv6-address prefix-length* [**not-advertise**]	必选 缺省情况下，ABR 没有对路由进行聚合


说明

路由聚合只有在 ABR 上配置才会有效。

1.6.3 配置 OSPFv3 对通过接收到的 LSA 计算出来的路由信息进行过滤

OSPFv3 接收到 LSA 后，可以根据一定的过滤条件来决定是否将计算后得到的路由信息加入到本地路由表中。

表1-8 配置 OSPFv3 对通过接收到的 LSA 计算出来的路由信息进行过滤

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
配置对接收的路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	必选 缺省情况下，没有对接收的路由信息过滤


说明

filter-policy import 命令只对 OSPFv3 计算出来的路由进行过滤，没有通过过滤的路由将不被加入到本地路由表中，从而不能用于转发报文。

1.6.4 配置 OSPFv3 接口的开销值

OSPFv3 有两种方式来配置接口的开销值，即从该接口发送出去的数据包的出站接口代价：

- 第一种方法是在接口视图下直接配置开销值；
- 第二种方法是配置接口的带宽参考值，OSPFv3 根据带宽参考值自动计算接口的开销值，计算公式为：接口开销＝带宽参考值÷接口带宽，当计算出来的开销值大于 65535，开销取最大值 65535。

如果没有在接口视图下显式的配置此接口的开销值，OSPFv3 会根据该接口的带宽自动计算其开销值。

1. 配置接口的开销值

表1-9 配置 OSPFv3 接口的开销值

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
设置 OSPFv3 接口的开销值	**ospfv3 cost** *value* [**instance** *instance-id*]	可选 缺省情况下，OSPFv3 根据接口的带宽自动计算链路开销；对于 Loopback 接口，缺省取值为 0

2. 配置带宽参考值

表1-10 配置带宽参考值

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
配置带宽参考值	**bandwidth-reference** *value*	可选 缺省情况下，带宽参考值为 100Mbits/s

1.6.5 配置 OSPFv3 最大等价路由条数

如果到一个目的地有几条开销相同的路径，可以实现等价路由负载分担，IPv6 报文在这几个链路上轮流发送，以提高链路利用率。该配置用以设置 OSPFv3 协议的最大等价路由条数。

表1-11 配置 OSPFv3 最大等价路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
配置最大等价路由条数	**maximum load-balancing** *maximum*	可选 缺省情况下，最大等价路由条数为 8

1.6.6 配置 OSPFv3 协议的优先级

由于路由器上可能同时运行多个动态路由协议，就存在各个路由协议之间路由信息共享和选择的问题。系统为每一种路由协议设置一个优先级，在不同协议发现同一条路由时，优先级高的路由将被优选。

表1-12 配置 OSPFv3 协议的优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id* |	-

操作	命令	说明
配置 OSPFv3 协议的优先级	**preference** [**ase**] [**route-policy** *route-policy-name*] *preference*	可选 缺省情况下，OSPFv3 内部路由的优先级为 10，OSPFv3 外部路由的优先级为 150

1.6.7 配置 OSPFv3 引入外部路由

表1-13 配置 OSPFv3 引入外部路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
配置外部路由的缺省开销值	**default cost** *value*	可选 缺省情况下，缺省开销值为 1
引入外部路由信息	**import-route** { **isisv6** *process-id* | **ospfv3** *process-id* | **ripng** *process-id* | **bgp4+** [**allow-ibgp**] | **direct** | **static** } [**cost** *value* | **type** *type* | **route-policy** *route-policy-name*] *	必选 缺省情况下，没有引入外部路由信息
配置 OSPFv3 引入缺省路由	**default-route-advertise** [**always** | **cost** *cost* | **type** *type* | **route-policy** *route-policy-name*] *	可选 缺省情况下，没有引入缺省路由
对引入的外部路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [**isisv6** *process-id* | **ospfv3** *process-id* | **ripng** *process-id* | **bgp4+** | **direct** | **static**]	可选 缺省情况下，没有对引入的路由信息过滤

说明

- 在 OSPFv3 路由器上配置 **import-route** 或 **default-route-advertise** 命令后，这台 OSPFv3 路由器就成为 ASBR。
- 只能通过 **default-route-advertise** 命令引入并发布缺省路由。
- 由于 OSPFv3 是基于链路状态的路由协议，不能直接对发布的 LSA 进行过滤，所以只能在 OSPFv3 引入路由时进行过滤，只有符合条件的路由才能转换成 LSA 发布出去。
- **filter-policy export** 命令只对本设备使用 **import-route** 引入的路由起作用。如果没有配置 **import-route** 命令来引入其它外部路由(包括不同进程的 OSPFv3 路由)，则 **filter-policy export** 命令失效。

1.7 调整和优化 OSPFv3 网络

本节主要介绍配置 OSPFv3 定时器、配置接口的 DR 优先级、忽略 DD 报文中的 MTU 检查和禁止接口收发 OSPFv3 报文。

OSPFv3 包括如下定时器：

- OSPFv3 的报文定时器，可以调整 OSPFv3 网络的收敛速度以及协议报文带来的网络负荷。
- LSA 的延迟时间：在一些低速链路上，需要考虑接口传送 LSA 的延迟时间。
- SPF 定时器：通过调整 SPF 计算间隔时间，可以抑制由于网络频繁变化带来的资源消耗问题。

对于广播网，通过配置接口的 DR 优先级来影响 DR/BDR 的选择。

如果要使 OSPFv3 路由信息不被某一网络中的路由器获得，可禁止在与该网络相连的接口上发送 OSPFv3 报文。

1.7.1 配置准备

在调整和优化 OSPFv3 网络之前，需完成以下任务：

- 使能 IPv6 报文转发功能
- 配置 OSPFv3 基本功能

1.7.2 配置 OSPFv3 定时器

表1-14 配置 OSPF 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口发送 hello 报文的时间间隔	**ospfv3 timer hello** *seconds* [**instance** *instance-id*]	可选 缺省情况下，P2P、Broadcast 网络类型接口发送 Hello 报文的时间间隔的值为 10 秒
配置 Poll 定时器	**ospfv3 timer poll** *seconds* [**instance** *instance-id*]	可选 缺省情况下，发送轮询 Hello 报文的时间间隔为 120 秒
配置相邻路由器间失效时间	**ospfv3 timer dead** *seconds* [**instance** *instance-id*]	可选 缺省情况下，P2P、Broadcast 网络类型接口的 OSPFv3 邻居失效时间为 40 秒
配置相邻路由器重传 LSA 的时间间隔	**ospfv3 timer retransmit** *interval* [**instance** *instance-id*]	可选 缺省情况下，LSA 的重传时间间隔为 5 秒
配置接口的 LSA 传输延迟时间	**ospfv3 trans-delay** *seconds* [**instance** *instance-id*]	可选 缺省情况下，接口的 LSA 传输延迟时间为 1 秒
退回系统视图	**quit**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
设置 SPF 定时器	**spf timers** *delay-interval hold-interval*	可选 缺省情况下，*delay-interval* 值为 5 秒，*hold-interval* 值为 10 秒


说明

- 相邻路由器间失效时间间隔的值不要设置得太小，否则邻居很容易失效。
- 相邻路由器重传 LSA 时间间隔的值不要设置得太小，否则将会引起不必要的重传。

1.7.3 配置接口的 DR 优先级

表1-15 配置接口的 DR 优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 DR 优先级	**ospfv3 dr-priority** *priority* [**instance** *instance-id*]	可选 缺省情况下，接口的 DR 优先级为 1

说明

路由器接口的 DR 优先级将影响接口在选举 DR 时所具有的资格，优先级为 0 的路由器不会被选举为 DR 或 BDR。

1.7.4 忽略 DD 报文中的 MTU 检查

在 LSA 数量不多的情况下，没有必要去检查 MTU 大小，可以设置忽略 DD 报文中的 MTU 检查，从而提高性能。

表1-16 忽略 DD 报文中的 MTU 检查

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
忽略 DD 报文中的 MTU 检查	**ospfv3 mtu-ignore** [**instance** *instance-id*]	必选 缺省情况下，不忽略 MTU 检查

1.7.5 禁止接口发送 OSPFv3 报文

表1-17 禁止接口收发 OSPFv3 报文

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
禁止接口发送 OSPFv3 报文	**silent-interface** { *interface-type interface-number* | **all** }	必选 缺省情况下，允许接口发送 OSPFv3 报文

说明

- 不同的进程可以对同一接口禁止发送 OSPFv3 报文，但 **silent-interface** 命令只对本进程已经使能的 OSPFv3 接口起作用，不对其它进程的接口起作用。
- 当运行 OSPFv3 协议的接口被配置为 Silent 状态后，该接口的直连路由仍可以由同一路由器的其他接口通过 Intra-Area-Prefix-LSA 发布，但 OSPFv3 报文将被阻塞，接口上不会建立 OSPFv3 邻居关系。这一特性可以增强 OSPFv3 的组网适应能力。

1.7.6 配置邻接状态变化输出开关

表1-18 配置邻接状态变化输出开关

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
配置邻接状态变化输出开关	**log-peer-change**	必选 缺省情况下，邻接状态变化输出开关处于打开状态

1.8 配置 OSPFv3 GR

说明

OSPFv3 GR 特性和 OSPFv3 虚连接特性不同时支持，在配置了虚连接后将不能再配置 GR 特性。

支持 OSPFv3 的 GR Restarter 能力的设备主备倒换后后，为了实现设备转发业务的不中断，它必须完成下列两项任务：

- 重启过程 GR Restarter 转发表项保持稳定；
- 重启流程结束后重建所有邻居关系，重新获取完整的网络拓扑信息。

主备倒换后，具有 GR Restarter 能力的设备首先向邻居发送 Grace LSA 通告邻居本设备进入 GR（该设备称为 Restarter），邻居收到 GR Restarter 发出的 Grace-LSA 后，如果支持 GR Helper 能力则进入 Helper 模式（此时该邻居称为 GR Helper）。之后 restarte 重新建立邻居。Helper 帮助 restarte 进行 LSDB 的同步。这样就能实现在主备倒换时保证数据转发的正常进行。

1.8.1 配置 GR Restarter

可以在 GR Restarter 设备上配置 GR Restarter 能力。

表1-19 配置 GR Restarter

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
使能 GR 能力	**graceful-restart enable**	必选 缺省情况下，OSPFv3 协议的 GR Restarter 能力处于关闭状态
配置 GR 重启时间间隔	**graceful-restart interval** *interval-value*	可选 缺省情况下，OSPFv3 协议的 GR 重启间隔时间为 120 秒

1.8.2 配置 GR Helper

可以在 GR Helper 设备上配置 GR Helper 能力。

表1-20 配置 GR Helper

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPFv3 视图	**ospfv3** [*process-id*]	-
使能 Helper 能力	**graceful-restart helper enable**	可选 缺省情况下，OSPFv3 的 GR Helper 能力处于打开状态
使能 LSA 严格检查能力	**graceful-restart helper strict-lsa-checking**	可选 缺省情况下，OSPFv3 协议的 GR Helper 严格 LSA 检查能力处于关闭状态

1.9 OSPFv3 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 OSPFv3 的运行情况，通过查看显示信息验证配置的效果。

表1-21 OSPFv3 显示和维护

操作	命令
显示 OSPFv3 的调试信息开关状态	**display debugging ospfv3**
显示 OSPFv3 进程的概要信息	**display ospfv3** [*process-id*]
显示 OSPFv3 接口信息	**display ospfv3 interface** [*interface-type interface-number* | **statistic**]
显示 OSPFv3 的 LSDB 信息	**display ospfv3** [*process-id*] **lsdb** [[**external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace**] [*link-state-id*] [**originate-router** *router-id*] | **total**]
显示 OSPFv3 LSDB 中 LSA 的统计信息	**display ospfv3 lsdb statistic**
显示 OSPFv3 邻居信息	**display ospfv3** [*process-id*] [**area** *area-id*] **peer** [[*interface-type interface-number*] [**verbose**] | *peer-router-id*]
显示 OSPFv3 邻居的统计信息	**display ospfv3 peer statistic**
显示 OSPFv3 路由表信息	**display ospfv3** [*process-id*] **routing** [*ipv6-address prefix-length* **|** *ipv6-address/prefix-length* | **abr-routes** | **asbr-routes** | **all | statistics**]
显示 OSPFv3 区域拓扑信息	**display ospfv3** [*process-id*] **topology** [**area** *area-id*]
显示 OSPFv3 虚连接信息	**display ospfv3** [*process-id*] **vlink**
显示 OSPFv3 下一跳信息	**display ospfv3** [*process-id*] **next-hop**
显示 OSPFv3 链路状态请求列表信息	**display ospfv3** [*process-id*] **request-list** [{ **external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace** } [*link-state-id*] [**originate-router** *ip-address*] | **statistics**]
显示 OSPFv3 链路状态重传列表信息	**display ospfv3** [*process-id*] **retrans-list** [{ **external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace** } [*link-state-id*] [**originate-router** *ip-address*] | **statistics**]
显示接口收发 OSPFv3 报文的统计信息	**display ospfv3 statistic**
显示 OSPFv3 进程的 GR 状态信息	**display ospfv3 [process-id] graceful-restart status**

1.10 OSPFv3 配置举例

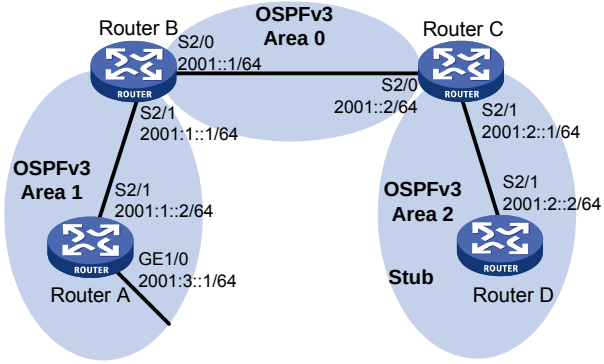
1.10.1 配置 OSPFv3 区域

1. 组网需求

- 所有的路由器都运行 OSPFv3，整个自治系统划分为 3 个区域。其中 Router B 和 Router C 作为 ABR 来转发区域之间的路由。
- 要求将 Area2 配置为 Stub 区域，减少通告到此区域内的 LSA 数量，但不影响路由的可达性。

2. 组网图

图1-2 配置 OSPFv3 区域组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 OSPFv3 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] ospfv3 1
[RouterA-ospfv3-1] router-id 1.1.1.1
[RouterA-ospfv3-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospfv3 1 area 1
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ospfv3 1 area 1
[RouterA-Serial2/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] ospfv3 1
[RouterB-ospf-1] router-id 2.2.2.2
[RouterB-ospf-1] quit
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ospfv3 1 area 0
```

```
[RouterB-Serial2/0] quit
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ospfv3 1 area 1
[RouterB-Serial2/1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] ospfv3 1
[RouterC-ospfv3-1] router-id 3.3.3.3
[RouterC-ospfv3-1] quit
[RouterC] interface serial 2/0
[RouterC-Serial2/0] ospfv3 1 area 0
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] ospfv3 1 area 2
[RouterC-Serial2/1] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] ipv6
[RouterD] ospfv3 1
[RouterD-ospfv3-1] router-id 4.4.4.4
[RouterD-ospfv3-1] quit
[RouterD] interface serial 2/1
[RouterD-Serial2/1] ospfv3 1 area 2
[RouterD-Serial2/1] quit
```

查看 Router B 的 OSPFv3 邻居状态。

```
[RouterB] display ospfv3 peer

            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
3.3.3.3         1     Full/Backup      00:00:34    S2/0        0

            OSPFv3 Area ID 0.0.0.1 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         1     Full/DR          00:00:35    S2/1        0
```

查看 Router C 的 OSPFv3 邻居状态。

```
[RouterC] display ospfv3 peer

            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         1     Full/DR          00:00:35    S2/0        0
```

```
            OSPFv3 Area ID 0.0.0.2 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
4.4.4.4         1     Full/Backup      00:00:36    S2/1        0
```

查看 Router D 的 OSPFv3 路由表信息。

```
[RouterD] display ospfv3 routing

E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route

            OSPFv3 Router with ID (4.4.4.4) (Process 1)
------------------------------------------------------------------------
 *Destination: 2001::/64
  Type       : IA                                    Cost     : 2
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001:1::/64
  Type       : IA                                    Cost     : 3
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001:2::/64
  Type       : I                                     Cost     : 1
  NextHop    : directly-connected                    Interface: S2/1

 *Destination: 2001:3::/64
  Type       : IA                                    Cost     : 4
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1
```

(3) 配置 Stub 区域

配置 Router D 的 Stub 区域。

```
[RouterD] ospfv3
[RouterD-ospfv3-1] area 2
[RouterD-ospfv3-1-area-0.0.0.2] stub
```

配置 Router C 的 Stub 区域，设置发送到 Stub 区域的缺省路由的开销为 10。

```
[RouterC] ospfv3
[RouterC-ospfv3-1] area 2
[RouterC-ospfv3-1-area-0.0.0.2] stub
[RouterC-ospfv3-1-area-0.0.0.2] default-cost 10
```

查看 Router D 的 OSPFv3 路由表信息，可以看到路由表中多了一条缺省路由，它的开销值为直连路由的开销和所配置的开销值之和。

```
[RouterD] display ospfv3 routing

E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route
```

```
            OSPFv3 Router with ID (4.4.4.4) (Process 1)
 ------------------------------------------------------------------------
 *Destination: ::/0
  Type       : IA                                    Cost     : 11
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001::/64
  Type       : IA                                    Cost     : 2
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001:1::/64
  Type       : IA                                    Cost     : 3
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001:2::/64
  Type       : I                                     Cost     : 1
  NextHop    : directly-connected                    Interface: S2/1

 *Destination: 2001:3::/64
  Type       : IA                                    Cost     : 4
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1
```

(4) 进一步减少 Stub 区域路由表规模，将 Area 2 配置为 Totally Stub 区域

配置 Router C，设置 Area2 为 Totally Stub 区域。

```
[RouterC-ospfv3-1-area-0.0.0.2] stub no-summary
```

查看 Router D 的 OSPFv3 路由表，可以发现路由表项数目减少了，其他非直连路由都被抑制，只有缺省路由被保留。

```
[RouterD] display ospfv3 routing

E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route

            OSPFv3 Router with ID (4.4.4.4) (Process 1)
 ------------------------------------------------------------------------
 *Destination: ::/0
  Type       : IA                                    Cost     : 11
  NextHop    : FE80::F40D:0:93D0:1                   Interface: S2/1

 *Destination: 2001:2::/64
  Type       : I                                     Cost     : 1
  NextHop    : directly-connected                    Interface: S2/1
```

1.10.2 配置 OSPFv3 的 DR 选择

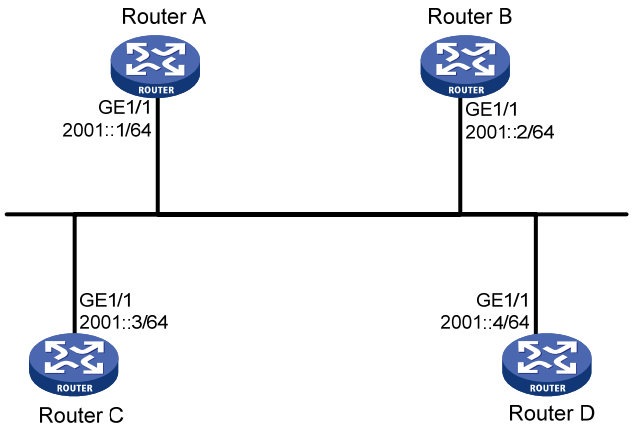
1. 组网需求

- Router A 的优先级为 100，它是网络上的最高优先级，所以 Router A 被选为 DR；
- Router C 的优先级为 2，它是优先级次高的，被选为 BDR；

- Router B 的优先级为 0，这意味着它将无法成为 DR；
- Router D 没有配置优先级，取缺省值 1。

2. 组网图

图1-3 配置 OSPFv3 的 DR 选择组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 OSPFv3 基本功能

配置 Router A，启动 OSPFv3，并设置其 Router ID 为 1.1.1.1。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] ospfv3
[RouterA-ospfv3-1] router-id 1.1.1.1
[RouterA-ospfv3-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospfv3 1 area 0
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B，启动 OSPFv3，并设置其 Router ID 为 2.2.2.2。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] ospfv3
[RouterB-ospfv3-1] router-id 2.2.2.2
[RouterB-ospfv3-1] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ospfv3 1 area 0
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C，启动 OSPFv3，并设置其 Router ID 为 3.3.3.3。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] ospfv3
[RouterC-ospfv3-1] router-id 3.3.3.3
[RouterC-ospfv3-1] quit
```

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ospfv3 1 area 0
[RouterC-GigabitEthernet1/1] quit
```

配置 Router D，启动 OSPFv3，并设置其 Router ID 为 4.4.4.4。

```
<RouterD> system-view
[RouterD] ipv6
[RouterD] ospfv3
[RouterD-ospfv3-1] router-id 4.4.4.4
[RouterD-ospfv3-1] quit
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] ospfv3 1 area 0
[RouterD-GigabitEthernet1/1] quit
```

查看 Router A 的邻居信息，可以看到 DR 优先级（缺省为 1）以及邻居状态。此时优先级相等，Router ID 大者被选为 DR，可以看到 Router D 为 DR，Router C 为 BDR。

```
[RouterA] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         1     2-Way/DROther    00:00:36    GE1/1          0
3.3.3.3         1     Full/Backup      00:00:35    GE1/1          0
4.4.4.4         1     Full/DR          00:00:33    GE1/1          0
```

查看 Router D 的邻居信息，可以看到 Router D 和其他邻居之间的邻居状态都为 Full。

```
[RouterD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         1     Full/DROther     00:00:30    GE1/1          0
2.2.2.2         1     Full/DROther     00:00:37    GE1/1          0
3.3.3.3         1     Full/Backup      00:00:31    GE1/1          0
```

(3) 配置接口的 DR 优先级

配置 Router A 的接口 GigabitEthernet1/1 的 DR 优先级为 100。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospfv3 dr-priority 100
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B 的接口 GigabitEthernet1/1 的 DR 优先级为 0。

```
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ospfv3 dr-priority 0
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C 的接口 GigabitEthernet1/1 的 DR 优先级为 2。

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ospfv3 dr-priority 2
[RouterC-GigabitEthernet1/1] quit
```

显示 Router A 的邻居信息，可以看到 DR 优先级已经更新，但 DR/BDR 并未改变。

```
[RouterA] display ospfv3 peer
```

```
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         0     2-Way/DROther    00:00:38    GE1/1          0
3.3.3.3         2     Full/Backup      00:00:32    GE1/1          0
4.4.4.4         1     Full/DR          00:00:36    GE1/1          0
```

显示 Router D 的邻居信息，可以看到 Router D 仍然为 DR。

```
[RouterD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         100   Full/DROther     00:00:33    GE1/1          0
2.2.2.2         0     Full/DROther     00:00:36    GE1/1          0
3.3.3.3         2     Full/Backup      00:00:40    GE1/1          0
```

(4) 重新进行 DR/BDR 选择

将接口进行一次 **shutdown** 和 **undo shutdown**，使 OSPFv3 进行 DR/BDR 的重新选举（略）。

查看 Router A 的邻居信息，可以看到 Router C 为 BDR。

```
[RouterA] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         0     Full/DROther     00:00:31    GE1/1          0
3.3.3.3         2     Full/Backup      00:00:39    GE1/1          0
4.4.4.4         1     Full/DROther     00:00:37    GE1/1          0
```

查看 Router D 的邻居信息，可以看到 Router A 为 DR。

```
[RouterD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ---------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         100   Full/DR          00:00:34    GE1/1          0
2.2.2.2         0     2-Way/DROther    00:00:34    GE1/1          0
3.3.3.3         2     Full/Backup      00:00:32    GE1/1          0
```

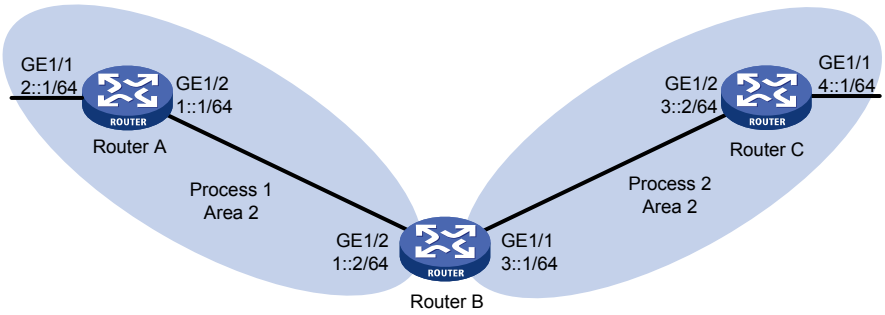
1.10.3 配置 OSPFv3 引入外部路由

1. 组网需求

- Router A、Router B 和 Router C 位于 Area 2 内；
- Router B 上运行两个 OSPFv3 进程：OSPFv3 1 和 OSPFv3 2。Router B 通过 OSPFv3 1 和 Router A 交换路由信息，通过 OSPFv3 2 和 Router C 交换路由信息；
- 在 Router B 上配置 OSPFv3 进程 2 引入外部路由，引入直连路由和 OSPFv3 进程 1 的路由，并将引入的外部路由的缺省度量值设置为 3，使得 Router C 能够学习到达 1::0/64 和 2::0/64 的路由，但 Router A 不能学习到达 3::0/64 和 4::0/64 的路由。

2. 组网图

图1-4 配置 OSPFv3 引入外部路由组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 OSPFv3

在 Router A 上启动 OSPFv3 进程 1。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] ospfv3 1
[RouterA-ospfv3-1] router-id 1.1.1.1
[RouterA-ospfv3-1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ospfv3 1 area 2
[RouterA-GigabitEthernet1/2] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospfv3 1 area 2
[RouterA-GigabitEthernet1/1] quit
```

在 Router B 上启动两个 OSPFv3 进程，进程号分别为 1 和 2。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] ospfv3 1
[RouterB-ospfv3-1] router-id 2.2.2.2
[RouterB-ospfv3-1] quit
[RouterB] interface GigabitEthernet 1/2
[RouterB-GigabitEthernet1/2] ospfv3 1 area 2
[RouterB-GigabitEthernet1/2] quit
[RouterB] ospfv3 2
[RouterB-ospfv3-2] router-id 3.3.3.3
[RouterB-ospfv3-2] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ospfv3 2 area 2
[RouterB-GigabitEthernet1/1] quit
```

在 Router C 上启动 OSPFv3 进程 2。

```
<RouterC> system-view
[RouterC] ipv6
```

```
[RouterC] ospfv3 2
[RouterC-ospfv3-2] router-id 4.4.4.4
[RouterC-ospfv3-2] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] ospfv3 2 area 2
[RouterC-GigabitEthernet1/2] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ospfv3 2 area 2
[RouterC-GigabitEthernet1/1] quit
```

查看 Router C 的路由表信息。

```
[RouterC] display ipv6 routing-table
Routing Table :
         Destinations : 6        Routes : 6

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 3::/64                                     Protocol  : Direct
NextHop    : 3::2                                       Preference: 0
Interface  : GE1/2                                      Cost      : 0

Destination: 3::2/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : Direct
NextHop    : 4::1                                       Preference: 0
Interface  : GE1/1                                      Cost      : 0

Destination: 4::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

(3) 配置 OSPFv3 引入外部路由

在 Router B 上配置 OSPFv3 进程 2 引入外部路由,，引入直连路由和 OSPFv3 进程 1 的路由。

```
[RouterB] ospfv3 2
[RouterB-ospfv3-2] default cost 3
[RouterB-ospfv3-2] import-route ospfv3 1
[RouterB-ospfv3-2] import-route direct
[RouterB-ospfv3-2] quit
```

查看路由引入后 Router C 的路由表信息。

```
[RouterC] display ipv6 routing-table
```

```
Routing Table :
         Destinations : 8        Routes : 8

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : OSPFv3
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : GE1/2                                      Cost      : 3

Destination: 2::/64                                     Protocol  : OSPFv3
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : GE/1/2                                     Cost      : 3

Destination: 3::/64                                     Protocol  : Direct
NextHop    : 3::2                                       Preference: 0
Interface  : GE1/2                                      Cost      : 0

Destination: 3::2/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : Direct
NextHop    : 4::1                                       Preference: 0
Interface  : GE1/1                                      Cost      : 0

Destination: 4::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

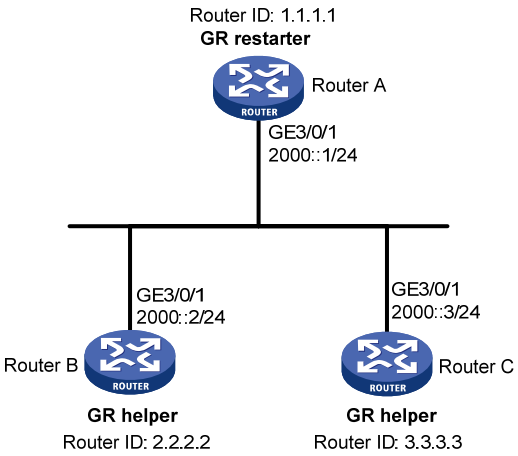
1.10.4 配置 OSPFv3 GR

1. 组网需求

- Router A、Router B 和 Router C 既属于同一自治系统，也属于同一 OSPFv3 区域，通过 OSPFv3 协议实现网络互连，并提供 GR 机制；
- Router A 作为 GR Restarter，Router B 和 Router C 作为 GR Helper 并且通过 GR 机制与 Router A 保持带外同步。

2. 组网图

图1-5 配置 OSPFv3 的 GR 组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 OSPFv3 基本功能

配置 Router A，启动 OSPFv3，并设置其 Router ID 为 1.1.1.1。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] ospfv3 1
[RouterA-ospfv3-1] router-id 1.1.1.1
[RouterA-ospfv3-1] graceful-restart enable
[RouterA-ospfv3-1] quit
[RouterA] interface GigabitEthernet 3/0/1
[RouterA-GigabitEthernet3/0/1] ospfv3 1 area 1
[RouterA-GigabitEthernet3/0/1] quit
```

配置 Router B，启动 OSPFv3，并设置其 Router ID 为 2.2.2.2。缺省情况下，Router B 的 GR Helper 能力处于开启状态。

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] ospfv3 1
[RouterB-ospfv3-1] router-id 2.2.2.2
[RouterB-ospfv3-1] quit
[RouterB] interface GigabitEthernet 3/0/1
[RouterB-GigabitEthernet3/0/1] ospfv3 1 area 1
[RouterB-GigabitEthernet3/0/1] quit
```

在 Router C，启动 OSPFv3，并设置其 Router ID 为 3.3.3.3。缺省情况下，Router C 的 GR Helper 能力处于开启状态。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] ospfv3 1
[RouterC-ospfv3-1] router-id 3.3.3.3
```

```
[RouterC-ospfv3-1] quit
[RouterC] interface GigabitEthernet 3/0/1
[RouterC-GigabitEthernet3/0/1] ospfv3 1 area 1
[RouterC-GigabitEthernet3/0/1] quit
```

(3) 检验配置效果

运行稳定后，在 Router A 上主备倒换进入 OSPFv3 协议的 GR 进程。

1.11 常见配置错误举例

1.11.1 OSPFv3 邻居无法建立

1. 故障现象

OSPFv3 邻居无法建立

2. 分析

如果物理连接和下层协议正常，则检查接口上配置的 OSPFv3 参数，必须保证与相邻路由器的参数一致，区域号相同。

相邻的两台路由器接口的网络类型必须一致。若网络类型为广播网，则至少有一个接口的 DR 优先级应大于零。

3. 处理过程

(1) 使用 **display ospfv3 peer** 命令查看 OSPFv3 邻居状态。

(2) 使用 **display ospfv3 interface** 命令查看 OSPFv3 接口的信息。

(3) 检查物理连接及下层协议是否正常运行，可通过 Ping 命令测试。若从本地路由器 Ping 对端路由器不通，则表明物理连接和下层协议有问题。

(4) 检查 OSPFv3 定时器，在同一接口上失效时间应至少为 Hello 间隔时间的 4 倍。

(5) 如果网络类型为广播网，则至少有一个接口的 DR 优先级大于零。

1.11.2 OSPFv3 路由信息不正确

1. 故障现象

OSPFv3 不能发现其他区域的路由

2. 分析

应保证骨干区域与所有的区域相连接。若一台路由器配置了两个以上的区域，则至少有一个区域应与骨干区域相连。骨干区域不能配置成 Stub 区域。

在 Stub 区域内的路由器不能接收外部 AS 的路由。如果一个区域配置成 Stub 区域，则与这个区域相连的所有路由器都应将此区域配置成 Stub 区域。

3. 处理过程

(1) 使用 **display ospfv3 peer** 命令查看 OSPFv3 邻居状态。

(2) 使用 **display ospfv3 interface** 命令查看 OSPFv3 接口的信息。

(3) 使用 **display ospfv3 lsdb** 查看 LSDB 的信息是否完整。

(4) 使用 **display current-configuration configuration** 命令查看区域是否配置正确。若配置了两个以上的区域，则至少有一个区域与骨干区域相连。

(5) 如果某区域是 Stub 区域，则该区域中的所有路由器都要配置 **stub** 命令。

(6) 如果配置了虚连接，使用 **display ospf vlink** 命令查看邻居状态是否正常。

目 录

1 RIPng配置 1-1
1.1 简介 1-1
1.1.1 RIPng工作机制 1-1
1.1.2 RIPng的报文格式 1-2
1.1.3 RIPng报文处理过程 1-3
1.1.4 协议规范 1-3
1.2 配置RIPng的基本功能 1-3
1.2.1 配置准备 1-3
1.2.2 配置RIPng的基本功能 1-4
1.3 配置RIPng路由特性 1-4
1.3.1 配置接口附加度量值 1-4
1.3.2 配置RIPng路由聚合 1-5
1.3.3 配置RIPng发布缺省路由 1-5
1.3.4 配置RIPng对接收/发布的路由进行过滤 1-5
1.3.5 配置RIPng协议优先级 1-6
1.3.6 配置RIPng引入外部路由 1-6
1.4 调整和优化RIPng网络 1-6
1.4.1 配置RIPng定时器 1-6
1.4.2 配置水平分割和毒性逆转 1-7
1.4.3 配置RIPng报文的零域检查 1-8
1.4.4 配置最大等价路由条数 1-8
1.5 RIPng显示和维护 1-8
1.6 RIPng典型配置举例 1-9
1.6.1 配置RIPng基本功能 1-9
1.6.2 配置RIPng引入外部路由 1-11

1 RIPng 配置

1.1 简介

RIPng 又称为下一代 RIP 协议（RIP next generation），它是对原来的 IPv4 网络中 RIP-2 协议的扩展。大多数 RIP 的概念都可以用于 RIPng。

为了在 IPv6 网络中应用，RIPng 对原有的 RIP 协议进行了如下修改：

- UDP 端口号：使用 UDP 的 521 端口发送和接收路由信息。
- 组播地址：使用 FF02::9 作为链路本地范围内的 RIPng 路由器组播地址。
- 前缀长度：目的地址使用 128 比特的前缀长度。
- 下一跳地址：使用 128 比特的 IPv6 地址。
- 源地址：使用链路本地地址 FE80::/10 作为源地址发送 RIPng 路由信息更新报文。

1.1.1 RIPng 工作机制

RIPng 协议是基于距离矢量（Distance-Vector）算法的协议。它通过 UDP 报文交换路由信息，使用的端口号为 521。

RIPng 使用跳数来衡量到达目的地址的距离（也称为度量值或开销）。在 RIPng 中，从一个路由器到其直连网络的跳数为 0，通过与其相连的路由器到达另一个网络的跳数为 1，其余以此类推。当跳数大于或等于 16 时，目的网络或主机就被定义为不可达。

RIPng 每 30 秒发送一次路由更新报文。如果在 180 秒内没有收到网络邻居的路由更新报文，RIPng 将从邻居学到的所有路由标识为不可达。如果再过 120 秒内仍没有收到邻居的路由更新报文，RIPng 将从路由表中删除这些路由。

为了提高性能并避免形成路由环路，RIPng 既支持水平分割也支持毒性逆转。此外，RIPng 还可以从其它的路由协议引入路由。

每个运行 RIPng 的路由器都管理一个路由数据库，该路由数据库包含了到所有可达目的地的路由项，这些路由项包含下列信息：

- 目的地址：主机或网络的 IPv6 地址。
- 下一跳地址：为到达目的地，需要经过的相邻路由器的接口 IPv6 地址。
- 出接口：转发 IPv6 报文通过的出接口。
- 度量值：本路由器到达目的地的开销。
- 路由时间：从路由项最后一次被更新到现在所经过的时间，路由项每次被更新时，路由时间重置为 0。
- 路由标记（Route Tag）：用于标识外部路由，以便在路由策略中根据 Tag 对路由进行灵活的控制。关于路由策略的详细信息，请参见“IP 路由分册”中的“路由策略配置”。

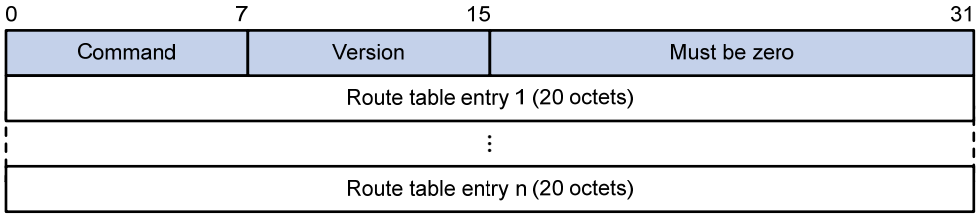
1.1.2 RIPng 的报文格式

1. 基本格式

RIPng 报文由头部（Header）和多个路由表项（RTEs）组成。在同一个 RIPng 报文中，RTE 的最大条数与发送接口的 IPv6 MTU 有关。

RIPng 报文基本格式如下图所示。

图1-1 RIPng 报文基本格式



各字段的含义如下：

- Command：定义报文的类型。0x01 表示 Request 报文，0x02 表示 Response 报文。
- Version：RIPng 的版本，目前其值只能为 0x01。
- RTE（Route Table Entry）：路由表项，每项的长度为 20 字节。

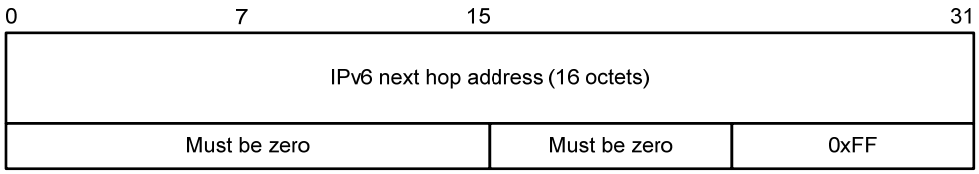
2. RTE 的格式

在 RIPng 里有两类 RTE，分别是：

- 下一跳 RTE：位于一组具有相同下一跳的“IPv6 前缀 RTE”的前面，它定义了下一跳的 IPv6 地址。
- IPv6 前缀 RTE：位于某个“下一跳 RTE”的后面。同一个“下一跳 RTE”的后面可以有多个不同的“IPv6 前缀 RTE”。它描述了 RIPng 路由表中的目的 IPv6 地址、路由标记、前缀长度以及度量值。

下一跳 RTE 的格式如下图所示。

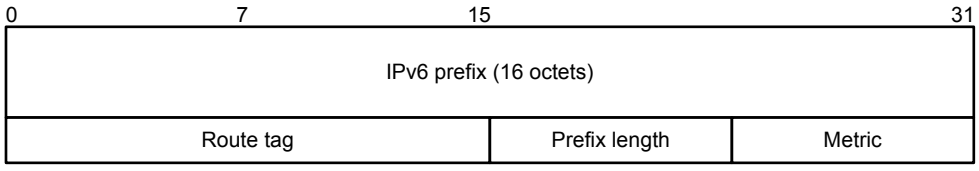
图1-2 下一跳 RTE 格式



其中，IPv6 next hop address 表示下一跳的 IPv6 地址。

IPv6 前缀 RTE 的格式如下图所示。

图1-3 IPv6 前缀 RTE 格式



各字段的解释如下：

- IPv6 prefix：目的 IPv6 地址的前缀。
- route tag：路由标记。
- prefix len：IPv6 地址的前缀长度。
- metric：路由的度量值。

1.1.3 RIPng 报文处理过程

1. Request 报文

当 RIPng 路由器启动后或者需要更新部分路由表项时，便会发出 Request 报文，向邻居请求需要的路由信息。通常情况下以组播方式发送 Request 报文。

收到 Request 报文的 RIPng 路由器会对其中的 RTE 进行处理。如果 Request 报文中只有一项 RTE，且 IPv6 前缀和前缀长度都为 0，度量值为 16，则表示请求邻居发送全部路由信息，被请求路由器收到后会把当前路由表中的全部路由信息，以 Response 报文形式发回给请求路由器。如果 Request 报文中有多项 RTE，被请求路由器将对 RTE 逐项处理，更新每条路由的度量值，最后以 Response 报文形式返回给请求路由器。

2. Response 报文

Response 报文包含本地路由表的信息，一般在下列情况下产生：

- 对某个 Request 报文进行响应
- 作为更新报文周期性地发出
- 在路由发生变化时触发更新

收到 Response 报文的路由器会更新自己的 RIPng 路由表。为了保证路由的准确性，RIPng 路由器会对收到的 Response 报文进行有效性检查，比如源 IPv6 地址是否是链路本地地址，端口号是否正确等，没有通过检查的报文会被忽略。

1.1.4 协议规范

与 RIPng 相关的规范有：

- RFC 2080：RIPng for IPv6
- RFC 2081：RIPng Protocol Applicability Statement

1.2 配置 RIPng 的基本功能

RIPng 基本功能的配置任务包括 RIPng 特性配置中涉及的基本配置，完成本节的配置就可以使用 RIPng 特性。

在 RIPng 的配置中，应该最先使能 RIPng。但是与 RIPng 功能特性相关的接口配置不受是否启动了 RIPng 的限制，比如说给接口配置一个 IPv6 地址。

1.2.1 配置准备

在配置 RIPng 基本功能之前，需完成以下任务：

- 启动 IPv6 报文转发功能
- 配置接口的网络层地址，使相邻节点的网络层可达

1.2.2 配置 RIPng 的基本功能

表1-1 配置 RIPng 的基本功能

操作	命令	说明
进入系统视图	**system-view**	-
创建 RIPng 进程并进入 RIPng 视图	**ripng** [*process-id*]	必选 缺省情况下，没有 RIPng 进程在运行
退回系统视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
在指定的网络接口上使能 RIPng	**ripng** *process-id* **enable**	必选 缺省情况下，接口禁用 RIPng

说明

如果接口没有使能 RIPng，那么 RIPng 进程在该接口上既不发送也不接收 RIPng 路由。

1.3 配置 RIPng 路由特性

在配置 RIPng 的路由特性之前，需完成以下任务：

- 配置接口的网络层地址
- 配置 RIPng 基本功能
- 如果使用 IPv6 ACL 进行路由过滤，需要定义 IPv6 ACL，相关配置请参见“安全分册”中的“ACL 配置”。
- 如果使用 IPv6 地址前缀列表进行路由过滤，需要定义 IPv6 地址前缀列表，相关配置请参见“IP 路由分册”中的“路由策略配置”。

1.3.1 配置接口附加度量值

附加度量值是附加在 RIPng 路由上的输入输出度量值，包括发送附加度量值和接收附加度量值。发送附加度量值不会改变路由表中的路由度量值，仅当接口发送 RIPng 路由信息时才会添加到发送路由上；接收附加度量值会影响接收到的路由度量值，接口接收到一条合法的 RIPng 路由时，在将其加入路由表前会把附加度量值加到该路由上。

表1-2 配置接口附加度量值

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
设置接口接收 RIPng 路由时的附加度量值	**ripng metricin** *value*	可选 缺省情况下，接口接收 RIPng 路由时的附加度量值为 0

操作	命令	说明
设置接口发送 RIPng 路由时的附加度量值	**ripng metricout** *value*	可选 缺省情况下，接口发送 RIPng 路由时的附加度量值为 1

1.3.2 配置 RIPng 路由聚合

表1-3 配置 RIPng 路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 RIPng 在接口发布聚合的 IPv6 前缀	**ripng summary-address** *ipv6-address prefix-length*	必选

1.3.3 配置 RIPng 发布缺省路由

表1-4 配置 RIPng 发布缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 RIPng 发布缺省路由	**ripng default-route** { **only** | **originate** } [**cost** *cost*]	必选 缺省情况下，RIPng 进程不发布缺省路由

说明

缺省路由将被强制通过指定接口的路由更新报文发布出去，该路由的发布不考虑其是否已经存在于本设备的 IPv6 路由表中。

1.3.4 配置 RIPng 对接收/发布的路由进行过滤

用户可通过使用 IPv6 ACL 和 IPv6 前缀列表对接收到的路由信息进行过滤，只有通过过滤的路由才能被加入到 RIPng 路由表；此外，还可对本机所有要发布的路由进行过滤，包括从其它路由协议引入的路由和从邻居学到的 RIPng 路由，只有通过过滤的路由才能被发布给 RIPng 邻居。

表1-5 配置 RIPng 对接收/发布的路由进行过滤

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-
对接收的路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	必选 缺省情况下，RIPng 不对接收的路由信息进行过滤

操作	命令	说明
对发布的路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol* [*process-id*]]	必选 缺省情况下，RIPng 不对发布的路由信息进行过滤

1.3.5 配置 RIPng 协议优先级

任何路由协议都具备特有的协议优先级，在设备进行路由选择时能够在不同的协议中选择最佳路由。可以手工设置 RIPng 协议的优先级，设置的值越小，其优先级越高。

表1-6 配置 RIPng 协议优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-
配置 RIPng 路由的优先级	**preference** [**route-policy** *route-policy-name*] *preference*	可选 缺省情况下，RIPng 路由的优先级为 100

1.3.6 配置 RIPng 引入外部路由

表1-7 配置 RIPng 引入外部路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-
配置引入路由的缺省度量值	**default cost** *cost*	可选 缺省情况下，引入路由的缺省度量值为 0
引入外部路由	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name*] *	必选 缺省情况下，RIPng 不引入其它路由

1.4 调整和优化 RIPng 网络

本节将介绍如何调整和优化 RIPng 网络的性能，以及在特殊网络环境中某些 RIPng 特性的应用，在调整和优化 RIPng 网络之前，需完成以下任务：

- 配置接口的网络层地址
- 配置 RIPng 基本功能

1.4.1 配置 RIPng 定时器

用户可通过调节 RIPng 定时器来调整 RIPng 路由协议的性能，以满足网络需要

表1-8 配置 RIPng 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-

操作	命令	说明
配置 RIPng 定时器的值	**timers** { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *	可选 缺省情况下，Update 定时器的值为 30 秒，Timeout 定时器的值为 180 秒，Suppress 定时器的值为 120 秒，Garbage-collect 定时器的值为 120 秒

说明

在配置 RIPng 定时器时需要注意，定时器值的调整应考虑网络的性能，并在所有运行 RIPng 的路由器上进行统一配置，避免增加不必要的网络流量。

1.4.2 配置水平分割和毒性逆转

说明

如果同时配置了水平分割和毒性逆转，则只有毒性逆转功能生效。

1. 配置水平分割

配置水平分割可以使得从一个接口学到的路由不能通过此接口向外发布，用于避免相邻路由器间的路由环路。

表1-9 配置水平分割

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能水平分割功能	**ripng split-horizon**	可选 缺省情况下，水平分割功能处于使能状态

说明

- 通常情况下，为了防止路由环路的出现，水平分割都是必要的，因此，建议不要关闭水平分割。
- 在帧中继和 X.25 等 NBMA（Non-Broadcast Multi-Access，非广播多点可达）网络中，当主接口和点到多点子接口配置了多条虚电路时，为了保证路由信息的正确传播，需要关闭水平分割功能。关于帧中继和 X.25 的详细信息，请参见“接入分册”中的“帧中继配置”和“LAPB 和 X.25 配置”。

2. 配置毒性逆转

配置毒性逆转可以使得从一个接口学到的路由还可以从这个接口向外发布，但此时这些路由的度量值已设置为 16，即不可达。

表1-10 配置毒性逆转

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能毒性逆转功能	**ripng poison-reverse**	必选 缺省情况下，毒性逆转功能处于关闭状态

1.4.3 配置 RIPng 报文的零域检查

RIPng 报文头部中的一些字段必须配置为 0，也称为零域。使能 RIPng 报文的零域检查功能后，如果报文头部零域中的值不为零，这些报文将被丢弃，不做处理。如果能确保所有报文都是可信任的，则不需要进行该项检查，以节省 CPU 处理时间。

表1-11 配置 RIPng 报文的零域检查

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-
使能对 RIPng 报文头部的零域检查操作	**checkzero**	可选 缺省情况下，RIPng 进行零域检查操作

1.4.4 配置最大等价路由条数

表1-12 配置最大等价路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIPng 视图	**ripng** [*process-id*]	-
配置 RIPng 最大等价路由条数	**maximum load-balancing** *number*	可选 缺省情况下，最大等价路由条数为 8

1.5 RIPng 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 RIPng 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以重启 RIPng 进程或清除指定 RIPng 进程的统计信息。

表1-13 RIPng 显示和维护

操作	命令
显示 RIPng 进程的配置信息	**display ripng** [*process-id*]
显示 RIPng 发布数据库中的路由	**display ripng** *process-id* **database**
显示指定 RIPng 进程的路由信息	**display ripng** *process-id* **route**
显示 RIPng 的接口信息	**display ripng** *process-id* **interface** [*interface-type interface-number*]

操作	命令
重启 RIPng 进程	reset ripng process-id process
清除 RIPng 进程的统计信息	reset ripng process-id statistics

1.6 RIPng 典型配置举例

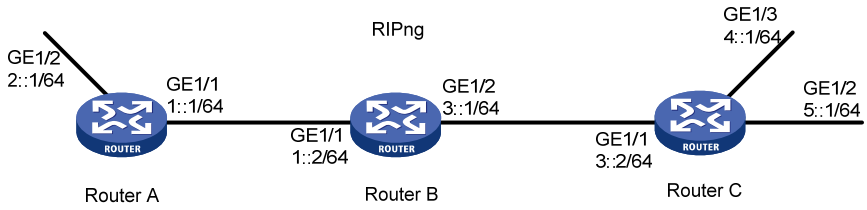
1.6.1 配置 RIPng 基本功能

1. 组网需求

- Router A、Router B 和 Router C 相连并通过 RIPng 来学习网络中的 IPv6 路由信息；
- 在 Router B 上对接收的 Router C 的路由（3::/64）进行过滤，使其不加入到 Router B 的 RIPng 进程的路由表中，也不发布给 Router A。

2. 组网图

图1-4 RIPng 配置举例组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 RIPng 的基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ripng 1
[RouterA-ripng-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ripng 1 enable
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ripng 1 enable
[RouterA-GigabitEthernet1/2] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ripng 1
[RouterB-ripng-1] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ripng 1 enable
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface GigabitEthernet 1/2
```

```
[RouterB-GigabitEthernet1/2] ripng 1 enable
[RouterB-GigabitEthernet1/2] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ripng 1
[RouterC-ripng-1] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ripng 1 enable
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] ripng 1 enable
[RouterC-GigabitEthernet1/2] quit
[RouterC] interface GigabitEthernet 1/3
[RouterC-GigabitEthernet1/3] ripng 1 enable
[RouterC-GigabitEthernet1/3] quit
```

查看 Router B 的 RIPng 路由表。

```
[RouterB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::20F:E2FF:FE23:82F5  on GigabitEthernet1/1
 Dest 1::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 6 Sec
 Dest 2::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 6 Sec

 Peer FE80::20F:E2FF:FE00:100  on GigabitEthernet1/2
 Dest 3::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
 Dest 4::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
 Dest 5::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
```

(3) 配置 Router B 对接收和发布的路由进行过滤

```
[RouterB] acl ipv6 number 2000
[RouterB-acl6-basic-2000] rule deny source 3::/64
[RouterB-acl6-basic-2000] rule permit
[RouterB-acl6-basic-2000] quit
[RouterB] ripng 1
[RouterB-ripng-1] filter-policy 2000 import
[RouterB-ripng-1] filter-policy 2000 export
```

查看 Router B 和 Router A 的 RIPng 路由表。

```
[RouterB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------
```

```
 Peer FE80::20F:E2FF:FE23:82F5  on GigabitEthernet1/1
 Dest 1::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 2 Sec
 Dest 2::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 2 Sec
 Peer FE80::20F:E2FF:FE00:100  on GigabitEthernet1/2
 Dest 4::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 5 Sec
 Dest 5::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 5 Sec
[RouterA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::20F:E2FF:FE00:1235  on GigabitEthernet1/1
 Dest 1::/64,
     via FE80::20F:E2FF:FE00:1235, cost  1, tag 0, A, 2 Sec
 Dest 4::/64,
     via FE80::20F:E2FF:FE00:1235, cost  2, tag 0, A, 2 Sec
 Dest 5::/64,
     via FE80::20F:E2FF:FE00:1235, cost  2, tag 0, A, 2 Sec
```

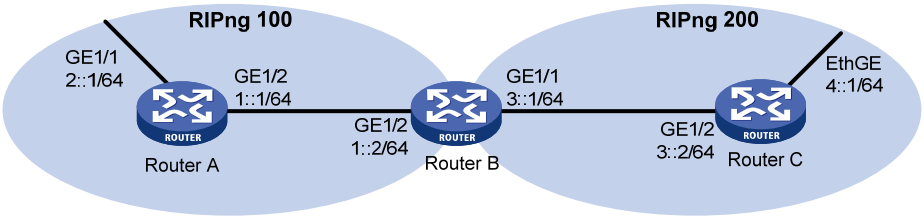
1.6.2 配置 RIPng 引入外部路由

1. 组网需求

- Router B 上运行两个 RIPng 进程：RIPng100 和 RIPng200。Router B 通过 RIPng100 和 Router A 交换路由信息，通过 RIPng200 和 Router C 交换路由信息。
- 要求在 Router B 上配置路由引入，将两个不同进程的 RIPng 路由相互引入到对方的 RIPng 进程中，将引入的 RIPng200 的路由缺省度量值设为 3。

2. 组网图

图1-5 配置 RIPng 引入外部路由组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 RIPng

在 Router A 上启动 RIPng 进程 100。

```
<RouterA> system-view
[RouterA] ripng 100
[RouterA-ripng-100] quit
```

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ripng 100 enable
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ripng 100 enable
```

在 Router B 上启动两个 RIPng 进程，进程号分别为 100 和 200。

```
<RouterB> system-view
[RouterB] ripng 100
[RouterB-ripng-100] quit
[RouterB] interface GigabitEthernet 1/2
[RouterB-GigabitEthernet1/2] ripng 100 enable
[RouterB-GigabitEthernet1/2] quit
[RouterB] ripng 200
[RouterB-ripng-200] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ripng 200 enable
```

在 Router C 上启动 RIPng 进程 200。

```
<RouterC> system-view
[RouterC] ripng 200
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ripng 200 enable
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] ripng 200 enable
[RouterC-GigabitEthernet1/2] quit
```

查看 Router A 的路由表信息。

```
[RouterA] display ipv6 routing-table
Routing Table :
         Destinations : 6        Routes : 6

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : Direct
NextHop    : 1::1                                       Preference: 0
Interface  : GE1/2                                      Cost      : 0

Destination: 1::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 2::/64                                     Protocol  : Direct
NextHop    : 2::1                                       Preference: 0
Interface  : GE1/1                                      Cost      : 0
```

```
Destination: 2::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

(3) 配置 RIPng 引入外部路由

在 Router B 上将两个不同 RIPng 进程的路由相互引入到对方的路由表中。

```
[RouterB] ripng 100
[RouterB-ripng-100] default cost 3
[RouterB-ripng-100] import-route ripng 200
[RouterB-ripng-100] quit
[RouterB] ripng 200
[RouterB-ripng-200] import-route ripng 100
[RouterB-ripng-200] quit
```

查看路由引入后 Router A 的路由表信息。

```
[RouterA] display ipv6 routing-table
Routing Table :
         Destinations : 7        Routes : 7

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : Direct
NextHop    : 1::1                                       Preference: 0
Interface  : GE1/2                                      Cost      : 0

Destination: 1::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 2::/64                                     Protocol  : Direct
NextHop    : 2::1                                       Preference: 0
Interface  : GE1/1                                      Cost      : 0

Destination: 2::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : RIPng
NextHop    : FE80::200:BFF:FE01:1C02                    Preference: 100
Interface  : GE1/2                                      Cost      : 4

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
```

```
Interface  : NULL0                                      Cost      : 0
```

目 录

1 IPv6 静态路由配置 1-1
1.1 IPv6 静态路由简介 1-1
1.1.1 IPv6 静态路由属性及功能 1-1
1.1.2 IPv6 缺省路由 1-1
1.2 配置IPv6 静态路由 1-1
1.2.1 配置准备 1-1
1.2.2 配置IPv6 静态路由 1-2
1.3 IPv6 静态路由显示和维护 1-2
1.4 IPv6 静态路由典型配置举例 1-2

1 IPv6 静态路由配置

1.1 IPv6 静态路由简介

静态路由是一种特殊的路由，它由管理员手工配置。当网络结构比较简单时，只需配置静态路由就可以使网络正常工作。恰当地设置和使用静态路由可以改进网络的性能，并可为重要的应用保证带宽。

静态路由的缺点在于：当网络发生故障或者拓扑发生变化后，可能会出现路由不可达，导致网络中断，此时必须由网络管理员手工修改静态路由的配置。

1.1.1 IPv6 静态路由属性及功能

IPv6 静态路由与 IPv4 静态路由类似，适合于一些结构比较简单的 IPv6 网络。

它们之间的主要区别是目的地址和下一跳地址有所不同，IPv6 静态路由使用的是 IPv6 地址，而 IPv4 静态路由使用 IPv4 地址。另外，IPv6 静态路由暂不支持 VPN 实例。

1.1.2 IPv6 缺省路由

在配置 IPv6 静态路由时，如果指定的目的地址为::/0（前缀长度为 0），则表示配置了一条 IPv6 缺省路由。如果报文的目的地址无法匹配路由表中的任何一项，路由器将选择 IPv6 缺省路由来转发 IPv6 报文。

1.2 配置 IPv6 静态路由

在小型 IPv6 网络中，可以通过配置 IPv6 静态路由达到网络互连的目的。相对使用动态路由来说，可以节省带宽。

1.2.1 配置准备

在配置 IPv6 静态路由之前，需完成以下任务：

- 配置相关接口的物理参数
- 配置相关接口的链路层属性
- 使能 IPv6 报文转发能力
- 相邻节点网络层（IPv6）可达

1.2.2 配置 IPv6 静态路由

表1-1 配置 IPv6 静态路由

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 静态路由（出接口类型为广播或者 NBMA）	**ipv6 route-static** *ipv6-address prefix-length* [*interface-type interface-number*] *nexthop-address* [**preference** *preference-value*]	必选 缺省情况下，IPv6 静态路由的优先级为 60
配置 IPv6 静态路由（出接口类型为点到点）	**ipv6 route-static** *ipv6-address prefix-length* { *interface-type interface-number* | *nexthop-address* } [**preference** *preference-value*]	

说明

在配置静态路由时，可以指定出接口（*interface-type interface-number*），也可指定下一跳地址（*nexthop-address*），具体采用哪种方法，需要根据实际情况而定：

- 如果出接口类型为广播（如以太网接口）或者 NBMA 类型（如封装帧中继的接口等），必须指定下一跳地址；
- 如果出接口类型为点到点类型（如串口等），可以指定出接口，也可以指定下一跳地址，但不能同时指定二者。

1.3 IPv6 静态路由显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令查看 IPv6 静态路由配置的运行情况并检验配置结果。

在系统视图下执行 **delete** 命令可以删除所有静态路由。

表1-2 IPv6 静态路由显示和维护

操作	命令
查看 IPv6 静态路由表信息	**display ipv6 routing-table protocol static** [**inactive** | **verbose**]
删除所有 IPv6 静态路由	**delete ipv6 static-routes all**

说明

使用 **undo ipv6 route-static** 命令可以删除一条 IPv6 静态路由，而使用 **delete ipv6 static-routes all** 命令可以删除包括缺省路由在内的所有 IPv6 静态路由。

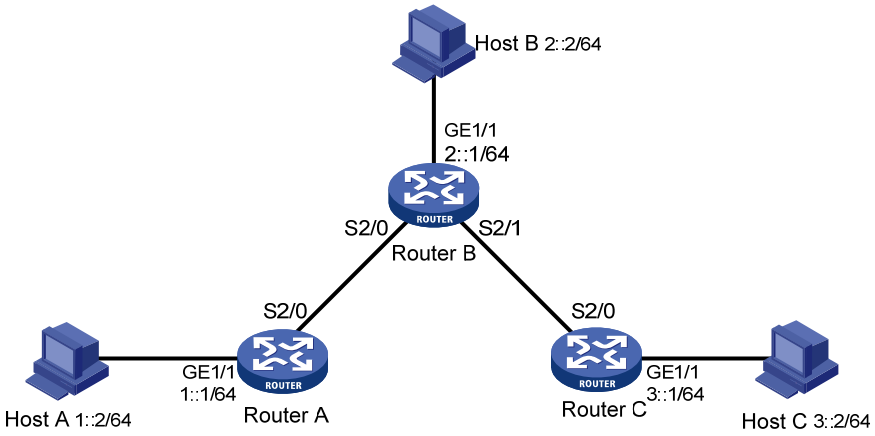
1.4 IPv6 静态路由典型配置举例

1. 组网要求

要求各路由器之间配置 IPv6 静态路由协议后，可以使所有主机和路由器之间互通。路由器的串口使用 IPv6 本地链路地址。

2. 组网图

图1-1 静态路由配置示例的组网图



3. 配置步骤

(1) 配置各接口的 IPv6 地址（略）

(2) 配置 IPv6 静态路由

在 Router A 上配置 IPv6 缺省路由。

```
<RouterA> system-view
[RouterA] ipv6 route-static :: 0 serial 2/0
```

在 Router B 上配置两条 IPv6 静态路由。

```
<RouterB> system-view
[RouterB] ipv6 route-static 1:: 64 serial 2/0
[RouterB] ipv6 route-static 3:: 64 serial 2/1
```

在 Router C 上配置 IPv6 缺省路由。

```
<RouterC> system-view
[RouterC] ipv6 route-static :: 0 serial 2/0
```

(3) 配置主机地址和网关

根据组网图配置好各主机的 IPv6 地址，并将 Host A 的缺省网关配置为 1::1，Host B 的缺省网关配置为 2::1，Host C 的缺省网关配置为 3::1。

(4) 查看配置结果

查看 Router A 的 IPv6 路由表。

```
[RouterA] display ipv6 routing-table
Routing Table :
         Destinations : 5        Routes : 5

 Destination  : ::/128                              Protocol     : Static
 NextHop      : FE80::510A:0:8D7:1                  Preference   : 60
 Interface    : S2/0                                Cost         : 0

 Destination  : ::/128                              Protocol     : Direct
 NextHop      : ::1                                 Preference   : 0
```

```
 Interface    : InLoop0                              Cost         : 0

 Destination  : 1::/64                               Protocol     : Direct
 NextHop      : 1::1                                 Preference   : 0
 Interface    : GE1/1                               Cost         : 0

 Destination  : 1::/128                              Protocol     : Direct
 NextHop      : ::1                                  Preference   : 0
 Interface    : InLoop0                              Cost         : 0

 Destination  : FE80::/10                            Protocol     : Direct
 NextHop      : ::                                   Preference   : 0
 Interface    : NULL0                                Cost         : 0
```

使用 Ping 进行验证。

```
[RouterA] ping ipv6 3::1
  PING 3::1 : 56  data bytes, press CTRL_C to break
    Reply from 3::1
    bytes=56 Sequence=1 hop limit=254  time = 63 ms
    Reply from 3::1
    bytes=56 Sequence=2 hop limit=254  time = 62 ms
    Reply from 3::1
    bytes=56 Sequence=3 hop limit=254  time = 62 ms
    Reply from 3::1
    bytes=56 Sequence=4 hop limit=254  time = 63 ms
    Reply from 3::1
    bytes=56 Sequence=5 hop limit=254  time = 63 ms

  --- 3::1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 62/62/63 ms
```

目 录

1 BFD配置 1-1
1.1 BFD简介 1-1
1.1.1 BFD工作机制 1-1
1.1.2 BFD报文格式 1-3
1.1.3 协议规范 1-4
1.2 BFD配置任务简介 1-4
1.3 配置BFD基本功能 1-5
1.3.1 配置准备 1-5
1.3.2 配置步骤 1-5
1.4 配置各协议与BFD联动 1-5
1.4.1 配置OSPF与BFD联动 1-5
1.4.2 配置IS-IS与BFD联动 1-6
1.4.3 配置RIP与BFD联动 1-6
1.4.4 配置BGP与BFD联动 1-7
1.4.5 配置MPLS与BFD联动 1-8
1.4.6 配置VRRP与BFD联动 1-8
1.4.7 配置静态路由与BFD联动 1-9
1.5 开启Trap功能 1-10
1.6 BFD显示和维护 1-11
1.7 BFD典型配置举例 1-11
1.7.1 配置OSPF与BFD联动 1-11
1.7.2 配置IS-IS与BFD联动 1-14
1.7.3 配置RIP运行BFD（echo报文单跳检测） 1-15
1.7.4 配置RIP运行BFD（control报文双向检测） 1-18
1.7.5 配置BGP与BFD联动 1-22
1.7.6 配置MPLS RSVP-TE与BFD联动 1-24
1.7.7 配置VRRP与BFD联动（Backup监视Master） 1-26
1.7.8 配置VRRP与BFD联动（监视上行链路） 1-29
1.7.9 配置静态路由与BFD联动（单跳检测） 1-31
1.7.10 配置静态路由与BFD联动（双向检测） 1-34

1 BFD 配置

说明

在以下路由协议的介绍中所指的路由器及路由器图标，代表了一般意义下的路由器以及运行了路由协议的以太网交换机。为提高可读性，在手册的描述中将不另行说明。

1.1 BFD 简介

BFD（Bidirectional Forwarding Detection，双向转发检测）是一套全网统一的检测机制，用于快速检测、监控网络中链路连通状况。为了提升现有网络性能，协议邻居之间必须能快速检测到通信故障，从而更快的建立起备用通道恢复通信。通常采用以下几种检测方法：

- 通过硬件检测信号（如 SDH 传输系统告警），快速检测到链路硬件上的故障。
- 在没有提供硬件检测信号，或不能通过硬件信号检测出故障时，网络通常采用路由协议中相对比较慢的 Hello 报文机制，检测到故障的时间超过 1 秒钟。当数据达到吉比特速率级时，这样长的检测到故障时间将导致大量的数据丢失。
- 用单一的机制对任何介质、任何协议层进行实时检测，并支持不同的检测时间与开销。

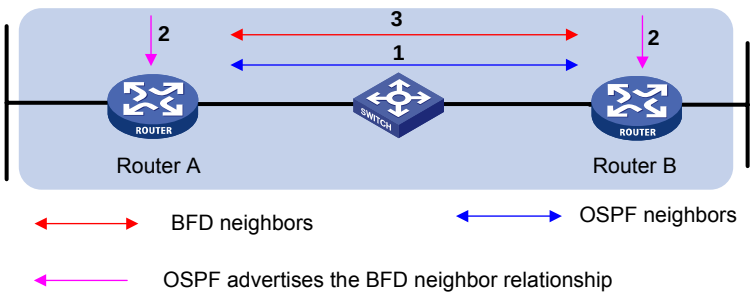
1.1.1 BFD 工作机制

BFD 提供了一个通用的、标准化的、介质无关、协议无关的快速故障检测机制，可以为各上层协议如路由协议、MPLS 等统一地快速检测两台路由器间双向转发路径的故障。

BFD 在两台路由器上建立会话，用来监测两台路由器间的双向转发路径，为上层协议服务。BFD 本身并没有发现机制，而是靠被服务的上层协议通知其该与谁建立会话，会话建立后如果在检测时间内没有收到对端的 BFD 控制报文则认为发生故障，通知被服务的上层协议，上层协议进行相应的处理。

1. BFD 工作流程

图1-1 BFD 建立流程图

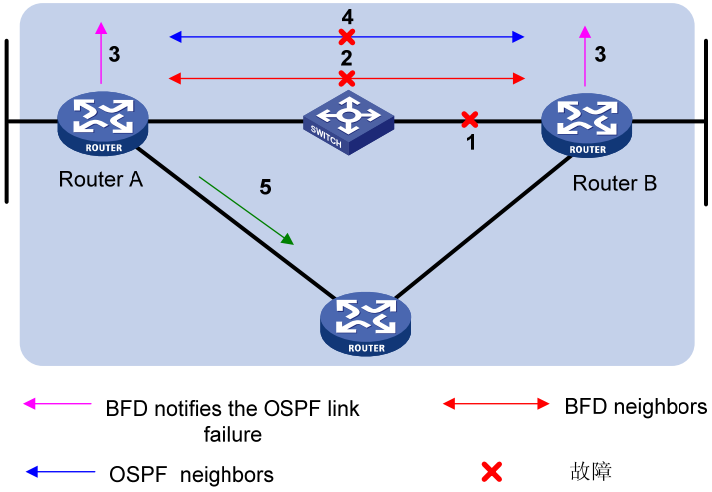


BFD 建立过程：

- 上层协议通过自己的 Hello 机制发现邻居并建立连接；
- 上层协议在建立了新的邻居关系时，将邻居的参数及检测参数都（包括目的地址和源地址等）通告给 BFD；

- BFD 根据收到的参数进行计算并建立邻居。

图1-2 BFD 发生故障处理流程图



当网络出现故障时：

- BFD 检测到链路/网络故障；
- 拆除 BFD 邻居会话；
- BFD 通知本地上层协议进程 BFD 邻居不可达；
- 本地上层协议中止上层协议邻居关系；
- 如果网络中存在备用路径，路由器将选择备用路径。

说明

BFD 草案中没有规定检测的时间精度，目前支持 BFD 的设备大多数提供的是毫秒级检测。

---

2. BFD 会话的工作方式

- 控制报文方式：链路两端会话通过控制报文交互监测链路状态。
- Echo 报文方式：链路某一端通过发送 Echo 报文由另一端转发回来，实现对链路的双向监测。

3. BFD 运行模式

BFD 会话建立前有两种模式：主动模式和被动模式。

- 主动模式：在建立会话前不管是否收到对端发来的 BFD 控制报文，都会主动发送 BFD 控制报文；
- 被动模式：在建立会话前不会主动发送 BFD 控制报文，直到收到对端发送来的控制报文；

在会话初始化过程中，通信双方至少要有一个运行在主动模式才能成功建立起会话。

BFD 会话建立后有两种模式：异步模式和查询模式。通信双方要求运行在相同的模式。

- 异步模式：以异步模式运行的设备周期性地发送 BFD 控制报文，如果在检测时间内没有收到 BFD 控制报文则将会话 down。
- 查询模式：假定每个协议都有一个独立的方法，确认自己连接到其他协议。这样，只要有一个 BFD 会话建立，协议停止发送 BFD 控制报文，除非某个协议需要显式地验证连接性。

说明

- 目前仅支持异步模式。
- 当 BFD 会话工作于 echo 报文方式时，不受运行模式控制。
- 在需要显式验证连接性的情况下，系统以协商的周期连续发送几个 P 比特位置 1 的 BFD 控制报文。如果在检测时间内没有收到返回的报文，就认为会话 down；如果认为连通，则不再发送报文，等待下一次查询的触发。

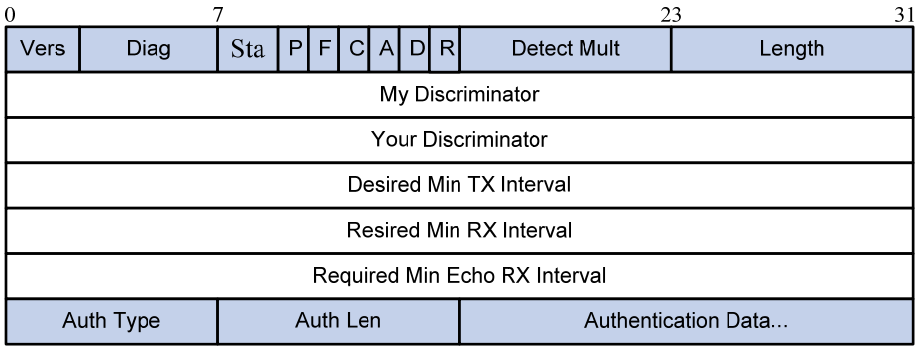
4. 动态改变 BFD 参数功能

会话建立后，可以动态协商 BFD 的相关参数（例如最小发送间隔、最小接收间隔、初始模式等），两端协议通过发送相应的协商报文后采用新的参数，不影响会话的当前状态。

1.1.2 BFD 报文格式

BFD控制报文为UDP报文，其端口号为 3784，如 图 1-3

图1-3 BFD 控制报文格式图



- Vers：协议的版本号，协议版本为 1；
- Diag：给出本地协议最后一次从up状态转换到其他状态的原因如 表 1-1；

表1-1 Diag 原因描述

Diag	描述
0	无诊断信息（No Diagnostic）
1	控制检测超时（Control Detection Time Expired）
2	回声功能失效（Echo Function Failed）
3	邻居通知会话 down（Neighbor Signaled Session Down）
4	转发平面重启（Forwarding Plane Reset）
5	通道失效（Path Down）
6	连接通道失效（Concatenated Path Down）
7	管理 down（Administratively Down）
8～31	保留位（Reserved for future use）

- State（Sta）：BFD 会话当前状态，取值为：0 代表 AdminDown，1 代表 Down，2 代表 Init，3 代表 Up。

- Demand（D）：设置为 1，表示发送协议希望操作在查询模式；设置为 0，表示发送协议不区分操作在查询模式，或者表示发送协议不能操作在查询模式；
- Poll（P）：设置为 1，表示发送协议请求进行连接确认，或者发送请求参数改变的确认；设置为 0，表示发送协议不请求确认；
- Final（F）：设置为 1，表示发送协议响应一个接收到 P 比特为 1 的 BFD 包；设置为 0，表示发送协议不响应一个 P 比特为 1 的 BFD 包；
- Control Plane Independent（C）：设置为 1，表示发送协议的 BFD 实现不依赖于它的控制平面（换句话说，BFD 在转发平面实施，即使控制平面失效了，BFD 仍然能够起作用）；设置为 0，表示 BFD 在控制平面实施；
- Authentication Present（A）：如果设置为 1，则表示控制包包含认证字段，并且会话是被认证的；
- Reserved（R）：在发送时设置为 0，在接收时忽略；
- Detect Mult：检测时间倍数。
- Length：BFD 控制包的长度，单位字节；
- My Discriminator：发送协议产生的一个唯一的、非 0 鉴别值，用来对两个协议之间的多个 BFD 会话进行分离；
- Your Discriminator：从远端协议接收到的鉴别值，这个域直接返回接收到的“My Discriminator”，如果不知道这个值就返回 0；
- Desired Min Tx Interval：本地协议发送 BFD 控制包时想要采用的最小间隔，单位毫秒；
- Required Min Rx Interval：本地协议能够支持的接收两个 BFD 控制包之间的间隔，单位毫秒；
- Required Min Echo Rx Interval：本地协议能够支持的接收两个 BFD 回声包之间的间隔，单位毫秒。如果这个值设置为 0，则发送协议不支持接收 BFD 回声包；
- Auth Type：BFD 控制包使用的认证类型；
- Auth Len：认证字段的长度，包括认证类型与认证长度字段。

1.1.3 协议规范

与 BFD 相关的协议规范有：

- draft-ietf-bfd-base-05：Protocol Independent Bidirectional Forwarding Detection
- draft-ietf-bfd-v4v6-1hop-05：BFD for IPv4 and IPv6 （Single Hop）

1.2 BFD 配置任务简介

在需要为网络提供检测机制的情况下，可以配置 BFD 的各项功能。

表1-2 BFD 配置任务简介

配置任务	说明	详细配置
配置 BFD 基本功能	可选 双向转发检测的基本配置，是其他配置任务的基础	1.3
配置各协议与 BFD 联动	必选 使能 BFD 应用，实现在运行 OSPF、IS-IS、RIP、BGP、MPLS、VRRP、静态路由协议的链路上进行转发连通状况检测	1.4

1.3 配置 BFD 基本功能

1.3.1 配置准备

在配置 BFD 检测方式之前，需完成以下任务：

- 配置接口的网络层地址，使相邻节点之间网络层可达
- 配置可支持 BFD 的路由协议

1.3.2 配置步骤

表1-3 配置 BFD 基本功能

操作	命令	说明
进入系统视图	**system-view**	-
使能会话建立模式	**bfd session init-mode { active | passive }**	可选 缺省情况下，会话模式为 **active**
进入接口视图	**interface** *interface-type interface-number*	-
配置接口最小报文发送时间间隔	**bfd min-transmit-interval** *value*	可选 缺省情况下，接口最小报文发送间隔为 400ms
配置接口 echo 报文最小接收时间间隔	**bfd min-echo-receive-interval** *value*	可选 缺省情况下，接口 echo 报文最小接收间隔为 400ms
配置接口最小报文接收时间间隔	**bfd min-receive-interval** *value*	可选 缺省情况下，接口最小报文接收间隔为 400ms
配置会话监测系数	**bfd detect-multiplier** *value*	可选 缺省情况下，会话监测系数值为 5

1.4 配置各协议与 BFD 联动

1.4.1 配置 OSPF 与 BFD 联动

OSPF 使用 BFD 来进行快速故障检测时，OSPF 可以通过 Hello 报文动态发现邻居，OSPF 将邻居地址通知 BFD 就开始建立会话。BFD 会话建立前处于 down 状态，此时 BFD 控制报文以不小于 1 秒的时间间隔周期发送以减少控制报文流量，直到会话建立以后才会以协商的时间间隔发送以实现快速检测。进行配置 BFD 之前，需要配置 OSPF 功能。

表1-4 配置 OSPF 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
指定接口上使能 BFD	**ospf bfd enable**	必选 缺省情况下，OSPF 下的接口不使用 BFD 提供的链路检测功能

说明

- 一个网段只能属于同一个区域，并且必须为每个运行 OSPF 协议的接口指明属于某一个特定的区域。
- 创建 BFD 会话的通信双方必须处于特定区域的同一网段。
- 有关 OSPF 协议的配置，请参见“IP 路由分册”中的“OSPF 配置”。

1.4.2 配置 IS-IS 与 BFD 联动

进行本配置之前，首先配置各节点 IS-IS 的功能。

表1-5 配置 IS-IS 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
指定接口上使能 IS-IS	**isis enable** [*process-id*]	必选 缺省情况下，接口上不使能 IS-IS 路由进程
指定接口上使能 BFD	**isis bfd enable**	必选 缺省情况下，运行 IS-IS 的接口不使用 BFD 提供的链路检测功能

说明

有关 IS-IS 的配置，请参见“IP 路由分册”中的“IS-IS 配置”。

1.4.3 配置 RIP 与 BFD 联动

RIP 协议依赖周期性发送路由更新请求作为检测机制，当在指定时间内没有收到路由更新回应时，认为此条路由不再生效。这种方式不能快速响应链路故障。当 BFD 检测到链路中断时，RIP 能快速撤销失效路由，而不需等待 Update 定时器超时。

RIP 支持 BFD 提供了两种检测方式：

- 直连邻居采用 echo 报文单跳检测方式，在对端有路由发送时才能建立 BFD Session。
- 非直连邻居采用 control 报文双向检测方式，当两端互有路由发送时，且使能 BFD 的接口与接收接口为同一接口，邻居之间才能建立 BFD Session。

1. echo 报文单跳检测

表1-6 配置 RIP 与 BFD 联动（echo 报文单跳检测）

操作	命令	说明
进入系统视图	**system-view**	-
配置 echo 报文源地址	bfd echo-source-ip ip-address	必选 缺省情况下，没有配置 echo 报文源地址
进入接口视图	**interface** *interface-type interface-number*	-
使能 BFD 功能	**rip bfd enable**	必选 缺省情况下，BFD 功能处于关闭状态

2. control 报文双向检测

表1-7 配置 RIP 与 BFD 联动（control 报文双向检测）

操作	命令	说明
进入系统视图	**system-view**	-
创建 RIP 进程并进入 RIP 视图	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，RIP 进程处于关闭状态 **vpn-instance** *vpn-instance-name* 参数的支持情况与设备的型号有关，请以设备的实际情况为准
配置 RIP 邻居	**peer** *ip-address*	必选 缺省情况下，RIP 不向任何定点地址发送更新报文
进入接口视图	**interface** *interface-type interface-number*	-
使能 BFD 功能	**rip bfd enable**	必选 缺省情况下，BFD 功能处于关闭状态

说明

- 当检测 RIP 的会话两端在直连网段（即 IP 报文的一跳），适合采用 BFD 的 echo 单向检测方式，但是，经过多跳到达邻居时 echo 方式则会失效。
- 由于 **peer** 命令与邻居之间没有对应关系，**undo peer** 操作并不能立刻删除邻居，因此不能立刻删除 BFD 会话。
- 有关 RIP 的配置，请参见“IP 路由分册”中的“RIP 配置”。

1.4.4 配置 BGP 与 BFD 联动

缺省情况下，BGP 协议的 keepalive 时间间隔为 60 秒，holdtime 时间间隔为 180 秒。如果使用命令行修改，当保持时间和存活时间间隔都不为 0 时，设置的保持时间应该至少为存活时间间隔的三倍。这样的配置要求使得 BGP 邻居关系的检测比较慢，对于报文收发速度快的接口会导致大量报文丢失。通过 BFD 进行快速故障检测，可以实现邻居关系的快速检测，加快协议收敛。

表1-8 配置 BGP 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
指定邻居上使能 BFD	**peer** *ip-address* **bfd**	必选 缺省情况下，BGP 不使用 BFD 提供的链路检测功能



说明

- 目前只支持在 BGP IPv4 邻居上配置 BFD。
- 在 BGP 使能 GR 能力后，请慎用 BFD。
- 有关 BGP 协议的配置，请参见“IP 路由分册”中的“BGP 配置”。

1.4.5 配置 MPLS 与 BFD 联动

MPLS 运行 BFD 用于邻居快速检测，检测到的故障一般用于主备转发表项的切换：即 MPLS 将业务从主路径切换到备份路径上。在使能 BFD 提供的链路检测功能之前，需要配置 MPLS 相关功能。

1. 配置 MPLS RSVP-TE 与 BFD 联动

表1-9 配置 MPLS RSVP-TE 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
进入已使能 MPLS RSVP-TE 的接口视图	**interface** *interface-type interface-number*	-
配置 MPLS RSVP-TE 运行 BFD	**mpls rsvp-te bfd enable**	必选 缺省情况下，已使能 RSVP 的接口下未使能 BFD 提供的链路检测功能



说明

有关 MPLS RSVP-TE 的配置请参考“MPLS 分册”中的“MPLS TE 配置”。

1.4.6 配置 VRRP 与 BFD 联动

VRRP 和 BFD session 联动时，必须先配置 Track 对象，通过创建 Track 对象和 BFD 互动，同时也需要配置 VRRP 和 Track 对象的绑定关系。

Track 负责启动和停止 BFD session，当 BFD 探测到邻居出现故障后通知 Track，再由 Track 通知 VRRP 根据配置实现 VRRP 快速切换。

在配置 VRRP 和 Track 的关联之前，需要在接口上创建了 VRRP 备份组，配置虚拟 IP 地址。

表1-10 配置 VRRP 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
配置 echo 报文源地址	**bfd echo-source-ip** *ip-address*	必选 缺省情况下，没有配置 echo 报文源地址
配置绑定 Track 对象和 BFD	**track** *track-entry-number* **bfd echo interface** *interface-type interface-number* **remote ip** *remote-ip* **local ip** *local-ip*	必选 缺省情况下，没有配置 Track 对象和 BFD 会话的关联关系
进入接口视图	**interface** *interface-type interface-number*	-
配置绑定 VRRP 和 Track 对象	**vrrp vrid** *virtual-router-id* **track** *track-entry-number* { **switchover** | [**reduced priority-reduced**] }	必选 缺省情况下，没有指定被监视的接口

注意

- Track 对象的 local ip 必须是 BFD 会话出接口上的某个 IP 地址，并且必须和 remote ip 在同一网段。
- 有关 VRRP 的配置，请参见“系统分册”中的“VRRP 配置”。
- 有关 Track 的配置，请参见“系统分册”中的“Track 配置”。

1.4.7 配置静态路由与 BFD 联动

BFD 是一个双向检测协议，需要检测两端建立会话。对于动态路由协议来说是有邻居概念的，因此在检测两端动态路由协议都会通知 BFD 会话邻居信息，从而检测两端的 BFD 任务可以通过向邻居发送 BFD 控制报文来建立会话。由于静态路由没有什么邻居的概念一般使用以下方法解决：

1. 双向检测

静态路由使用控制报文方式 BFD 功能时，对端也必须存在对应的 BFD 会话。检测两个方向上的链路状态，实现毫秒级别的链路故障检测。

表1-11 配置静态路由与 BFD 联动（双向检测）

操作	命令	说明
进入系统视图	**system-view**	-
配置静态路由运行 BFD	**ip route-static** *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	二者必选其一
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

2. 单跳检测

利用 BFD echo 报文，通过报文建立会话，echo 报文的目的地址为本设备接口地址，发送给下一跳设备后会直接转发回本设备而不经过 BFD 任务处理。这里所说的“单跳”是 IP 的一跳。

表1-12 配置静态路由与 BFD 联动（单跳检测）

操作	命令	说明
进入系统视图	**system-view**	-
配置 echo 报文源地址	**bfd echo-source-ip** *ip-address*	必选 缺省情况下，没有配置 echo 报文源地址
配置静态路由运行 BFD	**ip route-static** *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	二者必选其一
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* | *mask-length* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

注意

- 路由振荡时，使能 BFD 检测功能可能会加剧振荡，需谨慎使用。
- 当 BFD 会话工作于 echo 报文方式时，必须配置 echo 报文源地址。
- 配置静态路由与 BFD 联动时，必须同时指定出接口和路由的下一跳 IP 地址。
- 静态路由如果出接口为含 SPOOFING 属性的接口，不能使用 BFD 进行检测。
- 静态路由仅支持使用 BFD 检测直连的下一跳。如果静态路由配置的下一跳非直连，则不支持 BFD 检测。
- 在草案中 BFD echo 功能进行了修改，当使用 echo 报文方式时，仅在一端建立 BFD 会话。
- 有关静态路由协议的配置，请参见“IP 路由分册”中的“静态路由配置”。

1.5 开启 Trap 功能

开启 BFD 模块的 Trap 功能后，模块会生成级别为 notifications 的 Trap 报文，用于报告该模块的重要事件。生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-13 配置 BFD 的 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 BFD 的 Trap 功能	**snmp-agent trap enable bfd**	可选 缺省情况下，Trap 开关打开，BFD 发送 Trap 报文

说明

snmp-agent trap enable isdn 命令的详细介绍请参见“系统分册/SNMP 命令”中的 **snmp-agent trap enable** 命令。

1.6 BFD 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 BFD 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 BFD 会话的统计信息。

表1-14 BFD 显示和维护

操作		命令
显示使能的 BFD 接口信息		**display bfd interface** [**verbose**]
显示使能的 BFD 调试信息开关		**display bfd debugging-switches**
显示 BFD 会话信息	集中式设备	**display bfd session** [**verbose**]
	分布式设备	**display bfd session** [**verbose**] [**slot** *slot-number* [**all** | **verbose**]]
清除 BFD 会话统计信息	集中式设备	**reset bfd session statistics**
	分布式设备	**reset bfd session statistics** [**slot** *slot-number*]

1.7 BFD 典型配置举例

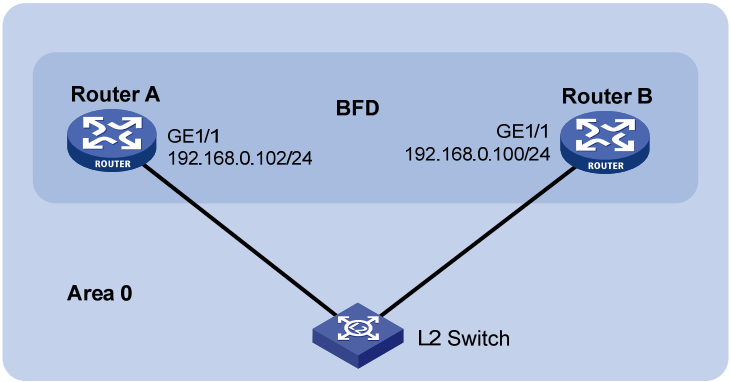
1.7.1 配置 OSPF 与 BFD 联动

1. 组网需求

- Router A、Router B 通过二层交换机互连，并且在双方接口上使能 BFD 应用，在设备上运行 OSPF 使网络层相互可达。
- 在 Router A 和二层交换机之间的链路发生故障后，BFD 能够快速检测并通告 OSPF 协议。

2. 组网图

图1-4 OSPF 与 BFD 联动配置组网图



3. 配置步骤

(1) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
```

```
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ospf bfd enable
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ospf bfd enable
[RouterB-GigabitEthernet1/1] quit
```

(2) 配置各路由器接口及 BFD 参数

配置 Router A 各路由器接口及 BFD 参数。

```
<RouterA> system-view
[RouterA] bfd session init-mode active
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.102 24
[RouterA-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 7
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bfd session init-mode active
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 192.168.0.100 24
[RouterB-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterB-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterB-GigabitEthernet1/1] bfd detect-multiplier 6
```

(3) 检查配置结果

打开 Router A 的调试开关。

```
<RouterA> debugging bfd scm
<RouterA> debugging bfd event
<RouterA> debugging ospf event bfd
<RouterA> terminal debugging
```

Router A 和二层交换机之间的链路发生故障后。可以看到 Router A 能够快速检测 Router B 的变化。

```
%Nov 12 18:34:48:823 2005 RouterA BFD/5/LOG: Sess[192.168.0.102/192.168.0.100, GE1/1], Sta :
UP->DOWN, Diag: 1
```

```
%Nov 12 18:34:48:824 2005 RouterA RM/4/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
192.168.0.102 (GigabitEthernet1/1) from Full to Down
*0.50673825 RouterA BFD/8/SCM:Sess[192.168.0.102/192.168.0.100,GE1/1],Oper: Reset
*0.50673825 RouterA BFD/8/EVENT:Send sess-down Msg, [Src：192.168.0.102, Dst：192.168.0.100,
GigabitEthernet1/1] Protocol: OSPF
*0.50673826 RouterA RM/7/RMDEBUG:OSPF-BFD: Message Type rcv BFD down, Connect Type
direct-connect, Src IP Address 192.168.0.102, Src IFIndex 5, Dst IP Address 192.168.0.100
*0.50673827 RouterA RM/7/RMDEBUG:OSPF-BFD: Message Type delete session, Connect Type
direct-connect, Src IP Address 192.168.0.102, Src IFIndex 5, Dst IP Address 192.168.0.100
OSPF 1: Nbr 192.168.0.100 Rcv KillNbr State Full -> Down.
*0.50673829 RouterA BFD/8/EVENT:Receive Delete-sess, [Src:192.168.0.102,
Dst:192.168.0.100,GigabitEthernet1/1], Direct, Proto:OSPF
*0.50673830 RouterA BFD/8/SCM:Sess[192.168.0.102/192.168.0.100,GE1/1], Oper: Del
application(OSPF)
*0.50673831 RouterA BFD/8/SCM:No application in session, delete
session[192.168.0.102/192.168.0.100,GE1/1]
*0.50673831 RouterA BFD/8/SCM:Sess[192.168.0.102/192.168.0.100,GE1/1], Oper: Delete
*0.50673832 RouterA BFD/8/SCM:Delete send-packet timer
*0.50673833 RouterA BFD/8/SCM:Delete session entry
*0.50673833 RouterA BFD/8/SCM:Delete session from IP hash table
*0.50673834 RouterA BFD/8/SCM:Delete session from bfd interface
*0.50673834 RouterA BFD/8/SCM:No session under bfd-int[GigabitEthernet1/1] with default
configuration, delete bfd-if
*0.50673835 RouterA BFD/8/SCM:Bfd-if[GigabitEthernet1/1], Oper: Delete
*0.50673840 RouterA BFD/8/SCM:No bfd session exists, stop receiving any bfd packets
```

显示 Router A 的 BFD 邻居详细信息。

```
<RouterA> display bfd session verbose

Total Session Num: 1            Init Mode: Active

 Session Working Under Ctrl Mode:

     Local Discr: 1                   Remote Discr: 1
       Source IP: 192.168.0.102     Destination IP: 192.168.0.100
   Session State: Up                      Interface: GigabitEthernet1/1
 Min Trans Inter: 400ms            Act Trans Inter: 400ms
  Min Recv Inter: 400ms           Act Detect Inter: 2800ms
    Recv Pkt Num: 362                 Send Pkt Num: 362
       Hold Time: 2600ms              Connect Type: Direct
  Running Up for: 00:02:20               Auth mode: None
        Protocol: OSPF
       Diag Info: No Diagnostic
```

当 Router A 和二层交换机之间的链路发生故障，BFD 快速检测到链路发生变化立刻通知 OSPF。使用 **display ospf peer verbose** 命令观察到 Router A 的 OSPF 邻居关系转变为 down 状态。

```
<RouterA> display ospf peer verbose
Last Neighbor Down Event:
 Router ID: 192.168.0.100
```

```
Local Address: 192.168.0.102
Remote Address: 192.168.0.100
Time: May 14 09:39:50 2007
Reason: BFD session down
```

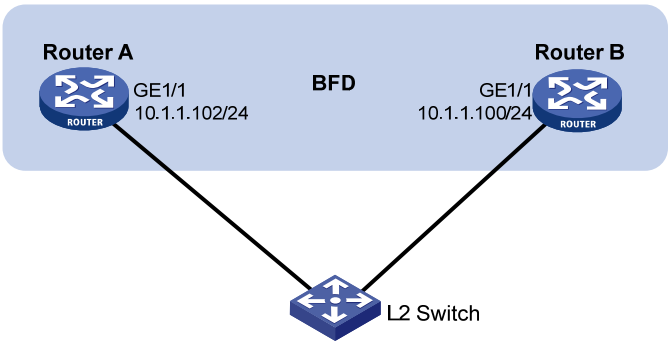
1.7.2 配置 IS-IS 与 BFD 联动

1. 组网需求

- Router A、Router B 通过二层交换机互连，并且在双方接口上使能 BFD 应用，之间运行 IS-IS，网络层相互可达。
- 当断开 Router A 和二层交换机之间的链路后，BFD 能够快速检测并通告 IS-IS 协议。

2. 组网图

图1-5 配置 IS-IS 与 BFD 联动组网图



3. 配置步骤

(1) 配置 IS-IS

配置 Router A。

```
<RouterA> system-view
[RouterA] isis
[RouterA-isis-1] network-entity 10.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface GigabitEthernet 1/1
[SwitchA-GigabitEthernet1/1] ip address 10.1.1.102 24
[RouterA-GigabitEthernet1/1] isis enable
[RouterA-GigabitEthernet1/1] isis bfd enable
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis
[RouterB-isis-1] network-entity 10.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface GigabitEthernet 1/1
[SwitchB-GigabitEthernet1/1] ip address 10.1.1.100 24
[RouterB-GigabitEthernet1/1] isis enable
[RouterB-GigabitEthernet1/1] isis bfd enable
```

```
[RouterB-GigabitEthernet1/1] quit
```

(2) 配置各路由器接口及 BFD 参数

配置 Router A。

```
<RouterA> system-view
[RouterA] bfd session init-mode active
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 7
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bfd session init-mode active
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterB-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterB-GigabitEthernet1/1] bfd detect-multiplier 8
```

(3) 检查配置结果

显示 Router B 的 BFD 邻居详细信息。

```
[RouterB-GigabitEthernet1/1] display bfd interface verbose

 Total Interface Num: 1

      Interface: GigabitEthernet1/1            Session Num: 1
Min Trans Inter: 400ms               Min Recv Inter: 400ms
     DetectMult: 8              Min Echo Recv Inter: 400ms
      Auth mode: None

 LD/RD         SourceAddr      DestAddr        ConnType  State     Mode
 7/7           10.1.1.100      10.1.1.102      Direct    Up        Ctrl
```

断开 Router B 和二层交换机之间的链路后使用 display bfd interface verbose 命令查看 GigabitEthernet1/1 接口下会话信息被删除。

```
[RouterB-GigabitEthernet1/1] display bfd interface verbose

 Total Interface Num: 1

      Interface: GigabitEthernet1/1            Session Num: 0
Min Trans Inter: 400ms               Min Recv Inter: 400ms
     DetectMult: 8              Min Echo Recv Inter: 400ms
      Auth mode: None
```

1.7.3 配置 RIP 运行 BFD（echo 报文单跳检测）

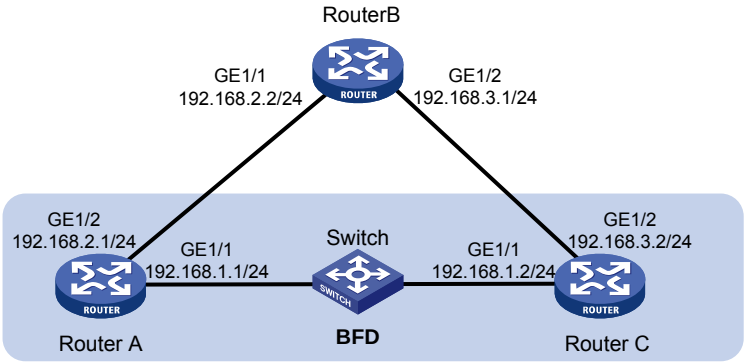
1. 组网需求

- Router A 和 Router C 通过二层交换机互连，它们的接口 GigabitEthernet1/1 都运行 RIP 进程 1。并且 Router A 的接口 GigabitEthernet1/1 上还使能了 BFD 检测功能。

- Router A 通过 Router B 与 Router C 互连，Rourer A 的接口 GigabitEthernet1/2 运行 RIP 进程 2。Rouer C 的接口 GigabitEthernet1/2、Rourer B 的接口 GigabitEthernet1/1 和 GigabitEthernet1/2 上都运行 RIP 进程 1。
- Router C 上配置静态路由，并将静态路由引入 RIP 进程中，使 Router C 有路由发送至 Router A。Router A 上学习到 Router C 发送的静态路由，出接口为与二层交换机相连的接口。
- 在 Router A 和交换机之间的链路发生故障后，BFD 能够快速检测链路中断并通告 RIP 协议。RIP 协议响应 BFD 会话 down，删除与 Router C 的邻居，并删除从 Router C 学习的路由。Router A 上学习到 Router C 发送的静态路由，出接口为与 Router B 连接的接口。

2. 组网图

图1-6 配置 RIP 运行 BFD 特性组网图（echo 报文单跳检测）



3. 配置步骤

(1) 配置 RIP 基本功能并且在接口上使能 BFD

配置 Router A。

```
<RouterA> system-view
[RouterA] rip 1
[RouterA-rip-1] network 192.168.1.0
[RouterA-rip-1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] rip bfd enable
[RouterA-GigabitEthernet1/1] quit
[RouterA] rip 2
[RouterA-rip-2] network 192.168.2.0
[RouterA- rip-2] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] rip 1
[RouterB-rip-1] network 192.168.2.0
[RouterB-rip-1] network 192.168.3.0
[RouterB-rip-1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] rip 1
```

```
[RouterC-rip-1] network 192.168.1.0
[RouterC-rip-1] network 192.168.3.0
[RouterC-rip-1] import-route static
[RouterC-rip-1] quit
```

(2) 配置接口 BFD 参数

配置 Router A。

```
[RouterA] bfd session init-mode active
[RouterA] bfd echo-source-ip 11.11.11.11
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 7
[RouterA-GigabitEthernet1/1] return
```

(3) Router C 配置静态路由

```
[RouterC] ip route-static 100.1.1.1 24 null 0
```

(4) 检查配置结果

显示 Router A 的 BFD 信息。

```
<RouterA> display bfd session
 Total Session Num: 1            Init Mode: Active
 Session Working Under Echo Mode:
 LD          SourceAddr      DestAddr        State   Holdtime    Interface
 3           192.168.1.1     192.168.1.2     Up      2000ms      Eth1/1
```

显示 Router A 上学到的路由 100.1.1.0/24。

```
<RouterA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 2
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 1
   Preference: 100                   Cost: 1
      NextHop: 192.168.1.2      Interface: GigabitEthernet1/1
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.1.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h00m47s
          Tag: 0
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
      NextHop: 192.168.2.2      Interface: GigabitEthernet1/2
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.2.2
    Tunnel ID: 0x0                  Label: NULL
        State: Inactive Adv           Age: 00h12m50s
          Tag: 0
```

打开 Router A 调试信息开关。

```
<RouterA> debugging rip 1 event
<RouterA> terminal debugging
```

Router C 和二层交换机之间的链路发生故障后。可以看到 Router A 能够快速检测与二层交换机之间的链路的变化。

```
%Jan 19 10:41:51:203 2008 RouterA BFD/4/LOG:Sess[192.168.1.1/192.168.1.2, Eth1/1,Ctrl], Sta: UP->DOWN, Diag: 1
*Jan 19 10:33:12:813 2008 RouterA RM/6/RMDEBUG: RIP-BFD: Message Type Disable, Connect Type Direct-connect, Pkt Type Echo, Src IP Address 192.168.1.1, Src IFIndex4, Nbr IP Address 192.168.1.2.
```

显示 Router A 的 BFD 信息。

查看 Router A 的 BFD 信息，Router A 已取消与 Router C 的邻居关系，没有任何输出信息。

```
<RouterA> display bfd session
```

显示 Router A 的 RIP 进程 1 路由信息。

查看 Router A 的 RIP 进程 1 的路由信息，没有从 Router C 学到的 RIP 路由。

```
<RouterA> display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
```

显示 Router A 上学到的路由 100.1.1.0/24。

```
<RouterA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 1
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
      NextHop: 192.168.2.2      Interface: GigabitEthernet1/2
    BkNextHop: 0.0.0.0        BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.2.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h18m40s
          Tag: 0
```

1.7.4 配置 RIP 运行 BFD（control 报文双向检测）

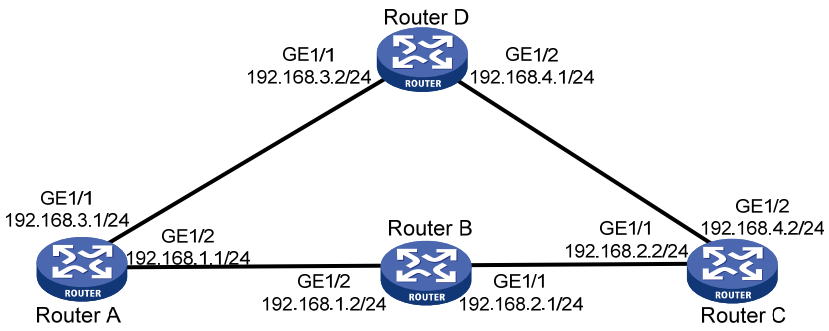
1. 组网需求

- Router A、Router B 与 Router C 互连。Rourer A 的接口 GigabitEthernet1/2、Rouer C 的接口 GigabitEthernet1/1、Rourer B 的接口 GigabitEthernet1/1 和 GigabitEthernet1/2 上都运行 RIP 进程 1。
- 分别在 Router A 和 Router C 上配置到达对端的静态路由，并在 Rourer A 的接口 GigabitEthernet1/2 和 Rouer C 的接口 GigabitEthernet1/1 上使能 BFD 检测功能。
- Router A 通过 Router D 与 Router C 互连。Rourer A 的接口 GigabitEthernet1/1 运行 RIP 进程 2。Rouer C 的接口 GigabitEthernet1/2、Rourer D 的接口 GigabitEthernet1/1 和 GigabitEthernet1/2 上运行 RIP 进程 1。

- 为使 Router A 与 Router C 互有路由发送，Router A 与 Router C 上的 RIP 协议都要配置引入静态路由。Router A 建立至 Router C 的会话。Router A 上学习到 Router C 发送的静态路由，出接口为与 Router B 连接的接口。
- 在 Router B 与 Router C 之间的链路发生故障后，BFD 能够快速检测链路中断并通告 RIP 协议。RIP 协议响应 BFD 会话 down，删除与 Router C 的邻居，并删除从 Router C 学习的路由。Router A 上学习到 Router C 发送的静态路由，出接口为与 Router D 连接的接口。

2. 组网图

图1-7 配置 RIP 运行 BFD 特性组网图（control 报文双向检测）



3. 配置步骤

(1) 配置 RIP 基本功能，并引入静态路由，使 Router A 与 Router C 互有路由发送

配置 Router A。

```
<RouterA> system-view
[RouterA] rip 1
[RouterA-rip-1] network 192.168.1.0
[RouterA-rip-1] peer 192.168.2.2
[RouterA-rip-1] undo validate-source-address
[RouterA-rip-1] import-route static
[RouterA-rip-1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] rip bfd enable
[RouterA-GigabitEthernet1/2] quit
[RouterA] rip 2
[RouterA-rip-2] network 192.168.3.0
[RouterA-rip-2] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] rip 1
[RouterC-rip-1] network 192.168.2.0
[RouterC-rip-1] network 192.168.4.0
[RouterC-rip-1] peer 192.168.1.1
[RouterC-rip-1] undo validate-source-address
[RouterC-rip-1] import-route static
[RouterC-rip-1] quit
```

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] rip bfd enable
[RouterC-GigabitEthernet1/1] quit
```

配置 Router D。

```
<RouterC> system-view
[RouterC] rip 1
[RouterC-rip-1] network 192.168.3.0
[RouterC-rip-1] network 192.168.4.0
[RouterC-rip-1] quit
```

(2) 配置接口及 BFD 参数

配置 Router A。

```
[RouterA] bfd session init-mode active
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.3.1 24
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ip address 192.168.1.1 24
[RouterA-GigabitEthernet1/2] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/2] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/2] bfd detect-multiplier 7
[RouterA-GigabitEthernet1/2] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/2
[RouterB-GigabitEthernet1/2] ip address 192.168.1.2 24
[RouterB-GigabitEthernet1/2] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 192.168.2.1 24
```

配置 Router C。

```
[RouterC] bfd session init-mode active
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ip address 192.168.2.2 24
[RouterC-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterC-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterC-GigabitEthernet1/1] bfd detect-multiplier 6
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface GigabitEthernet 1/2
[RouterC-GigabitEthernet1/2] ip address 192.168.4.2 24
[RouterC-GigabitEthernet1/2] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] interface GigabitEthernet 1/2
[RouterD-GigabitEthernet1/2] ip address 192.168.4.1 24
[RouterD-GigabitEthernet1/2] quit
[RouterD] interface GigabitEthernet 1/1
```

```
[RouterD-GigabitEthernet1/1] ip address 192.168.3.2 24
[RouterD-GigabitEthernet1/1] quit
```

(3) 配置静态路由

配置 Router A。

```
[RouterA] ip route-static 192.168.2.0 24 GigabitEthernet1/2 192.168.1.2
[RouterA] quit
```

配置 Router C。

```
[RouterC] ip route-static 192.168.1.0 24 GigabitEthernet1/1 192.168.2.1
[RouterC] ip route-static 100.1.1.1 24 null 0
```

(4) 检查配置结果

显示 Router A 的 BFD 信息。

```
<RouterA> display bfd session
Total Session Num: 1            Init Mode: Active
 Session Working Under Ctrl Mode:
 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 6/1           192.168.1.1     192.168.2.2     Up    1700ms   Eth1/2
```

显示 Router A 上学到的路由 100.1.1.0/24。

```
<RouterA> Routing Table : Public
Summary Count : 2
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 1
   Preference: 100                   Cost: 1
      NextHop: 192.168.2.2      Interface: GigabitEthernet1/2
    BkNextHop: 0.0.0.0        BkInterface:
  RelyNextHop: 192.168.1.2      Neighbor : 192.168.2.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv GotQ        Age: 00h04m02s
          Tag: 0
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
      NextHop: 192.168.3.2      Interface: GigabitGigabitEthernet1/1
    BkNextHop: 0.0.0.0        BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.3.2
    Tunnel ID: 0x0                  Label: NULL
        State: Inactive Adv           Age: 00h03m57s
          Tag: 0
```

打开 Router A 的调试开关。

```
<RouterA> debugging rip 1 event
<RouterA> terminal debugging
```

Router B 和 Router C 之间的链路发生故障后。可以看到 Router A 能够快速检测到链路的变化。

```
%Jan 19 10:41:51:203 2008 RouterA BFD/4/LOG:Sess[192.168.1.1/192.168.2.2, Eth1/2, Ctrl], Sta:
UP->DOWN, Diag: 1
*Jan 19 10:41:51:203 2008 RouterA RM/6/RMDEBUG: RIP-BFD: Message Type Disable, Connect Type
```

```
Indirect-connect, Pkt Type Control, Src IP Address 192.168.1.1, Src IFIndex 4, Nbr IP Address
192.168.2.2.
```

显示 Router A 的 BFD 信息。

查看 Router A 的 BFD 信息，Router A 已取消与 Router C 的邻居关系，没有任何输出信息。

```
<RouterA> display bfd session
```

显示 Router A 的 RIP 路由。

查看 Router A 的 RIP 路由信息，没有从 Router C 学到的 RIP 路由。

```
<RouterA> display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
```

显示 Router A 上学到的路由 100.1.1.0/24。

```
<RouterA> display ip routing-table 100.1.1.0 verbose
Routing Table : Public
Summary Count : 1
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
      NextHop: 192.168.3.2      Interface: GigabitEthernet1/2
    BkNextHop: 0.0.0.0        BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.3.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h10m35s
          Tag: 0
```

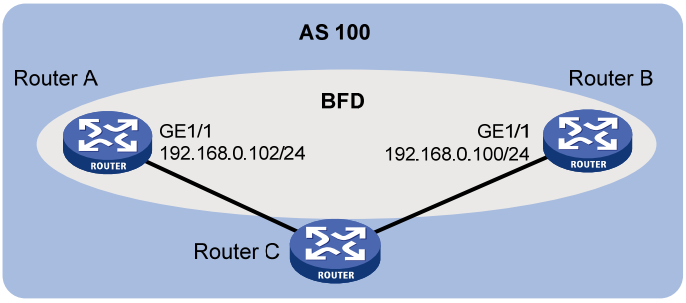
1.7.5 配置 BGP 与 BFD 联动

1. 组网需求

- 在 AS 100 内使用 OSPF 作为 IGP 协议；
- Router A 与 Router B 建立 IBGP 连接并使能 BFD；
- Router A 和 Router C 之间的链路发生故障后，BFD 能够快速检测并通告 BGP 协议。

2. 组网图

图1-8 BGP 与 BFD 联动配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 OSPF（略）

(3) 配置 BGP 连接

配置 Router A。

```
[RouterA] bgp 100
[RouterA-bgp] peer 192.168.0.100 as-number 100
[RouterA-bgp] peer 192.168.0.100 bfd
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bgp 100
[RouterB-bgp] peer 192.168.0.102 as-number 100
[RouterB-bgp] peer 192.168.0.102 bfd
```

(4) 配置各路由器接口及 BFD 参数

配置 Router A。

```
<RouterA> system-view
[RouterA] bfd session init-mode active
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.102 24
[RouterA-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 7
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bfd session init-mode active
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 192.168.0.100 24
[RouterB-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterB-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterB-GigabitEthernet1/1] bfd detect-multiplier 6
[RouterB-GigabitEthernet1/1] quit
```

(5) 检查配置结果

显示 Router A 的 BFD 邻居详细信息。

```
[RouterA] display bfd interface verbose

 Total Interface Num: 1

      Interface: GigabitEthernet1/1            Session Num: 1
Min Trans Inter: 300ms               Min Recv Inter: 300ms
     DetectMult: 6              Min Echo Recv Inter: 300ms
      Auth mode: None
```

```
 LD/RD          SourceAddr      DestAddr        ConnType  State     Mode
 7/7            192.168.0.100   192.168.0.102   Direct    Up        Ctrl
```

Router A 和 Router B 之间的链路发生故障后，使用 **display bfd interface verbose** 命令查看 GigabitEthernet1/1 接口下会话信息被删除。

```
<RouterA> display bfd interface verbose

 Total Interface Num: 1

      Interface: GigabitEthernet1/1            Session Num: 0
Min Trans Inter: 400ms               Min Recv Inter: 400ms
     DetectMult: 8              Min Echo Recv Inter: 400ms
      Auth mode: None
```

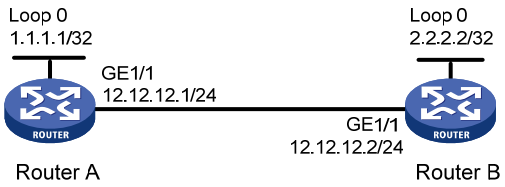
1.7.6 配置 MPLS RSVP-TE 与 BFD 联动

1. 组网需求

- Router A、Router B 物理直连，并且在双方直连口上使能 MPLS RSVP-TE BFD 应用，在设备上运行 OSPF，网络层相互可达。
- 当 Router A 和 Router B 之间的链路出现故障后，BFD 能够快速检测并通告 MPLS RSVP-TE 协议。

2. 组网图

图1-9 MPLS RSVP-TE 与 BFD 联动配置组网图



3. 配置步骤

(1) 配置 MPLS RSVP-TE 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] mpls lsr-id 1.1.1.1
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls rsvp-te
[RouterA-mpls] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mpls
[RouterA-GigabitEthernet1/1] mpls te
[RouterA-GigabitEthernet1/1] mpls rsvp-te
[RouterA-GigabitEthernet1/1] mpls rsvp-te bfd enable
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] mpls lsr-id 2.2.2.2
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls rsvp-te
[RouterB-mpls] quit
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] mpls
[RouterB-GigabitEthernet1/1] mpls te
[RouterB-GigabitEthernet1/1] mpls rsvp-te
[RouterB-GigabitEthernet1/1] mpls rsvp-te bfd enable
[RouterB-GigabitEthernet1/1] quit
```

(2) 配置 OSPF 基本功能

配置 Router A。

```
<RouterA> system-view
[RouterA] ospf
[Router-A-ospf-1] area 0
[Router-A-ospf-1-area-0.0.0.0] network 12.12.12.1 0.0.0.255
[Router-A-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[Router-A-ospf-1-area-0.0.0.0] quit
[Router-A-ospf-1] quit
```

配置 Router B。

```
<RouterB>system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 12.12.12.2 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
```

(3) 配置各路由器接口

配置 Router A。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 12.12.12.1 24
[RouterA-GigabitEthernet1/1] quit
```

配置 Router B。

```
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 12.12.12.2 24
```

配置 Router A 到 Router B 的一条触发 RSVP-TE 建立邻居隧道。

```
[RouterA] interface tunnel 1
[RouterA-Tunnel1] ip address 10.10.10.1 24
[RouterA-Tunnel1] tunnel-protocol mpls te
[RouterA-Tunnel1] destination 2.2.2.2
[RouterA-Tunnel1] mpls te signal-protocol rsvp-te
```

```
[RouterA-Tunnel1] mpls te commit
[RouterA-Tunnel1] quit
```

(4) 检查配置结果

显示 Router A 的 BFD 邻居详细信息。

```
<RouterA> display bfd session verbose

 Total Session Num: 1            Init Mode: Active

 Session Working Under Ctrl Mode:

     Local Discr: 19                 Remote Discr: 18
       Source IP: 12.12.12.1        Destination IP: 12.12.12.2
   Session State: Up                      Interface: GigabitEthernet1/1
 Min Trans Inter: 400ms             Act Trans Inter: 400ms
  Min Recv Inter: 400ms           Act Detect Inter: 2000ms
  Running Up for: 00:00:01               Auth mode: None
    Connect Type: Direct                 Board Num: 6
        Protocol: RSVP
       Diag Info: No Diagnostic
```

打开 RouterA 的调试开关。

```
<RouterA> debugging bfd scm
<RouterA> debugging bfd event
<RouterA> terminal debugging
```

断开 Router A 和 Router B 之间的链路。可以看到 Router A 能够快速检测 Router B 的变化。

```
%Dec 27 10:50:03:380 2006 RouterA RM/3/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
12.12.12.2(GigabitEthernet1/1) from Full to Down
%Dec 27 10:50:05:140 2006 RouterA BFD/4/LOG:Sess[12.12.12.1/12.12.12.2,Eth1/1,Ctrl], Sta:
UP->DOWN, Diag: 1
*Dec 27 10:50:05:140 2006 RouterA BFD/7/SCM:Sess[12.12.12.1/12.12.12.2,Eth1/1,Ctrl], Oper:
Reset
*Dec 27 10:50:05:140 2006 RouterA BFD/7/EVENT:Send sess-down Msg,
[Src:12.12.12.1,Dst:12.12.12.2,GigabitEthernet1/1,Ctrl],instance:0, Protocol: RSVP
```

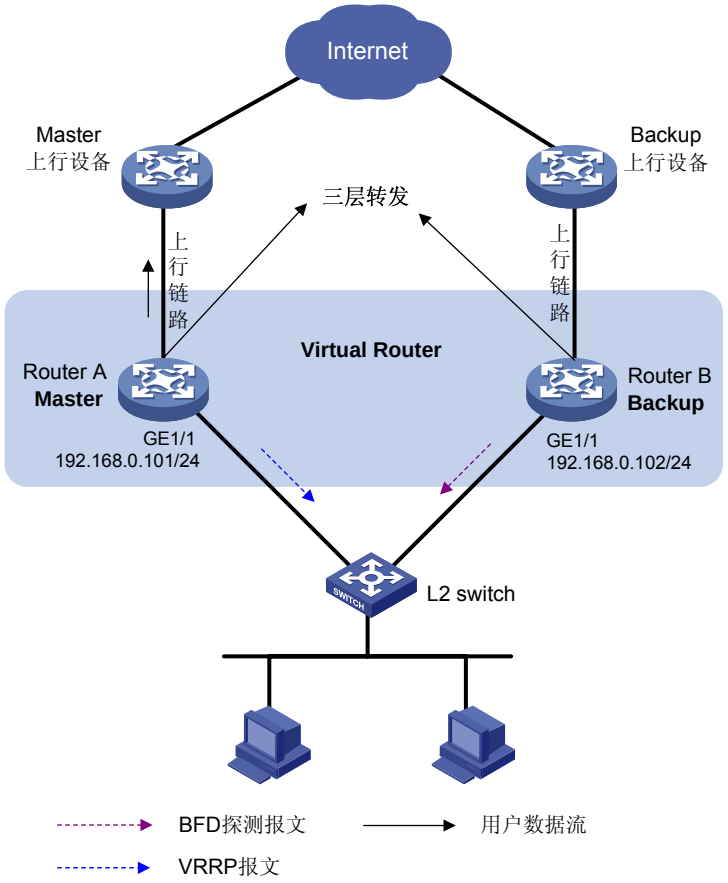
1.7.7 配置 VRRP 与 BFD 联动（Backup 监视 Master）

1. 组网需求

- Master 出现故障时，只能依赖于 Backup 设置的超时时间来判断是否应该抢占，切换时间一般在 3 秒～4 秒之间，无法达到秒级以下的切换速度。VRRP 必须依赖一种可靠的链路检测机制来探测 Master 的当前状态。Backup 上的探测协议快速检测 Master 的运行状态，当 Master 状态出现故障时，Backup 能够立即抢占成为 Master，切换时间在 100ms 以内。

2. 组网图

图1-10 BFD 与 VRRP 联动（Backup 监视 Master）配置组网图



3. 配置步骤

配置 Router A。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.101 24
[RouterA-GigabitEthernet1/1] vrrp vrid 1 virtual-ip 192.168.0.10
[RouterA-GigabitEthernet1/1] vrrp vrid 1 priority 110
```

配置 Router B。

```
<RouterB> system-view
[RouterB] bfd session init-mode active
[RouterB] bfd echo-source-ip 10.10.10.10
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 192.168.0.102 24
[RouterB-GigabitEthernet1/1] bfd min-echo-receive-interval 500
[RouterB-GigabitEthernet1/1] bfd detect-multiplier 3
[RouterB-GigabitEthernet1/1] quit
[RouterB] track 1 bfd echo interface GigabitEthernet 1/1 remote ip 192.168.0.101 local ip
192.168.0.102
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] vrrp vrid 1 virtual-ip 192.168.0.10
[RouterB-GigabitEthernet1/1] vrrp vrid 1 track 1 switchover
```

通过 display vrrp verbose 命令查看配置后的结果。

显示 Router A 上详细信息。

```
<RouterA> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
Total number of virtual routers: 1
 Interface         : GigabitEthernet1/1
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 110                  Run Pri        : 110
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.101
```

显示 Router B 上详细信息。

```
<RouterB> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : GigabitEthernet1/1
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Switchover
 Virtual IP        : 192.168.0.10
 Master IP         : 192.168.0.101
```

以上显示信息表示在 Router A 为 Master 路由器，Router B 为 Backup 路由器。当 Router A 出现故障时，通过 **display vrrp verbose** 命令查看备份组的信息。

显示 Router B 上的详细信息。

```
<RouterB> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
Total number of virtual routers: 1
 Interface         : GigabitEthernet1/1
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Switchover
 Virtual IP        : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.102
```

显示 Router B 上 Track 对象的详细信息。

```
<RouterB> display track 1
Track ID: 1
Status: Negative
Reference Object:
BFD Session:
Packet type: Echo
Interface  : GigabitEthernet1/1
Remote IP  : 192.168.0.101
Local IP   : 192.168.0.102
```

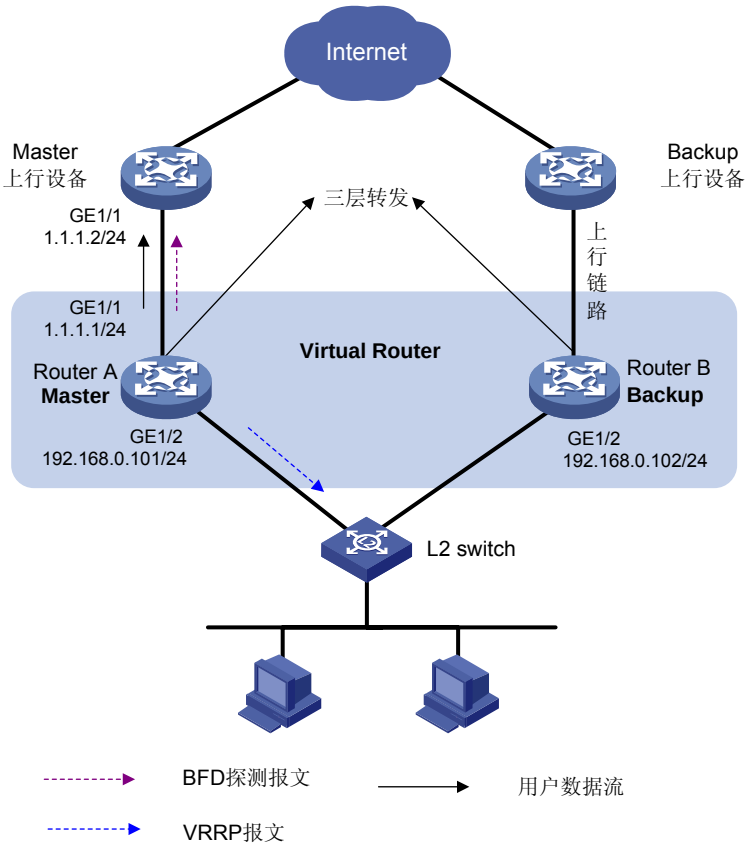
1.7.8 配置 VRRP 与 BFD 联动（监视上行链路）

1. 组网需求

Master 监视上行链路的状态，当上行链路 down 时，降低备份组的优先级并且立即以新的优先级发送 VRRP 报文；Backup 收到优先级比自己低的报文后，延迟很短的时间后就抢占成为 Master。

2. 组网图

图1-11 BFD 与 VRRP 联动（Master 监视上行链路）配置组网图



3. 配置步骤

配置 Router A，上行设备的 IP 地址为 1.1.1.2。

```
<RouterA> system-view
[RouterA] bfd session init-mode active
[RouterA] bfd echo-source-ip 10.10.10.10
[RouterA] interface GigabitEthernet 1/1
```

```
[RouterA-GigabitEthernet1/1] ip address 1.1.1.1 24
[RouterA-GigabitEthernet1/1] bfd min-echo-receive-interval 10
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 3
[RouterA-GigabitEthernet1/1] quit
[RouterA] track 1 bfd echo interface GigabitEthernet 1/1 remote ip 1.1.1.2 local ip 1.1.1.1
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] ip address 192.168.0.101 24
[RouterA-GigabitEthernet1/2] vrrp vrid 1 virtual-ip 192.168.0.10
[RouterA-GigabitEthernet1/2] vrrp vrid 1 priority 110
[RouterA-GigabitEthernet1/2] vrrp vrid 1 track 1 reduced 20
```

配置 Router B。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/2
[RouterB-GigabitEthernet1/2] ip address 192.168.0.102 24
[RouterB-GigabitEthernet1/2] vrrp vrid 1 virtual-ip 192.168.0.10
```

通过 display vrrp verbose 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
<RouterA> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
Total number of virtual routers: 1
 Interface         : GigabitEthernet1/2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 110                  Run Pri        : 110
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Pri Reduced    : 20
 Virtual IP        : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.101
```

显示 Router B 上的详细信息。

```
<RouterB> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : GigabitEthernet1/2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Master IP         : 192.168.0.101
```

以上显示信息表示在 Router A 为 Master 路由器，Router B 为 Backup 路由器。

当 Router A 监视的上行链路状态为 down 时，通过 **display vrrp verbose** 命令查看备份组的信息。

Router A 监视的上行链路状态为 down 时，显示 Router A 上备份组 1 的详细信息。

```
<RouterA> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : GigabitEthernet1/2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 110                  Run Pri        : 90
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Pri Reduced    : 20
 Virtual IP        : 192.168.0.10
 Master IP         : 192.168.0.102
```

显示 Router B 上的详细信息。

```
<RouterB> display vrrp verbose
 IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : GigabitEthernet1/2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.102
```

显示 Router A 上 Track 对象的详细信息。

```
<RouterA> display track 1
Track ID: 1
Status: Negative
Reference Object:
BFD Session:
Packet type: Echo
Interface  : GigabitEthernet1/1
Remote IP  : 1.1.1.2
Local IP   : 1.1.1.1
```

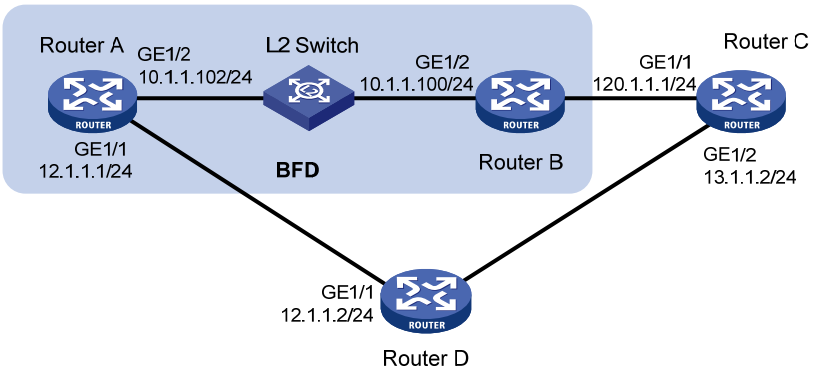
1.7.9 配置静态路由与 BFD 联动（单跳检测）

1. 组网需求

- 在 Router A 上配置静态路由可以到达 Router C，并使能 BFD 检测功能。
- 当 Router A 和 Router B 之间的链路出现故障时，Router A 选择经过 Router D 到达 Router C。

2. 组网图

图1-12 静态路由与 BFD 联动（单跳检测）配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址（略）

(2) 配置 BFD 和静态路由

在 Router A 上配置静态路由，并使能 BFD 检测功能，通过 BFD echo 报文方式实现 BFD 功能。

```
<RouterA> system-view
[RouterA] bfd echo-source-ip 123.1.1.1
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] bfd min-echo-receive-interval 500
[RouterA-GigabitEthernet1/2] bfd detect-multiplier 7
[RouterA-GigabitEthernet1/2] quit
[RouterA] ip route-static 120.1.1.1 24 GigabitEthernet1/2 10.1.1.100 bfd echo-packet
[RouterA] ip route-static 120.1.1.1 24 GigabitEthernet1/1 12.1.1.2 preference 65
[RouterA] quit
```

(3) 检查配置结果

显示 Router A 使能 BFD 信息。

```
<RouterA> display bfd Session

Total Session Num: 1            Init Mode: Active

 Session Working Under Echo Mode:

 LD            SourceAddr      DestAddr        State Holdtime Interface
 7             10.1.1.102       10.1.1.100      Up    1700ms   GigabitEthernet1/2
```

显示 Router A 路由表详细信息。

```
<RouterA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 2

Static Routing table Status : <Active>
Summary Count : 1
```

```
Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24        Static 60   0            10.1.1.100      GigabitEthernet1/2

Direct Routing table Status : <Inactive>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24         Static 65   0            12.1.1.2        GigabitEthernet1/1
```

在 Router A 上打开 BFD 功能调试信息开关。

```
<RouterA> debugging bfd event
<RouterA> debugging bfd scm
<RouterA> terminal debugging
```

Router B 和二层交换机之间链路发生故障时，可以看到 Router A 能够快速感知 Router B 的变化。

```
%Nov 12 19:28:28:592 2005 RouterA BFD/5/LOG:Sess[123.1.1.1/10.1.1.100, Eth1/2], Sta:
UP->DOWN, Diag: 1
*0.53892593 RouterA BFD/8/SCM:Sess[123.1.1.1/10.1.1.100, Eth1/2], Oper: Reset
*0.53892593 RouterA BFD/8/EVENT:Send sess-down Msg, [Src：123.1.1.1, Dst：10.1.1.100, Eth1/2]
Protocol: STATIC
*0.53892595                RouterA                RM/7/LOG:static                route
[Dest:120.1.1.1/24,Nexthop:10.1.1.100,ExitIf:Eth1/2] became invalid
```

当 Router A 到 Router B 链路发生故障时，通过 **display ip routing-table protocol static** 命令查看路表信息。Router A 选择经过 Router D 静态路由到达 Router C。

```
<RouterA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 2

Static Routing table Status : < Active>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24         Static 65   0            12.1.1.2        GigabitEthernet1/1

Static Routing table Status : < Inactive>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24        Static 60   0            10.1.1.100      GigabitEthernet1/2
```

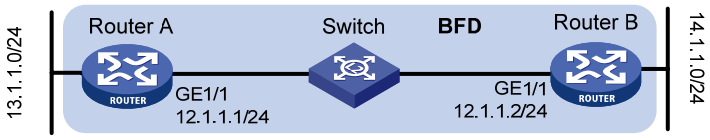
1.7.10 配置静态路由与 BFD 联动（双向检测）

1. 组网需求

- 在 Router A 上配置静态路由可以到达 14.1.1.0/24 网段路由，在 Router B 上配置静态路由可以到达 13.1.1.0/24 网段路由，并使能 BFD 检测功能；
- 当 Router A 和 Router B 链路出现故障时 BFD 能够快速感知。

2. 组网图

图1-13 静态路由与 BFD 联动（双向检测）配置组网图



3. 配置步骤

(1) 配置 BFD

配置 Router A。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 12.1.1.1 24
[RouterA-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterA-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterA-GigabitEthernet1/1] bfd detect-multiplier 9
[RouterA-GigabitEthernet1/1] quit
[RouterA] ip route-static 14.1.1.0 24 GigabitEthernet1/1 12.1.1.2 bfd control-packet
[RouterA] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 12.1.1.2 24
[RouterB-GigabitEthernet1/1] bfd min-transmit-interval 500
[RouterB-GigabitEthernet1/1] bfd min-receive-interval 500
[RouterB-GigabitEthernet1/1] bfd detect-multiplier 9
[RouterB-GigabitEthernet1/1] quit
[RouterB] ip route-static 13.1.1.0 24 12.1.1.1 bfd control-packet
[RouterB] quit
```

(2) 检查配置结果

```
<RouterA> display bfd session

 Total Session Num: 1            Init Mode: Active

 Session Working Under Ctrl Mode:

 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 4/7           12.1.1.1        12.1.1.2        Up    2000ms   GigabitEthernet1/1
```

在 Router A 上查看静态路由。

```
<RouterA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 1

Static Routing table Status : < Active>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

14.1.1.0/24         Static 60   0            12.1.1.2        GigabitEthernet 1/1
```

在 Router A 上打开 BFD 功能调试信息开关。

```
<RouterA> debugging bfd event
<RouterA> debugging bfd scm
<RouterA> terminal debugging
```

Router A 和交换机之间链路发生故障时，可以看到 Router A 能够快速感知 Router B 的变化。

```
%Jul     27     10:18:18:672     2007     RouterA     BFD/4/LOG:Sess[12.1.1.1/12.1.1.2,
GigabitEthernet1/1,Ctrl],
 Sta: UP->DOWN, Diag: 1
*Jul 27 10:18:18:672 2007 RouterA BFD/7/EVENT:Send sess-down Msg, [Src:12.1.1.1,
Dst:12.1.1.2,GigabitEthernet1/1,Ctrl], instance:0, protocol:STATIC
*Jul 27 10:18:19:172 2007 RouterA BFD/7/EVENT:Receive Delete-sess, [Src:12.1.1.1
,Dst:12.1.1.2,GigabitEthernet1/1,Ctrl], Direct, Instance:0x0, Proto:STATIC
*Jul 27 10:18:19:172 2007 RouterA BFD/7/EVENT:Notify driver to stop receiving bf
```

此时查看静态路由，路由处于 Inactive 状态。

```
<RouterA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 1

Static Routing table Status : < Inactive>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

14.1.1.0/24         Static 60   0            12.1.1.2        GigabitEthernet1/1
```

目 录

1 MPLS基本配置 1-1
1.1 MPLS简介 1-1
1.1.1 MPLS基本概念 1-1
1.1.2 MPLS体系结构 1-4
1.1.3 MPLS与路由协议 1-5
1.1.4 MPLS的应用 1-5
1.2 MPLS基本配置简介 1-7
1.2.1 标签的发布和管理 1-7
1.2.2 倒数第二跳弹出 1-8
1.2.3 MPLS MTU 1-8
1.2.4 MPLS对TTL的处理 1-9
1.2.5 检测MPLS LSP 1-9
1.3 LDP协议简介 1-10
1.3.1 LDP基本概念 1-10
1.3.2 LDP标签分发 1-11
1.3.3 LDP基本操作 1-12
1.3.4 LDP环路检测 1-13
1.3.5 LDP 标签过滤 1-14
1.3.6 LDP GR 1-15
1.4 配置MPLS基本能力 1-15
1.4.1 配置准备 1-15
1.4.2 配置MPLS基本能力 1-16
1.5 配置PHP特性 1-16
1.5.1 配置准备 1-16
1.5.2 配置PHP特性 1-16
1.6 配置MPLS MTU 1-17
1.6.1 配置准备 1-17
1.6.2 配置MPLS MTU 1-17
1.7 配置静态LSP 1-17
1.7.1 配置准备 1-17
1.7.2 配置静态LSP 1-18
1.8 配置MPLS LDP 1-18
1.8.1 配置准备 1-18
1.8.2 MPLS LDP配置任务简介 1-18
1.8.3 配置MPLS LDP能力 1-19
1.8.4 配置LDP本地会话的参数 1-19
1.8.5 配置LDP远端会话的参数 1-20
1.8.6 配置LSP触发策略 1-20
1.8.7 配置LDP标签分配控制方式 1-21
1.8.8 配置LDP环路检测 1-22
1.8.9 配置LDP MD5 认证 1-22
1.8.10 配置LDP标签过滤 1-22

1.9 配置LDP实例 1-23
1.9.1 配置准备 1-23
1.9.2 配置LDP实例 1-23
1.10 配置LDP GR 1-24
1.11 配置MPLS的TTL处理 1-25
1.11.1 配置准备 1-25
1.11.2 配置MPLS的IP TTL复制功能 1-25
1.11.3 配置ICMP响应报文使用的路径 1-25
1.12 配置MPLS统计功能 1-26
1.12.1 使能MPLS统计功能 1-26
1.13 检测MPLS LSP 1-26
1.14 开启MPLS的Trap功能 1-27
1.15 MPLS显示和维护 1-27
1.15.1 重启LDP会话 1-27
1.15.2 显示MPLS运行状态 1-27
1.15.3 显示MPLS LDP运行状态 1-28
1.15.4 清除MPLS相关信息 1-29
1.16 MPLS配置举例 1-29
1.16.1 配置LDP会话示例 1-29
1.16.2 使用LDP建立LSP示例 1-33

1 MPLS 基本配置

说明

- 有关 VPN 的详细介绍请参见“MPLS 分册”中的“MPLS L2VPN 配置”和“MPLS L3VPN 配置”；
- 有关 MPLS TE 的详细介绍请参见“MPLS 分册”中的“MPLS TE 配置”；
- 有关 QoS 的详细介绍和说明请参见“QoS 分册”。

1.1 MPLS 简介

MPLS（Multiprotocol Label Switching，多协议标签交换）起源于 IPv4（Internet Protocol version 4，因特网协议版本 4），最初是为了提高转发速度而提出的，其核心技术可扩展到多种网络协议，包括 IPv6（Internet Protocol version 6，因特网协议版本 6）、IPX（Internet Packet Exchange，网际报文交换）和 CLNP（Connectionless Network Protocol，无连接网络协议）等。MPLS 中的“M”指的就是支持多种网络协议。

MPLS 技术集二层的快速交换和三层的路由转发于一体，可以满足各种新应用对网络的要求。

说明

MPLS 结构的详细介绍可参考 RFC 3031（Multiprotocol Label Switching Architecture）。

1.1.1 MPLS 基本概念

1. 转发等价类

MPLS 作为一种分类转发技术，将具有相同转发处理方式的分组归为一类，称为 FEC（FEC，Forwarding Equivalence Class，转发等价类）。相同 FEC 的分组在 MPLS 网络中将获得完全相同的处理。

FEC 的划分方式非常灵活，可以是以源地址、目的地址、源端口、目的端口、协议类型或 VPN 等为划分依据的任意组合。例如，在传统的采用最长匹配算法的 IP 转发中，到同一个目的地址的所有报文就是一个 FEC。

2. 标签

标签是一个长度固定、只具有本地意义的短标识符，用于唯一标识一个分组所属的 FEC。在某些情况下，例如要进行负载分担，对应一个 FEC 可能会有多个标签，但是一个标签只能代表一个 FEC。

标签由报文的头部所携带，不包含拓扑信息，只具有局部意义。标签的长度为 4 个字节（32bits），封装结构如 图 1-1所示。

图1-1 标签的封装结构



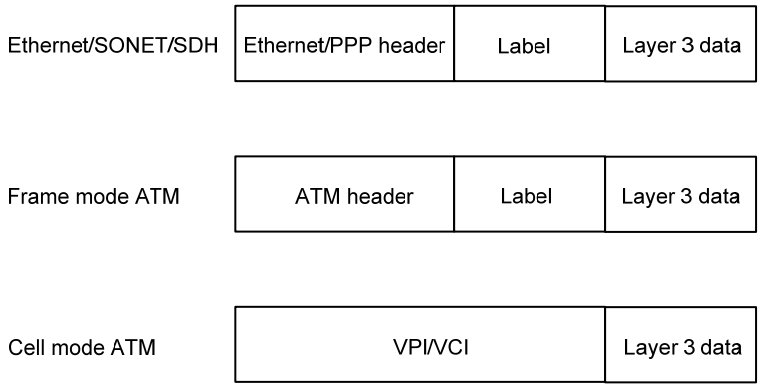
标签共有 4 个域：

- Label：标签值字段，长度为 20bits，用于转发的指针；
- Exp：3bits，用于 QoS；
- S：1bit，用于标识该标签是否是栈底标签，值为 1 时表明为最底层标签。主要应用于 MPLS 标签的多重嵌套；
- TTL：8bits，和 IP 分组中的 TTL（Time To Live，生存时间）意义相同。

标签与 ATM 的 VPI/VCI 以及 Frame Relay 的 DLCI 类似，是一种连接标识符。如果链路层协议具有标签域，如 ATM 的 VPI/VCI 或 Frame Relay 的 DLCI，则标签封装在这些域中；如果不支持，则标签封装在链路层和 IP 层之间的一个垫层中。这样，标签能够被任意的链路层所支持。

标签在分组中的封装位置如 图 1-2所示：

图1-2 标签在分组中的封装位置



Label：标签，Frame mode：帧模式，Cell mode：信元模式



说明

有关信元模式的实现，目前暂不支持。

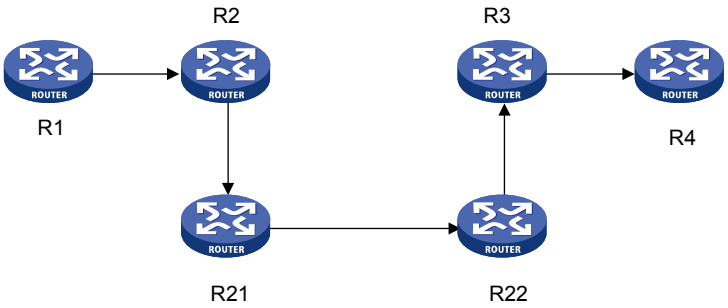
3. 标签交换路由器

LSR（Label Switching Router，标签交换路由器）是 MPLS 网络中的基本元素，所有 LSR 都支持 MPLS 技术。

4. 标签交换路径

一个转发等价类在MPLS网络中经过的路径称为LSP（Label Switched Path，标签交换路径）。在一条LSP上，沿数据传送的方向，相邻的LSR分别称为上游LSR和下游LSR。如 图 1-3中，R2 为R1 的下游LSR，相应的，R1 为R2 的上游LSR。

图1-3 标签交换路径 LSP



LSP 在功能上与 ATM 和帧中继（Frame Relay）的虚电路相同，是从 MPLS 网络的入口到出口的一个单向路径。LSP 中的每个节点由 LSR 组成。

5. 标签分发协议

LDP（Label Distribution Protocol，标签分发协议）是 MPLS 的控制协议，它相当于传统网络中的信令协议，负责 FEC 的分类、标签的分配以及 LSP 的建立和维护等一系列操作。

MPLS 可以使用多种标签发布协议，包括专为标签发布而制定的协议，例如：LDP、CR-LDP（Constraint-Based Routing using LDP，基于约束路由的 LDP）；也包括现有协议扩展后支持标签发布的，例如：BGP（Border Gateway Protocol，边界网关协议）、RSVP（Resource Reservation Protocol，资源预留协议）。同时，还可以手工配置静态 LSP。

说明

- 有关 CR-LDP 和 RSVP 的介绍请参见“MPLS 分册”中的“MPLS TE 配置”；
- 有关 BGP 的介绍请参见“IP 路由分册”的“BGP 配置”。

---

6. LSP 隧道技术

MPLS 支持 LSP 隧道技术。

一条LSP的上游LSR和下游LSR，尽管它们之间的路径可能并不在路由协议所提供的路径上，但是MPLS允许在它们之间建立一条新的LSP，这样，上游LSR和下游LSR分别就是这条LSP的起点和终点。这时，上游LSR和下游LSR间的LSP就是LSP隧道，它避免了采用传统的网络层封装隧道。如 图1-3中LSP<R2→R21→R22→R3>就是R2、R3 间的一条隧道。

如果隧道经由的路由与逐跳从路由协议中取得的路由一致，这种隧道就称为逐跳路由隧道（Hop-by-Hop Routed Tunnel）；否则称为显式路由隧道（Explicitly Routed Tunnel）。

7. 多层标签栈

如果分组在超过一层的 LSP 隧道中传送，就会有多层标签，形成标签栈（Label Stack）。在每一隧道的入口和出口处，进行标签的入栈（PUSH）和出栈（POP）操作。

标签栈按照“后进先出”（Last-In-First-Out）方式组织标签，MPLS 从栈顶开始处理标签。

MPLS 对标签栈的深度没有限制。若一个分组的标签栈深度为 m，则位于栈底的标签为 1 级标签，位于栈顶的标签为 m 级标签。未压入标签的分组可看作标签栈为空（即标签栈深度为零）的分组。

1.1.2 MPLS 体系结构

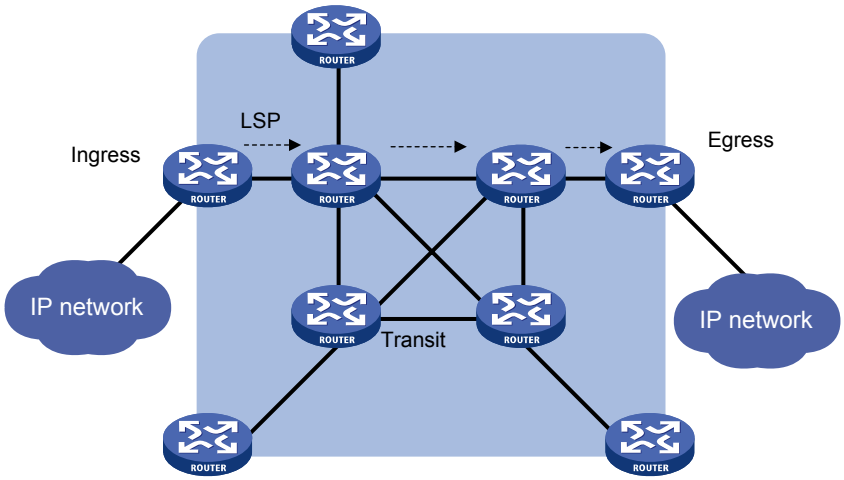
1. MPLS 网络结构

如 图 1-4所示，MPLS网络的基本构成单元是LSR，由LSR构成的网络称为MPLS域。

位于 MPLS 域边缘、连接其它用户网络的 LSR 称为 LER（Label Edge Router，边缘 LSR），区域内部的 LSR 称为核心 LSR。核心 LSR 可以是支持 MPLS 的路由器，也可以是由 ATM 交换机等升级而成的 ATM-LSR。域内部的 LSR 之间使用 MPLS 通信，MPLS 域的边缘由 LER 与传统 IP 技术进行适配。

分组在入口 LER 被压入标签后，沿着由一系列 LSR 构成的 LSP 传送，其中，入口 LER 被称为 Ingress，出口 LER 被称为 Egress，中间的节点则称为 Transit。

图1-4 MPLS 网络结构



结合上图简要介绍 MPLS 的基本工作过程：

(1) 首先，LDP 和传统路由协议（如 OSPF、ISIS 等）一起，在各个 LSR 中为有业务需求的 FEC 建立路由表和 LIB（Label Information Base，标签信息表）；

(2) 入口 LER 接收分组，完成第三层功能，判定分组所属的 FEC，并给分组加上标签，形成 MPLS 标签分组；

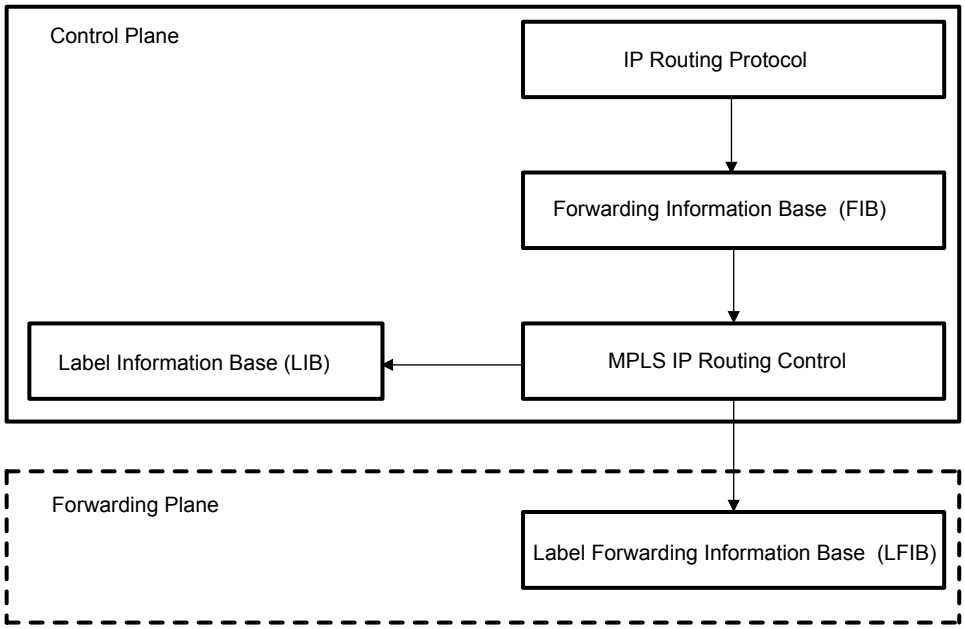
(3) 接下来，在 LSR 构成的网络中，LSR 根据分组上的标签以及 LFIB（Label Forwarding Information Base，标签转发表）进行转发，不对标签分组进行任何第三层处理；

(4) 最后，在 MPLS 出口 LER 去掉分组中的标签，继续进行后面的 IP 转发。

由此可以看出，MPLS 并不是一种业务或者应用，它实际上是一种隧道技术，也是一种将标签交换转发和网络层路由技术集于一身的路由与交换技术平台。这个平台不仅支持多种高层协议与业务，而且，在一定程度上可以保证信息传输的安全性。

2. MPLS 节点结构

图1-5 MPLS 节点结构示意图



如 图 1-5所示，MPLS节点由两部分组成：

- 控制平面（Control Plane）：负责标签的分配、路由的选择、标签转发表的建立、标签交换路径的建立、拆除等工作；
- 转发平面（Forwarding Plane）：依据标签转发表对收到的分组进行转发。

对于普通的 LSR，在转发平面只需要进行标签分组的转发，需要使用到 LFIB（Label Forwarding Information Base，标签转发表）。对于 LER，在转发平面不仅需要进行标签分组的转发，也需要进行 IP 分组的转发，所以既会使用到 LFIB，也会使用到 FIB（Forwarding Information Base，转发信息表）。

1.1.3 MPLS 与路由协议

LDP 通过逐跳方式建立 LSP 时，利用沿途各 LSR 路由转发表中的信息来确定下一跳，而路由转发表中的信息一般是通过 IGP、BGP 等路由协议收集的。LDP 并不直接和各种路由协议关联，只是间接使用路由信息。另一方面，通过对 BGP、RSVP 等已有协议进行扩展，也可以支持标签的分发。

在 MPLS 的应用中，也可能需要对某些路由协议进行扩展。例如，基于 MPLS 的 VPN 应用需要对 BGP 进行扩展，使 BGP 能够传播 VPN（Virtual Private Network，虚拟专用网）的路由信息；基于 MPLS 的 TE（Traffic Engineering，流量工程）需要对 OSPF 或 IS-IS 协议进行扩展，以携带链路状态信息。

1.1.4 MPLS 的应用

最初，MPLS 技术结合了二层交换技术和三层路由技术，提高了路由查找速度。但是，随着 ASIC（Application-Specific Integrated Circuit，专用集成电路）技术的发展，路由查找速度已经不成为阻碍网络发展的瓶颈。这使得 MPLS 在提高转发速度方面不具备明显的优势。

但由于 MPLS 结合了 IP 网络强大的三层路由功能和传统二层网络高效的转发机制，在转发平面采用面向连接方式，与现有二层网络转发方式非常相似，这些特点使得 MPLS 能够很容易地实现 IP

与 ATM、帧中继等二层网络的无缝融合，并为 QoS（Quality of Service，服务质量）、TE、VPN 等应用提供更好的解决方案。

1. 基于 MPLS 的 VPN

传统的 VPN 一般是通过 GRE、L2TP、PPTP 等隧道协议来实现私有网络间数据流在公网上的传送，LSP 本身就是公网上的隧道，因此，用 MPLS 来实现 VPN 有天然的优势。

基于 MPLS 的 VPN 就是通过 LSP 将私有网络的不同分支连接起来，形成一个统一的网络。基于 MPLS 的 VPN 还支持对不同 VPN 间的互通控制。

图1-6 基于 MPLS 的 VPN

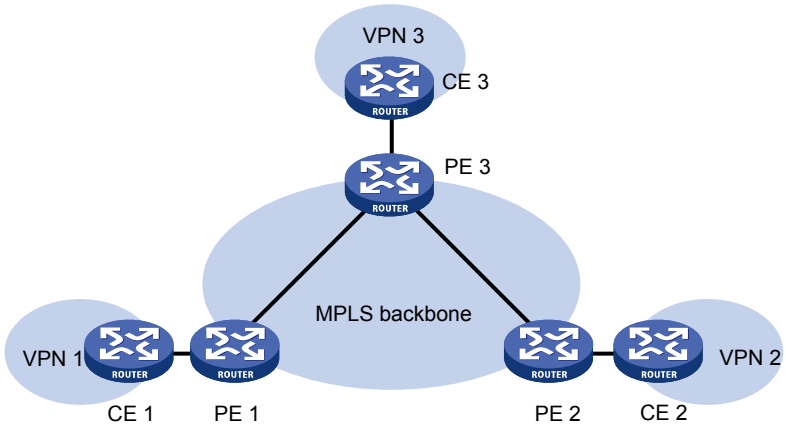


图 1-6是基于MPLS的VPN的基本结构：CE（Customer Edge，用户边缘设备）可以是路由器，也可以是交换机或主机；PE（Provider Edge，服务商边缘路由器）位于骨干网络。

PE 负责对 VPN 用户进行管理、建立各 PE 间 LSP 连接、同一 VPN 用户各分支间路由分派。PE 间的路由分派通常是用 LDP 或扩展的 BGP 协议实现。

基于 MPLS 的 VPN 支持不同分支间 IP 地址复用，并支持不同 VPN 间互通。与传统的路由相比，VPN 路由中需要增加分支和 VPN 的标识信息，这就需要对 BGP 协议进行扩展，以携带 VPN 路由信息。

2. 基于 MPLS 的流量工程

基于 MPLS 的 TE 和差分服务 Diff-serv 特性，在保证网络高利用率的同时，可以根据不同数据流的优先级实现差别服务，从而为语音、视频等数据流提供有带宽保证的低延时、低丢包率的服务。

由于全网实施流量工程的难度比较大，因此，在实际的组网方案中往往通过差分服务模型来实施 QoS。

Diff-Serv 的基本机制是在网络边缘，根据业务的服务质量要求将该业务映射到一定的业务类别中，利用 IP 分组中的 DS 字段（由 ToS 域而来）唯一的标记该类业务，然后，骨干网络中的各节点根据该字段对各种业务采取预先设定的服务策略，保证相应的服务质量。

Diff-Serv 的这种对服务质量的分类和 MPLS 的标签分配机制十分相似，事实上，基于 MPLS 的 Diff-Serv 就是通过将 DS 的分配与 MPLS 的标签分配过程结合来实现的。

1.2 MPLS 基本配置简介

1.2.1 标签的发布和管理

在 MPLS 体系中，由下游 LSR 决定将标签分配给特定 FEC，再通知上游 LSR。即，标签由下游指定，标签的分配按从下游到上游的方向。

1. 标签发布方式（Label Advertisement Mode）

标签发布方式分为两种：

- 下游按需方式 DoD（Downstream On Demand）：对于一个特定的 FEC，LSR 从上游获得标签请求消息之后才进行标签分配与分发；
- 下游自主方式 DU（Downstream Unsolicited）：对于一个特定的 FEC，LSR 无须从上游获得标签请求消息即进行标签分配与分发。

具有标签分发邻接关系的上游LSR和下游LSR之间必须使用相同的标签发布方式，否则LSP无法正常建立。详细的介绍请参见 1.3.2 LDP标签分发中的描述。

注意

目前，设备只支持 DU 标签发布方式。

---

2. 标签分配控制方式（Label Distribution Control Mode）

标签分配控制方式分为两种：

- 独立标签分配控制（Independent）：LSR 可以在任意时间向与它连接的 LSR 通告标签映射。这种方式可能导致在收到下游标签之前就向上游发布了标签。
- 有序标签控制方式（Ordered）：对于 LSR 上某个 FEC 的标签映射，只有当该 LSR 已经具有此 FEC 下一跳的标签映射消息或者该 LSR 就是此 FEC 的出口节点时，该 LSR 才可以向上游发送此 FEC 的标签映射。

3. 标签保持方式（Label Retention Mode）

标签保持方式是指 LSR 对收到的、但目前暂时用不到的标签—FEC 绑定的处理方式。

标签保持方式也分为两种：

- 自由标签保持方式（Liberal）：对于从邻居 LSR 收到的标签映射，无论邻居 LSR 是不是自己的下一跳都保留。
- 保守标签保持方式（Conservative）：对于从邻居 LSR 收到的标签映射，只有当邻居 LSR 是自己的下一跳时才保留。

使用自由标签保持方式，LSR 能够迅速适应路由变化；而使用保守标签保持方式，LSR 可以分配和保存较少的标签数量。

保守标签保持方式通常与 DoD 方式一起，用于对于标签空间有限的 LSR。

注意

目前，设备只支持 Liberal 标签保持方式。

4. 标签交换中的几个基本概念

- NHLFE（Next Hop Label Forwarding Entry）：下一跳标签转发项。用于描述对标签执行的操作，包括入栈（Push）操作和交换（Swap）操作。
- FTN（FEC to NHLFE map）：在 Ingress 节点将转发等价类 FEC 映射到 NHLFE 的过程。
- ILM（Incoming Label Map）：入标签映射。对于接收的标签分组，LSR 将标签映射到 NHLFE 的过程，包括空（Null）操作、出栈（Pop）操作。

5. 标签交换的过程

入口 LER（Ingress）将进入网络的分组划分成转发等价类 FEC。属于相同 FEC 的分组在 MPLS 域中将经过相同的路径，即使用同一条 LSP。LSR 对到来的 FEC 分组分配一个标签，然后从相应的接口转发出去。

标签交换的具体操作描述如下：

(1) LSP 沿途的所有 LSR 都先建立 ILM，入标签是根据 ILM 表项进行映射的；

(2) 对于接收到的标签分组，LSR 将标签映射到 NHLFE；

(3) LSR 只需根据标签从标签信息表中找到相应的 NHLFE，用新的标签替换原来的标签，然后对标签分组继续转发。

1.2.2 倒数第二跳弹出

在介绍 MPLS 的基本工作过程时提到：在 MPLS 网络中，核心 LSR 根据分组上的标签进行转发；在 Egress 节点（出口 LER）去掉分组中的标签，继续进行 IP 转发。

实际上，在比较简单的 MPLS 应用中，Egress 节点只需要进行 IP 转发，标签已经没有使用价值。这种情况下，为了减轻 Egress 节点的负担，提高 MPLS 网络对报文的处理能力，可以利用 PHP（Penultimate Hop Popping，倒数第二跳弹出）特性，在倒数第二个节点处将标签弹出，Egress 节点就不再进行标签操作了。

1.2.3 MPLS MTU

MPLS 标签栈象“垫层”一样，位于二层数据帧头和数据之间。在 MPLS 转发过程中，虽然网络层报文长度小于接口的 MTU，但是增加 MPLS 标签后，报文长度可能超过链路层允许发送的范围，从而导致报文无法正常转发。为此，设备上定义了 MPLS MTU，MPLS 转发时将增加标签后的报文长度与 MPLS MTU 比较。报文长度大于 MPLS MTU 时，如果允许分片，则将报文分片后再进行转发；如果不允许分片，则直接丢弃。

说明

如果 MPLS 报文内封装的是 L2VPN 报文，则即使报文长度大于 MPLS MTU，也会发送该报文，由接口的实际发送能力决定报文能否发送成功。

1.2.4 MPLS 对 TTL 的处理

MPLS 的 TTL 处理包括 IP TTL 复制和 ICMP 响应报文返回路径这两个方面。

1. IP TTL 复制

MPLS 标签中包含一个 8 位的 TTL 域，其含义与 IP 头中的 TTL 域相同。

根据 RFC 3031 中的描述，LSR 节点在对分组压入标签时，需要将原 IP 分组或下层标签中的 TTL 值拷贝到新增加的标签的 TTL 域。LSR 在转发标签分组时，对栈顶标签的 TTL 值做减一操作。标签出栈时，再将栈顶的 TTL 值拷贝回 IP 分组或下层标签。

TTL 除了用于防止产生路由环路外，也用于实现 tracert 功能：

- 如果在 Ingress 节点使能 IP TTL 复制功能，报文在 LSP 中经过的每一跳都体现为 IP TTL 逐跳递减，tracert 的结果将反映报文实际经过的路径；
- 如果不在 Ingress 节点使能 IP TTL 复制功能，则报文在 LSP 中经过的跳数不会导致 IP TTL 递减，tracert 的结果不包括 MPLS 骨干网络中每一跳，就好像 Ingress 路由器与 Egress 路由器是直连的。

注意

- 在 MPLS 域内部，MPLS 报文多层标签之间的 TTL 值总是互相复制。
- MPLS 的 IP TTL 复制功能对本地发送报文没有影响，本地发送报文都将进行 IP TTL 复制，从而保证本地管理员能够使用 tracert 检测网络。

在 MPLS VPN 应用中，出于网络安全的考虑，需要隐藏 MPLS 骨干网络的结构，这种情况下，对于私网报文，Ingress 节点上不能使用 TTL 的复制功能。

2. ICMP 响应报文

在 MPLS VPN 网络中，P 路由器无法对 MPLS 承载的 VPN 报文进行路由。当 MPLS 报文的 TTL 超时时，ICMP 响应报文将按照 LSP 继续传送，到达 LSP 终点路由器后，再根据 IP 路由转发 ICMP 响应报文。这种处理方式增加了网络流量和报文转发时延。

说明

有关 P 路由器的介绍和配置请参见“MPLS 分册”中的“MPLS L3VPN 配置”和“MPLS L2VPN 配置”。

对于仅有一层标签的 MPLS 报文，当 TTL 超时时，直接使用 IP 路由返回 ICMP 响应报文。

1.2.5 检测 MPLS LSP

在 MPLS 中，如果 LSP 转发数据失败，负责建立 LSP 的 MPLS 控制平面将无法检测到这种错误，这会给网络维护带来困难。

MPLS LSP Ping/Traceroute 为用户提供了发现 LSP 错误、并及时定位失效节点的机制。类似于普通 IP 的 Ping/Traceroute，MPLS LSP Ping/Traceroute 使用 MPLS Echo Request 报文和 MPLS Echo Reply 报文检测 LSP 的可用性。MPLS Echo Request 中携带需要检测的 FEC 信息，和其他属于此 FEC 的报文一样沿 LSP 发送，从而实现对 LSP 的检测。

- MPLS LSP Ping 是用于对 LSP 的有效性、可达性进行检测的工具。采取方法是通过发送一个叫做 MPLS Echo Request 的报文，通过 LSP 的数据转发，到达出口后，在 MPLS 域的 Egress，由 Egress 节点的控制平面确认本 LSR 是否为该 FEC 的出口，返回一个叫做 MPLS Echo Reply 的报文，如果发送方收到该报文，则说明这条 LSP 可以正确用于数据转发。
- MPLS LSP Traceroute 是对 LSP 的错误进行定位的工具。采取方法是 Echo Request 数据包被发送到每一个中间 LSR 的控制平面，以确定本 LSR 是否是此路径的中间节点。


说明

为了防止消息到达 Egress 节点后又被转发给其他节点，Echo Request 消息的 IP 头中目的地址需要设置为 127.0.0.0/8 网段的任意地址（本机环回地址），IP 头中的 TTL 值设置为 1。

1.3 LDP 协议简介

1.3.1 LDP 基本概念

LDP 协议规定标签分发过程中的各种消息以及相关的处理进程。

通过 LDP，LSR 可以把网络层的路由信息直接映射到数据链路层的交换路径上，进而建立起 LSP。LSP 既可以建立在两个相邻的 LSR 之间，也可以建立在两个非直连的 LSR 之间，从而在网络中所有中间节点上都使用标签交换。


说明

关于 LDP 的详细介绍可以参考 RFC 3036（LDP Specification）。

1. LDP 对等体

LDP 对等体是指相互之间存在 LDP 会话、使用 LDP 来交换标签/FEC 映射关系的两个 LSR。LDP 对等体通过它们之间的 LDP 会话获得对方的标签映射消息。

2. LDP 会话

LDP 会话用于在 LSR 之间交换标签映射、释放等消息。LDP 会话可以分为两种类型：

- 本地 LDP 会话（Local LDP Session）：建立会话的两个 LSR 之间是直连的；
- 远端 LDP 会话（Remote LDP Session）：建立会话的两个 LSR 之间是非直连的。

3. LDP 消息类型

LDP 协议主要使用四类消息：

- 发现（Discovery）消息：用于通告和维护网络中 LSR 的存在；
- 会话（Session）消息：用于建立、维护和终止 LDP 对等体之间的会话；
- 通告（Advertisement）消息：用于创建、改变和删除标签—FEC 绑定；
- 通知（Notification）消息：用于提供建议性的消息和差错通知。

为保证 LDP 消息的可靠发送，除了发现阶段使用 UDP 传输外，LDP 的 Session 消息、Advertisement 消息和 Notification 消息都使用 TCP 传输。

4. 标签空间与 LDP 标识符

LDP 对等体之间分配标签的范围称为标签空间（Label space）。可以为 LSR 的每个接口指定一个标签空间（per-interface label space），也可以整个 LSR 使用一个标签空间（per-platform label space）。

LDP 标识符（LDP Identifier）用于标识特定 LSR 的标签空间，是一个六字节的数值，格式如下：

<LSR ID>：<标签空间序号>

其中，LSR ID 占四字节，标签空间序号占两字节。标签空间序号取值为 1 时表示每个接口指定一个标签空间；取值为 0 时表示整个 LSR 使用一个标签空间。

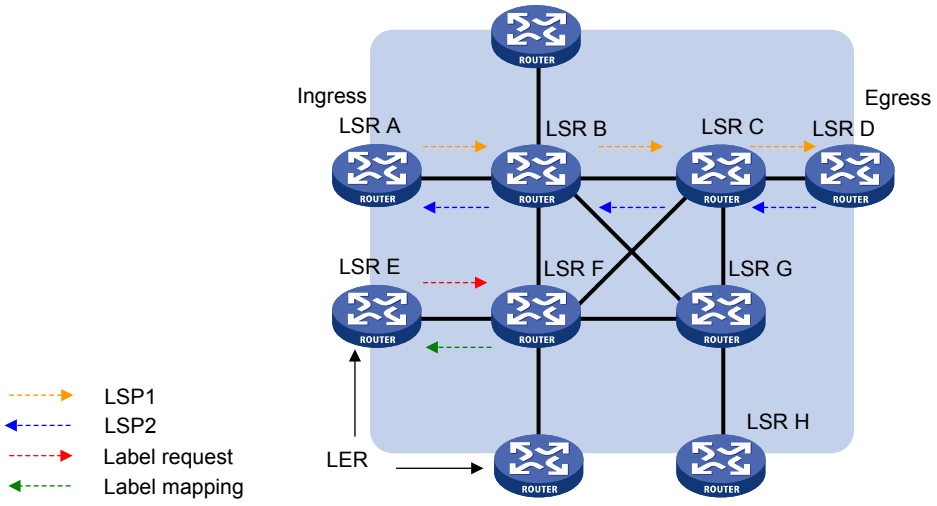
说明

目前设备上只实现了整个 LSR 使用一个标签空间。

1.3.2 LDP 标签分发

图 1-7为LDP标签分发示意图。

图1-7 标签分发过程



例如，图 1-7中的LSP1 上，LSR B为LSR C的上游LSR。

本章前面提到，标签的分发过程有两种模式，主要区别在于标签映射的发布是上游请求（DoD）还是下游主动发布（DU）。

下面分别详细描述这两种模式的标签分发过程：

1. DoD（downstream-on-demand）模式

上游 LSR 向下游 LSR 发送标签请求消息（Label Request Message），其中包含 FEC 的描述信息。下游 LSR 为此 FEC 分配标签，并将绑定的标签通过标签映射消息（Label Mapping Message）反馈给上游 LSR。

下游 LSR 何时反馈标签映射消息，取决于该 LSR 采用的标签分配控制方式。

- 采用 Ordered 方式时，只有收到它的下游返回的标签映射消息后，才向其上游发送标签映射消息；

- 采用 Independent 方式时，不管有没有收到它的下游返回的标签映射消息，都立即向其上游发送标签映射消息。

上游LSR一般是根据其路由表中的信息来选择下游LSR。在 图 1-7中，LSP1 沿途的LSR都采用Ordered方式，LSP2 上的LSR F则采用Independent方式。

2. DU（downstream unsolicited）模式

下游 LSR 在 LDP 会话建立成功后，主动向其上游 LSR 发布标签映射消息。上游 LSR 保存标签映射信息，并根据路由表信息来处理收到的标签映射信息。

1.3.3 LDP 基本操作

按照先后顺序，LDP 的操作主要包括以下四个阶段：

- 发现阶段
- 会话建立与维护
- LSP 建立与维护
- 会话撤销

1. 发现阶段

在这一阶段，希望建立会话的 LSR 向相邻 LSR 周期性地发送 Hello 消息，通知相邻节点自己的存在。通过这一过程，LSR 可以自动发现它的 LDP 对等体，而无需进行手工配置。

LDP 有两种发现机制：

- 基本发现机制

基本发现机制用于发现本地的 LDP 对等体，即通过链路层直接相连的 LSR，建立本地 LDP 会话。

这种方式下，LSR 周期性以 UDP 报文形式从接口发送 LDP 链路 Hello 消息（LDP Link Hello），发往标识“子网内所有路由器”的组播地址。

LDP 链路 Hello 消息带有接口的 LDP 标识符及其他相关信息，如果 LSR 在某个接口收到了 LDP 链路 Hello 消息，则表明在该接口（链路层）存在 LDP 对等体。

- 扩展发现机制

扩展发现机制用于发现远端的 LDP 对等体，即不通过链路层直接相连的 LSR，建立远端 LDP 会话。

这种方式下，LSR 周期性以 UDP 报文形式向指定的 IP 地址发送 LDP 目标 Hello 消息（LDP Targeted Hello）。

LDP 目标 Hello 消息带有 LSR 的 LDP 标识符及其他相关信息，如果 LSR 收到 LDP 目标 Hello 消息，则表明在网络层存在 LDP 对等体。

2. 会话建立与维护

发现邻居之后，LSR 开始建立会话。这一过程又可分为两步：

(1) 建立传输层连接，即，在 LSR 之间建立 TCP 连接；

(2) 随后对 LSR 之间的会话进行初始化，协商会话中涉及的各种参数，如 LDP 版本、标签分发方式、定时器值、标签空间等。

会话建立后，通过不断地发送 Hello 消息和 Keepalive 消息来维护这个会话。

3. LSP 建立与维护

LSP 的建立过程实际就是将 FEC 和标签进行绑定，并将这种绑定通告 LSP 上相邻 LSR。这个过程是通过 LDP 实现的，以 DU 模式、有序标签控制方式为例，主要步骤如下：

(1) 当网络的路由改变时，如果有一个边缘节点发现自己的路由表中出现了新的目的地址，并且这一地址不属于任何现有的 FEC，则该边缘节点需要为这一目的地址建立一个新的 FEC。边缘 LSR 决定该 FEC 将要使用的路由，向其下游 LSR 发起标签请求消息，并指明是要为哪个 FEC 分配标签；

(2) 如果此边缘节点存在上游 LSR，并且尚有可供分配的标签，则该节点为 FEC 分配标签，并向上游发出标签映射消息，其中包含分配的标签等信息。

(3) 收到标签映射消息的 LSR 记录相应的标签映射信息，若消息由对应 FEC 的下一跳发送，则将在其标签转发表中增加相应的条目。此 LSR 为它的上游 LSR 分配标签，并继续向上游 LSR 发送标签映射消息。

(4) 当入口 LSR 收到标签映射消息时，它也需要在标签转发表中增加相应的条目。这时，就完成了 LSP 的建立，接下来就可以对该 FEC 对应的数据分组进行标签转发了。

4. 会话撤销

LDP 通过检测 Hello 消息来判断邻接关系；通过检测 Keepalive 消息来判断会话的完整性。

LDP 在维持邻接关系和 LDP 会话时使用不同的定时器：

- Hello 保持定时器：LDP 对等体之间，通过周期性发送 Hello 消息表明自己希望继续维持这种邻接关系。如果 Hello 保持定时器超时仍没有收到新的 Hello 消息，则删除 Hello 邻接关系。
- Keepalive 定时器：LDP 对等体之间通过 LDP 会话连接上传送的 Keepalive 消息来维持 LDP 会话。如果会话保持定时器超时仍没有收到任何 Keepalive 消息，则关闭连接，结束 LDP 会话。

1.3.4 LDP 环路检测

在 MPLS 域中建立 LSP 也要防止产生环路，LDP 环路检测机制可以检测 LSP 环路的出现，并避免发生环路。

如果对 MPLS 域进行环路检测，则必须在所有 LSR 上都配置环路检测。但在建立 LDP 会话时，并不要求双方的环路检测配置一致。

LDP 环路检测有两种方式：

1. 最大跳数

在传递标签绑定（或者标签请求）的消息中包含跳数信息，每经过一跳该值就加一。当该值超过规定的最大值时即认为出现环路，LSP 建立失败。

2. 路径向量

在传递标签绑定（或者标签请求）的消息中记录路径信息，每经过一跳，相应的设备就检查自己的 LSR ID 是否在此记录中。在以下条件之一时认为出现环路，LSP 建立失败：

- 路径向量记录表中已有本 LSR 的记录；
- 路径的跳数超过设置的最大值。

如果记录中没有自己的 LSR ID，就会将其添加到该记录中。

1.3.5 LDP 标签过滤

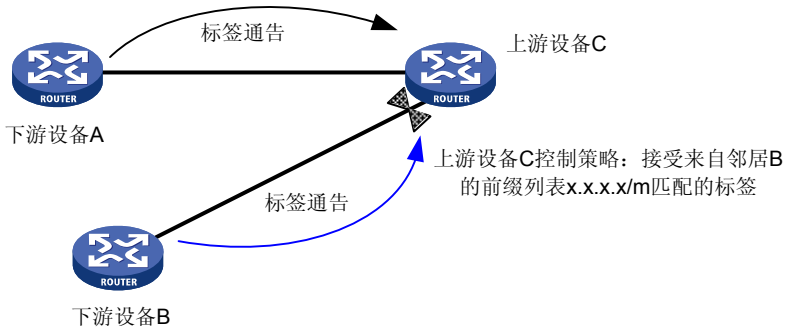
LDP 协议在缺省的情况下，从下游邻居接收到的所有 FEC 的标签映射（标签－FEC 绑定）都将接受；并按照水平分割的原则向各个上游邻居通告 FEC 的标签映射。

LDP 标签过滤提供了两种机制，可以有选择的接受从指定 LDP 下游邻居接收到的标签映射、也可以选择性地向指定 LDP 上游邻居通告指定地址前缀的标签映射。

1. 标签接受控制

标签接受控制（Label Acceptance Control）或入站标签过滤（Inbound Filtering），选择性地接受指定下游设备通告过来的、指定地址前缀的标签映射。

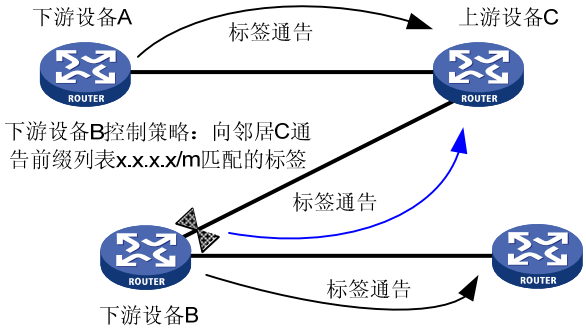
图1-8 标签接受控制示意图



2. 标签通告控制

标签通告控制（Label Advertisement Control）或出站标签过滤（Outbound Filtering），选择性地向指定上游设备通告指定地址前缀的标签映射。

图1-9 标签通告控制示意图



说明

在同一个 LDP 管理域内，推荐使用出站标签过滤策略，入站标签过滤策略可以用一个出站标签过滤策略来等价替代，这种情况下，为了减轻下游 LSR 设备的处理压力。

1.3.6 LDP GR

说明

有关 GR（Graceful Restart，平滑重启）的原理介绍请参见“系统分册”中的“GR 概述”。

在 MPLS LDP 会话建立过程中，LDP 设备需要进行 FT（Fault Tolerance，容错）和 GR 能力协商。如果双方都是 GR 设备，建立的会话就具有 FT/GR 感知能力。但是如果有一方是非 GR 设备，建立的会话就不具备 FT/GR 感知能力。为了支持 GR 能力，GR 设备需要备份 FEC 和标签信息。

假设 LDP 会话具有 GR 能力，当 GR Restarter 发生重启时，GR Helper 邻居在检测到对应的 LDP 会话进入 down 状态后，将继续保持与 GR Restarter 的邻居关系以及会话信息，直到重连定时器（Reconnect Timer）超时。在重连定时器超时前，如果邻居收到了该 GR Restarter 的会话建立请求，它将保留该会话的 LSP（Label Switching Path，标签交换路径）和标签信息，并恢复与该 GR Restarter 的会话连接；否则，将删除与该会话有关的所有 LSP 和标签信息。

会话恢复后，GR Restarter 与其邻居会分别启动各自的邻居存活状态定时器（Neighbor Liveness Timer）和恢复定时器（Recovery Timer），同时恢复该会话的所有 LSP 值，然后互相发送标签映射和标签请求消息。GR Restarter 与其邻居收到映射消息后会删除 LSP Stale 标记，并在邻居存活状态定时器和恢复定时器超时后删除该会话的所有带有 LSP Stale 标记的 LSP 信息。

在重启过程中，数据层面上的 LSP 信息将被保留，因此 MPLS 报文转发将不会中断。

1.4 配置 MPLS 基本能力

MPLS 域中参与 MPLS 转发的路由器都需要配置 MPLS 基本能力，并且，只有配置了 MPLS 基本能力后，才能进行 MPLS 其他特性的配置。

说明

目前支持 MPLS 能力的接口类型为同步串口、三层以太网接口、POS 接口、虚模板（Virtual-Template）、Mp-group 接口、MFR 接口和隧道接口（Tunnel）、ATM 接口、三层虚拟以太网接口（Virtual-Ethernet）。

1.4.1 配置准备

在配置 MPLS 基本能力之前，需要完成以下任务：

- 配置相关接口的物理参数
- 配置相关接口的链路层属性
- 配置相关接口的 IP 地址
- 配置静态路由或 IGP 协议保证各 LSR 之间可达

说明

各 LSR 之间也可以在不可达的条件下配置 MPLS 能力，这就需要根据命令 **mpls ldp transport-address** 配置，具体命令的介绍请参见“MPLS 分册”中的“MPLS 基本配置命令”。

1.4.2 配置 MPLS 基本能力

表1-1 配置 MPLS 基本能力

操作	命令	说明
进入系统视图	**system-view**	-
配置本节点的 LSR ID	**mpls lsr-id** *lsr-id*	必选 缺省情况下，未配置 LSR ID
使能本节点的 MPLS 能力，并进入 MPLS 视图	**mpls**	必选 缺省情况下，未使能本节点的 MPLS 能力
退回系统视图	**quit**	-
进入需要转发 MPLS 报文的接口视图	**interface** *interface-type interface-number*	-
使能接口的 MPLS 能力	**mpls**	必选 缺省情况下，未使能接口的 MPLS 能力



说明

LSR ID 使用 IP 地址格式，在 MPLS 域内唯一。推荐使用 Loopback 接口的 IP 地址作为 LSR ID。

1.5 配置 PHP 特性

倒数第二跳弹出特性 PHP 在 Egress 节点上配置，根据倒数第二跳节点对 PHP 特性的支持情况选择标签分配方式。

1.5.1 配置准备

在配置 PHP 特性之前，需完成以下任务：在各节点上配置 MPLS 基本能力。

1.5.2 配置 PHP 特性

根据 RFC 3032（MPLS Label Stack Encoding）中的定义：

- 标签值 0 表示 IPv4 显式空标签（Explicit-null），这个值只有出现在标签栈底时才有效。
- 标签值 3 表示隐式空标签（Implicit-null），这个值不会出现在标签栈中。当一个 LSR 发现自己被分配了隐式空标签时，它并不用这个值替代栈顶原来的标签，而是直接执行弹出标签操作。

表1-2 配置 PHP 特性

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-

操作	命令	说明
配置 Egress 节点向倒数第二跳弹出分配标签值	**label advertise { explicit-null | implicit-null | non-null }**	可选 缺省情况下，支持 PHP 特性，Egress 节点向倒数第二跳分配隐式空标签（**implicit-null**） 会话建立后，需重启 LDP 会话本命令的配置才能生效

1.6 配置 MPLS MTU

1.6.1 配置准备

配置接口的 MPLS MTU 之前，必须先使能接口的 MPLS 能力。

1.6.2 配置 MPLS MTU

表1-3 配置 MPLS MTU

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 MPLS MTU	**mpls mtu** *value*	必选 缺省情况下，未配置接口的 MPLS MTU 值

说明

配置的 MPLS MTU 值大于接口 MTU 时，有可能导致数据转发失败。如果没有配置接口的 MPLS MTU 值，则根据接口的 MTU 值进行分片，分片的长度不包含 MPLS 标签长度，即分片后 MPLS 报文的长度可能会大于接口 MTU 的值。

1.7 配置静态 LSP

标签交换路径 LSP 分为静态 LSP 和动态 LSP 两种。其中，静态 LSP 由手工配置的，动态 LSP 则利用 LDP 协议动态产生。

必须对静态 LSP 沿途的 LSR 均进行相应配置后，这条 LSP 才能正常工作。

静态 LSP 可以用于 MPLS L2VPN。

说明

有关 MPLS L2VPN 的配置请参见“MPLS 分册”中的“MPLS L2VPN 配置”。

1.7.1 配置准备

在配置静态 LSP 之前，需完成以下任务：

- 确定静态 LSP 的入节点、中间节点和出节点
- 在各节点上配置 MPLS 基本能力

1.7.2 配置静态 LSP

表1-4 配置静态 LSP

操作	命令	说明
进入系统视图	**system-view**	-
为 Ingress 节点配置静态 LSP	**static-lsp ingress** *lsp-name* **destination** *dest-addr* { *mask* | *mask-length* } { **nexthop** *next-hop-addr* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*	可选
为 Transit 节点配置静态 LSP	**static-lsp transit** *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label* { **nexthop** *next-hop-addr* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*	可选
为 Egress 节点配置静态 LSP	**static-lsp egress** *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label*	可选

说明

- 如果在配置静态 LSP 时指定了下一跳，并且下一跳地址存在于路由信息表中，则在配置 IP 静态路由时也必须指定下一跳。
- 如果在配置静态 LSP 时指定了出接口，则在配置 IP 静态路由时也必须指定出接口。
- 配置 Ingress 和 Transit 时，本地的公网地址不能被指定为下一跳。
- 有关配置 IP 静态路由的介绍请参见“IP 路由分册”中的“静态路由配置”。

1.8 配置 MPLS LDP

1.8.1 配置准备

在配置 LDP 之前，需完成以下任务：

配置 MPLS 基本能力

1.8.2 MPLS LDP 配置任务简介

表1-5 MPLS LDP 配置任务简介

配置任务	说明	详细配置
配置 MPLS LDP 能力	必选	1.8.3
配置 LDP 本地会话的参数	可选	1.8.4
配置 LDP 远端会话的参数	可选	1.8.5
配置 LSP 触发策略	可选	1.8.6
配置 LDP 标签分配控制方式	可选	1.8.7
配置 LDP 环路检测	可选	1.8.8
配置 LDP MD5 认证	可选	1.8.9

1.8.3 配置 MPLS LDP 能力

表1-6 配置 MPLS LDP 能力

操作	命令	说明
进入系统视图	**system-view**	-
使能本节点的 LDP 能力，并进入 MPLS-LDP 视图	**mpls ldp**	必选 缺省情况下，未使能 LDP 能力
配置 LDP 的 LSR ID	**lsr-id** *lsr-id*	可选 缺省情况下，LDP 的 LSR ID 等于 MPLS LSR ID
退回系统视图	**quit**	-
进入建立 LDP 会话的接口视图	**interface** *interface-type interface-number*	-
使能接口的 LDP 能力	**mpls ldp**	必选 缺省情况下，在接口上未使能 LDP 能力

说明

- 目前支持 LDP 能力的接口类型为同步串口（Serial）、三层以太网接口、POS 接口、虚模板（Virtual-Template）、Mp-group 接口、MFR 接口、隧道接口（Tunnel）、ATM 接口、三层虚拟以太网接口（Virtual-Ethernet）。
- 如果取消接口的 LDP 能力将会导致接口下的所有 LDP 会话中断，基于这些会话的所有 LSP 也将被删除。
- 通常情况下 LDP 使用缺省的 MPLS LSR ID 即可，在某些使用 VPN 实例，例如 MPLS L3VPN 的组网方案中，如果 VPN 与公网地址空间重叠，则需要为 LDP 另外配置 LSR ID，以保证 TCP 连接能够正常建立。

1.8.4 配置 LDP 本地会话的参数

配置 LDP 本地会话传输地址时，既可以选择本接口的 IP 地址为传输地址，也可以选择指定 IP 地址为传输地址。

表1-7 配置 LDP 本地会话的参数

操作	命令	说明
进入系统视图	**system-view**	-
进入建立 LDP 会话的接口视图	**interface** *interface-type interface-number*	-
配置链路 Hello 保持定时器的值	**mpls ldp timer hello-hold** *value*	可选 缺省情况下，链路 Hello 保持定时器为 15 秒
配置链路 Keepalive 保持定时器的值	**mpls ldp timer keepalive-hold** *value*	可选 缺省情况下，链路 Keepalive 保持定时器为 45 秒
配置 LDP 传输地址	**mpls ldp transport-address** { *ip-address* | **interface** }	可选 缺省情况下，传输地址是本 LSR 的 LSR ID

1.8.5 配置 LDP 远端会话的参数

配置 LDP 远端会话传输地址时，选择指定接口的 IP 地址为传输地址。

表1-8 配置 LDP 远端会话的参数

操作	命令	说明
进入系统视图	**system-view**	-
创建远端对等体实体并进入 MPLS-LDP 远端对等体视图	**mpls ldp remote-peer** *remote-peer-name*	必选
指定 LDP 远端对等体的 IP 地址	**remote-ip** *ip-address*	必选
配置通过远端会话通告基于地址前缀的标签	**prefix-label advertise**	可选 缺省情况下，不会通过远端会话通告基于地址前缀的标签
配置目标 Hello 保持定时器的值	**mpls ldp timer hello-hold** *value*	可选 缺省情况下，目标 Hello 保持定时器为为 45 秒
配置目标 Keepalive 保持定时器	**mpls ldp timer keepalive-hold** *value*	可选 缺省情况下，目标 Keepalive 保持定时器为 45 秒
配置 LDP 传输地址	**mpls ldp transport-address** *ip-address*	可选 缺省情况下，传输地址是本 LSR 的 LSR ID

说明

- 缺省情况下，LDP 不会通过远端会话通告基于地址前缀的标签映射消息，远端会话只用来为 L2VPN 传送消息。有关远端会话的应用请参见“MPLS 分册”中的“MPLS L2VPN 配置”中的“配置 Martini 方式 MPLS L2VPN”部分。
- 通过远端会话通告基于地址前缀的标签主要用来实现 MPLS LDP over MPLS TE 功能。有关 MPLS LDP over MPLS TE 功能的应用请参见“MPLS 分册”中的“MPLS TE 配置”。

注意

- 如果对等体与指定的远端对等体之间已经存在本地邻接关系，则远程邻接关系将不能建立；如果已经存在远程邻接关系，又为远端对等体创建了本地邻接关系，则在本地对等体和远端对等体的 LDP 传输地址、Keepalive 保持时间配置一致时，则将建立本地邻接关系，远端对等体将被删除。；本地对等体和远端对等体的 LDP 传输地址或 Keepalive 保持时间配置不一致时，将无法建立本地邻接关系，继续保持远程邻接关系。即两个 LSR 之间只能存在一个远程会话或一个本地会话，并且，本地会话的优先级高于远程会话。
- 即将配置的远端对等体的 IP 地址不能与已经存在的远端对等体 IP 地址重复；否则配置不成功，将返回出错消息。
- 配置的 LDP 传输地址应为设备接口的 IP 地址，否则 LDP 会话将无法建立。

1.8.6 配置 LSP 触发策略

LSP 的触发策略包括：

- 所有静态路由和 IGP 路由项都触发 LDP 建立 LSP；

- 通过 IP 地址前缀列表过滤的静态路由和 IGP 路由项触发 LDP 建立 LSP，被 IP 地址前缀列表拒绝的静态路由和 IGP 路由项不能触发建立 LSP。

表1-9 配置 LSP 触发策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置 LSP 的触发策略	**lsp-trigger** [**vpn-instance** *vpn-instance-name*] { **all** | **ip-prefix** *prefix-name* }	可选 缺省情况下，只有本地 32 位掩码的环回地址能触发 LDP 来建立 LSP

说明

- 如果指定 **vpn-instance** *vpn-instance-name* 参数，则配置指定 VPN 实例的 LSP 触发策略；如果未指定 **vpn-instance** *vpn-instance-name* 参数，则配置公网路由的 LSP 触发策略。
- LSP 的建立需要 LSR 上有精确匹配的路由项，如果使用 32 位掩码的 Loopback 接口，则必须有精确匹配的主机路由才能触发 LSP 的建立。
- 有关 IP 前缀列表的介绍请参见“IP 路由分册”中的“路由策略配置”。

1.8.7 配置 LDP 标签分配控制方式

表1-10 配置 LDP 标签分配控制方式

操作	命令	说明
进入系统视图	**system-view**	-
使能本节点的 LDP 能力，并进入 MPLS-LDP 视图	**mpls ldp**	必选
配置标签分配控制方式	**label-distribution** { **independent** | **ordered** }	可选 缺省情况下，标签分配控制方式为有序方式（**ordered**） 会话建立后，需重启 LDP 会话本命令的配置才能生效
配置标签保持方式	**label-retention** { **liberal** | **conservative** }	可选 缺省情况下，标签保持方式为自由方式（**liberal**） 会话建立后，需重启 LDP 会话本命令的配置才能生效
使能 DU 模式下标签重发布特性	**du-readvertise**	可选 缺省情况下，DU 模式下已使能标签重发布特性
设置 DU 模式下定期重发布标签的定时器值	**du-readvertise timer** *value*	可选 缺省情况下，DU 模式下定期重发布标签的定时器值为 30 秒

1.8.8 配置 LDP 环路检测

表1-11 配置 LDP 环路检测

操作	命令	说明
进入系统视图	**system-view**	-
使能本节点的 LDP 能力，并进入 MPLS-LDP 视图	**mpls ldp**	必选
开启环路检测功能	**loop-detect**	必选 缺省情况下，环路检测功能是关闭的
设置环路检测的最大跳数	**hops-count** *hop-number*	可选 缺省情况下，环路检测最大跳数为 32
设置路径向量的最大跳数	**path-vectors** *pv-number*	可选 缺省情况下，路径向量值为 32



说明

修改环路检测配置不影响已经建立的 LSP。



注意

环路检测功能需要在所有接口 LDP 使能之前进行配置。

1.8.9 配置 LDP MD5 认证

为了提高 LDP 会话连接的安全性，可以对 LDP 使用的 TCP 连接配置 MD5 认证。

表1-12 配置 LDP MD5 认证

操作	命令	说明
进入系统视图	**system-view**	-
使能本节点的 LDP 能力，并进入 MPLS-LDP 视图	**mpls ldp**	必选
使能 LDP 的 MD5 认证	**md5-password** { **cipher** | **plain** } *peer-lsr-id password*	必选 缺省情况下，未使能 LDP 的 MD5 认证

1.8.10 配置 LDP 标签过滤

为了保证 LDP 创建的标签映射数目不会超过系统的最大规格限制，可以配置 LDP 标签过滤策略。

表1-13 配置 LDP 标签过滤

操作	命令	说明
进入系统视图	**system-view**	-
使能本节点的 LDP 能力，并进入 MPLS-LDP 视图	**mpls ldp**	必选

操作	命令	说明
设置标签接受控制策略	**accept-label peer** *peer-id* **ip-prefix** *ip-prefix-name*	可选 缺省情况下，未设置标签接受控制策略
设置标签通告控制策略	**advertise-label ip-prefix** *ip-prefix-name* [**peer** *peer-ip-prefix-name*]	必选 缺省情况下，未设置标签通告控制策略

1.9 配置 LDP 实例

LDP 实例用于 MPLS L3VPN 运营商的运营商（Carriers' carriers）组网模型中。配置时需要将已创建的 VPN 实例使能 LDP 能力。

除了 LDP 的 GR 特性外，MPLS LDP 视图下的命令都可以在 MPLS-LDP-VPN 实例视图下使用。

1.9.1 配置准备

在配置 LDP 实例之前，需完成以下任务：

- 配置 VPN 实例
- 配置 MPLS 基本能力
- 配置 MPLS LDP 能力

1.9.2 配置 LDP 实例

通常情况下，LDP 使用缺省的 MPLS LSR ID 即可。在某些使用 VPN 实例（例如 MPLS L3VPN）的组网方案中，如果 VPN 与公网地址空间重叠，则需要为 LDP 另外配置 LSR ID，以保证 TCP 连接能够正常建立。

表1-14 配置 LDP 实例

操作	命令	说明
进入系统视图	**system-view**	-
对指定的 VPN 实例使能 LDP 能力，并进入 MPLS-LDP-VPN 实例视图	**mpls ldp vpn-instance** *vpn-instance-name*	必选
配置 LDP 实例的 LSR ID	**lsr-id** *lsr-id*	可选 缺省情况下，LDP 实例的 LSR ID 等于本 LSR 的 LSR ID

注意

- 对于使能 LDP 能力的接口，MPLS-LDP-VPN 实例视图下的配置只影响绑定 VPN 实例的接口；而 MPLS LDP 视图下的配置对绑定到 VPN 实例的接口没有影响。为 LDP 实例配置传输地址时，也必须使用绑定到 VPN 实例的接口的 IP 地址。
- 对于私网侧的 LDP 邻居关系，缺省情况下是使用使能 LDP 能力的接口地址建立的；而公网上的 LDP 邻居关系，缺省情况下是使用 LDP 实例的 LSR ID 建立的。

1.10 配置 LDP GR

1. 配置准备

在配置 LDP GR 之前，需完成以下任务：

作为 GR Restarter 的设备上和 GR Helper 的设备均配置 MPLS LDP 能力。

2. 配置 LDP GR

说明

设备即可以作为 GR Restarter，又可以作为 GR Helper。

表1-15 配置 LDP GR

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS-LDP 视图	**mpls ldp**	-
使能 MPLS LDP 协议的 GR 能力	**graceful-restart**	必选 缺省情况下，MPLS LDP 协议的 GR 能力处于关闭状态
配置 FT 重连定时器值	**graceful-restart timer reconnect** *timer*	可选 缺省情况下，FT 重连定时器值为 300 秒
配置 LDP 邻居存活定时器值	**graceful-restart timer neighbor-liveness** *timer*	可选 缺省情况下，LDP 邻居存活定时器的值为 120 秒
配置 LDP 恢复时间	**graceful-restart timer recovery** *timer*	可选 缺省情况下，LDP 恢复时间为 300 秒

说明

在 MPLS LDP GR 过程中，GR Helper 将本设备配置的 LDP 邻居存活定时器值和对端 GR Restarter 设备配置的 FT 重连定时器值中的较小者，作为本设备上的 FT 重连定时器的值；GR Helper 将本设备配置的 LDP 恢复时间值和对端 GR Restarter 设备配置的 LDP 恢复时间值中的较小者，作为本设备上的 LDP 恢复时间值。

3. 平滑重启 MPLS LDP

当需要在不发生主备倒换的情况下测试 MPLS LDP GR 时，可以对 MPLS LDP 进行平滑重启。正常情况下，请不要执行此操作。

表1-16 平滑重启 MPLS LDP

操作	命令	说明
平滑重启 MPLS LDP	**graceful-restart mpls ldp**	必选 请在用户视图下执行此命令

1.11 配置 MPLS 的 TTL 处理

1.11.1 配置准备

在配置 MPLS 的 IP TTL 复制功能之前，需完成以下任务：配置 MPLS 基本能力。

1.11.2 配置 MPLS 的 IP TTL 复制功能

表1-17 配置 MPLS 的 IP TTL 复制功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能 MPLS 的 IP TTL 复制功能	**ttl propagate { public | vpn }**	可选 缺省情况下，只对公网报文使能 TTL 复制功能

注意

- **ttl propagate** 命令只影响从 IP 的 TTL 复制到 MPLS 标签的 TTL。在 Egress 节点，系统将从 IP 头的 TTL 域和 MPLS 标签的 TTL 域中选取值较小者，作为 IP 头的 TTL 值，并执行减一操作。
- 如果配置 **ttl propagate vpn** 命令使能对 VPN 报文的 IP TTL 复制功能，则建议在所有相关 PE 上都使能此功能，以保证从不同的 PE 执行 tracert 得到的结果一致。

1.11.3 配置 ICMP 响应报文使用的路径

ICMP 响应报文使用的路径有两种：IP 路由和 LSP。

对于仅有一层标签的 MPLS 报文，当 TTL 超时时，用户可以配置直接使用 IP 路由而不是 LSP 返回 ICMP 响应报文。

在 MPLS 业务中，一般情况下 IP 路由器仅有公网路由，而一层标签的 MPLS 报文承载的也是公网报文，可以配置此功能。

在 MPLS VPN 中，对于 ASBR（Autonomous System Boundary Router，自治系统边界路由器）以及 HoVPN 组网应用中的 SPE（包括嵌套应用中的 SPE），其承载 VPN 报文的 MPLS 报文可能只有一层标签，这种情况下，如果要对 VPN 进行 tracert 操作查看公网路由器的转发路径，需要进行两部分配置：

- 在所有相关 PE 上配置 **ttl propagate vpn** 命令，对 VPN 报文进行 IP TTL 复制；
- 在 ASBR 和 SPE 上配置 **undo ttl expiration pop** 命令，保证 ICMP 响应报文按照原 LSP 转发。

说明

SPE 是指上层 PE（Superstratum PE）或运营商侧 PE（Sevice Provider-end PE），有关 HoVPN 的详细介绍请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

表1-18 配置 ICMP 响应报文使用的路径

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能 MPLS TTL 超时发送 ICMP 差错报文功能	**ttl expiration enable**	可选 缺省情况下，MPLS TTL 超时发送 ICMP 差错报文功能处理开启状态
配置 MPLS TTL 超时时，沿本地 IP 路由返回 ICMP 报文	**ttl expiration pop**	可选 缺省情况下，对于一层标签的 MPLS TTL 超时时，沿本地 IP 路由返回 ICMP 报文 对于多层标签，不受该命令控制，当 MPLS TTL 超时时，沿 LSP 路径转发 ICMP 响应报文
配置 MPLS TTL 超时后，沿 LSP 返回 ICMP 报文	**undo ttl expiration pop**	可选

说明

实际配置时，根据希望的 ICMP 响应报文返回路径，选择 ttl expiration pop 或者 undo ttl expiration pop。

1.12 配置 MPLS 统计功能

1.12.1 使能 MPLS 统计功能

缺省情况下，MPLS 统计功能处于关闭状态，只有使能该功能后，才能通过显示命令查看 MPLS 的统计信息。

表1-19 使能 MPLS 统计功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 MPLS 统计功能	**mpls statistics enable** { **interface** *interface-type interface-number* | **lsp in-label** *in-label* | **tunnel token** *token* }	必选 缺省情况下，MPLS 的统计功能处于关闭状态

1.13 检测 MPLS LSP

下列操作可在任意视图下执行。

表1-20 检测 MPLS LSP

操作	命令
检测 MPLS LSP 的有效性和可达性	**ping lsp** [**-a** *source-ip* | **-c** *count* | **-exp** *exp-value* | **-h** *ttl-value* | **-m** *wait-time* | **-r** *reply-mode* | **-s** *packet-size* | **-t** *time-out* | **-v**] * { **ipv4** *dest-addr mask-length* [*destination-ip-addr-header*] | **te** *interface-type interface-number* }

操作	命令
对 MPLS LSP 的错误进行定位	**tracert lsp** [**-a** *source-ip* | **-exp** *exp-value* | **-h** *ttl-value* | **-r** *reply-mode* |**-t** *time-out*] * { **ipv4** *dest-addr mask-length* [*destination-ip-addr-header*] | **te** *interface-type interface-number* }

1.14 开启 MPLS 的 Trap 功能

开启 MPLS 模块的 Trap 功能后，该模块会生成级别为 notifications 的 Trap 报文，用于报告该模块的重要事件。生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-21 开启 MPLS 的 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 MPLS 的 Trap 功能	**snmp-agent trap enable mpls**	必选 缺省情况下，MPLS 的 Trap 功能处于关闭状态

说明

snmp-agent trap enable mpls 命令的详细介绍请参见“系统分册/SNMP 命令”中的 **snmp-agent trap enable** 命令。

1.15 MPLS 显示和维护

1.15.1 重启 LDP 会话

LDP 会话状态为 up 后，如果修改 LDP 会话的任何参数，将会引起 LDP 会话的不能正常进行。此时，需要重启 LDP 会话，重新协商各种参数，建立 LDP 会话。

在用户视图下执行 **reset** 命令可以重启 LDP 会话。

表1-22 重启 LDP 会话

操作	命令
重启 LDP 会话	**reset mpls ldp** [**all** | [**vpn-instance** *vpn-instance-name*] [*fec mask* | **peer** *peer-id*]]

1.15.2 显示 MPLS 运行状态

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MPLS 的运行情况，用户可以通过查看显示信息验证配置的效果。

表1-23 MPLS 运行状态显示

操作		命令
显示使能了 MPLS 能力的接口的信息		**display mpls interface** [*interface-type interface-number*] [**verbose**]
显示ILM表的信息	集中式设备	**display mpls ilm** [*label*] [**include** *text*]
	分布式设备	**display mpls ilm** [*label*] [**slot** *slot-number*] [**include** *text*]
显示 MPLS 标签信息		**display mpls label** { *label-value1* [**to** *label-value2*] | **all** }
显示 LSP 信息		**display mpls lsp** [**incoming-interface** *interface-type interface-number*] [**outgoing-interface** *interface-type interface-number*] [**in-label** *in-label-value*] [**out-label** *out-label-value*] [**asbr** | [**vpn-instance** *vpn-instance-name*] [**protocol** { **bgp** | **bgp-ipv6** | **crldp** | **ldp** | **rsvp-te** | **static** | **static-cr** }]] [**egress** | **ingress** | **transit**] [{ **exclude** | **include** } *dest-addr mask-length*] [**verbose**]
显示 LSP 统计信息		**display mpls lsp statistics**
显示 NHLFE 表的信息	集中式设备	**display mpls nhlfe** [*token*] [**include** *text*]
	分布式设备	**display mpls nhlfe** [*token*] [**slot** *slot-number*] [**include** *text*]
显示 NHLFE 表项的使用情况		**display mpls nhlfe reflist** *token* [**slot** *slot-number*]
显示静态 LSP 信息		**display mpls static-lsp** [**lsp-name** *lsp-name*] [{ **include** | **exclude** } *dest-addr mask-length*] [**verbose**]
显示路由的 LSP 相关信息		**display mpls route-state** [**vpn-instance** *vpn-instance-name*] [*dest-addr mask-length*]
根据 LSP 显示 MPLS 统计信息		**display mpls statistics lsp** { **all** | *index* | *lsp-name* }
根据使能 MPLS 能力的接口显示 MPLS 统计信息		**display mpls statistics interface** { *interface-type interface-number* | **all** }
根据 LSP 的入标签显示 LSP 统计信息		**display mpls statistics lsp** [**in-label** *in-label*]
显示指定 LSP Token 对应的公网隧道统计信息		**display mpls statistics tunnel** [**token** *token*]

1.15.3 显示 MPLS LDP 运行状态

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MPLS LDP 的运行情况，用户可以通过查看显示信息验证配置的效果。

表1-24 显示 MPLS LDP 运行状态

操作	命令
显示 LDP 信息	**display mpls ldp** [**all** [**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]]
显示指定 FEC 的标签通告信息	**display mpls ldp fec** [**vpn-instance** *vpn-instance-name*] *dest-addr mask-length*
显示使能了 LDP 能力的接口信息	**display mpls ldp interface** [**all** [**verbose**] | [**vpn-instance** *vpn-instance-name*] [*interface-type interface-number* | **verbose**]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示对等体的信息	**display mpls ldp peer** [**all** [**verbose**] | [**vpn-instance** *vpn-instance-name*] [*peer-id* | **verbose**]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示远端对等体信息	**display mpls ldp remote-peer** [**remote-name** *remote-peer-name*] [**|** { **begin** | **exclude** | **include** } *regular-expression*]

操作	命令
显示对等体间会话信息	**display mpls ldp session** [**all** [**verbose** | **statistics**] | [**vpn-instance** *vpn-instance-name*] [*peer-id* | **verbose**]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示 LDP 创建的 LSP 相关信息	**display mpls ldp lsp** [**all** | [**vpn-instance** *vpn-instance-name* [*dest-addr mask-length*]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示 CR-LDP 创建的 CR-LSP 信息	**display mpls ldp cr-lsp** [**lspid** *lsr-id lsp-id*] [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示指定 LDP 实例的信息	**display mpls ldp vpn-instance** *vpn-instance-name* [**|** { **begin** | **exclude** | **include** } *regular-expression*]

1.15.4 清除 MPLS 相关信息

在完成上述配置后，在用户视图下执行 **reset** 命令可以清除 MPLS 的相关信息。

表1-25 MPLS LDP 统计信息清除

操作	命令
清除 MPLS 接口统计信息	**reset mpls statistics interface** { *interface-type interface-number* | **all** }
清除 LSP 统计信息	**reset mpls statistics lsp** { index | **all** | **name** *lsp-name* }
清除指定入标签对应的 LSP 统计信息	**reset mpls statistics lsp** [**in-label** *in-label*]
清除指定 LSP Token 对应的公网隧道统计信息	**reset mpls statistics tunnel** [**token** *token*]

1.16 MPLS 配置举例

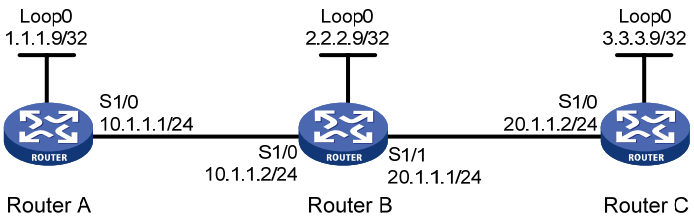
1.16.1 配置 LDP 会话示例

1. 组网需求

- Router A、Router B 和 Router C 均支持 MPLS，运行 OSPF 作为 MPLS 骨干网上的 IGP。
- Router A 和 Router B、Router B 和 Router C 之间建立本地 LDP 会话；Router A 和 Router C 之间建立远端 LDP 会话。

2. 组网图

图1-10 配置 LDP 会话组网图



3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-10配置各接口IP地址和掩码，包括Loopback接口，以Router A为例，其他设备配置略。

配置 Router A。

```
<Sysname> system-view
[Sysname] sysname RouterA
[RouterA] interface loopback 0
[RouterA-LoopBack0] ip address 1.1.1.9 32
[RouterA-LoopBack0] quit
[RouterA] interface serial 1/0
[RouterA-Serial1/0] ip address 10.1.1.1 24
[RouterA-Serial1/0] quit
```

(2) 配置 OSPF 协议发布 LSR ID 的主机路由

配置 Router A。

```
[RouterA] ospf
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[RouterA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] ospf
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[RouterB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[RouterB-ospf-1-area-0.0.0.0] return
```

配置 Router C。

```
<RouterC> system-view
[RouterC] ospf
[RouterC-ospf-1] area 0
[RouterC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[RouterC-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[RouterC-ospf-1-area-0.0.0.0] return
```

配置完成后，在各设备上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到对方的主机路由。以 Router A 为例：

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 9        Routes : 9
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
      1.1.1.9/32  Direct 0    0         127.0.0.1       InLoop0
      2.2.2.9/32  OSPF   10   1563      10.1.1.2        S1/0
      3.3.3.9/32  OSPF   10   3125      10.1.1.2        S1/0
     10.1.1.0/24  Direct 0    0         10.1.1.1        S1/0
     10.1.1.1/32  Direct 0    0         127.0.0.1       InLoop0
     10.1.1.2/32  Direct 0    0         10.1.1.2        S1/0
```

```
     20.1.1.0/24  OSPF   10   3124      10.1.1.2      S1/0
    127.0.0.0/8   Direct 0    0         127.0.0.1     InLoop0
    127.0.0.1/32  Direct 0    0         127.0.0.1     InLoop0
```

Router A 和 Router B，Router B 和 Router C 之间应建立起 OSPF 邻居关系，执行 **display ospf peer verbose** 命令可以看到邻居达到 FULL 状态。以 Router A 为例：

```
[RouterA] display ospf peer verbose
          OSPF Process 1 with Router ID 1.1.1.9
                  Neighbors
 Area 0.0.0.0 interface 10.1.1.1(Serial1/0)'s neighbors
 Router ID: 2.2.2.9      Address: 10.1.1.2        GR State: Normal
   State: Full  Mode:Nbr is  Master  Priority: 1
   DR: None   BDR: None   MTU: 1500
   Dead timer due in 39  sec
   Neighbor is up for 00:02:13
   Authentication Sequence: [ 0 ]
```

(3) 配置 MPLS 基本能力，并使能 LDP

配置 Router A。

```
[RouterA] mpls lsr-id 1.1.1.9
[RouterA] mpls
[RouterA-mpls] quit
[RouterA] mpls ldp
[RouterA-mpls-ldp] quit
[RouterA] interface serial 1/0
[RouterA-Serial1/0] mpls
[RouterA-Serial1/0] mpls ldp
[RouterA-Serial1/0] quit
```

配置 Router B。

```
[RouterB] mpls lsr-id 2.2.2.9
[RouterB] mpls
[RouterB-mpls] quit
[RouterB] mpls ldp
[RouterB-mpls-ldp] quit
[RouterB] interface serial 1/0
[RouterB-Serial1/0] mpls
[RouterB-Serial1/0] mpls ldp
[RouterB-Serial1/0] quit
[RouterB] interface serial 1/1
[RouterB-Serial1/1] mpls
[RouterB-Serial1/1] mpls ldp
[RouterB-Serial1/1] quit
```

配置 Router C。

```
[RouterC] mpls lsr-id 3.3.3.9
[RouterC] mpls
[RouterC-mpls] quit
[RouterC] mpls ldp
```

```
[RouterC-mpls-ldp] quit
[RouterC] interface serial 1/0
[RouterC-Serial1/0] mpls
[RouterC-Serial1/0] mpls ldp
[RouterC-Serial1/0] quit
```

完成上述配置后，Router A 和 Router B、Router B 和 Router C 之间的本地 LDP 会话建立成功。

在各设备上执行 **display mpls ldp session** 命令，可以看到 LDP 会话的建立情况；执行 **display mpls ldp peer** 命令，可以看到 LDP 的对等体情况。以 Router A 为例：

```
[RouterA] display mpls ldp session
               LDP Session(s) in Public Network
Total number of sessions: 1
 ------------------------------------------------------------------
 Peer-ID       Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 2.2.2.9:0     Operational   DU   Passive  Off  Off  5/5
 ------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
[RouterA] display mpls ldp peer
         LDP Peer Information in Public network
Total number of peers: 1
------------------------------------------------------------------
 Peer-ID                Transport-Address  Discovery-Source
 ------------------------------------------------------------------
 2.2.2.9:0              2.2.2.9            Serial1/0
 ------------------------------------------------------------------
```

(4) 配置远端 LDP 会话

配置 Router A。

```
[RouterA] mpls ldp remote-peer peerc
[RouterA-mpls-ldp-remote-peerc] remote-ip 3.3.3.9
[RouterA-mpls-ldp-remote-peerc] quit
```

配置 Router C。

```
[RouterC] mpls ldp remote-peer peera
[RouterC-mpls-ldp-remote-peera] remote-ip 1.1.1.9
[RouterC-mpls-ldp-remote-peera] quit
```

完成上述配置后，在 Router A 上查看 LDP 会话的建立情况和 LDP 的对等体情况，可以看到增加了与 Router C 的远端 LDP 会话。

```
[RouterA] display mpls ldp session
               LDP Session(s) in Public Network
Total number of sessions: 2
 ------------------------------------------------------------------
 Peer-ID       Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 2.2.2.9:0     Operational   DU   Passive  Off  Off  35/35
 3.3.3.9:0     Operational   DU   Passive  Off  Off  8/8
 ------------------------------------------------------------------
```

```
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
[RouterA] display mpls ldp peer
         LDP Peer Information in Public network
Total number of peers: 2
 ------------------------------------------------------------------
 Peer-ID                Transport-Address  Discovery-Source
 ------------------------------------------------------------------
 2.2.2.9:0              2.2.2.9            Serial1/0
 3.3.3.9:0              3.3.3.9            Remote Peer : peerc
 ------------------------------------------------------------------
```

1.16.2 使用 LDP 建立 LSP 示例

1. 组网需求

在 图 1-10的网络中，要求从Router A到Router C之间使用LDP建立一条LSP，并检测LSP的有效性和可达性。

2. 组网图

参见 图 1-10。

3. 配置步骤

(1) 配置LDP会话，请参见“1.16.1 配置LDP会话示例”

(2) 配置 LSP 的触发建立策略



说明

实际配置LDP LSP时，Router A和Router C之间不必建立远端LDP会话，只需要在LDP LSP沿途的设备之间存在LDP会话。对于 图 1-10的情况，Router A和Router B之间、Router B和Router C之间存在本地LDP会话，就可以建立LDP LSP。

配置 Router A。

```
[RouterA] mpls
[RouterA-mpls] lsp-trigger all
[RouterA-mpls] quit
```

配置 Router B。

```
[RouterB] mpls
[RouterB-mpls] lsp-trigger all
[RouterB-mpls] quit
```

配置 Router C。

```
[RouterC] mpls
[RouterC-mpls] lsp-trigger all
[RouterC-mpls] quit
```

配置完成后，在各设备上执行 **display mpls ldp lsp** 命令，可以看到 LDP LSP 的建立情况。以 Router A 为例。

```
[RouterA] display mpls ldp lsp
                         LDP LSP Information
```

```
-------------------------------------------------------------------
SN  DestAddress/Mask   In/OutLabel   Next-Hop     In/Out-Interface
-------------------------------------------------------------------
1    1.1.1.9/32         3/NULL       127.0.0.1     Ser1/0/InL0
2    2.2.2.9/32         NULL/3       10.1.1.2      -------/Ser1/0
3    3.3.3.9/32         NULL/1025    10.1.1.2      -------/Ser1/0
4    20.1.1.0/24        NULL/3       10.1.1.2      -------/Ser1/0
-------------------------------------------------------------------
A '*' before an LSP means the LSP is not established
A '*' before a Label means the USCB or DSCB is stale
```

检测 MPLS LSP 的有效性和可达性。

```
<RouterA> ping lsp ipv4 3.3.3.9 32
 LSP PING FEC: LDP IPV4 PREFIX 3.3.3.9/32 : 100  data bytes, press CTRL_C to break
 Reply from 20.1.1.2: bytes=100 Sequence=1 time = 2 ms
 Reply from 20.1.1.2: bytes=100 Sequence=2 time = 1 ms
 Reply from 20.1.1.2: bytes=100 Sequence=3 time = 2 ms
 Reply from 20.1.1.2: bytes=100 Sequence=4 time = 1 ms
 Reply from 20.1.1.2: bytes=100 Sequence=5 time = 2 ms

 --- FEC: LDP IPV4 PREFIX 3.3.3.9/32 ping statistics ---
 5 packet(s) transmitted
 5 packet(s) received
 0.00% packet loss
 round-trip min/avg/max = 1/1/2 ms
```

目 录

1 MPLS TE配置 1-1
1.1 MPLS TE简介 1-1
1.1.1 流量工程与MPLS TE 1-1
1.1.2 MPLS TE的基本概念 1-2
1.1.3 MPLS TE的实现 1-2
1.1.4 CR-LSP 1-3
1.1.5 CR-LDP 1-4
1.1.6 RSVP-TE 1-4
1.1.7 流量转发 1-8
1.1.8 自动带宽调整 1-9
1.1.9 CR-LSP备份 1-10
1.1.10 快速重路由 1-10
1.1.11 DiffServ-Aware TE 1-11
1.1.12 MPLS LDP over MPLS TE 1-12
1.1.13 协议规范 1-12
1.2 MPLS TE配置任务简介 1-13
1.3 配置MPLS TE基本能力 1-13
1.3.1 配置准备 1-13
1.3.2 配置MPLS TE基本能力 1-14
1.4 使用静态CR-LSP配置MPLS TE隧道 1-14
1.4.1 配置准备 1-14
1.4.2 使用静态CR-LSP配置MPLS TE隧道 1-15
1.5 使用动态信令协议配置MPLS TE隧道 1-15
1.5.1 配置准备 1-16
1.5.2 使用动态信令协议配置MPLS TE隧道 1-16
1.6 配置RSVP-TE高级特性 1-20
1.6.1 配置准备 1-20
1.6.2 配置RSVP-TE高级特性 1-21
1.7 调整CR-LSP的建立 1-24
1.7.1 配置准备 1-24
1.7.2 调整CR-LSP的建立 1-24
1.8 调整MPLS TE隧道的建立 1-26
1.8.1 配置准备 1-26
1.8.2 调整MPLS TE隧道的建立 1-26
1.9 配置流量转发 1-28
1.9.1 配置准备 1-28
1.9.2 配置流量转发 1-28
1.10 配置影响流量转发的参数 1-30
1.10.1 配置准备 1-31
1.10.2 配置影响流量转发的参数 1-31
1.11 配置自动带宽调整过程 1-32

1.11.1 配置准备 1-32
1.11.2 配置自动带宽调整过程 1-32
1.12 配置CR-LSP备份 1-33
1.12.1 配置准备 1-33
1.12.2 配置CR-LSP备份 1-33
1.13 配置MPLS TE快速重路由 1-34
1.13.1 配置准备 1-34
1.13.2 配置MPLS TE快速重路由 1-34
1.14 MPLS TE显示和维护 1-36
1.14.1 MPLS TE显示和维护 1-36
1.14.2 重启自动带宽调整功能 1-38
1.15 MPLS TE典型组网举例 1-38
1.15.1 使用静态CR-LSP配置MPLS TE隧道示例 1-38
1.15.2 使用RSVP-TE配置MPLS TE隧道示例 1-43
1.15.3 配置RSVP-TE GR示例 1-49
1.15.4 使用CR-LDP配置MPLS TE隧道示例 1-51
1.15.5 配置CR-LSP备份示例 1-59
1.15.6 配置快速重路由示例 1-63
1.15.7 配置MPLS LDP over MPLS TE示例 1-72
1.15.8 配置MPLS L3VPN中的MPLS TE示例 1-79
1.16 MPLS TE常见配置错误举例 1-87
1.16.1 不能产生TE LSA 1-87

1 MPLS TE 配置

1.1 MPLS TE 简介

1.1.1 流量工程与 MPLS TE

1. 流量工程

(1) 流量工程的作用

网络拥塞是影响骨干网络性能的主要问题。拥塞的原因可能是网络资源不足，也可能网络资源负载不均衡导致的局部拥塞。TE（Traffic Engineering，流量工程）解决的是由于负载不均衡导致的拥塞。

流量工程通过实时监控网络的流量和网络单元的负载，动态调整流量管理参数、路由参数和资源约束参数等，使网络运行状态迁移到理想状态，优化网络资源的使用，避免负载不均衡导致的拥塞。

总的来说，流量工程的性能指标包括两个方面：

- 面向业务的性能指标：增强业务的 QoS（Quality of Service，服务质量）性能，例如对分组丢失、时延、吞吐量以及 SLA（Service Level Agreement，服务等级协定）的影响。
- 面向资源的性能指标：优化资源利用。带宽是一种重要的资源，对带宽资源进行高效管理是流量工程的一项中心任务。

(2) 流量工程的解决方案

现有的 IGP 协议都是拓扑驱动的，只考虑网络的连接情况，不能灵活反映带宽和流量特性这类动态状况。

解决 IGP 上述缺点的方法之一是使用重叠模型（Overlay），如 IP over ATM、IP over FR 等。重叠模型在网络的物理拓扑结构之上提供了一个虚拟拓扑结构，从而扩展了网络设计的空间，为支持流量与资源控制提供了许多重要功能，可以实现多种流量工程策略。然而，由于协议之间往往存在很大差异，重叠模型在可扩展性方面存在不足。

为了在大型骨干网络中部署流量工程，必须采用一种可扩展性好、简单的解决方案。MPLS TE 就是为这一需求而提出的。

2. MPLS TE

MPLS 本身具有一些不同于 IGP 的特性，其中就有实现流量工程所需要的，例如：

- MPLS 支持显式 LSP 路由；
- LSP 较传统单个 IP 分组转发更便于管理和维护；
- CR-LDP（Constraint-based Routed Label Distribution Protocol，基于约束路由的 LDP）可以实现流量工程的各种策略；
- 基于 MPLS 的流量工程的资源消耗较其它实现方式更低。

MPLS TE 结合了 MPLS 技术与流量工程，通过建立到达指定路径的 LSP 隧道进行资源预留，使网络流量绕开拥塞节点，达到平衡网络流量的目的。

在资源紧张的情况下，MPLS TE 能够抢占低优先级 LSP 隧道带宽资源，满足大带宽 LSP 或重要用户的需求。

同时，当 LSP 隧道故障或网络的某一节点发生拥塞时，MPLS TE 可以通过备份路径和 FRR（Fast ReRoute，快速重路由）提供保护。

使用 MPLS TE，网络管理员只需要建立一些 LSP 和旁路拥塞节点，就可以消除网络拥塞。随着 LSP 数量的增长，还可以使用专门的离线工具进行业务量分析。

1.1.2 MPLS TE 的基本概念

1. LSP 隧道

对于一条 LSP，一旦在 Ingress 节点为报文分配了标签，流量的转发就完全由标签决定了。流量对 LSP 的中间节点是透明的，从这个意义上来说，一条 LSP 可以看作是一条 LSP 隧道。

2. MPLS TE 隧道

在部署重路由（Reroute）或需要将流量通过多条路径传输时，可能需要用到多条 LSP 隧道。在 TE 中，这样的一组 LSP 隧道称为 TE 隧道（Traffic Engineered Tunnel）。

1.1.3 MPLS TE 的实现

MPLS TE 主要实现两类功能：

- 静态 CR-LSP（Constraint-based Routed Label Switched Paths，基于约束路由的 LSP）的处理：创建和删除静态 CR-LSP。这些 LSP 有带宽需求，需要通过手工配置。
- 动态 CR-LSP 处理：包括对三种不同类型 CR-LSP 的处理：基本 CR-LSP、备份 CR-LSP 和快速重路由 CR-LSP。

静态 CR-LSP 的处理比较简单。对于动态 CR-LSP，MPLS TE 在实现上主要包括四个部分。

1. 发布含 TE 属性的信息

MPLS TE 需要了解每条链路的动态 TE 相关属性，这可以通过对现有的使用链路状态算法的 IGP 协议进行扩展来实现，比如 OSPF 协议和 IS-IS 协议的扩展。

扩展后的 OSPF 和 IS-IS 协议在链路连接状态中增加了链路带宽、着色等 TE 相关属性，其中，链路的最大可预留带宽和每个优先级的链路未被预留带宽尤为重要。

每台设备收集本区域或本级别所有设备每条链路的 TE 相关信息，生成 TEDB（TE DataBase，流量工程数据库）。

2. 计算路径

使用链路状态算法的路由协议通过 SPF（Shortest Path First，最短路径优先）算法计算出到达网络各个节点的最短路径。

MPLS TE 使用 CSPF（Constraint-based Shortest Path First，基于约束的最短路径优先）算法计算出到达某个节点的最短路径。

CSPF 算法是从 SPF 算法衍生来的，CSPF 有两个输入条件：

- 需要建立的 LSP 的带宽、着色、抢占/保持优先级、显式路径等约束条件，这些都在 LSP 的入口处配置；
- 流量工程数据库 TEDB。

CSPF 的计算过程就是针对 LSP 要求，先对 TEDB 中的链路进行剪切，把不满足 TE 属性要求的链路剪掉；再采用 SPF 算法，寻找一条到 LSP 出口的最短路径。

3. 建立路径

支持建立 LSP 隧道的信令包括 CR-LDP 和 RSVP-TE。它们都能够携带 LSP 的带宽、部分显式路由、着色等约束参数，两者完成的功能是一样的。

从内部实现来看，CR-LDP 通过 TCP 建立 LSP，RSVP-TE 则通过 Raw IP 建立 LSP 连接。

RSVP 技术经历了多年的发展，其体系结构、协议规程与对各种业务的支持机制相对比较成熟；CR-LDP 则是新技术，在可扩展性方面优势明显。

目前设备支持使用 CR-LDP 和 RSVP-TE 两种信令协议。

4. 转发报文

使用建立的隧道转发报文。

1.1.4 CR-LSP

基于一定约束条件建立的 LSP 称为 CR-LSP，与普通 LSP 不同，CR-LSP 的建立不仅依赖路由信息，还需要满足其他一些条件，比如指定的带宽、选定的路径或 QoS 参数。

建立和管理约束条件的机制称为 CR（Constraint-based Routing，基于约束的路由）。

下面对 CR 的主要内容进行简单介绍。

1. 严格显式路由与松散显式路由

- 如果约束信息是对沿途 LSR 的精确指定，建立的 LSP 称为严格的显式路由（Strict Explicit Route）；
- 如果约束信息是对选择下游 LSR 时的模糊限制，建立的 LSP 称为松散的显式路由（Loose Explicit Route）。

2. 流量参数

路径的流量参数有三个：峰值速率（peak rate）和承诺速率（committed rate），描述路径本身对带宽的约束；另外一个是服务粒度（service granularity）。

3. 抢占

如果在建立 CR-LSP 的过程中，无法找到满足所需带宽要求的路径，一种解决方法是拆除另外一条已经建立的路径，占用为它分配的带宽资源，这种处理方式称为抢占（Preemption）。

CR-LSP 使用两个优先级属性来决定是否可以进行抢占：建立优先级（Setup Priority）和保持优先级（Holding Priority）。建立优先级和保持优先级的取值范围都是 0～7，数值越小则优先级越高。

抢占由 RSVP-TE 的 Resv 消息发起。当新建一条路径 Path1 时，如果需要与已建立的路径 Path2 争夺资源，只有当 Path1 的建立优先级高于 Path2 的保持优先级时，Path1 才能抢占成功。

因此，为保证 CR-LSP 能够正确建立，建立优先级不能高于保持优先级，否则可能会导致 LSP 间无穷尽的互相抢占，造成振荡。

4. 路由固定（Route Pinning）

CR-LSP 创建成功后，不随路由变化而变化的特性叫做路由固定。

当某个网络未运行 IGP TE 时，网络管理员不能确定网络上的哪些地方可以获得带宽，这时需要选择具有所需带宽的松散 ER-hop（Explicit Route）来创建 CR-LSP，但这些 CR-LSP 将会随路由变化而变化。当路由变化时，比如出现了一个更好的下一跳，已建立的 CR-LSP 也将会随之改变。

如果不希望使用松散路由建立的 CR-LSP 随路由变化而改变，网络管理员可以在 CR-LSP 创建成功时把这些 CR-LSP 配置成永久性的，不随路由变化而变化。

5. 管理组和亲和属性

MPLS TE 隧道的亲和属性决定隧道使用的链路属性，亲和属性与链路管理组配合，确定隧道可以使用哪些链路。

6. 重优化

流量工程是系统规划网络资源使用的过程。根据用户需求可以配置流量工程，提供要求的 QoS。

服务提供商通常利用一定的机制去优化 CR-LSP，以优化网络资源使用。一种方法是人工配置，但是需要服务提供商进行测量和对 CR-LSP 微调。使用 MPLS TE 则能够动态优化 CR-LSP，从而节省人力。

动态优化 CR-LSP 即定期重计算 CR-LSP 穿越的路由。如果重计算的路由优于当前路由，则创建一条新的 CR-LSP，为之分配新路由，并将业务从旧的 CR-LSP 切换至新的 CR-LSP，删除旧 CR-LSP。

1.1.5 CR-LDP

CR-LDP（Constraint-based Routed Label Distribution Protocol，基于约束路由的 LDP）是对普通 LDP 的扩展，在 MPLS TE 中用于从 Ingress 节点到 Egress 节点建立一条显式路径，并在此路径上进行资源的预留。

所谓 CR-LDP，是指入口节点在发起建立 LSP 时，在标签请求消息中对 LSP 路由附加一定的约束信息。

1.1.6 RSVP-TE

1. RSVP-TE 概述

现在使用两种 QoS 体系：IntServ（Integrated Service，综合业务模型）和 DiffServ（Differentiated Service，区分业务模型）。

RSVP（Resource Reservation Protocol，资源预留协议）是为 IntServ（Integrated Service，综合业务模型）而设计的，用于在一条路径的各节点上进行资源预留。RSVP 工作在传输层，但不参与应用数据的传送，是一种 Internet 上的控制协议，类似于 ICMP。

简单来说，RSVP 具有以下几个主要特点：

- 单向；
- 面向接收者，由接收者发起对资源预留的请求，并维护资源预留信息；
- 使用“软状态”（soft state）机制维护资源预留信息。

RSVP 经扩展后可以支持 MPLS 标签的分发，并在传送标签绑定消息的同时携带资源预留信息，这种扩展后的 RSVP 称为 RSVP-TE，作为一种信令协议用于在 MPLS TE 中建立 LSP 隧道。

2. RSVP-TE 基本概念

(1) 软状态

“软状态”是指在 RSVP-TE 中，通过消息的定时刷新来维持节点上的资源预留状态。

资源预留状态包括由 Path 消息创建的路径状态（path state）和由 Resv 消息创建的预留状态（reservation state）。这两种状态分别由 Path 消息和 Resv 消息定时刷新。对于某个状态，如果连续没有收到刷新消息，这个状态将被删除。

(2) 资源预留类型

使用 RSVP-TE 建立的 LSP 都具有某种资源预留类型（reservation style），在建立 RSVP 会话时，由接收者决定此会话使用哪种预留类型，从而决定可以使用哪些 LSP。

目前设备支持以下两种预留类型：

- FF（Fixed-Filter style）：固定过滤器类型。为每个发送者单独预留资源，不能与同一会话中其他发送者共享资源。
- SE（Shared-Explicit style）：共享显式类型。为同一个会话的发送者建立一个预留，可以共享资源。

说明

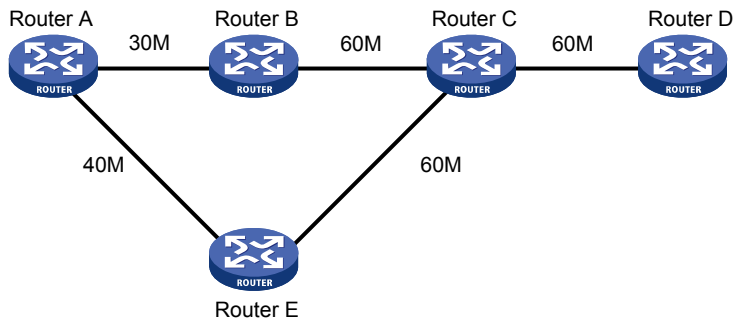
由于目前同一会话不能同时存在多条 LSP，SE 资源预留方式主要用于中断前建立（make-before-break）。

---

3. make-before-break

make-before-break 是指一种可以在尽可能不丢失数据，也不占用额外带宽的前提下改变 MPLS TE 隧道属性的机制。

图1-1 make-before-break 示意图



在 图 1-1中，假设需要建立一条Router A到Router D的路径，保留 30M带宽，开始建立的路径是 Router A→Router B→Router C→Router D。

现在希望将带宽增大为 40M，Router A→Router B→Router C→Router D 路径不能满足要求。而如果选择 Router A→Router E→Router C→Router D，则 Router C→Router D 也存在带宽不够的问题。

采用 make-before-break 机制，新建立的路径在 Router C→Router D 可以共享原路径的带宽，新路径建立成功后，流量转到新路径上，之后拆除原路径。

4. RSVP-TE 消息类型

RSVP-TE 使用 RSVP 的消息类型，并进行了扩展。RSVP 使用以下消息类型：

- Path 消息：由发送者沿数据报文传输的方向向下游发送，在沿途所有节点上保存路径状态（path state）。
- Resv 消息：由接收者沿数据报文传输的方向逆向发送，在沿途所有节点上进行资源预留，并创建和维护预留状态（reservation state）。
- PathTear 消息：此消息产生后马上向下游发送，并立即删除沿途节点的路径状态和相关的预留状态。
- ResvTear 消息：此消息产生后马上向上游发送，并立即删除沿途节点的预留状态。
- PathErr 消息：如果在处理 Path 消息的过程中发生了错误，就会向上游发送 PathErr 消息，PathErr 消息不影响沿途节点的状态，只是把错误报告给发送者。
- ResvErr 消息：如果在处理 Resv 消息的过程中发生了错误，或者由于抢占导致预留被破坏，就会向下游节点发送 ResvErr 消息。
- ResvConf 消息：该消息发往接收者，用于对预留消息进行确认。
- Hello 消息：在两个直连的 RSVP 邻居之间建立和维持链路局部的邻居关系。

RSVP 的 TE 扩展主要是在其 Path 消息和 Resv 消息中增加新的对象，新增对象除了可以携带标签绑定信息外，还可以携带对 LSR 在沿途寻找路径时的限制信息，从而支持 CR-LSP 的功能，并支持 FRR。

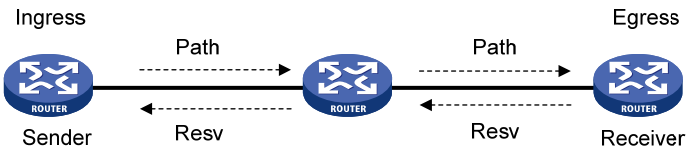
- Path 消息新增的对象包括：LABEL_REQUEST、EXPLICIT_ROUTE、RECORD_ROUTE 和 SESSION_ATTRIBUTE。
- Resv 消息新增的对象包括：LABEL 和 RECORD_ROUTE。

LABEL_REQUEST 对象包含在 Path 消息中，为 LSP 请求标签绑定，该对象也保存在路径状态块 PSB（Path State Block）中。接收到该对象的节点将分配的标签通过 Resv 消息中的 LABEL 对象通知上游节点，从而完成标签的发布和传递。

5. 建立 LSP 隧道

图 1-2是使用RSVP建立LSP隧道的示意图。

图1-2 建立 LSP 隧道



使用 RSVP 建立 LSP 隧道的过程可以简单描述为：

(1) Ingress LSR 产生携带标签请求信息的 Path 消息，沿着通过 CSPF 计算出的路径逐跳发送给 Egress LSR；

(2) Egress LSR 收到 Path 消息后，产生携带预留信息和标签的 Resv 消息，沿着 Path 消息发送的相反路径逐跳返回 Ingress LSR，同时，Resv 消息在沿途的 LSR 上进行资源预留；

(3) 当 Ingress LSR 收到 Resv 消息时，LSP 建立成功。

采用 RSVP-TE 建立的 LSP 具有资源预留功能，沿途的 LSR 可以为该 LSP 分配一定的资源，使在此 LSP 上传送的业务得到保证。

6. RSVP 刷新机制

RSVP 通过 Refresh 消息来维护路径和预留状态，Refresh 消息不仅用于在 RSVP 邻居节点进行状态同步，也用于恢复丢失的 RSVP 消息。

Refresh 消息并不是一种新的消息，它是以前发布过的消息的再次传送，Refresh 消息中携带的主要信息和传送时使用的路径都与它要刷新的消息完全一致。只有 Path 消息和 Resv 消息才可能是 Refresh 消息。

由于 Refresh 消息是定时发送的，当网络中的 RSVP 会话比较多时，Refresh 消息会加重网络负载；而对于时延敏感的应用，当消息丢失时，等待通过 Refresh 消息恢复的时间可能无法接受。简单地调整刷新间隔并不能同时解决这两类问题。

RFC 2961（RSVP Refresh Overhead Reduction Extensions）定义了几种新的扩展机制，用于解决 Refresh 消息带来的上述问题。

(1) Message_ID 扩展

RSVP 本身使用 Raw IP 发送消息，RFC 2961 中定义的 Message_ID 扩展机制增加了可以在 RSVP 消息中携带的对象，其中，Message_ID 和 Message_ID_ACK 对象用于 RSVP 消息确认，从而提高 RSVP 消息发送的可靠性。

在接口使能 Message_ID 机制后，可以配置重传功能，设定 RSVP 消息的重传参数。如果在重传初始时间间隔内（假设为 Rf 秒），没有收到应答消息 ACK，经过（1＋Delta）×Rf 秒后，将重传此消息。

(2) 摘要刷新扩展

摘要刷新 Srefresh（Summary Refresh）可以不传送标准的 Path 或 Resv 消息，而仍能实现对 RSVP 的状态刷新，从而可以减少网络上的 Refresh 消息流量，并加快节点对这类消息的处理速度。

摘要刷新扩展需要与 Message_ID 扩展配合使用。只有那些已经被包含 Message_ID 对象的 Path 和 Resv 消息发布过的状态才能使用摘要刷新扩展机制刷新。

7. PSB、RSB 与 BSB 的超时

为建立 LSP，发送者在 Path 消息中携带 LABEL_REQUEST 对象，接收者收到带有 LABEL_REQUEST 对象的 Path 消息后，就会分配一个标签，并将标签放在 Resv 消息的 LABEL 对象中。

LABEL_REQUEST 对象保存在上游节点的 PSB（Path State Block，路径状态块）中，LABEL 对象则保存在下游节点的 RSB（Reservation State Block，预留状态块）中。当连续未收到刷新消息的次数超过 PSB 或 RSB 的超时倍数（当达到此数值时即为超时）时，PSB 或 RSB 中相应的状态将被删除。

假设有一个资源预留请求，在某些节点上没有通过准入控制，有时可能不希望立即删除这个请求的状态，但这个请求也不应该阻止其他请求使用它预留的资源。这种情况下，节点将进入阻塞状态（Blockade State），在下游节点生成 BSB（Blockade State Block，阻塞状态块）。当连续未收到刷新消息的次数超过阻塞状态超时倍数时，BSB 中相应的状态被删除。

8. RSVP-TE GR

说明

有关 GR（Graceful Restart，平滑重启）的原理介绍请参见“系统分册”中的“GR 概述”。

RSVP-TE GR 功能依赖于 RSVP-TE 的 Hello 扩展能力，通过扩展的 RSVP Hello 报文向邻居通告自己的 GR 能力和相关时间参数。设备和邻居如果都具备 RSVP GR 能力，那么在完成 GR 参数的交互后，就可以在检测到对方发生 GR 重启时，充当对方的 GR Helper，保证在 GR Restarter 重启的过程中，数据转发不会中断。

当 GR Restarter 发生重启时，GR Helper 连续丢失的 Hello 报文次数超过了配置的值，由此判定 GR Restarter 发生了重启。此时 GR Helper 会保留与该邻居相关的软状态信息，并保持向对方周期性发送 Hello 报文，直到重启定时器（Restart Timer）超时。

在重启定时器超时前，如果 GR Helper 邻居和 GR Restarter 重新建立了 Hello 会话协商，那么启动恢复定时器，并触发信令报文交互以恢复原有的软状态；否则，将删除与该邻居相关的所有 RSVP 软状态信息和转发表项。如果恢复定时器超时，则删除那些在 GR 恢复过程中没有恢复的软状态和表项信息。

说明

- 分布式且支持主备倒换的设备在使能 RSVP-TE GR 能力之后，可以同时充当 GR Restarter 和 GR Helper;
- 集中式设备以及分布式但无备用主控板的设备在使能 GR 能力之后，只能充当 GR Helper。

1.1.7 流量转发

当 MPLS TE 隧道建立之后，如果不配置流量沿隧道转发，缺省的情况下依然会沿 IP 路由转发。

配置流量沿隧道转发有如下三种方法：

1. 静态路由

使用静态路由转发流量，是最简便的方法，因为 Tunnel 的接口地址通常情况下不会发布到 IGP 中。这时候通过定义一条通过 Tunnel 接口到达目的网络地址的静态路由，就把流量引入到 MPLS TE 隧道上进行转发。

说明

有关静态路由的介绍请参见“IP 路由分册”中的“静态路由配置”。

2. 策略路由

使用基于策略的路由（Policy-based routing，PBR），通过 Tunnel 接口的流量需要通过 ACL 定义策略，如果匹配该流量，将下一跳的接口指向 Tunnel，在流量的入接口应用策略路由，就把流量引入到 MPLS TE 隧道上进行转发。

说明

有关策略路由的介绍请参见“IP 业务分册”中的“IP 单播策略路由配置”。

3. 自动路由发布

自动路由发布会将 Tunnel 的接口发布到 IGP 路由中，这样流量都会通过 MPLS TE 隧道转发。

自动路由发布包括两种：IGP Shortcut 与转发邻接。

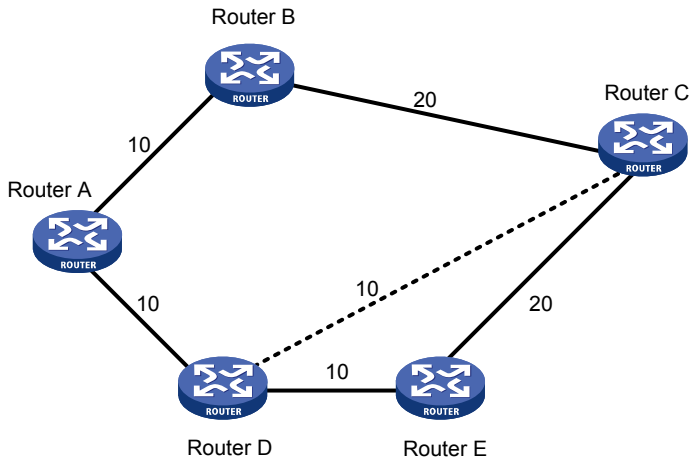
OSPF 和 IS-IS 支持 IGP Shortcut 和转发邻接特性，可以使用 TE Tunnel 作为出接口。在这种应用中，TE Tunnel 被看做点到点链路。

IGP Shortcut 特性也称为自动路由宣告（AutoRoute Announce），该特性将 TE Tunnel 看作直接与目的地址相连的逻辑接口，计算该 TE Tunnel 隧道入口设备的 IGP 路由。

IGP Shortcut 和转发邻接的区别在于：

- 在 IGP Shortcut 应用中，使能此特性的设备使用 TE Tunnel 作为出接口，但它不将这条路由发布给邻居设备，因此，其他设备不能使用此 TE Tunnel。
- 如果配置了转发邻接，则使能此特性的设备在使用 TE Tunnel 作为出接口的同时，也将这条 TE Tunnel 发布给邻居设备，因此，其他设备能够使用此 TE Tunnel。

图1-3 IGP Shortcut 与转发邻接示意图



在 图 1-3中，Router D到Router C之间有一条TE Tunnel，IGP Shortcut只能使入节点Router D在计算IGP路由时利用这条隧道，Router A并不能利用这条隧道到达Router C。如果配置了转发邻接特性，则Router A也能够知道这条TE Tunnel的存在，从而可以利用该隧道将到Router C的流量转发到Router D上。

IGP Shortcut 和转发邻接包括 Tunnel 上的配置和 IGP 本身的配置两部分。

Tunnel 接口上的配置需要注意：

- Tunnel 接口的目的地址应该属于使能相应特性的区域内；
- Tunnel 接口的目的地址可通过区域内路由到达。

1.1.8 自动带宽调整

流量工程要求在环境发生变化时能够动态分配资源，并且不中断业务。

这通常是由于：用户最初不能确定有多少业务需要通过服务提供商的网络传输，他们更愿意为已经使用的带宽付费。因此，服务提供商需要具备这样一种功能：CR-LSP 能在最初时为用户请求带宽建立流量工程隧道；当用户业务增多时，自动调整分配给这些 CR-LSP 的带宽。

MPLS TE 的自动带宽调整特性可以实现此功能，这一特性基于测量的业务量动态调整为流量工程隧道分配的带宽。

1.1.9 CR-LSP 备份

CR-LSP 备份是一种端到端的路径保护（Path Protection，end-to-end protection），对整条 LSP 提供保护，而 FRR 则是一种局部保护措施，只能保护 LSP 中的某条链路和某个节点。并且，FRR 是一种快速响应的临时性保护措施，对于切换时间有严格要求，LSP 备份则没有时间要求。

同一条隧道下对主 LSP 进行路径备份的 LSP 称为备份路径。当 Ingress 感知到主 LSP 不可用时，将流量切换到备份路径上，当主 LSP 路径恢复后再将流量切换回来，以实现对主 LSP 路径的备份保护。

有两种备份方法：

- 热备份：创建主 CR-LSP 后随即创建备份 CR-LSP。主 CR-LSP 失效时，通过 MPLS TE 直接将业务切换至备份 CR-LSP。
- 普通备份：指主 CR-LSP 失效后创建备份 CR-LSP。

1.1.10 快速重路由

1. 快速重路由概述

快速重路由 FRR（Fast ReRoute），是 MPLS TE 中实现网络局部保护的技术。FRR 的切换速度可以达到 50ms，能够最大程度减少网络故障时数据的丢失。

对 LSP 配置 FRR 功能后，当 LSP 上的某条链路或某个节点失效时，流量会被切换到保护链路上，同时 LSP 头节点尝试建立新的 LSP。

2. 基本概念

下面介绍 FRR 中的几个概念：

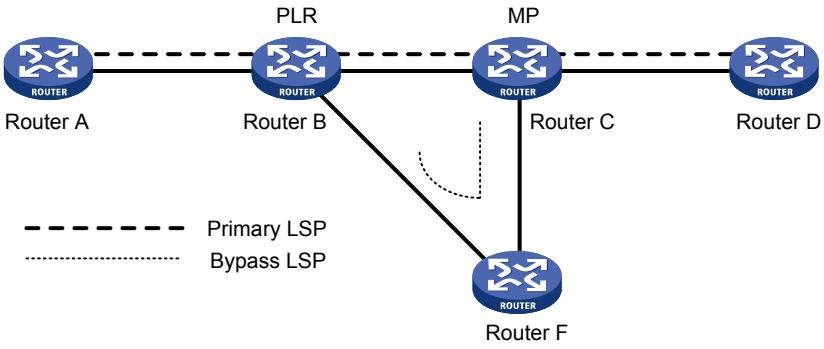
- 主 LSP：被保护的 LSP。
- Bypass LSP：旁路 LSP，保护主 LSP 的 LSP。
- PLR（Point of Local Repair）：本地修复节点。Bypass LSP 的头节点，必须在主 LSP 的路径上，并且不能是主 LSP 的尾节点。
- MP（Merge Point）：汇聚点。Bypass LSP 的尾节点，必须在主 LSP 的路径上，并且不能是主 LSP 的头节点。

3. 保护方式

根据保护的对象不同，FRR 分为两类：

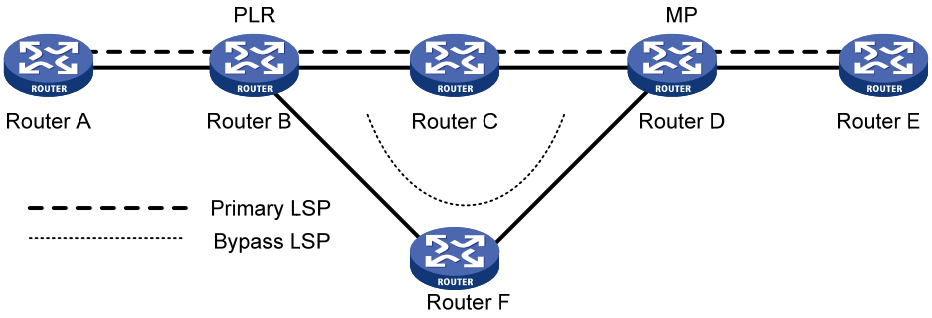
- 链路保护：PLR和MP之间有直接链路连接，主LSP经过这条链路。当这条链路失效时，流量可以切换到Bypass LSP上。如 图 1-4所示，主LSP是Router A→Router B→Router C→Router D，Bypass LSP是Router B→Router F→Router C。

图1-4 FRR 链路保护示意图



- 节点保护：PLR和MP之间通过一台设备连接，主LSP经过这台设备。当这台设备失效时，流量可以切换到Bypass LSP上。如 图 1-5所示，主LSP是Router A→Router B→Router C→Router D→Router E，Bypass LSP是Router B→Router F→Router D，Router C是被保护的设备。

图1-5 FRR 节点保护示意图



4. 部署快速重路由

在配置 Bypass LSP 时，应该规划好它所保护的链路或节点，并确保该 Bypass LSP 不会经过它所保护的链路或节点，否则不能真正起到保护作用。

另外，由于 Bypass 隧道需要预先建立，快速重路由会占用额外的带宽。在网络带宽余量不多的情况下，只能对关键的接口或链路进行快速重路由保护。

1.1.11 DiffServ-Aware TE

Diff-Serv 作为一种 QoS 解决方案，其主要实现机制是对流量按照服务类型（class of service）进行划分，基于服务类型提供不同的 QoS 保证。

而 MPLS TE 作为流量工程解决方案，主要用于对网络资源的使用进行优化。

DiffServ-Aware TE 结合上述两者的优势，能够基于按服务类型划分的流量进行网络资源优化，即，对不同的服务类型进行不同的带宽约束。

概括来说，DiffServ-Aware TE 将不同服务类型的流量与 LSP 进行映射，使流量经过的路径符合对其服务类型的流量工程约束条件。

DiffServ-Aware TE 涉及下面两个概念：

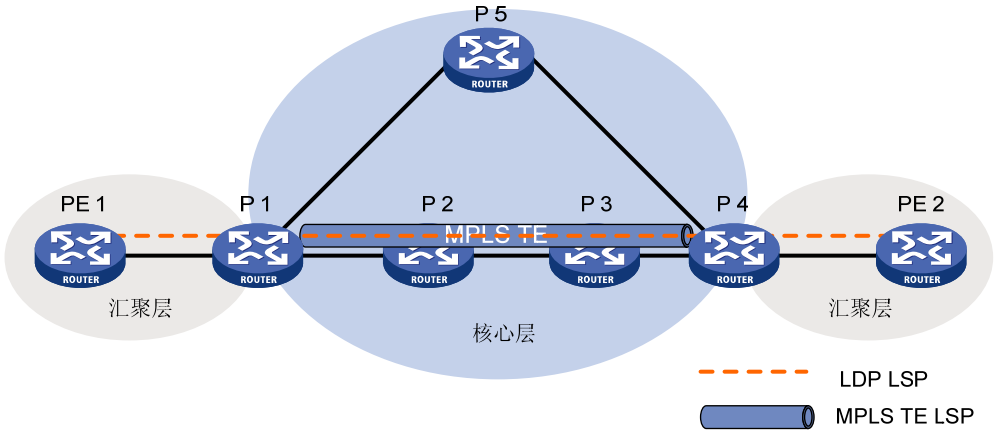
- 服务类型 CT（Class Type）：CT 指的是满足一定带宽约束的链路的集合，用于分配链路带宽、实施约束路由及进行准入控制。对于一个给定的 Traffic Trunk，其经过的链路都属于相同的 CT。
- 带宽约束 BC（Bandwidth Constraints）：为了控制 CT，可以构造不同的带宽约束模型。带宽约束模型由两部分内容决定：最大 BC 数目（MaxBC）、BC 与 CT 的对应关系。

说明

目前设备暂不支持 DiffServ-Aware TE。

1.1.12 MPLS LDP over MPLS TE

图1-6 MPLS LDP over MPLS TE 典型应用



如 图 1-6所示，分层网络中，一般只在核心层部署MPLS TE，汇聚层MPLS网络一般采用LDP作为标签分发信令，不会部署MPLS TE。汇聚层网络上的LDP LSP隧道穿越核心层网络时，可以利用核心层的MPLS TE隧道，即在MPLS TE隧道之上承载LDP LSP，不需要在核心层的每台LSR之间建立LDP LSP，这就是MPLS LDP over MPLS TE。

为了实现 MPLS LDP over MPLS TE，需要在 MPLS TE 隧道的头节点和尾节点之间建立 LDP 远端会话，通过远端会话发送 Label Mapping 消息，从而在 MPLS TE 隧道的头节点和尾节点之间建立 LDP LSP，这条 LDP LSP 隧道承载于 MPLS TE 隧道之上，形成了分层 LSP。

说明

缺省情况下，LDP 不会通过远端会话通告基于地址前缀的标签映射消息，需要通过 **prefix-label advertise** 命令配置 LDP 通过远端会话通告基于地址前缀的标签。**prefix-label advertise** 命令的详细介绍请参见“MPLS 分册”中的“MPLS 基本配置”。

1.1.13 协议规范

与 MPLS TE 相关的协议规范有：

- RFC 2702：Requirements for Traffic Engineering Over MPLS
- RFC 3212：Constraint-Based LSP Setup using LDP
- RFC 2205：Resource ReSerVation Protocol

- RFC 3209：RSVP-TE: Extensions to RSVP for LSP Tunnels
- RFC 2961：RSVP Refresh Overhead Reduction Extensions
- RFC 3564：Requirements for Support of Differentiated Service-aware MPLS Traffic Engineering

1.2 MPLS TE 配置任务简介

表1-1 MPLS TE 配置过程

配置任务		说明	详细配置
配置 MPLS TE 基本能力		必选	1.3
配置 MPLS TE 隧道	使用静态 CR-LSP 配置 MPLS TE 隧道	二者必选其一	1.4
	使用动态信令协议配置 MPLS TE 隧道		1.5
配置 RSVP-TE 高级特性		可选	1.6
调整 CR-LSP 的建立		可选	1.7
调整 MPLS TE 隧道的建立		可选	1.8
配置流量转发	使用静态路由沿 MPLS TE 隧道转发流量	三者必选其一	1.9
	使用策略路由沿 MPLS TE 隧道转发流量		
	使用自动路由发布沿 MPLS TE 隧道转发流量		
配置影响流量转发的参数		可选	1.10
配置自动带宽调整过程		可选	1.11
配置 CR-LSP 备份		可选	1.12
配置 MPLS TE 快速重路由		可选	1.13

1.3 配置 MPLS TE 基本能力

MPLS TE 基本能力的配置任务包括各种 MPLS TE 特性配置中都涉及的基本配置，仅完成本节的配置并不能使用 MPLS TE 特性，还需要根据更具体的要求进行进一步的配置。

1.3.1 配置准备

在配置 MPLS TE 的基本能力之前，需要完成以下任务：

- 配置静态路由或 IGP 协议保证各 LSR 之间可达
- 配置 MPLS 基本能力


说明

有关 MPLS 基本能力的配置请参见“MPLS 分册”中的“MPLS 配置”。

1.3.2 配置 MPLS TE 基本能力

表1-2 配置 MPLS TE 基本能力

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 MPLS TE 能力	**mpls te**	必选 缺省情况下，未使能本节点的 MPLS TE 能力
退回到系统视图	**quit**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能接口的 MPLS TE 能力	**mpls te**	必选 缺省情况下，未使能接口的 MPLS TE 能力
退回系统视图	**quit**	-
创建 Tunnel 接口，并进入 Tunnel 接口视图	**interface tunnel** *tunnel-number*	必选
配置隧道接口的 IP 地址	**ip address** *ip-address netmask*	必选
配置隧道协议为 MPLS TE	**tunnel-protocol mpls te**	必选
配置隧道的目的地址	**destination** *ip-address*	必选
配置隧道的 Tunnel ID	**mpls te tunnel-id** *tunnel-id*	必选
提交隧道当前配置	**mpls te commit**	必选



说明

根据使用的 MPLS TE 信令协议不同，隧道的配置也不同。本小节描述的是 MPLS TE 隧道基本配置，如果完成这些配置后隧道还不能工作，需要根据使用的信令协议做进一步的配置。

1.4 使用静态 CR-LSP 配置 MPLS TE 隧道

使用静态 CR-LSP 建立 MPLS TE 隧道的过程非常简单，不需要配置隧道的约束条件，也不涉及 IGP TE 扩展或 CSPF。只需要建立一条静态 CR-LSP 和使用静态信令的 TE 隧道，并将二者关联起来即可。

静态 CR-LSP 建立的 TE 隧道不能配置带宽约束，也不能根据网络的变化动态调整，因此实际应用非常有限。

静态 CR-LSP 作为一种特殊的静态 LSP，其配置限制条件与静态 LSP 相同，使用相同的标签空间。

1.4.1 配置准备

在使用静态 CR-LSP 建立 MPLS TE 隧道之前，需完成以下任务：

- 配置静态路由或 IGP 协议保证各 LSR 之间可达
- 配置 MPLS 基本能力

- 配置 MPLS TE 基本能力

1.4.2 使用静态 CR-LSP 配置 MPLS TE 隧道

表1-3 使用静态 CR-LSP 配置 MPLS TE 隧道

操作		命令	说明
进入系统视图		**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图		**interface tunnel** *tunnel-number*	-
配置隧道使用静态 CR-LSP		**mpls te signal-protocol static**	必选
提交隧道当前配置		**mpls te commit**	必选
退回到系统视图		**quit**	-
配置指定静态 CR-LSP 的节点时根据本节点在网络中的位置	为 Ingress 节点配置静态 CR-LSP	**static-cr-lsp ingress** *tunnel-name* **destination** *dest-addr* { **nexthop** *next-hop-addr* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label-value* [**bandwidth** [**bc0** | **bc1**] *bandwidth-value*]	必选 三者选择其一
	为 Transit 节点配置静态 CR-LSP	**static-cr-lsp transit** *tunnel-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label-value* { **nexthop** *next-hop-addr* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label-value* [**bandwidth** [**bc0** | **bc1**] *bandwidth-value*]	
	为 Egress 节点配置静态 CR-LSP	**static-cr-lsp egress** *tunnel-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label-value* [**lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id*]	

说明

- 参数 *tunnel-name* 是使用此静态 CR-LSP 的 MPLS TE 隧道的名称。
- **static-cr-lsp ingress** 命令的参数 *tunnel-name* 大小写敏感。假设使用命令 **interface tunnel** 2 创建了一个 Tunnel 接口，则其静态 CR-LSP 入节点中的参数应该写作“Tunnel2”，否则隧道将不能正确建立。中间节点和出节点无此限制。
- 配置 ingress 和 Transit 时，下一跳地址不能为本地的公网地址。

1.5 使用动态信令协议配置 MPLS TE 隧道

动态信令协议可以根据网络的动态变化调整 TE 隧道的路径，并能够应用备份、快速重路由等高级特性。

使用动态信令协议建立 MPLS TE 隧道的过程可以简要描述为以下几个部分：

- 配置链路的 MPLS TE 属性，并通过 IGP 的 TE 扩展发布，形成 TEDB；
- 配置隧道的约束条件；
- 使用 CSPF 算法，通过 TEDB 和隧道约束条件计算出符合要求的路径；
- 利用动态信令协议 RSVP-TE 或 CR-LDP 建立路径。

说明

如果不配置 IGP 的 TE 扩展，网络中就不会有 TE LSA，不能形成 TEDB。这种情况下生成的 CR-LSP 是由 IGP 路由得到的，而不是 CSPF 计算出来的。

1.5.1 配置准备

在使用动态信令协议建立 MPLS TE 隧道之前，需完成以下任务：

- 配置静态路由或 IGP 协议保证各 LSR 之间可达
- 配置 MPLS 基本能力
- 配置 MPLS TE 基本能力

1.5.2 使用动态信令协议配置 MPLS TE 隧道

1. 使用动态信令协议配置 MPLS TE 隧道任务简介

表1-4 使用动态信令协议配置 MPLS TE 隧道过程

配置任务	说明	详细配置
配置链路的 MPLS TE 属性	可选	2.
配置 CSPF	可选	3.
配置 OSPF TE	配置 CSPF 后，二者必选其一 根据 IGP 协议的类型选择	4.
配置 IS-IS TE		5.
配置 MPLS TE 显示路径	可选	6.
配置 MPLS TE 隧道的约束条件	可选	7.
使用 CR-LDP 建立 MPLS TE 隧道	二者可选其一 缺省情况下，使用 RSVP-TE 建立 MPLS TE 隧道	8.
使用 RSVP-TE 建立 MPLS TE 隧道		9.

2. 配置链路的 MPLS TE 属性

链路的 TE 属性将通过 IGP 的 TE 扩展发布到网络中，形成 TEDB。

表1-5 配置链路的 MPLS TE 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
配置链路最大带宽	**mpls te max-link-bandwidth** *bandwidth-value* [**bc1** *bc1-bandwidth*]	可选
配置 MPLS TE 链路最大可预留带宽	**mpls te max-reservable-bandwidth** *bandwidth-value* [**bc1** *bc1-bandwidth*]	可选

3. 配置 CSPF

表1-6 配置 CSPF

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 CSPF 功能	**mpls te cspf**	必选 缺省情况下，未使能本节点的 CSPF 功能

4. 配置 OSPF TE

当路由协议使用 OSPF 并使用动态信令协议配置 MPLS TE 隧道时，需要配置 OSPF TE 功能。

OSPF TE 扩展使用 Opaque Type 10 LSA 携带链路的 TE 属性信息，因此，配置 OSPF TE 时必须先使能 OSPF 的 Opaque 能力。并且，只有当至少有一个邻居处于 FULL 状态时，才会产生 TE LSA。

表1-7 配置 OSPF TE

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 协议视图	**ospf** [*process-id*]	-
使能 OSPF 的 Opaque LSA 能力	**opaque-capability enable**	必选 缺省情况下，未使能 OSPF 的 Opaque LSA 能力
进入 OSPF 的区域视图	**area** *area-id*	必选
在当前 OSPF 区域下使能 MPLS TE 能力	**mpls-te enable**	必选 缺省情况下，OSPF 区域不支持 MPLS TE 能力
退回 OSPF 协议视图	**quit**	-



说明

- 有关 OSPF Opaque 能力的介绍请参见“IP 路由分册”中的“OSPF 配置”。
- 由于 MPLS TE 无法在 OSPF 虚连接上预留资源和分配标签，即 MPLS TE 无法通过 OSPF 虚连接建立 LSP 隧道，因此，配置 OSPF TE 时，OSPF 路由域内不能存在虚连接。

5. 配置 IS-IS TE

当路由协议使用 IS-IS 并使用动态信令协议配置 MPLS TE 隧道时，需要配置 IS-IS TE 功能。当 OSPF TE 和 IS-IS TE 共存时，优选 OSPF 计算。

IS-IS TE 扩展使用 IS 可达性 TLV（类型为 22）的子 TLV 携带 TE 属性信息。因此，配置 IS-IS TE 时必须先使能 IS-IS 的 WideMetric 特性，可以配置 wide、compatible 或 wide-compatible。

注意

- RFC 3784 规定，在某些情况下 IS 可达性 TLV（22）的长度将会达到其最大长度 255 个字节。
- 为了使 IS-IS LSP 能顺利携带此类 TLV 并且在使能 IS-IS 的所有接口上进行正确泛洪，要求使能 IS-IS 的所有接口的 MTU 不小于 284 字节（其中包含 LSP 头长度 27 字节以及 TLV 头长度 2 字节）。如果 LSP 同时需要携带认证信息，则最小 MTU 需要根据报文结构另行推算。

综合考虑以上因素，在使用 TE 特性的情况下使能 IS-IS 的所有接口的 MTU 不要小于 512 字节，以确保 IS-IS LSP 能在网络上正确泛洪。

表1-8 配置 IS-IS TE

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 协议视图	**isis** [*process-id*]	-
配置 IS-IS 的 Wide Metric 属性	**cost-style** { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }	必选 缺省情况下，IS-IS 只收发采用 Narrow 方式表示路由权值的报文
使能 IS-IS TE 功能	**traffic-eng** [**level-1** | **level-1-2** | **level-2**]	必选 缺省情况下，未使能 IS-IS TE 功能
配置携带 DS-TE 参数的子 TLV 的 TLV 类型	**te-set-subtlv** { **bw-constraint** *value* | **lo-multiplier** *value* | **unreserved-bw-sub-pool** *value* }	可选 缺省情况下，**bw-constraint**（带宽约束）的子 TLV 为 252；**lo-multiplier**（本地过预订倍数）的子 TLV 为 253；**unreserved-bw-sub-pool**（子池未预订带宽）的子 TLV 为 251

说明

- 有关 IS-IS 的介绍请参见“IP 路由分册”中的“IS-IS 配置”。
- IS-IS TE 不支持发布从 IP 地址功能。在配置了多个 IP 地址的接口上使能 IS-IS TE 后，IS-IS TE 在其 LSP 的 IS 可达性 TLV（22）中只发布该接口的主 IP 地址。建议避免在配置了从 IP 地址的接口上配置 IS-IS TE 特性。

6. 配置 MPLS TE 显式路径

显式路径由一系列节点构成，一条显式路径上的两个相邻节点之间存在两种关系：

- 严格下一跳（**strict**）：两个节点必须直接相连；
- 松散下一跳（**loose**）：两个节点之间可以存在其他设备。

在向显式路径中增加或修改节点时，参数 **include** 表示建立的 LSP 必须经过指定节点；参数 **exclude** 表示建立的 LSP 不能经过指定节点。

表1-9 配置 MPLS TE 显式路径

操作	命令	说明
进入系统视图	**system-view**	-
创建隧道的显式路径，进入显式路径视图	**explicit-path** *path-name* [**disable** | **enable**]	必选
向显式路径中插入一个节点	**add hop** *ip-address1* [**include** [**loose** | **strict**] | **exclude**] { **after** | **before** } *ip-address2*	可选 缺省情况下，显式路径包含指定的节点（**include**），加入的节点是严格下一跳（**strict**）
在显式路径中指定下一跳 IP 地址	**next hop** *ip-address* [**include** [**loose** | **strict**] | **exclude**]	必选 缺省情况下，显式路径中指定的下一跳为严格（**strict**）下一跳
修改显式路径中的节点 IP 地址	**modify hop** *ip-address1 ip-address2* [[**include** [**loose** | **strict**]] | **exclude**]	可选 缺省情况下，显式路径包含指定的节点（**include**），修改的节点是严格下一跳（**strict**）
在显式路径中删除一个节点	**delete hop** *ip-address*	可选
显示显式路径节点信息	**list hop** [*ip-address*]	可选

7. 配置 MPLS TE 隧道的约束条件

表1-10 配置 MPLS TE 隧道的约束条件

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置隧道使用的带宽	**mpls te bandwidth** [**bc0** | **bc1**] *bandwidth*	可选 缺省情况下，未配置 MPLS TE 隧道所需的带宽，即带宽为 0
配置隧道应用的路径及路径的优先级	**mpls te path** { **dynamic** | **explicit-path** *pathname* } **preference** *value*	可选 缺省情况下，隧道应用动态自动计算的路径
提交隧道当前配置	**mpls te commit**	必选

8. 使用 CR-LDP 建立 MPLS TE 隧道

表1-11 使用 CR-LDP 建立 MPLS TE 隧道

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-

操作	命令	说明
配置 MPLS TE 建立隧道使用的信令协议为 CR-LDP	**mpls te signal-protocol crldp**	必选 缺省情况下，配置 MPLS TE 建立隧道使用 RSVP-TE 信令协议
提交隧道当前配置	**mpls te commit**	必选

9. 使用 RSVP-TE 建立 MPLS TE 隧道

表1-12 使用 RSVP-TE 建立 MPLS TE 隧道

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 RSVP-TE 能力	**mpls rsvp-te**	必选 缺省情况下，未使能的本节点 RSVP-TE 能力
退回系统视图	**quit**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能接口的 RSVP-TE 能力	**mpls rsvp-te**	必选 缺省情况下，未使能接口的 RSVP-TE 能力
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置隧道使用 RSVP-TE 作为信令协议	**mpls te signal-protocol rsvp-te**	可选 缺省情况下，隧道使用 RSVP-TE
提交隧道当前配置	**mpls te commit**	必选

注意

如果使用 RSVP-TE 作为 MPLS TE 隧道的信令协议，那么必须在隧道经过的路径上，使能接口的 MPLS TE 能力的同时使能接口的 RSVP-TE 能力。

1.6 配置 RSVP-TE 高级特性

RSVP-TE 主要是在 Path 消息和 Resv 消息中增加了新的对象，以支持建立 CR-LSP。

RSVP-TE 提供了丰富的配置选项，能够满足用户对于可靠性、网络资源、以及一些 MPLS TE 高级特性的应用。

在实施本节介绍的配置任务之前，需要仔细了解各配置任务的目的和可能对网络造成的影响。

1.6.1 配置准备

在配 RSVP-TE 高级特性之前，需完成以下任务：

- 配置 MPLS 基本能力
- 配置 MPLS TE 基本能力

- 使用 RSVP-TE 建立 MPLS TE 隧道

1.6.2 配置 RSVP-TE 高级特性

1. 配置 RSVP 资源预留风格

表1-13 配置 RSVP 资源预留风格

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置隧道的资源预留风格	**mpls te resv-style { ff | se }**	可选 缺省情况下，资源预留风格为 **se**
提交隧道当前配置	**mpls te commit**	必选



说明

在目前的 MPLS TE 应用中，**se** 方式主要用于中断前建立（make-before-break），而 **ff** 方式则很少使用。

2. 配置 RSVP 的状态定时器

表1-14 配置 RSVP 的状态定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置节点的路径/预留消息的刷新时间	**mpls rsvp-te timer refresh** *timevalue*	可选 缺省情况下，节点的路径/预留消息刷新周期为 30 秒
配置 PSB 和 RSB 的超时倍数	**mpls rsvp-te keep-multiplier** *number*	可选 缺省情况下，PSB 和 RSB 的超时倍数为 3
配置阻塞状态的超时倍数	**mpls rsvp-te blockade-multiplier** *number*	可选 缺省情况下，阻塞状态的超时倍数为 4

3. 配置 RSVP 刷新机制

Message_ID 扩展机制（可靠性机制）用于 RSVP 消息确认，提高 RSVP 消息发送的可靠性。

在接口上使能可靠性功能后，可以使能重传功能。

摘要刷新 Srefresh（Summary Refresh）需要与 Message_ID 扩展配合使用。只有那些已经被包含 Message_ID 对象的 Path 和 Resv 消息发布过的状态才能使用摘要刷新扩展机制刷新。

表1-15 配置 RSVP 刷新机制

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能可靠性功能	**mpls rsvp-te reliability**	可选
使能重传功能	**mpls rsvp-te timer retransmission** { **increment-value** [*increment-value*] | **retransmit-value** [*retrans-timer-value*] } *	可选 缺省情况下，未使能重传功能
使能摘要刷新功能	**mpls rsvp-te srefresh**	可选 缺省情况下，未使能摘要刷新功能

4. 配置 RSVP 的 Hello 扩展

表1-16 配置 RSVP 的 Hello 扩展

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能本节点的 RSVP 的 Hello 扩展
配置允许 Hello 消息丢失的最大次数	**mpls rsvp-te hello-lost** *times*	可选 缺省情况下，最多允许连续收不到 3 个 Hello 报文
配置 Hello 消息发送时间间隔	**mpls rsvp-te timer hello** *timevalue*	可选 缺省情况下，Hello 消息的发送时间间隔为 3 秒
退回系统视图	**quit**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能接口的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能接口 RSVP 的 Hello 扩展



说明

有关 RSVP 的 Hello 扩展机制用于检测 RSVP 邻居节点的可达性，在 RFC 3209 中有详细的介绍。

5. 配置 RSVP-TE 预留确认

表1-17 配置 RSVP-TE 预留确认

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-

操作	命令	说明
使能节点的预留确认机制	**mpls rsvp-te resvconfirm**	必选 缺省情况下，节点上不使能预留确认机制

说明

- 预留确认机制由接收者发起，接收者在发送 Resv 消息时，携带一个要求对预留进行确认的对象。
- 收到 ResvConf 消息并不表示成功建立了资源预留，只能说明在此 Resv 消息到达的最远上游节点处成功保留了资源，这些资源接下来仍可能被其它应用抢占。

6. 配置 RSVP 验证

为防止伪造的资源预留请求非法占用网络资源，RSVP 采用逐跳验证机制。

一条链路两端的接口需要配置相同的验证密钥，只有这样，它们才可以交换 RSVP 消息。

表1-18 配置 RSVP 验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能验证功能	**mpls rsvp-te authentication** { **cipher** | **plain** } *auth-key*	必选

说明

快速重路由特性和 RSVP 验证不能同时运行。

7. 配置 RSVP-TE GR

RSVP-TE GR 功能依赖于 RSVP-TE 的 Hello 扩展能力，因此在配置 RSVP-TE GR 功能之前必须使能 RSVP-TE 的 Hello 扩展能力。

分别在作为 GR Restarter 和 GR Helper 的设备上进行以下配置。

表1-19 配置 RSVP-TE GR

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能本节点的 RSVP 的 Hello 扩展
使能 MPLS RSVP-TE 协议的 GR 能力	**mpls rsvp-te graceful-restart**	必选 缺省情况下，MPLS RSVP-TE 协议的 GR 能力处于关闭状态

操作	命令	说明
配置 RSVP-TE GR 的重启定时器	**mpls rsvp-te timer graceful-restart restart** *restart-time*	可选 缺省情况下，RSVP-TE GR 重启定时器的值为 120 秒
配置 RSVP-TE GR 的恢复定时器	**mpls rsvp-te timer graceful-restart recovery** *recovery-time*	可选 缺省情况下，RSVP-TE GR 恢复定时器的值为 300 秒
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
使能接口的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能接口 RSVP 的 Hello 扩展

1.7 调整 CR-LSP 的建立

CSPF 使用 TEDB 和约束条件计算出符合要求的路径，并通过信令协议建立 CR-LSP。MPLS TE 提供多种方式影响 CSPF 的计算，从而调整 CR-LSP 的建立。

1.7.1 配置准备

本节的配置需要与 CSPF、动态信令协议（CR-LDP 或 RSVP-TE）配合使用。

本节介绍的配置任务是 MPLS TE 中 CSPF 的一些专门配置，在实施本节的配置任务之前，需要明确理解这些配置对系统可能造成的影响。

1.7.2 调整 CR-LSP 的建立

1. 配置 CSPF 的仲裁方法

CSPF 只计算到达隧道终点的最短路径，当在计算路径的过程中遇到多条权值相同的路径时，只需要选择其中一条就可以了。

可供选择的仲裁（tie-breaking）方法包括优先选择带宽使用率（已用带宽/链路最大可预留带宽）最小的路径、带宽使用率（已用带宽/链路最大可预留带宽）最大的路径、或随机选择路径。

表1-20 配置 CSPF 的仲裁方法

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置本节点的 CR-LSP 仲裁方法	**mpls te tie-breaking { least-fill | most-fill | random }**	可选 缺省情况下，采用随机选择方式（**random**）
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置当前隧道的 CR-LSP 仲裁方法	**mpls te tie-breaking { least-fill | most-fill | random }**	可选 缺省情况下，采用随机选择方式（random）
提交隧道当前配置	**mpls te commit**	必选

说明

隧道优先使用本隧道视图下配置的仲裁方式，如果隧道视图下没有配置，将使用 MPLS 视图下的配置。

2. 配置路由固定

如果使用路由固定特性，则不能同时使用 MPLS TE 重优化或自动带宽调整。

表1-21 配置路由固定

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
使能路由固定功能	**mpls te route-pinning**	必选 缺省情况下，未使能路由固定功能
提交隧道当前配置	**mpls te commit**	必选

3. 配置管理组和亲和属性

MPLS TE 隧道的亲和属性决定隧道使用的链路属性，亲和属性与链路管理组配合，确定隧道可以使用哪些链路。管理组、亲和属性、掩码都是 32 位数值。

如果希望某条链路能够被隧道所用，则在所有掩码为 1 的位中，管理组中至少有 1 位与亲和属性中的相应位都为 1。对于掩码为 0 的位，则不对管理组的相应位进行检查。

举例说明：假设亲和属性为 0xFFFFFFFF，掩码为 0x0000FFFF，则可用链路的管理组属性高 16 位可以任意取 0 或 1，低 16 位则至少有 1 位为 1。

在隧道首节点配置隧道亲和属性后，亲和属性将通过信令协议（CR-LDP 或 RSVP-TE）传递给各节点。

说明

不同设备制造商实现的管理组和亲和属性的关系可能有所不同，当在同一网络中使用不同设备制造商的设备时，需要事先了解各自的实现方式。

表1-22 配置管理组和亲和属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
配置链路的管理组	**mpls te link administrative group** *value*	可选 缺省情况下，链路管理组属性值为 0x00000000
退回系统视图	**quit**	-

操作	命令	说明
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置隧道的亲和属性	**mpls te affinity property** *properties* [**mask** *mask-value*]	可选 缺省情况下，隧道的亲和属性为 0x00000000，掩码为 0x00000000
提交隧道当前配置	**mpls te commit**	必选

4. 配置 CR-LSP 重优化

动态优化 CR-LSP 即定期重计算 CR-LSP 穿越的路径。如果重计算的路径优于当前路径，则创建一条新的 CR-LSP，为之分配新路径，并将业务从旧的 CR-LSP 切换至新的 CR-LSP，删除旧 CR-LSP。

表1-23 配置 CR-LSP 重优化

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
使能隧道重优化功能	**mpls te reoptimization** [**frequency** *seconds*]	必选 缺省情况下，未使能重优化功能
提交隧道当前配置	**mpls te commit**	必选
退到用户视图	**return**	-
立即对所有使能了重优化功能的 TE 隧道进行重优化	**mpls te reoptimization**	可选

1.8 调整 MPLS TE 隧道的建立

在建立 MPLS TE 隧道的过程中，根据实际应用，可能还需要进行一些专门的配置。本节介绍这些特殊配置的使用。

1.8.1 配置准备

本节介绍的任务是 MPLS TE 中的一些专门配置，在实施本节的配置任务之前，需要明确理解这些配置对系统可能造成的影响。

本节介绍的配置需要与动态信令协议（CR-LDP 或 RSVP-TE）配合使用。

1.8.2 调整 MPLS TE 隧道的建立

1. 配置环路检测

表1-24 配置环路检测

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-

操作	命令	说明
配置隧道建立时进行环路检测	**mpls te loop-detection**	必选 缺省情况下，不进行环路检测
提交隧道当前配置	**mpls te commit**	必选

2. 配置记录路由和标签

表1-25 配置记录路由和标签

操作		命令	说明
进入系统视图		**system-view**	-
进入MPLS TE隧道的Tunnel接口视图		**interface tunnel** *tunnel-number*	-
配置隧道支持路由记录或标签记录	建立隧道时记录路由	**mpls te record-route**	二者必选其一 缺省情况下，不进行路由记录和标签记录
	建立隧道时记录路由和标签	**mpls te record-route label**	
提交隧道当前配置		**mpls te commit**	必选

说明

当MPLS TE隧道的动态协议类型为CR-LDP时，不支持命令**mpls te record-route label**。

3. 配置隧道重建

配置隧道重建时间隔和重建次数后，可以在隧道建立不成功，会每隔一定时间自动尝试重建，直到达到允许的最大重建次数。

表1-26 配置隧道重建

操作	命令	说明
进入系统视图	**system-view**	-
进入MPLS TE隧道的Tunnel接口视图	**interface tunnel** *tunnel-number*	-
配置隧道重建的次数	**mpls te retry** *times*	可选 缺省情况下，隧道重建的次数为5次
配置隧道重建的间隔时间	**mpls te timer retry** *seconds*	可选 缺省情况下，重建隧道的时间间隔为2秒
提交隧道当前配置	**mpls te commit**	必选

4. 配置隧道的优先级

表1-27 配置隧道的优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置隧道的优先级	**mpls te priority** *setup-priority* [*hold-priority*]	可选 缺省情况下，建立优先级和保持优先级的值都为 7
提交隧道当前配置	**mpls te commit**	必选

1.9 配置流量转发

1.9.1 配置准备

在配置流量转发之前，需完成以下任务：

- 配置 MPLS 基本能力
- 配置 MPLS TE 基本能力
- 配置 MPLS TE 隧道

1.9.2 配置流量转发

1. 使用静态路由沿 MPLS TE 隧道转发流量

表1-28 使用静态路由沿 MPLS TE 隧道转发流量

操作	命令	说明
进入系统视图	**system-view**	-
使用静态路由沿 MPLS TE 隧道转发流量	**ip route-static** *dest-address* { *mask* | *mask-length* } *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	必选



说明

- *interface-type* 为 Tunnel 接口，并且必须设置 **preference** 值。
- 有关配置静态路由的命令请参见“IP 路由分册”中的“静态路由命令”。

2. 使用策略路由沿 MPLS TE 隧道转发流量

表1-29 使用策略路由沿 MPLS TE 隧道转发流量

操作	命令	说明
进入系统视图	**system-view**	-
创建并进入高级 IPv4 ACL 视图	**acl number** *acl-number* [**match-order** { **auto** | **config** }]	必选 缺省情况下，匹配顺序为 **config**
定义规则	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [**destination** { *dest-addr dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **established** | **fragment** | **icmp-type** { *icmp-type icmp-code* | *icmp-message* } | **logging** | **precedence** *precedence* | **reflective** | **source** { *sour-addr sour-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-name* | **tos** *tos* | **vpn-instance** *vpn-instance-name*] *	必选
退回系统视图	**quit**	-
创建策略或一个策略节点，并进入该策略路由视图	**policy-based-route** *policy-name* { **permit** | **deny** } **node** *node-number*	必选
设置 IP 地址匹配条件	**if-match acl** *acl-number*	必选
设置报文的发送接口为 Tunnel 接口	**apply output-interface tunnel** *number1* [**tunnel** *number2*]	必选
退回系统视图	**quit**	-
使能本地策略路由	**ip local policy-based-route** *policy-name*	必选

说明

有关配置策略路由的命令请参见“安全分册”中的“ACL 命令”和“IP 业务分册”中的“IP 单播策略路由命令”。

3. 使用自动路由发布沿 MPLS TE 隧道转发流量

配置自动路由发布前，需要首先在 MPLS TE 隧道的 Tunnel 接口使能 OSPF 或 ISIS 路由协议。

在 IGP Shortcut 和转发邻接特性中计算路径时，可以指定 TE 隧道使用的度量值。如果使用绝对度量（**absolute**），则 TE 隧道的度量值就是配置的值；如果使用相对度量（**relative**），则 TE 隧道的度量值是相应 IGP 路径度量值加上相对度量值。

(1) 配置 IGP Shortcut

表1-30 配置 IGP Shortcut

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-

操作	命令	说明
配置 IGP 在进行增强的 SPF 计算时，使用处于 up 状态的 MPLS TE 隧道	**mpls te igp shortcut** [**isis** | **ospf**]	必选 缺省情况下，IGP 在进行增强的 SPF 计算时不使用 MPLS TE 隧道 如果配置 IGP Shortcut 时不指定 IGP 类型，则缺省为 **ospf** 和 **isis** 都支持
配置 MPLS TE 隧道的度量值	**mpls te igp metric** { **absolute** *value* | **relative** *value* }	可选 缺省情况下，TE 隧道的度量等于其 IGP 度量
提交隧道当前配置	**mpls te commit**	必选
退回系统视图	**quit**	-
进入 OSPF 协议视图	**ospf** [*process-id*]	
使能 IGP Shortcut 特性	**enable traffic-adjustment**	必选 缺省情况下，不使能 IGP Shortcut 特性

(2) 配置转发邻接

要想使转发邻接功能生效，需要创建双向 MPLS TE 隧道，并在隧道的两端同时配置转发邻接功能。

表1-31 配置转发邻接

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
将 MPLS TE 隧道作为链路发布到 IGP 网络	**mpls te igp advertise** [**hold-time** *value*]	必选 缺省情况下，MPLS TE 隧道不被作为链路发布到 IGP 网络
配置 MPLS TE 隧道的度量值	**mpls te igp metric** { **absolute** | **relative** } *value*	可选 缺省情况下，TE 隧道的度量等于其 IGP 度量
提交隧道当前配置	**mpls te commit**	必选
退回系统视图	**quit**	-
进入 OSPF 协议视图	**ospf** [*process-id*]	-
使能转发邻接特性	**enable traffic-adjustment advertise**	必选 缺省情况下，未使能转发邻接特性



说明

使用自动路由发布特性时，必须将 TE 隧道的目的地址指定为对端节点的 LSR ID。

1.10 配置影响流量转发的参数

在 MPLS TE 中，影响流量转发是指可能导致 IP 流量或 MPLS 流量经过的路径发生变化，限制 TE 隧道流量类型的相关配置。本节介绍 MPLS TE 中影响流量转发的几种主要措施。

1.10.1 配置准备

本节介绍的配置需要与 CSPF、动态信令协议（CR-LDP 或 RSVP-TE）配合使用。

1.10.2 配置影响流量转发的参数

1. 配置失效链路定时器

一旦一条链路状态变为 down，失效链路定时器就会启动。如果 IGP 在定时器超时之前删除或修改此链路，IGP 将会把删除或修改情况通知 CSPF。CSPF 在 TEDB 中更新链路，并停止定时器。如果定时器超时时，IGP 还没有删除链路，链路状态将被更新为 up。

表1-32 配置失效链路定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置失效链路定时器值	**mpls te cspf timer failed-link** *timer-interval*	可选 缺省情况下，失效链路定时器的值为 10 秒

2. 配置泛洪阈值

当 MPLS TE 相关链路的带宽发生变化时，CSPF 可能需要重新计算路径。

为防止链路带宽变化导致的 CSPF 计算占用过多资源，可以规定当带宽变化到达一定限度时才进行泛洪。

用户可以进行两种配置：

- 当链路的可用带宽超过上限阈值时泛洪；
- 当链路的可预留带宽低于下限阈值时泛洪。

表1-33 配置泛洪阈值

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
配置带宽变化 IGP 泛洪阈值	**mpls te bandwidth change thresholds** { **down** | **up** } *percent*	可选 缺省情况下，带宽变化 IGP 泛洪阈值为 10
提交隧道当前配置	**mpls te commit**	必选

3. 配置选路使用的度量

表1-34 配置选路使用的度量

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置未配置度量类型的隧道选路时使用的链路度量类型	**mpls te path metric-type** { **igp** | **te** }	可选 缺省情况下，选路时使用链路的 TE 度量

操作	命令	说明
退回系统视图	**quit**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置当前隧道选路时使用的度量类型	**mpls te path metric-type** { **igp** | **te** }	可选 缺省情况下，选路时使用链路的 TE 度量
提交隧道当前配置	**mpls te commit**	必选
退回系统视图	**quit**	-
进入 MPLS TE 链路的接口视图	**interface** *interface-type interface-number*	-
配置链路的 TE 度量	**mpls te metric** *value*	可选 缺省情况下，链路使用其 IGP 度量作为 TE 的度量

说明

如果在隧道视图下没有配置 **mpls te path metric-type**，则使用 MPLS 视图下的配置。

4. 配置隧道转发的流量类型

表1-35 配置隧道转发的流量类型

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置 TE 隧道转发的流量类型	**mpls te vpn-binding** { **acl** *acl-number* | **vpn-instance** *vpn-instance-name* }	可选 缺省情况下，TE 隧道不限制可以转发的流量类型
提交隧道当前配置	**mpls te commit**	必选

1.11 配置自动带宽调整过程

1.11.1 配置准备

本节介绍的配置需要与 CSPF、动态信令协议（CR-LDP 或 RSVP-TE）配合使用。

1.11.2 配置自动带宽调整过程

表1-36 配置自动带宽调整

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-

操作	命令	说明
使能自动带宽调整，并对为自动带宽调整配置的隧道进行出口速率采样	**mpls te timer auto-bandwidth** [*seconds*]	必选 缺省情况下，未使能带宽自动调整
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
使能隧道的自动带宽调整功能	**mpls te auto-bandwidth adjustment** [**frequency** *seconds*] [**max-bw** *max-bandwidth* | **min-bw** *min-bandwidth*] *	必选 缺省情况下，未使能隧道的自动带宽调整功能
使能隧道收集出口带宽的频率功能	**mpls te auto-bandwidth collect-bw** [**frequency** *seconds*] [**max-bw** *max-bandwidth* | **min-bw** *min-bandwidth*]*	必选 缺省情况下，未使能隧道收集出口带宽的频率功能
提交隧道当前配置	**mpls te commit**	必选

说明

- 采样周期在 MPLS 视图下配置，对所有 MPLS TE 隧道有效。所有 MPLS TE 隧道的出口速率每经过一个采样周期就会被记录下来，这样可以获得一条 MPLS TE 隧道在一个采样周期内的实际平均带宽。
- 在隧道视图下启用了自动带宽调整后，当调整频率定时器（**mpls te auto-bandwidth adjustment frequency**）超时时，将这段时间内多次采样的最大值作为带宽约束，发起建立一条新的 LSP 隧道。建立成功后，将流量切换到新的 LSP 隧道上，同时拆除原来的 LSP 隧道；如果建立不成功，则流量继续走原来的 LSP 隧道。

1.12 配置 CR-LSP 备份

CR-LSP 备份用于端到端的路径保护，对整条 LSP 提供保护。

1.12.1 配置准备

在配置 CR-LSP 备份之前，需完成以下任务：

- 配置 MPLS 基本能力
- 配置 MPLS TE 基本能力
- 配置 MPLS TE 隧道

1.12.2 配置 CR-LSP 备份

表1-37 配置 CR-LSP 备份

操作	命令	说明
进入隧道头节点的系统视图	**system-view**	-
进入 MPLS TE 隧道的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-
配置当前隧道使用的备份模式	**mpls te backup** { **hot-standby** | **ordinary** }	必选 缺省情况下，隧道不进行备份
提交隧道当前配置	**mpls te commit**	必选

说明

CR-LSP 备份方式在隧道头节点进行配置，系统自动选择主 LSP 和备份 LSP 的路径，用户不需要分别配置主 LSP 和备份 LSP。

1.13 配置 MPLS TE 快速重路由

说明

当 MPLS TE 隧道的动态协议类型为 CR-LDP 时，不支持本特性。

FRR 是 MPLS TE 中的临时性局部保护技术。

由于 FRR 使用的 Bypass 隧道需要预先建立，占用额外的带宽，因此，在网络带宽余量不多的情况下，应该只对关键的接口或链路进行快速重路由保护。

用户可以指定哪种类型的 LSP 可以使用 Bypass 隧道，Bypass 隧道是否提供带宽保护以及带宽保护的总量。

Bypass 隧道的带宽用于保护主 LSP，用户在配置时应保证其带宽不小于被保护的所有 LSP 所需带宽之和，否则可能有主 LSP 不能绑定到 Bypass 隧道上。

Bypass 隧道一般不转发数据。如果 Bypass 隧道在保护主 LSP 的同时转发流量，需要为 Bypass 隧道提供足够的带宽。

Bypass 隧道不能同时做为 VPN 等业务的隧道使用。

1.13.1 配置准备

在配置 MPLS TE 快速重路由之前，需完成以下任务：

- 配置 IGP 协议保证各 LSR 之间可达
- 配置 MPLS 基本能力
- 配置 MPLS TE 基本能力
- 使用 RSVP-TE 建立 MPLS TE 隧道
- 建立主 LSP

1.13.2 配置 MPLS TE 快速重路由

1. 在主 LSP 的头节点使能快速重路由

表1-38 在主 LSP 的头节点使能快速重路由

操作	命令	说明
进入系统视图	**system-view**	-
进入主 LSP 的 Tunnel 接口视图	**interface tunnel** *tunnel-number*	-

操作	命令	说明
使能快速重路由功能	**mpls te fast-reroute**	必选 缺省情况下，未使能快速重路由功能
提交隧道配置	**mpls te commit**	必选

2. 在 PLR 上配置 Bypass 隧道

当一条隧道被指定保护一个接口以后，它所对应的 LSP 就成为 Bypass 隧道。Bypass 隧道的建立通过在 PLR 手工配置触发。它的配置与普通 LSP 相同，但不能配置快速重路由属性。也就是说，Bypass 隧道不能同时作为主 LSP，LSP 不能被嵌套保护。

为接口指定 Bypass 隧道，需要满足以下条件：

- Bypass 隧道必须是 up 状态的；
- 被保护的接口不是本身的出接口。

设备最多可以为被保护的接口指定 3 条 Bypass 隧道，将使用最佳适配（best-fit）算法来决定使用哪一条 Bypass 隧道。

设备对使用 Bypass 隧道的链路进行限制，使这些隧道的带宽总量不超过指定的带宽。

表1-39 在 PLR 上配置 Bypass 隧道

操作	命令	说明
进入系统视图	**system-view**	-
进入 Bypass 隧道的接口视图	**interface tunnel** *tunnel-number*	-
配置 Bypass 隧道的目的地址	**destination** *ip-address*	必选 • 配置节点保护时，目的地址为 PLR 的下下一跳路由器 LSR ID • 配置链路保护时，目的地址为 PLR 的下一跳设备的 LSR ID
配置 Bypass 隧道可保护的带宽和 LSP 类型	**mpls te backup bandwidth** { *bandwidth* | { **bc0** | **bc1** } { *bandwidth* | **un-limited** } }	必选 缺省情况下，不对带宽进行保护
提交隧道配置	**mpls te commit**	必选
退回系统视图	**quit**	-
进入被保护的 LSP 出接口的接口视图	**interface** *interface-type interface-number*	-
将 Bypass 隧道绑定到被保护的接口	**mpls te fast-reroute bypass-tunnel tunnel** *tunnel-number*	必选

注意

对于 Bypass 隧道，必须使用 **mpls te backup bandwidth** 命令配置可保护的带宽，否则，由于缺省情况是不对带宽进行保护，将导致主 LSP 不能绑定到 Bypass 隧道。

3. 配置节点保护

如果使用 FRR 进行节点保护，则需要在 PLR 和被保护节点上进行本小节的配置；如果只是进行链路保护，则不必进行本小节的配置。

表1-40 配置节点保护

操作	命令	说明
进入系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
使能本节点的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能本节点的 RSVP 的 Hello 扩展
退回系统视图	**quit**	-
进入 PLR 与被保护节点直连接口的接口视图	**interface** *interface-type interface-number*	-
使能接口的 RSVP 的 Hello 扩展	**mpls rsvp-te hello**	必选 缺省情况下，未使能接口的 RSVP 的 Hello 扩展

说明

对于链路失效引发的节点失效，不需要使用 RSVP 的 Hello 扩展机制来进行邻居检测。该配置主要用于在链路正常，但信令协议故障的特殊情况下检测节点失效。

4. 配置快速重路由扫描定时器

FRR 切换后，一旦被保护的 LSP 恢复正常或建立了新的 LSP，流量仍会切换回原 LSP 或新建立的 LSP。切换回来后，PLR 定时扫描，如果发现保护隧道，将刷新绑定关系。

表1-41 配置快速重路由扫描定时器

操作	命令	说明
进入 PLR 节点的系统视图	**system-view**	-
进入 MPLS 视图	**mpls**	-
配置快速重路由扫描间隔	**mpls te timer fast-reroute** [*second*]	可选 缺省情况下，扫描间隔为 300 秒

1.14 MPLS TE 显示和维护

1.14.1 MPLS TE 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MPLS TE 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 MPLS TE 统计信息。

表1-42 MPLS TE 的显示和维护

操作	命令
显示显式路径信息	**display explicit-path** [*pathname*]
显示静态 CR-LSP 信息	**display mpls static-cr-lsp** [**lsp-name** *lsp-name*] [{ **include** | **exclude** } *ip-address prefix-length*] [**verbose**]
显示 RSVP-TE 的配置信息	**display mpls rsvp-te** [**interface** [*interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]]
显示全局或接口的 RSVP-TE 信息	**display mpls rsvp-te established** [**interface** *interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE 邻居信息	**display mpls rsvp-te peer** [**interface** *interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE 请求消息信息	**display mpls rsvp-te request** [**interface** *interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP 资源预留信息	**display mpls rsvp-te reservation** [**interface** *interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE PSB 信息	**display mpls rsvp-te psb-content** { *ingress-lsr-id lspid tunnel-id egress-lsr-id* } [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE RSB 信息	**display mpls rsvp-te rsb-content** { *ingress-lsr-id lspid tunnel-id egress-lsr-id nexthop-address* } [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE 发送方消息信息	**display mpls rsvp-te sender** [**interface** *interface-type interface-number*] [**|** { **begin | exclude | include** } *regular-expression*]
显示 RSVP-TE 的统计信息	**display mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface-number*] }
显示满足指定条件的基于 CSPF 的 TEDB 信息	**display mpls te cspf tedb** { **all** | **area** *area-id* | **interface** *ip-address* | **network-lsa** | **node** [*mpls-lsr-id*] } [**|** { **begin | exclude | include** } *regular-expression*]
显示所有链路接纳的 CR-LSP 信息	**display mpls te link-administration admission-control** [**interface** *interface-type interface-number*]
显示使能了 MPLS TE 的接口分配的带宽	**display mpls te link-administration bandwidth-allocation** [**interface** *interface-type interface-number*]
显示 MPLS TE 隧道信息	**display mpls te tunnel** [**destination** *dest-addr*] [**lsp-id** *lsr-id lsp-id*] [**lsr-role** { **all** | **egress** | **ingress** | **remote** | **transit** }] [**name** *name*] [{ **incoming-interface** | **outgoing-interface** | **interface** } *interface-type interface-number*] [**verbose**]
显示本地节点的 MPLS TE 隧道的路径属性信息	**display mpls te tunnel path** [**tunnel-name** *tunnel-name*] [**lsp-id** *lsr-id lsp-id*]
显示 MPLS TE 隧道统计信息	**display mpls te tunnel statistics**
显示本地节点的 MPLS TE 隧道接口信息	**display mpls te tunnel-interface tunnel** *number*
显示与流量调节（IGP shortcut 及转发邻接）相关的 OSPF 进程的隧道信息	**display ospf** [*process-id*] **traffic-adjustment**
显示 OSPF TE 信息	**display ospf** [*process-id*] **mpls-te** [**area** *area-id*] [**self-originated**]
显示 IS-IS TE 最近一次发布的 TE 信息	**display isis traffic-eng advertisements** [**level-1** | **level-1-2** | **level-2**] [**lsp-id** *lsp-id* | **local**] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS TE 的链路信息	**display isis traffic-eng link** [**level-1** | **level-1-2** | **level-2**] [**verbose**] [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS TE 的网络信息	**display isis traffic-eng network** [**level-1** | **level-1-2** | **level-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]

操作	命令
显示 IS-IS TE 的统计信息	**display isis traffic-eng statistics** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示 IS-IS TE 扩展的子 TLV 信息	**display isis traffic-eng sub-tlvs** [*process-id* | **vpn-instance** *vpn-instance-name*]
显示隧道信息	**display tunnel-info** { *tunnel-id* | **all** | **statistics** }
清除 RSVP-TE 的统计信息	**reset mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface-number*]

1.14.2 重启自动带宽调整功能

在用户视图下执行 **reset** 命令可以重启自动带宽调整功能。

重启时，会清除出口速率采样信息，清除距下一次带宽调整剩余的时间。

表1-43 重启初始化自动带宽调整功能

操作	命令
重启自动带宽调整功能	**reset mpls te auto-bandwidth adjustment timers**

1.15 MPLS TE 典型组网举例

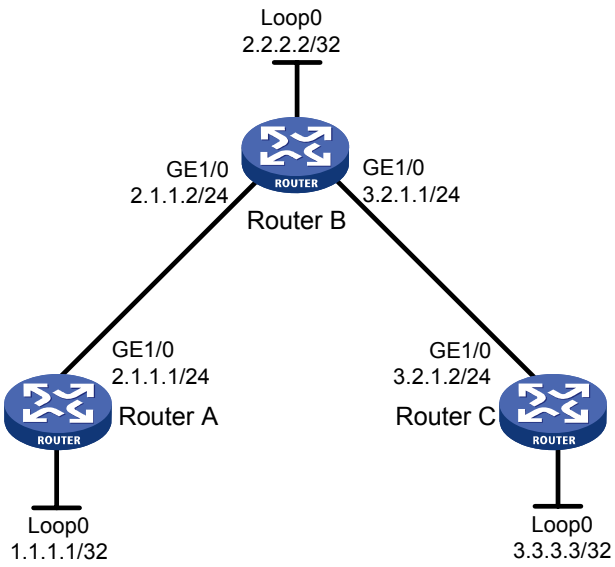
1.15.1 使用静态 CR-LSP 配置 MPLS TE 隧道示例

1. 组网需求

- 设备 Router A、Router B 和 Router C 运行 IS-IS；
- 使用静态 CR-LSP 建立一条 Router A 到 Router C 的 TE 隧道。

2. 组网图

图1-7 静态 CR-LSP 配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-7配置各接口的IP地址和掩码，具体配置过程略。

(2) 配置 IS-IS 协议发布 LSR ID 的主机路由

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] network-entity 00.0005.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] isis enable 1
[RouterA-GigabitEthernet1/0] quit
[RouterA] interface loopback 0
[RouterA-LoopBack0] isis enable 1
[RouterA-LoopBack0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] network-entity 00.0005.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] isis enable 1
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] isis enable 1
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface loopback 0
[RouterB-LoopBack0] isis enable 1
[RouterB-LoopBack0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 00.0005.0000.0000.0003.00
[RouterC-isis-1] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] isis enable 1
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface loopback 0
[RouterC-LoopBack0] isis enable 1
[RouterC-LoopBack0] quit
```

配置完成后，在各设备上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到对方的 LSR ID 的主机路由。以 Router A 为例：

```
[RouterA] display ip routing-table
Routing Tables: Public
```

```
         Destinations : 8        Routes : 8
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
1.1.1.1/32          Direct 0    0         127.0.0.1       InLoop0
2.1.1.0/24          Direct 0    0         2.1.1.1         GE1/0
2.1.1.1/32          Direct 0    0         127.0.0.1       InLoop0
2.2.2.2/32          ISIS   15   10        2.1.1.2         GE1/0
3.2.1.0/24          ISIS   15   20        2.1.1.2         GE1/0
3.3.3.3/32          ISIS   15   20        2.1.1.2         GE1/0
127.0.0.0/8         Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0         127.0.0.1       InLoop0
```

(3) 配置 MPLS TE 基本能力

配置 Router A。

```
[RouterA] mpls lsr-id 1.1.1.1
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B。

```
[RouterB] mpls lsr-id 2.2.2.2
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls
[RouterB-GigabitEthernet1/0] mpls te
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls
[RouterB-GigabitEthernet1/1] mpls te
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
[RouterC] mpls lsr-id 3.3.3.3
[RouterC] mpls
[RouterC-mpls] mpls te
[RouterC-mpls] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] mpls
[RouterC-GigabitEthernet1/1] mpls te
[RouterC-GigabitEthernet1/1] quit
```

(4) 配置 MPLS TE 隧道

在 Router A 上配置 MPLS TE 隧道。

```
[RouterA] interface tunnel 0
```

```
[RouterA-Tunnel0] ip address 6.1.1.1 255.255.255.0
[RouterA-Tunnel0] tunnel-protocol mpls te
[RouterA-Tunnel0] destination 3.3.3.3
[RouterA-Tunnel0] mpls te tunnel-id 10
[RouterA-Tunnel0] mpls te signal-protocol static
[RouterA-Tunnel0] mpls te commit
[RouterA-Tunnel0] quit
```

(5) 创建静态 CR-LSP

配置 Router A 为静态 CR-LSP 的入口节点：

```
[RouterA] static-cr-lsp ingress tunnel0 destination 3.3.3.3 nexthop 2.1.1.2 out-label 20
```

配置 Router B 为静态 CR-LSP 的中间节点：

```
[RouterB] static-cr-lsp transit tunnel0 incoming-interface gigabitethernet 1/0 in-label 20
nexthop 3.2.1.2 out-label 30
```

配置 Router C 为静态 CR-LSP 的出节点：

```
[RouterC] static-cr-lsp egress tunnel0 incoming-interface gigabitethernet 1/1 in-label 30
```

(6) 配置完成后的检验

配置完成后，在 Router A 上执行 **display interface tunnel** 命令，可以看到 Tunnel 接口的状态为 up。

```
[RouterA] display interface tunnel
Tunnel0 current state: UP
Line protocol current state: UP
Description: Tunnel0 Interface
The Maximum Transmit Unit is 64000
Internet Address is 6.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 3.3.3.3
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    0 packets output,  0 bytes
    0 output error
```

在各设备上执行 **display mpls te tunnel** 命令，可以看到 MPLS TE 隧道的建立情况。

```
[RouterA] display mpls te tunnel
LSP-Id          Destination     In/Out-If              Name
1.1.1.1:1       3.3.3.3         -/GE1/0                Tunnel0
[RouterB] display mpls te tunnel
LSP-Id          Destination     In/Out-If              Name
-               -               GE1/0/GE1/1            Tunnel0
[RouterC] display mpls te tunnel
LSP-Id          Destination     In/Out-If              Name
```

```
-                -                GE1/1/-                  Tunnel0
```

在各设备上执行 **display mpls lsp** 或 **display mpls static-cr-lsp** 命令，可以看到静态 CR-LSP 的建立情况。

```
[RouterA] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: STATIC CRLSP
-------------------------------------------------------------------
FEC              In/Out Label  In/Out IF              Vrf Name
3.3.3.3/32       NULL/20       -/GE1/0
[RouterB] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: STATIC CRLSP
-------------------------------------------------------------------
FEC              In/Out Label  In/Out IF               Vrf Name
-/-              20/30         GE1/0/GE1/1
[RouterC] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: STATIC CRLSP
-------------------------------------------------------------------
FEC              In/Out Label  In/Out IF               Vrf Name
-/-              30/NULL       GE1/0/-
[RouterA] display mpls static-cr-lsp
total statics-cr-lsp : 1
Name             FEC              I/O Label  I/O If             Stat
Tunnel0          3.3.3.3/32       NULL/20    -/GE1/0            Up
[RouterB] display mpls static-cr-lsp
total statics-cr-lsp : 1
Name             FEC              I/O Label  I/O If             Stat
Tunnel0          -/-              20/30      GE1/0/GE1/1       Up
[RouterC] display mpls static-cr-lsp
total statics-cr-lsp : 1
Name             FEC              I/O Label  I/O If             Stat
Tunnel0          -/-              30/NULL    GE1/1/-            Up
```



说明

使用静态 CR-LSP 建立 MPLS TE 隧道时，Transit 节点和 Egress 节点直接根据配置的入标签和出标签进行转发，因此，Router B 和 Router C 的相关显示信息中，FEC 等内容为空。

(7) 使用静态路由沿 MPLS TE 隧道转发流量

```
[RouterA] ip route-static 3.2.1.2 24 tunnel 0 preference 1
```

在 RouterA 上执行 **display ip routing-table** 命令，可以看到路由表中有以 Tunnel0 为出接口的静态路由信息。

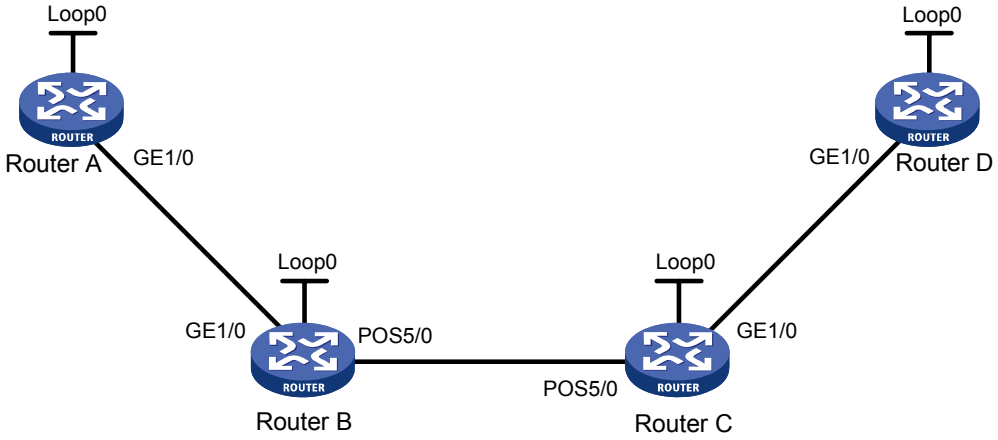
1.15.2 使用 RSVP-TE 配置 MPLS TE 隧道示例

1. 组网需求

- 设备 Router A、Router B、Router C 和 Router D 运行 IS-IS，都是 Level-2 设备；
- 使用 RSVP-TE 从 Router A 到 Router D 建立一条 TE 隧道，带宽为 2000Kbit/s；
- 隧道沿途的链路最大带宽为 10000Kbit/s，最大可预留带宽为 5000Kbit/s。

2. 组网图

图1-8 RSVP-TE 配置 MPLS TE 隧道组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	Loop0	1.1.1.9/32	Router C	Loop0	3.3.3.9/32
	GE1/0	10.1.1.1/24		GE1/0	30.1.1.1/24
Router B	Loop0	2.2.2.9/32		POS5/0	20.1.1.2/24
	GE1/0	10.1.1.2/24	Router D	Loop0	4.4.4.9/32
	POS5/1	20.1.1.1/24		GE1/0	30.1.1.2/24

3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-8配置各接口的IP地址和掩码，具体配置过程略。

(2) 配置 IS-IS 协议发布 LSR ID 的主机路由

配置 Router A。

```
<RouterA> system-view
[RouterA] isis 1
[RouterA-isis-1] network-entity 00.0005.0000.0000.0001.00
[RouterA-isis-1] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] isis enable 1
[RouterA-GigabitEthernet1/0] isis circuit-level level-2
[RouterA-GigabitEthernet1/0] quit
[RouterA] interface loopback 0
[RouterA-LoopBack0] isis enable 1
[RouterA-LoopBack0] isis circuit-level level-2
[RouterA-LoopBack0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] isis 1
[RouterB-isis-1] network-entity 00.0005.0000.0000.0002.00
[RouterB-isis-1] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] isis enable 1
[RouterB-GigabitEthernet1/0] isis circuit-level level-2
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface pos 5/0
[RouterB-POS5/0] isis enable 1
[RouterB-POS5/0] isis circuit-level level-2
[RouterB-POS5/0] quit
[RouterB] interface loopback 0
[RouterB-LoopBack0] isis enable 1
[RouterB-LoopBack0] isis circuit-level level-2
[RouterB-LoopBack0] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] isis 1
[RouterC-isis-1] network-entity 00.0005.0000.0000.0003.00
[RouterC-isis-1] quit
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] isis enable 1
[RouterC-GigabitEthernet1/0] isis circuit-level level-2
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface pos 5/0
[RouterC-POS5/0] isis enable 1
[RouterC-POS5/0] isis circuit-level level-2
[RouterC-POS5/0] quit
[RouterC] interface loopback 0
[RouterC-LoopBack0] isis enable 1
[RouterC-LoopBack0] isis circuit-level level-2
[RouterC-LoopBack0] quit
```

配置 Router D。

```
<RouterD> system-view
[RouterD] isis 1
[RouterD-isis-1] network-entity 00.0005.0000.0000.0004.00
[RouterD-isis-1] quit
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] isis enable 1
[RouterD-GigabitEthernet1/0] isis circuit-level level-2
[RouterD-GigabitEthernet1/0] quit
[RouterD] interface loopback 0
[RouterD-LoopBack0] isis enable 1
[RouterD-LoopBack0] isis circuit-level level-2
[RouterD-LoopBack0] quit
```

配置完成后，在各设备上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到对方的 LSR ID 的主机路由。以 Router A 为例：

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10
Destination/Mask  Proto  Pre  Cost     NextHop         Interface
      1.1.1.9/32  Direct 0    0        127.0.0.1       InLoop0
      2.2.2.9/32  ISIS   15   10       10.1.1.2        GE1/0
      3.3.3.9/32  ISIS   15   20       10.1.1.2        GE1/0
      4.4.4.9/32  ISIS   15   30       10.1.1.2        GE1/0
     10.1.1.0/24  Direct 0    0        10.1.1.1        GE1/0
     10.1.1.1/32  Direct 0    0        127.0.0.1       InLoop0
     20.1.1.0/24  ISIS   15   20       10.1.1.2        GE1/0
     30.1.1.0/24  ISIS   15   30       10.1.1.2        GE1/0
    127.0.0.0/8   Direct 0    0        127.0.0.1       InLoop0
    127.0.0.1/32  Direct 0    0        127.0.0.1       InLoop0
```

(3) 配置 MPLS TE 基本能力，并使能 RSVP-TE 和 CSPF

配置 Router A。

```
[RouterA] mpls lsr-id 1.1.1.9
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls rsvp-te
[RouterA-mpls] mpls te cspf
[RouterA-mpls] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] mpls rsvp-te
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B。

```
[RouterB] mpls lsr-id 2.2.2.9
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls rsvp-te
[RouterB-mpls] mpls te cspf
[RouterB-mpls] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls
[RouterB-GigabitEthernet1/0] mpls te
[RouterB-GigabitEthernet1/0] mpls rsvp-te
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface pos 5/0
[RouterB-POS5/0] mpls
[RouterB-POS5/0] mpls te
[RouterB-POS5/0] mpls rsvp-te
[RouterB-POS5/0] quit
```

配置 Router C。

```
[RouterC] mpls lsr-id 3.3.3.9
[RouterC] mpls
[RouterC-mpls] mpls te
[RouterC-mpls] mpls rsvp-te
[RouterC-mpls] mpls te cspf
[RouterC-mpls] quit
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] mpls
[RouterC-GigabitEthernet1/0] mpls te
[RouterC-GigabitEthernet1/0] mpls rsvp-te
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface pos 5/0
[RouterC-POS5/0] mpls
[RouterC-POS5/0] mpls te
[RouterC-POS5/0] mpls rsvp-te
[RouterC-POS5/0] quit
```

配置 Router D。

```
[RouterD] mpls lsr-id 4.4.4.9
[RouterD] mpls
[RouterD-mpls] mpls te
[RouterD-mpls] mpls rsvp-te
[RouterD-mpls] mpls te cspf
[RouterD-mpls] quit
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] mpls
[RouterD-GigabitEthernet1/0] mpls te
[RouterD-GigabitEthernet1/0] mpls rsvp-te
[RouterD-GigabitEthernet1/0] quit
```

(4) 配置 IS-IS TE

配置 Router A。

```
[RouterA] isis 1
[RouterA-isis-1] cost-style wide
[RouterA-isis-1] traffic-eng level-2
[RouterA-isis-1] quit
```

配置 Router B。

```
[RouterB] isis 1
[RouterB-isis-1] cost-style wide
[RouterB-isis-1] traffic-eng level-2
[RouterB-isis-1] quit
```

配置 Router C。

```
[RouterC] isis 1
[RouterC-isis-1] cost-style wide
[RouterC-isis-1] traffic-eng level-2
[RouterC-isis-1] quit
```

配置 Router D。

```
[RouterD] isis 1
[RouterD-isis-1] cost-style wide
[RouterD-isis-1] traffic-eng level-2
[RouterD-isis-1] quit
```

(5) 配置链路的 MPLS TE 属性

在 Router A 上配置链路的最大带宽和最大可预留带宽。

```
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterA-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterA-GigabitEthernet1/0] quit
```

在 Router B 上配置链路的最大带宽和最大可预留带宽。

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterB-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface pos 5/0
[RouterB-POS5/0] mpls te max-link-bandwidth 10000
[RouterB-POS5/0] mpls te max-reservable-bandwidth 5000
[RouterB-POS5/0] quit
```

在 Router C 上配置链路的最大带宽和最大可预留带宽。

```
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterC-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface pos 5/0
[RouterC-POS5/0] mpls te max-link-bandwidth 10000
[RouterC-POS5/0] mpls te max-reservable-bandwidth 5000
[RouterC-POS5/0] quit
```

在 Router D 上配置链路的最大带宽和最大可预留带宽。

```
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterD-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterD-GigabitEthernet1/0] quit
```

(6) 配置 MPLS TE 隧道

在 Router A 上配置 MPLS TE 隧道。

```
[RouterA] interface tunnel 1
[RouterA-Tunnel1] ip address 7.1.1.1 255.255.255.0
[RouterA-Tunnel1] tunnel-protocol mpls te
[RouterA-Tunnel1] destination 4.4.4.9
[RouterA-Tunnel1] mpls te tunnel-id 10
[RouterA-Tunnel1] mpls te signal-protocol rsvp-te
[RouterA-Tunnel1] mpls te bandwidth 2000
[RouterA-Tunnel1] mpls te commit
```

```
[RouterA-Tunnel1] quit
```

(7) 配置完成后的检验

配置完成后，在 Router A 上执行 **display interface tunnel** 命令可以看到隧道接口状态为 up。

```
[RouterA] display interface tunnel
Tunnel1 current state: UP
Line protocol current state: UP
Description: Tunnel1 Interface
The Maximum Transmit Unit is 64000
Internet Address is 7.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 4.4.4.9
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    0 packets output,  0 bytes
    0 output error
```

在 Router A 上执行 **display mpls te tunnel-interface** 命令可以看到隧道的详细信息。

```
[RouterA] display mpls te tunnel-interface
  Tunnel Name       :  Tunnel1
  Tunnel Desc       :  Tunnel1 Interface
  Tunnel State Desc :  CR-LSP is Up
  Tunnel Attributes :
    LSP ID              :  1.1.1.9:3
    Session ID          :  10
    Admin State         :  UP                Oper State   :  UP
    Ingress LSR ID      :  1.1.1.9           Egress LSR ID:  4.4.4.9
    Signaling Prot      :  RSVP              Resv Style   :  SE
    Class Type          :  CLASS 0           Tunnel BW    :  2000 kbps
    Reserved BW         :  2000 kbps
    Setup Priority      :  7                 Hold Priority:  7
    Affinity Prop/Mask  :  0x0/0x0
    Explicit Path Name  :  -
    Tie-Breaking Policy :  None
    Metric Type         :  None
    Record Route        :  Disabled    Record Label :  Disabled
    FRR Flag            :  Disabled    BackUpBW Flag:  Not Supported
    BackUpBW Type       :  -           BackUpBW     :  -
    Route Pinning       :  Disabled
    Retry Limit         :  5           Retry Interval:  10 sec
    Reopt               :  Disabled    Reopt Freq   :  -
    Back Up Type        :  None
    Back Up LSPID       :  -
```

```
Auto BW             :  Disabled    Auto BW Freq :  -
Min BW              :  -           Max BW       :  -
Current Collected BW:  -
Interfaces Protected:  -
VPN Bind Type       :  NONE
VPN Bind Value      :  -
Car Policy          :  Disabled
Tunnel Group        :  Primary
Primary Tunnel      :  -
Backup Tunnel       :  -
Group Status        :  -
```

在 Router A 上执行 **display mpls te cspf tedb all** 命令查看 TEDB 中的链路信息。

```
[RouterA] display mpls te cspf tedb all
Maximum Node Supported: 128              Maximum Link Supported: 256
Current Total Node Number: 4             Current Total Link Number: 6
Id      MPLS LSR-Id  IGP      Process-Id   Area            Link-Count
1       3.3.3.9      ISIS     1            Level-2         2
2       2.2.2.9      ISIS     1            Level-2         2
3       4.4.4.9      ISIS     1            Level-2         1
4       1.1.1.9      ISIS     1            Level-2         1
```

(8) 使用静态路由沿 MPLS TE 隧道转发流量

```
[RouterA] ip route-static 30.1.1.2 24 tunnel 1 preference 1
```

在 Router A 上执行 **display ip routing-table** 命令，可以看到路由表中有以 Tunnel1 为出接口的静态路由信息。

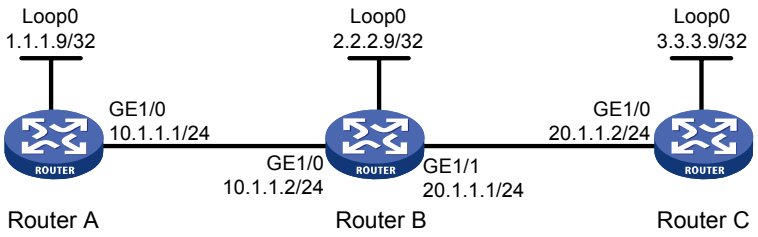
1.15.3 配置 RSVP-TE GR 示例

1. 组网需求

- 设备 Router A、Router B 和 Router C 运行 IS-IS，都是 Level-2 设备；
- 使用 RSVP-TE 建立从 Router A 到 Router C 的 TE 隧道；
- 设备 Router A、Router B 和 Router C 支持 RSVP 的 Hello 扩展能力；
- Router A、Router B 和 Router C 为 RSVP-TE 邻居，通过使能 GR 能力，邻居之间可以在对方发生 GR Restart 的时候提供 GR Helper 支持。

2. 组网图

图1-9 RSVP-TE GR 配置组网图



3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-9配置各接口的IP地址和掩码，具体配置过程略。

(2) 配置 IS-IS 协议发布 LSR ID 的主机路由

具体配置略。

(3) 配置 MPLS TE 基本能力，并使能 RSVP-TE 和节点的 RSVP 的 Hello 扩展

配置 Router A

```
<RouterA> system-view
[RouterA] mpls lsr-id 1.1.1.9
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls rsvp-te
[RouterA-mpls] mpls rsvp-te hello
[RouterA-mpls] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] mpls rsvp-te
[RouterA-GigabitEthernet1/0] mpls rsvp-te hello
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B

```
<RouterB> system-view
[RouterB] mpls lsr-id 2.2.2.9
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls rsvp-te
[RouterB-mpls] mpls rsvp-te hello
[RouterB-mpls] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls
[RouterB-GigabitEthernet1/0] mpls te
[RouterB-GigabitEthernet1/0] mpls rsvp-te
[RouterB-GigabitEthernet1/0] mpls rsvp-te hello
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls
[RouterB-GigabitEthernet1/1] mpls te
[RouterB-GigabitEthernet1/1] mpls rsvp-te
[RouterB-GigabitEthernet1/1] mpls rsvp-te hello
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C

```
<RouterC> system-view
[RouterC] mpls lsr-id 3.3.3.9
[RouterC] mpls
[RouterC-mpls] mpls te
[RouterC-mpls] mpls rsvp-te
```

```
[RouterC-mpls] mpls rsvp-te hello
[RouterC-mpls] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] mpls
[RouterC-GigabitEthernet1/0] mpls te
[RouterC-GigabitEthernet1/0] mpls rsvp-te
[RouterC-GigabitEthernet1/0] mpls rsvp-te hello
[RouterC-GigabitEthernet1/0] quit
```

(4) 配置 IS-IS TE

具体配置略。

(5) 配置 MPLS TE 隧道

具体配置略。

(6) 配置 RSVP-TE GR

配置 Router A

```
<RouterA> system-view
[RouterA] mpls
[RouterA-mpls] mpls rsvp-te graceful-restart
```

配置 Router B

```
<RouterB> system-view
[RouterB] mpls
[RouterB-mpls] mpls rsvp-te graceful-restart
```

配置 Router C

```
<RouterC> system-view
[RouterC] mpls
[RouterC-mpls] mpls rsvp-te graceful-restart
```

(7) 检验配置效果

在 Router A 和 Router C 之间创建隧道且运行稳定后，通过命令可以看到邻居的 GR 状态已经处于 Ready 状态。

```
<RouterA> display mpls rsvp-te peer
Interface GE1/0
 Neighbor Addr: 10.1.1.2
 SrcInstance: 880              NbrSrcInstance: 5017
 PSB Count: 0                  RSB Count: 1
 Hello Type Sent: REQ          Neighbor Hello Extension: ENABLE
 SRefresh Enable: NO
 Graceful Restart State: Ready
 Restart Time: 120 Sec         Recovery Time: 300 Sec
```

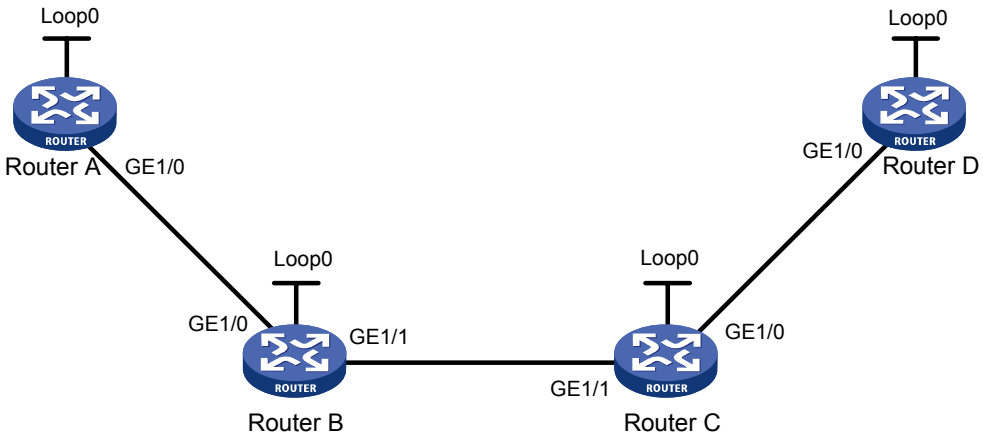
1.15.4 使用 CR-LDP 配置 MPLS TE 隧道示例

1. 组网需求

- 设备 Router A、Router B、Router C 和 Router D 运行 OSPF，都在区域 0 内；
- 使用 CR-LDP 从 Router A 到 Router D 建立一条 TE 隧道，带宽为 2000Kbit/s；
- 隧道沿途的链路最大带宽为 10000Kbit/s，最大可预留带宽为 5000Kbit/s。

2. 组网图

图1-10 使用 CR-LDP 配置 MPLS TE 隧道组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	Loop0	1.1.1.9/32	Router D	Loop0	4.4.4.9/32
	GE1/0	10.1.1.1/24		GE1/0	30.1.1.2/24
Router B	Loop0	2.2.2.9/32	Router C	Loop0	3.3.3.9/32
	GE1/0	10.1.1.2/24		GE1/0	30.1.1.1/24
	GE1/1	20.1.1.1/24		GE1/1	20.1.1.2/24

3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-10配置各接口的IP地址和掩码，具体配置过程略。

(2) 配置 OSPF 协议发布 LSR ID 的主机路由

具体配置过程略。配置完成后，在各设备上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到对方 LSR ID 的主机路由。

(3) 配置 MPLS TE 基本能力，并使能 CSPF

配置 Router A。

```
[RouterA] mpls lsr-id 1.1.1.9
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls te cspf
[RouterA-mpls] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B。

```
[RouterB] mpls lsr-id 2.2.2.9
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls te cspf
[RouterB-mpls] quit
[RouterB] interface gigabitethernet 1/0
```

```
[RouterB-GigabitEthernet1/0] mpls
[RouterB-GigabitEthernet1/0] mpls te
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls
[RouterB-GigabitEthernet1/1] mpls te
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
[RouterC] mpls lsr-id 3.3.3.9
[RouterC] mpls
[RouterC-mpls] mpls te
[RouterC-mpls] mpls te cspf
[RouterC-mpls] quit
[RouterC] interface GigabitEthernet 1/0
[RouterC-GigabitEthernet1/0] mpls
[RouterC-GigabitEthernet1/0] mpls te
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] mpls
[RouterC-GigabitEthernet1/1] mpls te
[RouterC-GigabitEthernet1/1] quit
```

配置 Router D。

```
[RouterD] mpls lsr-id 4.4.4.9
[RouterD] mpls
[RouterD-mpls] mpls te
[RouterD-mpls] mpls te cspf
[RouterD-mpls] quit
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] mpls
[RouterD-GigabitEthernet1/0] mpls te
[RouterD-GigabitEthernet1/0] quit
```

(4) 配置 OSPF TE

配置 Router A。

```
[RouterA] ospf
[RouterA-ospf-1] opaque-capability enable
[RouterA-ospf-1] area 0
[RouterA-ospf-1-area-0.0.0.0] mpls-te enable
[RouterA-ospf-1-area-0.0.0.0] quit
[RouterA-ospf-1] quit
```

配置 Router B。

```
[RouterB] ospf
[RouterB-ospf-1] opaque-capability enable
[RouterB-ospf-1] area 0
[RouterB-ospf-1-area-0.0.0.0] mpls-te enable
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
```

配置 Router C。

```
[RouterC] ospf
[RouterC-ospf-1] opaque-capability enable
[RouterC-ospf-1] area 0
[RouterC-ospf-1-area-0.0.0.0] mpls-te enable
[RouterC-ospf-1-area-0.0.0.0] quit
[RouterC-ospf-1] quit
```

配置 Router D。

```
[RouterD] ospf
[RouterD-ospf-1] opaque-capability enable
[RouterD-ospf-1] area 0
[RouterD-ospf-1-area-0.0.0.0] mpls-te enable
[RouterD-ospf-1-area-0.0.0.0] quit
[RouterD-ospf-1] quit
```

(5) 配置链路的 MPLS TE 属性

在 Router A 上配置链路的最大带宽和最大可预留带宽。

```
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterA-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterA-GigabitEthernet1/0] quit
```

在 Router B 上配置链路的最大带宽和最大可预留带宽。

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterB-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls te max-link-bandwidth 10000
[RouterB-GigabitEthernet1/1] mpls te max-reservable-bandwidth 5000
[RouterB-GigabitEthernet1/1] quit
```

在 Router C 上配置链路的最大带宽和最大可预留带宽。

```
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterC-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] mpls te max-link-bandwidth 10000
[RouterC-GigabitEthernet1/1] mpls te max-reservable-bandwidth 5000
[RouterC-GigabitEthernet1/1] quit
```

在 Router D 上配置链路的最大带宽和最大可预留带宽。

```
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] mpls te max-link-bandwidth 10000
[RouterD-GigabitEthernet1/0] mpls te max-reservable-bandwidth 5000
[RouterD-GigabitEthernet1/0] quit
```

配置完成后，在各设备上执行 **display mpls te cspf tedb all** 命令可以看到 TEDB 中的链路信息。以 Router A 的显示为例：

```
[RouterA] display mpls te cspf tedb all
Maximum Node Supported: 128           Maximum Link Supported: 256
Current Total Node Number: 4          Current Total Link Number: 6
Id     MPLS LSR-Id  IGP      Process-Id     Area         Link-Count
1      2.2.2.9      OSPF     1              0            2
2      3.3.3.9      OSPF     1              0            2
3      4.4.4.9      OSPF     1              0            1
4      1.1.1.9      OSPF     1              0            1
```

(6) 配置 LDP

配置 Router A。

```
[RouterA] mpls ldp
[RouterA-mpls-ldp] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls ldp
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B。

```
[RouterB] mpls ldp
[RouterB-mpls-ldp] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls ldp
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls ldp
[RouterB-GigabitEthernet1/1] quit
```

配置 Router C。

```
[RouterC] mpls ldp
[RouterC-mpls-ldp] quit
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] mpls ldp
[RouterC-GigabitEthernet1/0] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] mpls ldp
[RouterC-GigabitEthernet1/1] quit
```

配置 Router D。

```
[RouterD] mpls ldp
[RouterD-mpls-ldp] quit
[RouterD] interface gigabitethernet 1/0
[RouterD-GigabitEthernet1/0] mpls ldp
[RouterD-GigabitEthernet1/0] quit
```

配置完成后，在各设备上执行 **display mpls ldp session** 命令，可以看到 LDP 会话已经建立，会话状态为 Operational。以 RouterA 为例：

```
[RouterA] display mpls ldp session
```

```
              LDP Session(s) in Public Network
 ------------------------------------------------------------------
 Peer-ID         Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 2.2.2.9:0       Operational   DU   Passive  Off  Off  11/11
------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
```

(7) 配置 MPLS TE 隧道

在 Router A 上配置 MPLS TE 隧道。

```
[RouterA] interface tunnel 2
[RouterA-Tunnel2] ip address 8.1.1.1 255.255.255.0
[RouterA-Tunnel2] tunnel-protocol mpls te
[RouterA-Tunnel2] destination 4.4.4.9
[RouterA-Tunnel2] mpls te tunnel-id 10
[RouterA-Tunnel2] mpls te signal-protocol crldp
[RouterA-Tunnel2] mpls te bandwidth 2000
[RouterA-Tunnel2] mpls te commit
[RouterA-Tunnel2] quit
```

配置完成后，在 Router A 上执行 **display interface tunnel** 命令可以看到隧道接口状态为 up。

```
[RouterA] display interface tunnel
Tunnel2 current state: UP
Line protocol current state: UP
Description: Tunnel2 Interface
The Maximum Transmit Unit is 64000
Internet Address is 8.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 4.4.4.9
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    0 packets output,  0 bytes
    0 output error
```

在 Router A 上执行 **display mpls te tunnel-interface** 命令可以看到隧道的详细信息。

```
[RouterA] display mpls te tunnel-interface
  Tunnel Name       :  Tunnel2
  Tunnel Desc       :  Tunnel2 Interface
  Tunnel State Desc :  CR-LSP is Up
  Tunnel Attributes :
    LSP ID              :  1.1.1.9:2
    Session ID          :  10
    Admin State         :  UP                Oper State   :  UP
```

```
   Ingress LSR ID      :  1.1.1.9           Egress LSR ID:  4.4.4.9
   Signaling Prot      :  CR LDP            Resv Style   :  -
   Class Type          :  CLASS 0           Tunnel BW    :  2000 kbps
   Reserved BW         :  2000 kbps
   Setup Priority      :  7                 Hold Priority:  7
   Affinity Prop/Mask  :  0x0/0x0
   Explicit Path Name  :  -
   Tie-Breaking Policy :  None
   Metric Type         :  None
   Record Route        :  Disabled     Record Label :  Disabled
   FRR Flag            :  Disabled     BackUpBW Flag:  Not Supported
   BackUpBW Type       :  -            BackUpBW     :  -
   Route Pinning       :  Disabled
   Retry Limit         :  5            Retry Interval:  10 sec
   Reopt               :  Disabled     Reopt Freq   :  -
   Back Up Type        :  None
   Back Up LSPID       :  -
   Auto BW             :  Disabled      Auto BW Freq :  -
   Min BW              :  -             Max BW       :  -
   Current Collected BW:  -
   Interfaces Protected:  -
   VPN Bind Type       :  NONE
   VPN Bind Value      :  -
   Car Policy          :  Disabled
   Tunnel Group        :  Primary
   Primary Tunnel      :  -
   Backup Tunnel       :  -
   Group Status        :  -
```

在 Router A 上执行 **display ospf mpls-te** 命令可以查看 OSPF TE 的 LSAs。

```
[RouterA] display ospf mpls-te area 0 self-originated

          OSPF Process 100 with Router ID 1.1.1.9

 Area ID                  : 0.0.0.0

 Traffic Engineering LSA's of the database

 ------------------------------------------------

 LSA [ 1 ]

 ------------------------------------------------
   LSA  Type               : Opq-Area
   Opaque Type             : 1
   Opaque ID               : 1
   Advertising Router ID   : 1.1.1.9
   LSA  Age                : 811
   Length                  : 200
```

```
  LSA  Options             : E O
  LS Seq Number            : 8000000D
  CheckSum                 : B1C4

  Link Type                : MultiAccess
  Link ID                  : 10.1.1.2
  Local Interface Address  : 10.1.1.1
  Remote Interface Address : 0.0.0.0
  TE Metric                : 1
  Maximum Bandwidth        : 1250000 bytes/sec
  Maximum Reservable BW    : 625000 bytes/sec
  Admin Group              : 0X0

  Global Pool:
      Unreserved BW [ 0] =625000  bytes/sec
      Unreserved BW [ 1] =625000  bytes/sec
      Unreserved BW [ 2] =625000  bytes/sec
      Unreserved BW [ 3] =625000  bytes/sec
      Unreserved BW [ 4] =625000  bytes/sec
      Unreserved BW [ 5] =625000  bytes/sec
      Unreserved BW [ 6] =625000  bytes/sec
      Unreserved BW [ 7] =625000  bytes/sec
  Sub Pool :
      Unreserved BW [ 8] =   0  bytes/sec
      Unreserved BW [ 9] =   0  bytes/sec
      Unreserved BW [10] =   0  bytes/sec
      Unreserved BW [11] =   0  bytes/sec
      Unreserved BW [12] =   0  bytes/sec
      Unreserved BW [13] =   0  bytes/sec
      Unreserved BW [14] =   0  bytes/sec
      Unreserved BW [15] =   0  bytes/sec

  Bandwidth Constraints:
      BC [ 0] =625000 bytes/sec  BC [ 1] =   0 bytes/sec

  Local OverBooking Multipliers:
      LOM [ 0] =   1     LOM [ 1] =   1


-----------------------------------------------

LSA [ 2 ]

-----------------------------------------------
  LSA  Type                : Opq-Area
  Opaque Type              : 1
  Opaque ID                : 0
  Advertising Router ID    : 1.1.1.9
```

```
LSA  Age                 : 1118
Length                   : 28
LSA  Options             : E O
LS Seq Number            : 8000000B
CheckSum                 : ECBF

MPLS TE Router ID        : 1.1.1.9
```

(8) 使用静态路由沿 MPLS TE 隧道转发流量

```
[RouterA] ip route-static 4.4.4.9 32 tunnel 2 preference 1
```

在 Router A 上执行 **display ip routing-table** 命令，可以看到路由表中有以 Tunnel2 为出接口的静态路由信息。

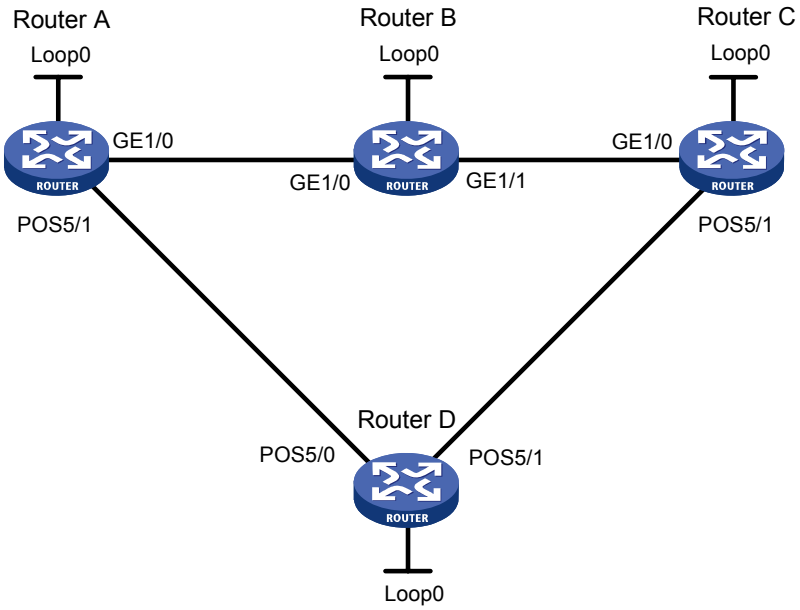
1.15.5 配置 CR-LSP 备份示例

1. 组网需求

要求建立一条 MPLS TE 隧道，使用 CR-LSP 的热备份。隧道头节点为 Router A，目的节点为 Router C。

2. 组网图

图1-11 CR-LSP 备份组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	Loop0	1.1.1.9/32	Router D	Loop0	4.4.4.9/32
	GE1/0	10.1.1.1/24		POS5/0	30.1.1.2/24
	POS5/1	30.1.1.1/24		POS5/1	40.1.1.1/24
Router B	Loop0	2.2.2.9/32	Router C	Loop0	3.3.3.9/32
	GE1/0	10.1.1.2/24		GE1/0	20.1.1.2/24
	GE1/1	20.1.1.1/24		POS5/1	40.1.1.2/24

3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-11配置各接口的IP地址和掩码，包括各Loopback接口，具体配置过程略。

(2) 配置 IGP 协议

在所有设备上配置 IS-IS 协议，发布 LSR ID 的主机路由，具体配置过程略。

配置完成后，在各设备上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到对方 LSR ID 的主机路由。

(3) 配置 MPLS TE 基本能力，并使能 RSVP-TE 和 CSPF

配置 Router A。

```
<RouterA> system-view
[RouterA] mpls lsr-id 1.1.1.9
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls rsvp-te
[RouterA-mpls] mpls te cspf
[RouterA-mpls] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] mpls rsvp-te
```

```
[RouterA-GigabitEthernet1/0] quit
[RouterA] interface pos 5/1
[RouterA-POS5/1] mpls
[RouterA-POS5/1] mpls te
[RouterA-POS5/1] mpls rsvp-te
[RouterA-POS5/1] quit
```



说明

Router B、Router C 和 Router D 的配置与 Router A 相似，此处不再赘述。

(4) 在 RouterA 上建立 MPLS TE Tunnel

配置主 LSP 的 MPLS TE 隧道。

```
[RouterA] interface tunnel 3
[RouterA-Tunnel3] ip address 9.1.1.1 255.255.255.0
[RouterA-Tunnel3] tunnel-protocol mpls te
[RouterA-Tunnel3] destination 3.3.3.9
[RouterA-Tunnel3] mpls te tunnel-id 10
```

使用热备份。

```
[RouterA-Tunnel3] mpls te backup hot-standby
[RouterA-Tunnel3] mpls te commit
[RouterA-Tunnel3] quit
```

配置完成后，在 Router A 上执行 **display interface tunnel** 命令，可以看到 Tunnel3 的状态为 up。

```
[RouterA] display interface tunnel
Tunnel3 current state: UP
Line protocol current state: UP
Description: Tunnel3 Interface
The Maximum Transmit Unit is 64000
Internet Address is 9.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 3.3.3.9
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    0 packets output,  0 bytes
    0 output error
```

(5) 配置完成后的检验

在 Router A 上执行 **display mpls te tunnel** 命令，可以看到存在两条 Tunnel，出接口分别是 GigabitEthernet1/0 和 POS5/1，即，在主 CR-LSP 创建后，备份 CR-LSP 也建立了。

```
[RouterA] display mpls te tunnel
```

```
LSP-Id            Destination     In/Out-If             Name
1.1.1.9:6         3.3.3.9         -/GE1/0               Tunnel3
1.1.1.9:2054      3.3.3.9         -/POS5/1              Tunnel3
```

在 Router A 上执行 **display mpls te tunnel path** 命令，可以看到这两条 Tunnel 使用的路径：

```
[RouterA] display mpls te tunnel path
 Tunnel Interface Name : Tunnel3
 Lsp ID : 1.1.1.9 :6
 Hop Information
  Hop 0   10.1.1.1
  Hop 1   10.1.1.2
  Hop 2   2.2.2.9
  Hop 3   20.1.1.1
  Hop 4   20.1.1.2
  Hop 5   3.3.3.9
 Tunnel Interface Name : Tunnel3
 Lsp ID : 1.1.1.9 :2054
 Hop Information
  Hop 0   30.1.1.1
  Hop 1   30.1.1.2
  Hop 2   4.4.4.9
  Hop 3   40.1.1.1
  Hop 4   40.1.1.2
  Hop 5   3.3.3.9
```

对隧道目的地址进行 tracert 操作：

```
[RouterA] tracert –a 1.1.1.9 3.3.3.9
 traceroute to  3.3.3.9(3.3.3.9) 30 hops max,40 bytes packet
 1 10.1.1.2 25 ms 30.1.1.2 25 ms 10.1.1.2 25 ms
 2 40.1.1.2 45 ms 20.1.1.2 29 ms 40.1.1.2 54 ms
```

从以上显示可以判断出目前使用的 LSP 经过 Router B，没有使用 Router D。

对 Router B 的接口 GigabitEthernet1/1 执行 shutdown，然后再对目的地址进行 tracert 操作，可以看到报文经过了 Router D：

```
[RouterA] tracert –a 1.1.1.9 3.3.3.9
 traceroute to  3.3.3.9(3.3.3.9) 30 hops max,40 bytes packet
 1 30.1.1.2 28 ms  27 ms  23 ms
 2 40.1.1.2 50 ms  50 ms  49 ms
```

在 Router A 上执行 **display mpls te tunnel** 命令，可以看剩一条经过 RouterD 的 Tunnel 了：

```
[RouterA] display mpls te tunnel
LSP-Id            Destination     In/Out-If             Name
1.1.1.9:2054      3.3.3.9         -/POS5/1              Tunnel3
```



说明

普通 CR-LSP 备份与热备份在配置上的唯一区别在于隧道接口上的 **mpls te backup** 命令。对于本配置例，如果采用普通备份方式，则备份 CR-LSP 只有在主 LSP 变为 Down 后才会创建，不会同时存在两条隧道。

---

(6) 使用静态路由沿 MPLS TE 隧道转发流量

```
[RouterA] ip route-static 3.3.3.9 32 tunnel 3 preference 1
```

在 Router A 上执行 **display ip routing-table** 命令，可以看到路由表中有以 Tunnel3 为出接口的静态路由信息。

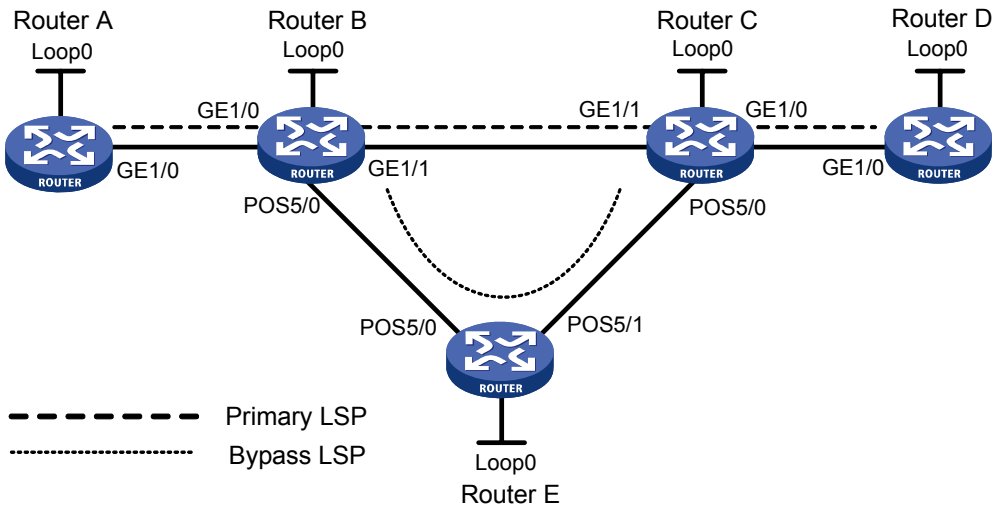
1.15.6 配置快速重路由示例

1. 组网需求

- 主 LSP 是 Router A→Router B→Router C→Router D，要求对 Router B→Router C 这段链路通过 FRR 进行链路保护；
- 建立一条 Bypass LSP，使用路径 Router B→Router E→Router C（Router B 是本地修复节点 PLR，Router C 是汇聚点 MP）；
- 使用显式路径方式建立 MPLS TE 的主隧道和 Bypass 隧道。使用的信令协议是 RSVP-TE。

2. 组网图

图1-12 MPLS TE 快速重路由配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	Loop0	1.1.1.1/32	Router E	Loop0	5.5.5.5/32
	GE1/0	2.1.1.1/24		POS5/0	3.2.1.2/24
Router B	Loop0	2.2.2.2/32		POS5/1	3.3.1.1/24
	GE1/0	2.1.1.2/24	Router C	Loop0	3.3.3.3/32
	GE1/1	3.1.1.1/24		GE1/0	4.1.1.1/24
	POS5/0	3.2.1.1/24		GE1/1	3.1.1.2/24
Router D	Loop0	4.4.4.4/32		POS5/1	3.3.1.2/24
	GE1/0	4.1.1.2/24			

3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-12配置各接口的IP地址和掩码，包括各Loopback接口，具体配置过程略。

(2) 配置 IGP 协议

在所有设备上配置 IS-IS 协议，发布 LSR ID 的主机路由，具体配置过程略。

配置完成后，在各设备上执行 **display ip routing-table** 命令，应可以看到相互之间都学到了到对方 LSR ID 的主机路由。

以 Router A 的显示为例：

```
<RouterA> display ip routing-table
Routing Tables: Public
         Destinations : 13       Routes : 13
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
      1.1.1.1/32  Direct 0    0         127.0.0.1       InLoop0
      2.1.1.0/24  Direct 0    0         2.1.1.1         GE1/0
      2.1.1.1/32  Direct 0    0         127.0.0.1       InLoop0
      2.2.2.2/32  ISIS   15   10        2.1.1.2         GE1/0
      3.1.1.0/24  ISIS   15   20        2.1.1.2         GE1/0
      3.2.1.0/24  ISIS   15   20        2.1.1.2         GE1/0
      3.3.1.0/24  ISIS   15   30        2.1.1.2         GE1/0
      3.3.3.3/32  ISIS   15   20        2.1.1.2         GE1/0
      4.1.1.0/24  ISIS   15   30        2.1.1.2         GE1/0
      4.4.4.4/32  ISIS   15   30        2.1.1.2         GE1/0
      5.5.5.5/32  ISIS   15   20        2.1.1.2         GE1/0
    127.0.0.0/8   Direct 0    0         127.0.0.1       InLoop0
    127.0.0.1/32  Direct 0    0         127.0.0.1       InLoop0
```

(3) 配置 MPLS TE 基本能力，并使能 RSVP-TE 和 CSPF

配置 Router A。

```
<RouterA> system-view
[RouterA] mpls lsr-id 1.1.1.1
[RouterA] mpls
[RouterA-mpls] mpls te
[RouterA-mpls] mpls rsvp-te
[RouterA-mpls] mpls te cspf
[RouterA-mpls] quit
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] mpls
[RouterA-GigabitEthernet1/0] mpls te
[RouterA-GigabitEthernet1/0] mpls rsvp-te
[RouterA-GigabitEthernet1/0] quit
```

配置 Router B。

```
<RouterB> system-view
[RouterB] mpls lsr-id 2.2.2.2
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls rsvp-te
[RouterB-mpls] mpls te cspf
[RouterB-mpls] quit
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] mpls
[RouterB-GigabitEthernet1/0] mpls te
[RouterB-GigabitEthernet1/0] mpls rsvp-te
[RouterB-GigabitEthernet1/0] quit
[RouterB] interface gigabitethernet 1/1
```

```
[RouterB-GigabitEthernet1/1] mpls
[RouterB-GigabitEthernet1/1] mpls te
[RouterB-GigabitEthernet1/1] mpls rsvp-te
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface pos 5/0
[RouterB-POS5/0] mpls
[RouterB-POS5/0] mpls te
[RouterB-POS5/0] mpls rsvp-te
[RouterB-POS5/0] quit
```



说明

Router C、Router D、Router E 的配置与 Router A 和 Router B 的配置相同，此处不再赘述。

(4) 在主 LSP 的头节点 Router A 上建立 MPLS TE Tunnel

配置主 LSP 的显式路径。

```
[RouterA] explicit-path pri-path
[RouterA-explicit-path-pri-path] next hop 2.1.1.2
[RouterA-explicit-path-pri-path] next hop 3.1.1.2
[RouterA-explicit-path-pri-path] next hop 4.1.1.2
[RouterA-explicit-path-pri-path] next hop 4.4.4.4
[RouterA-explicit-path-pri-path] quit
```

配置主 LSP 的 MPLS TE 隧道。

```
[RouterA] interface tunnel 4
[RouterA-Tunnel4] ip address 10.1.1.1 255.255.255.0
[RouterA-Tunnel4] tunnel-protocol mpls te
[RouterA-Tunnel4] destination 4.4.4.4
[RouterA-Tunnel4] mpls te tunnel-id 10
[RouterA-Tunnel4] mpls te path explicit-path pri-path preference 1
```

使能 FRR。

```
[RouterA-Tunnel4] mpls te fast-reroute
[RouterA-Tunnel4] mpls te commit
[RouterA-Tunnel4] quit
```

配置完成后，在 Router A 上执行 **display interface tunnel** 命令，可以看到 Tunnel4 的状态为 up。

```
[RouterA] display interface tunnel
Tunnel4 current state: UP
Line protocol current state: UP
Description: Tunnel4 Interface
The Maximum Transmit Unit is 64000
Internet Address is 10.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 4.4.4.4
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
```

```
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  0 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    0 packets output,  0 bytes
    0 output error
```

在 Router A 上执行 **display mpls te tunnel-interface** 命令，可以看到隧道接口的详细信息。

```
[RouterA] display mpls te tunnel-interface
  Tunnel Name         :  Tunnel4
  Tunnel Desc         :  Tunnel4 Interface
  Tunnel State Desc   :  CR-LSP is Up
  Tunnel Attributes   :
    LSP ID              :  1.1.1.1:1
    Session ID          :  10
    Admin State         :  UP               Oper State   :  UP
    Ingress LSR ID      :  1.1.1.1          Egress LSR ID:  4.4.4.4
    Signaling Prot      :  RSVP             Resv Style   :  SE
    Class Type          :  CLASS 0          Tunnel BW    :  0 kbps
    Reserved BW         :  0 kbps
    Setup Priority      :  7                Hold Priority:  7
    Affinity Prop/Mask  :  0x0/0x0
    Explicit Path Name  :  pri-path
    Tie-Breaking Policy :  None
    Metric Type         :  None
    Record Route        :  Enabled      Record Label :  Enabled
    FRR Flag            :  Enabled      BackUpBW Flag:  Not Supported
    BackUpBW Type       :  -            BackUpBW     :  -
    Route Pinning       :  Disabled
    Retry Limit         :  5                 Retry Interval:  10 sec
    Reopt               :  Disabled          Reopt Freq   :  -
    Back Up Type        :  None
    Back Up LSPID       :  -
    Auto BW             :  Disabled          Auto BW Freq :  -
    Min BW              :  -                 Max BW       :  -
    Current Collected BW:  -
    Interfaces Protected:  -
    VPN Bind Type       :  NONE
    VPN Bind Value      :  -
    Car Policy          :  Disabled
    Tunnel Group        :  Primary
    Primary Tunnel      :  -
    Backup Tunnel       :  -
    Group Status        :  -
```

(5)　在作为 PLR 的 Router B 上配置 Bypass Tunnel

配置 Bypass LSP 的显式路径。

```
[RouterB] explicit-path by-path
[RouterB-explicit-path-by-path] next hop 3.2.1.2
```

```
[RouterB-explicit-path-by-path] next hop 3.3.1.2
[RouterB-explicit-path-by-path] next hop 3.3.3.3
[RouterB-explicit-path-by-path] quit
```

配置 Bypass Tunnel。

```
[RouterB] interface tunnel 5
[RouterB-Tunnel5] ip address 11.1.1.1 255.255.255.0
[RouterB-Tunnel5] tunnel-protocol mpls te
[RouterB-Tunnel5] destination 3.3.3.3
[RouterB-Tunnel5] mpls te tunnel-id 15
[RouterB-Tunnel5] mpls te path explicit-path by-path
```

配置 Bypass Tunnel 可保护的带宽。

```
[RouterB-Tunnel5] mpls te backup bandwidth 10000
[RouterB-Tunnel5] mpls te commit
[RouterB-Tunnel5] quit
```

将 Bypass Tunnel 绑定到被保护的接口。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] mpls te fast-reroute bypass-tunnel tunnel 5
[RouterB-GigabitEthernet1/1] quit
```

配置完成后，在 Router B 上执行 **display interface tunnel** 命令可以看到接口 Tunnel5 的状态为 up。

在所有设备上执行 **display mpls lsp** 命令，可以看到 LSP 表项，并且，Router B 和 Router C 上有两条 LSP 经过。

```
[RouterA] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: RSVP LSP
-------------------------------------------------------------------
FEC                In/Out Label  In/Out IF               Vrf Name
4.4.4.4/32         NULL/1024     -/GE1/0
[RouterB] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: RSVP LSP
-------------------------------------------------------------------
FEC             In/Out Label  In/Out IF                Vrf Name
4.4.4.4/32      1024/1024     GE1/0/GE1/1
3.3.3.3/32      NULL/1024     -/POS5/0
[RouterC] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: RSVP LSP
-------------------------------------------------------------------
FEC               In/Out Label  In/Out IF               Vrf Name
4.4.4.4/32        1024/3        GE1/1/GE1/0
3.3.3.3/32        3/NULL        POS5/0/-
[RouterD] display mpls lsp
-------------------------------------------------------------------
                 LSP Information: RSVP LSP
```

```
------------------------------------------------------------------------
FEC              In/Out Label  In/Out IF                Vrf Name
4.4.4.4/32       3/NULL        GE1/0/-
[RouterE] display mpls lsp
------------------------------------------------------------------------
                 LSP Information: RSVP LSP
------------------------------------------------------------------------
FEC              In/Out Label  In/Out IF                 Vrf Name
3.3.3.3/32       1024/3        POS5/0/POS5/1
```

在所有设备上执行 **display mpls te tunnel** 命令，可以看到隧道的建立情况，Router B 和 Router C 上有两条隧道经过。

```
[RouterA] display mpls te tunnel
LSP-Id           Destination     In/Out-If           Name
1.1.1.1:1        4.4.4.4         -/GE1/0            Tunnel4
[RouterB] display mpls te tunnel
LSP-Id           Destination     In/Out-If           Name
1.1.1.1:1        4.4.4.4         GE1/0/GE1/1       Tunnel4
2.2.2.2:1        3.3.3.3         -/POS5/0            Tunnel5
[RouterC] display mpls te tunnel
LSP-Id           Destination     In/Out-If            Name
1.1.1.1:1        4.4.4.4         GE1/1/GE1/0        Tunnel4
2.2.2.2:1        3.3.3.3         POS5/0/-             Tunnel5
[RouterD] display mpls te tunnel
LSP-Id           Destination     In/Out-If            Name
1.1.1.1:1        4.4.4.4         GE1/0/-             Tunnel4
[RouterE] display mpls te tunnel
LSP-Id           Destination     In/Out-If            Name
2.2.2.2:1        3.3.3.3         POS5/0/POS5/1        Tunnel5
```

在 Router B 上执行 **display mpls lsp verbose** 命令，可以看到 Bypass 隧道绑定到出接口 GigabitEthernet1/1，目前未被使用。

```
[RouterB] display mpls lsp verbose
------------------------------------------------------------------------
                 LSP Information: RSVP LSP
------------------------------------------------------------------------
  No                  :  1
  IngressLsrID        :  1.1.1.1
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel4
  Fec                 :  4.4.4.4/32
  Nexthop             :  3.1.1.2
  In-Label            :  1024
  Out-Label           :  1024
  In-Interface        :  GE1/0
  Out-Interface       :  GE1/1
  LspIndex            :  4097
  Tunnel ID           :  0x22001
  LsrType             :  Transit
```

```
  Bypass In Use       :  Not Used
  BypassTunnel        :  Tunnel Index[Tunnel5], InnerLabel[1024]
  Mpls-Mtu            :  1500

  No                  :  2
  IngressLsrID        :  2.2.2.2
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel5
  Fec                 :  3.3.3.3/32
  Nexthop             :  3.2.1.2
  In-Label            :  NULL
  Out-Label           :  1024
  In-Interface        :  ----------
  Out-Interface       :  POS5/0
  LspIndex            :  4098
  Tunnel ID           :  0x22002
  LsrType             :  Ingress
  Bypass In Use       :  Not Exists
  BypassTunnel        :  Tunnel Index[---]
  Mpls-Mtu            :  1500
```

(6) 配置完成后的检验

使 PLR 上被保护的出接口失效。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] shutdown
%Sep    7  08:53:34  2004  RouterB  IFNET/5/UPDOWN:Line  protocol  on  the  interface
GigabitEthernet1/1 turns into DOWN state
```

在 Router A 上查看主 LSP 的状态，执行命令 **display interface tunnel** 4，可以看到 Tunnel 接口仍然处于 up 状态。

在 Router A 上执行 **display mpls te tunnel-interface** 命令，可以看到隧道接口的详细信息。

```
[RouterA] display mpls te tunnel-interface
  Tunnel Name         : Tunnel4
  Tunnel Desc         : Tunnel4 Interface
  Tunnel State Desc   : CR-LSP is up
  Tunnel Attributes   :
    LSP ID              :  1.1.1.1:1
    Session ID          :  10
    Admin State         :  UP              Oper State    :  UP
    Ingress LSR ID      :  1.1.1.1         Egress LSR ID:  4.4.4.4
    Signaling Prot      :  RSVP            Resv Style    :  SE
    Class Type          :  CLASS 0         Tunnel BW     :  0 kbps
    Reserved BW         :  0 kbps
    Setup Priority      :  7               Hold Priority:  7
    Affinity Prop/Mask  :  0/0
    Explicit Path Name  :  pri-path
    Tie-Breaking Policy :  None
    Metric Type         :  None
```

```
  Record Route        :  Enabled      Record Label :  Enabled
  FRR Flag            :  Enabled      BackUpBW Flag:  Not Supported
  BackUpBW Type       :  -            BackUpBW     :  -
  Route Pinning       :  Disabled
  Retry Limit         :  5                 Retry Interval:  10 sec
  Reopt               :  Disabled          Reopt Freq   :  -
  Back Up Type        :  None
  Back Up LSPID       :  -
  Auto BW             :  Disabled             Auto BW Freq :  -
  Min BW              :  -                    Max BW       :  -
  Current Collected BW:  -
  Interfaces Protected:  -
  VPN Bind Type       :  NONE
  VPN Bind Value      :  -
  Car Policy          :  Disabled
  Tunnel Group        :  Primary
  Primary Tunnel      :  -
  Backup Tunnel       :  -
  Group Status        :  -
Tunnel Name         :  Tunnel4
Tunnel Desc         :  Tunnel4 Interface
Tunnel State Desc   :  CR-LSP is up
Tunnel Attributes   :
  LSP ID              :  1.1.1.1:1025
  Session ID          :  10
  Admin State         :                Oper State   :  Modified
  Ingress LSR ID      :  1.1.1.1       Egress LSR ID:  4.4.4.4
  Signaling Prot      :  RSVP          Resv Style   :  SE
  Class Type          :  CLASS 0       Tunnel BW    :  0 kbps
  Reserved BW         :  0 kbps
  Setup Priority      :  7             Hold Priority:  7
  Affinity Prop/Mask  :  0/0
  Explicit Path Name  :  pri-path
  Tie-Breaking Policy :  None
  Metric Type         :  None
  Record Route        :  Enabled      Record Label :  Enabled
  FRR Flag            :  Enabled      BackUpBW Flag:  Not Supported
  BackUpBW Type       :  -            BackUpBW     :  -
  Route Pinning       :  Disabled
  Retry Limit         :  5                 Retry Interval:  10 sec
  Reopt               :  Disabled          Reopt Freq   :  -
  Back Up Type        :  None
  Back Up LSPID       :  -
  Auto BW             :  Disabled          Auto BW Freq :  -
  Min BW              :  -                 Max BW       :  -
  Current Collected BW:  -
  Interfaces Protected:  -
  VPN Bind Type       :  NONE
```

```
  VPN Bind Value      :  -
  Car Policy          :  Disabled
  Tunnel Group        :  Primary
  Primary Tunnel      :  -
  Backup Tunnel       :  -
  Group Status        :  -
```



说明

如果在 FRR 切换后马上执行 **display mpls te tunnel-interface** 命令查看隧道接口的详细信息，会看到两条处于 up 状态的 CR-LSP。这是因为 FRR 采用 **make-before-break** 方式建立新的 LSP，旧的 LSP 在新 LSP 建立成功后过一段时间才被删除。

在 Router B 上执行 **display mpls lsp verbose** 命令，可以看到 Bypass 隧道被使用。

```
[RouterB] display mpls lsp verbose
------------------------------------------------------------------------
                 LSP Information: RSVP LSP
------------------------------------------------------------------------
  No                  :  1
  IngressLsrID        :  1.1.1.1
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel4
  Fec                 :  4.4.4.4/32
  Nexthop             :  3.1.1.2
  In-Label            :  1024
  Out-Label           :  1024
  In-Interface        :  GE1/0
  Out-Interface       :  GE1/1
  LspIndex            :  4097
  Tunnel ID           :  0x22001
  LsrType             :  Transit
  Bypass In Use       :  In Use
  BypassTunnel        :  Tunnel Index[Tunnel5], InnerLabel[1024]
  Mpls-Mtu            :  1500

  No                  :  2
  IngressLsrID        :  2.2.2.2
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel5
  Fec                 :  3.3.3.3/32
  Nexthop             :  3.2.1.2
  In-Label            :  NULL
  Out-Label           :  1024
  In-Interface        :  ----------
  Out-Interface       :  POS5/0
  LspIndex            :  4098
  Tunnel ID           :  0x22002
  LsrType             :  Ingress
```

```
  Bypass In Use       :  Not Exists
  BypassTunnel        :  Tunnel Index[---]
  Mpls-Mtu            :  1500
```

在 PLR 上配置 FRR 的扫描定时器为 5 秒。

```
[RouterB] mpls
[RouterB-mpls] mpls te timer fast-reroute 5
[RouterB-mpls] quit
```

使 PLR 上被保护的出接口重新生效。

```
[RouterB] intrface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] undo shutdown
%Sep    7  09:01:31  2004  RouterB  IFNET/5/UPDOWN:Line  protocol  on  the  interface
GigabitEthernet1/1 turns into UP state
```

在 Router A 上查看主 LSP 的状态，执行命令 **display interface Tunnel4**，可以看到 Tunnel 接口处于 up 状态。

等待约 5 秒后，在 Router B 上执行 **display mpls lsp verbose** 命令，可以看到 Tunnel5 仍绑定到出接口 GigabitEthernet1/1，未被使用。

(7) 使用静态路由沿 MPLS TE 隧道转发流量

```
[RouterA] ip route-static 4.1.1.2 24 tunnel 4 preference 1
```

在 Router A 上执行 **display ip routing-table** 命令，可以看到路由表中有以 Tunnel3 为出接口的静态路由信息。

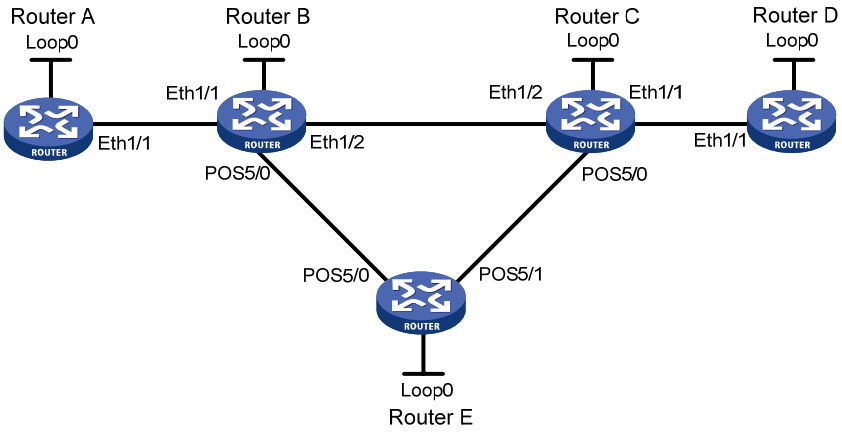
1.15.7 配置 MPLS LDP over MPLS TE 示例

1. 组网需求

- Router A、Route B、Route C、Route D 和 Route E 均支持 MPLS，运行 OSPF 作为 MPLS 网络中的 IGP；
- Route A 和 Route B、Route C 和 Route D 之间建立本地 LDP 会话；
- Route B 和 Route C 之间建立远端 LDP 会话；
- 使用 RSVP-TE 信令协议在 Router B→Router E→Router C 上建立 MPLS TE 的隧道。

2. 组网图

图1-13 MPLS LDP over MPLS TE 组网图（路由应用）



设备	接口	IP 地址	设备	接口	IP 地址
Router A	Loop0	1.1.1.1/32	Router E	Loop0	5.5.5.5/32
	Eth1/1	2.1.1.1/24		POS5/0	3.2.1.2/24
Router B	Loop0	2.2.2.2/32		POS5/1	3.3.1.1/24
	Eth1/1	2.1.1.2/24	Router C	Loop0	3.3.3.3/32
	Eth1/2	3.1.1.1/24		Eth1/1	4.1.1.1/24
	POS5/0	3.2.1.1/24		Eth1/2	3.1.1.2/24
Router D	Loop0	4.4.4.4/32		POS5/1	3.3.1.2/24
	Eth1/1	4.1.1.2/24			

3. 配置步骤

(1) 配置各接口的 IP 地址

按照 图 1-13配置各接口的IP地址和掩码，包括各Loopback接口，具体配置过程略。

(2) 配置 IGP 协议

在所有设备上配置 OSPF 协议，发布 LSR ID 的主机路由和各接口网段路由，具体配置过程略。

(3) 配置 MPLS LDP 基本能力

在 Router A 和 Router D 上配置 MPLS LDP 基本能力，具体配置过程略。

(4) 配置 MPLS TE

在 Router B、Router C、Router E 上配置 MPLS TE 基本能力，并使能 RSVP-TE 和 CSPF。

配置 Router B。

```
<RouterB> system-view
[RouterB] mpls lsr-id 2.2.2.2
[RouterB] mpls
[RouterB-mpls] mpls te
[RouterB-mpls] mpls rsvp-te
[RouterB-mpls] mpls te cspf
[RouterB-mpls] quit
[RouterB] interface pos 5/0
[RouterB-POS5/0] mpls
[RouterB-POS5/0] mpls te
[RouterB-POS5/0] mpls rsvp-te
[RouterB-POS5/0] quit
```

配置 Router E。

```
<RouterE> system-view
[RouterE] mpls lsr-id 5.5.5.5
[RouterE] mpls
[RouterE-mpls] mpls te
[RouterE-mpls] mpls rsvp-te
[RouterE-mpls] mpls te cspf
[RouterE-mpls] quit
[RouterE] interface pos 5/0
[RouterE-POS5/0] mpls
[RouterE-POS5/0] mpls te
[RouterE-POS5/0] mpls rsvp-te
[RouterE-POS5/0] quit
[RouterE] interface pos 5/1
[RouterE-POS5/1] mpls
[RouterE-POS5/1] mpls te
[RouterE-POS5/1] mpls rsvp-te
[RouterE-POS5/1] quit
```

配置 Router C。

```
<RouterC> system-view
[RouterC] mpls lsr-id 3.3.3.3
[RouterC] mpls
[RouterC-mpls] mpls te
[RouterC-mpls] mpls rsvp-te
[RouterC-mpls] mpls te cspf
[RouterC-mpls] quit
[RouterC] interface pos 5/0
[RouterC-POS5/0] mpls
[RouterC-POS5/0] mpls te
[RouterC-POS5/0] mpls rsvp-te
[RouterC-POS5/0] quit
```

(5) 在 Router B 上建立 MPLS TE 隧道

配置 MPLS TE 隧道。

```
[RouterB] interface tunnel 4
[RouterB-Tunnel4] ip address 10.1.1.1 255.255.255.0
[RouterB-Tunnel4] tunnel-protocol mpls te
[RouterB-Tunnel4] destination 3.3.3.3
[RouterB-Tunnel4] mpls te tunnel-id 10
```

配置 IGP Shortcut。

```
[RouterB-Tunnel4] mpls te igp shortcut
[RouterB-Tunnel4] mpls te igp metric relative -1
[RouterB-Tunnel4] mpls te commit
```

使能 MPLS。

```
[RouterB-Tunnel4] mpls
[RouterB-Tunnel4] quit
```

配置 OSPF TE。

```
[RouterB] ospf
[RouterB-ospf-1] opaque-capability enable
[RouterB-ospf-1] enable traffic-adjustment
[RouterB-ospf-1] area 0.0.0.0
[RouterB-ospf-1-area-0.0.0.0] mpls-te enable
[RouterB-ospf-1-area-0.0.0.0] quit
[RouterB-ospf-1] quit
```

配置完成后，在 Router B 上执行 **display interface tunnel** 命令，可以看到隧道接口 Tunnel4 的状态为 up。

```
[RouterB] display interface tunnel
Tunnel4 current state: UP
Line protocol current state: UP
Description: Tunnel4 Interface
The Maximum Transmit Unit is 64000
Internet Address is 10.1.1.1/24 Primary
Encapsulation is TUNNEL, service-loopback-group ID not set.
Tunnel source unknown, destination 3.3.3.3
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  0 bytes/sec, 0 packets/sec
    Last 300 seconds output:  8 bytes/sec, 0 packets/sec
    0 packets input,  0 bytes
    0 input error
    47 packets output,  2740 bytes
    0 output error
```

在 Router B 上执行 **display ip routing-table** 命令，可以看到从 Router B 到 Router C 的路由出接口为隧道接口 Tunnel4。

```
[RouterB] display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

1.1.1.1/32          OSPF   10   1            2.1.1.1         Eth1/1
2.1.1.0/24          Direct 0    0            2.1.1.2         Eth1/1
2.1.1.2/32          Direct 0    0            127.0.0.1       InLoop0
2.2.2.2/32          Direct 0    0            127.0.0.1       InLoop0
3.1.1.0/24          Direct 0    0            3.1.1.1         Eth1/2
3.1.1.1/32          Direct 0    0            127.0.0.1       InLoop0
3.3.3.3/32          OSPF   10   1            10.1.1.1        Tun4
4.1.1.0/24          OSPF   10   2            10.1.1.1        Tun4
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

注意

- 必须为 MPLS TE 隧道配置 IGP Shortcut 或配置静态路由，使得 IP 路由优选 MPLS TE 隧道接口为出接口。
- MPLS TE 隧道接口上必须使能 MPLS。
- MPLS TE 隧道的目的地址必须配置为隧道尾节点的 LSR ID。

(6) 在 Router B 和 Router C 上配置 LDP 能力

配置 Router B。

```
[RouterB] mpls
[RouterB-mpls] mpls ldp
[RouterB-mpls] quit
[RouterB] interface ethernet 1/1
[RouterB-Ethernet1/1] mpls
[RouterB-Ethernet1/1] mpls ldp
[RouterB-Ethernet1/1] quit
[RouterB] mpls ldp remote-peer C
[RouterB-mpls-ldp-remote-c] remote-ip 3.3.3.3
[RouterB-mpls-ldp-remote-c] quit
```

配置 Router C。

```
[RouterC] mpls
[RouterC-mpls] mpls ldp
[RouterC-mpls] quit
[RouterC] interface ethernet 1/1
[RouterC-Ethernet1/1] mpls
[RouterC-Ethernet1/1] mpls ldp
[RouterC-Ethernet1/1] quit
[RouterC] mpls ldp remote-peer b
[RouterC-mpls-ldp-remote-b] remote-ip 2.2.2.2
[RouterC-mpls-ldp-remote-b] prefix-label advertise
[RouterC-mpls-ldp-remote-b] quit
```

在 Router B 上执行 **display mpls ldp session** 命令，可以看到 Router B 和 Router C 之间建立了远端会话。

```
[RouterB] display mpls ldp session verbose

               LDP Session(s) in Public Network
 ------------------------------------------------------------------------
 Peer LDP ID     : 1.1.1.1:0            Local LDP ID   : 2.2.2.2:0
 TCP Connection  : 2.2.2.2 -> 1.1.1.1
 Session State   : Operational          Session Role   : Active
 Session FT Flag : Off                  MD5 Flag       : Off
 Reconnect Timer : ---                  Recovery Timer : ---

 Negotiated Keepalive Timer        : 45 Sec
 Keepalive Message Sent/Rcvd       : 437/437 (Message Count)
```

```
 Label Advertisement Mode          : Downstream Unsolicited
 Label Resource Status(Peer/Local) : Available/Available
 Peer Discovery Mechanism          : Basic
 Session existed time              : 000:01:48  (DDD:HH:MM)
 LDP Basic Discovery Source        : Ethernet1/1

 Addresses received from peer: (Count: 2)
 2.1.1.1            1.1.1.1

 ------------------------------------------------------------------------
 Peer LDP ID     : 3.3.3.3:0            Local LDP ID   : 2.2.2.2:0
 TCP Connection  : 2.2.2.2 <- 3.3.3.3
 Session State   : Operational          Session Role   : Passive
 Session FT Flag : Off                  MD5 Flag       : Off
 Reconnect Timer : ---                  Recovery Timer : ---

 Negotiated Keepalive Timer        : 45 Sec
 Keepalive Message Sent/Rcvd       : 411/411 (Message Count)
 Label Advertisement Mode          : Downstream Unsolicited
 Label Resource Status(Peer/Local) : Available/Available
 Peer Discovery Mechanism          : Extended
 Session existed time              : 000:01:42  (DDD:HH:MM)
 LDP Extended Discovery Source     : Remote peer: 1

 Addresses received from peer: (Count: 2)
 4.1.1.1            3.3.3.3

 ------------------------------------------------------------------------
```

在 Router B 上执行 **display mpls ldp lsp** 命令，可以看到 Router C 通过远端会话给 Router B 发送了标签映射消息，并与 Router B 建立了 LDP LSP。

```
[RouterB] display mpls ldp lsp

                          LDP LSP Information
 ------------------------------------------------------------------------
 SN     DestAddress/Mask   In/OutLabel   Next-Hop        In/Out-Interface
 ------------------------------------------------------------------------
 1      1.1.1.1/32         NULL/3        2.1.1.1         -------/Eth1/1
 2      2.2.2.2/32         3/NULL        127.0.0.1       -------/InLoop0
 3      3.3.3.3/32         NULL/3        10.1.1.1        -------/Tun4
 4      3.3.3.3/32         1024/3        10.1.1.1        -------/Tun4
 ------------------------------------------------------------------------
A '*' before an LSP means the LSP is not established
A '*' before a Label means the USCB or DSCB is stale
```



注意

LDP 远端对等体的 remote-ip 必须配置为 MPLS TE 隧道尾节点的 LSR ID。

(7) 配置完成后的检验

在 Router B 上执行 **display mpls lsp** 命令，可以看到从 Router B 到 Router C 的 LDP LSP 和 MPLS TE 隧道已经嵌套关联，即 LDP LSP 的出接口为 MPLS TE 隧道接口。

```
[RouterB] display mpls lsp include 3.3.3.3 32 verbose
------------------------------------------------------------------------------
                 LSP Information: RSVP LSP
------------------------------------------------------------------------------
  No                  :  1
  IngressLsrID        :  2.2.2.2
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel4
  Fec                 :  3.3.3.3/32
  Nexthop             :  3.2.1.2
  In-Label            :  NULL
  Out-Label           :  1024
  In-Interface        :  ----------
  Out-Interface       :  POS5/0
  LspIndex            :  3073
  Tunnel ID           :  0x11000c
  LsrType             :  Ingress
  Bypass In Use       :  Not Exists
  BypassTunnel        :  Tunnel Index[---]

------------------------------------------------------------------------------
                 LSP Information: LDP LSP
------------------------------------------------------------------------------
  No                  :  2
  VrfIndex            :
  Fec                 :  3.3.3.3/32
  Nexthop             :  10.1.1.1
  In-Label            :  NULL
  Out-Label           :  3
  In-Interface        :  ----------
  Out-Interface       :  Tunnel4
  LspIndex            :  6147
  Tunnel ID           :  0x11000e
  LsrType             :  Ingress
  Outgoing Tunnel ID  :  0x15000d
  Label Operation     :  PUSH

  No                  :  3
  VrfIndex            :
  Fec                 :  3.3.3.3/32
  Nexthop             :  10.1.1.1
  In-Label            :  1024
  Out-Label           :  3
  In-Interface        :  ----------
```

```
Out-Interface       :  Tunnel4
LspIndex            :  6148
Tunnel ID           :  0x11000f
LsrType             :  Transit
Outgoing Tunnel ID  :  0x15000d
Label Operation     :  SWAP
```

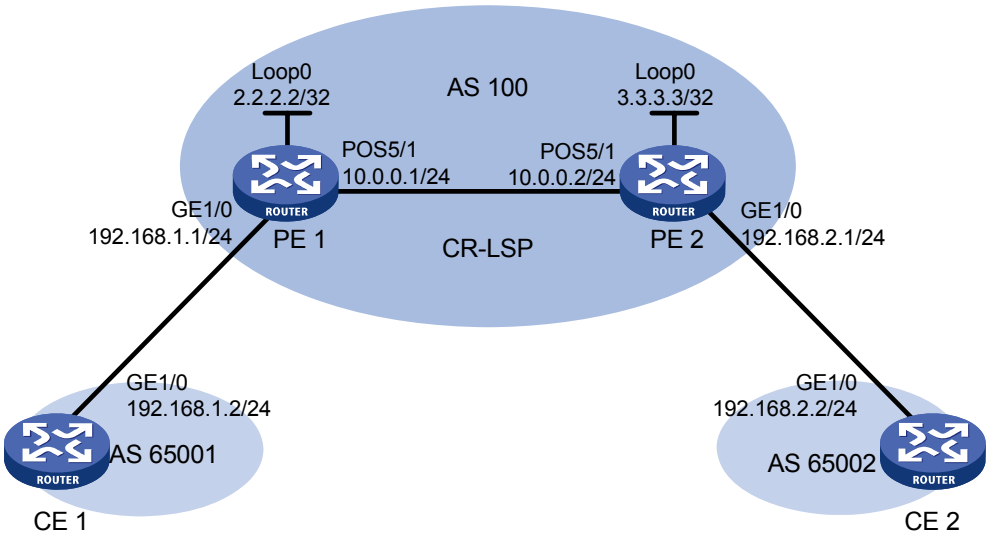
1.15.8 配置 MPLS L3VPN 中的 MPLS TE 示例

1. 组网需求

- CE 1 和 CE 2 都属于 VPN 1，分别通过 PE 1 和 PE 1 接入 MPLS 骨干网。MPLS 骨干网使用 OSPF 作为 IGP 协议；
- 配置 MPLS TE 隧道转发 PE 1 到 PE 2 的 VPN 流量；
- 为了使 TE 隧道能够转发 MPLS L3VPN 的流量，需要在建立 VPN 时配置隧道策略，使用 CR-LSP 作为 VPN 隧道。

2. 组网图

图1-14 在 VPN 中配置 MPLS TE 组网图



3. 配置步骤

(1) 配置 IGP 协议 OSPF，使 PE 1 和 PE 2 能学习到对方的路由。

配置 PE 1。

```
[PE1] system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.2 255.255.255.255
[PE1-LoopBack0] quit
[PE1] interface pos 5/1
[PE1-POS5/1] ip address 10.0.0.1 255.255.255.0
[PE1-POS5/1] quit
[PE1] ospf
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 10.0.0.0 0.0.0.255
```

```
[PE1-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

配置 PE 2。

```
[PE2] system-view
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 3.3.3.3 255.255.255.255
[PE2-LoopBack0] quit
[PE2] interface pos 5/1
[PE2-POS5/1] ip address 10.0.0.2 255.255.255.0
[PE2-POS5/1] quit
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 10.0.0.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

上述配置完成后，PE 之间应能建立 OSPF 邻居，执行 **display ospf peer** 命令可以看到邻居达到 FULL 状态。执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback 路由。以 PE 1 为例：

```
[PE1] display ospf peer
          OSPF Process 1 with Router ID 2.2.2.2
                  Neighbors
 Area 0.0.0.0 interface 10.0.0.1(POS5/1)'s neighbors
 Router ID: 3.3.3.3          Address: 10.0.0.2         GR State: Normal
   State: Full  Mode:Nbr is  Master  Priority: 1
   DR: None   BDR: None
   Dead timer due in 30  sec
   Neighbor is up for 00:01:00
   Authentication Sequence: [ 0 ]
[PE1] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7
Destination/Mask  Proto  Pre  Cost     NextHop         Interface
      2.2.2.2/32  Direct 0    0        127.0.0.1       InLoop0
      3.3.3.3/32  OSPF   10   1563     10.0.0.2        POS5/1
     10.0.0.0/24  Direct 0    0        10.0.0.1        POS5/1
     10.0.0.1/32  Direct 0    0        127.0.0.1       InLoop0
     10.0.0.2/32  Direct 0    0        10.0.0.2        POS5/1
    127.0.0.0/8   Direct 0    0        127.0.0.1       InLoop0
    127.0.0.1/32  Direct 0    0        127.0.0.1       InLoop0
```

(2) 配置 MPLS 基本能力和 LDP

配置 PE 1。

```
[PE1] mpls lsr-id 2.2.2.2
[PE1] mpls
[PE1-mpls] lsp-trigger all
```

```
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface pos 1/1
[PE1-POS5/1] mpls
[PE1-POS5/1] mpls ldp
[PE1-POS5/1] quit
```

配置 PE2。

```
[PE2] mpls lsr-id 3.3.3.3
[PE2] mpls
[PE2-mpls] lsp-trigger all
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface pos 1/1
[PE2-POS5/1] mpls
[PE2-POS5/1] mpls ldp
[PE2-POS5/1] quit
```

上述配置完成后，PE 之间应能建立 LDP 会话，执行 **display mpls ldp session** 命令可以看到显示结果中 Session State 项为 Operational。

以 PE 1 的显示为例：

```
[PE1] display mpls ldp session
               LDP Session(s) in Public Network
 ------------------------------------------------------------------
 Peer-ID         Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 3.3.3.3:0       Operational   DU   Passive  Off  Off  2/2
 ------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
```

(3) 使能 MPLS TE、CSPF 和 OSPF TE

配置 PE 1。

```
[PE1] mpls
[PE1-mpls] mpls te
[PE1-mpls] mpls te cspf
[PE1-mpls] quit
[PE1] interface pos 5/1
[PE1-POS5/1] mpls te
[PE1-POS5/1] quit
[PE1] ospf
[PE1-ospf-1] opaque-capability enable
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] mpls-te enable
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

配置 PE 2。

```
[PE2] mpls
[PE2-mpls] mpls te
[PE2-mpls] mpls te cspf
[PE2-mpls] quit
[PE2] interface pos 5/1
[PE2-POS5/1] mpls te
[PE2-POS5/1] quit
[PE2] ospf
[PE2-ospf-1] opaque-capability enable
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] mpls-te enable
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

(4) 配置 MPLS TE 隧道

以 PE 1 为首端创建一条 TE 隧道，尾端为 PE 2，信令协议为 CR-LDP。

```
[PE1] interface tunnel 1
[PE1-Tunnel1] ip address 12.1.1.1 255.255.255.0
[PE1-Tunnel1] tunnel-protocol mpls te
[PE1-Tunnel1] destination 3.3.3.3
[PE1-Tunnel1] mpls te tunnel-id 10
[PE1-Tunnel1] mpls te signal-protocol crldp
[PE1-Tunnel1] mpls te commit
[PE1-Tunnel1] quit
```

上述配置完成后，在 PE 1 上执行 **display interface tunnel** 命令可以看到隧道接口状态为 up。

(5) 在 PE 设备上配置 VPN 实例，并绑定到连接 CE 的接口上。

配置 CE 1。

```
[CE1] interface gigabitethernet 1/0
[CE1-GigabitEthernet1/0] ip address 192.168.1.2 255.255.255.0
[CE1-GigabitEthernet1/0] quit
```

在 PE 1 上配置 VPN 实例，使用 CR-LSP 建立 VPN，并将此实例绑定到连接 CE 1 的接口。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 100:1 both
[PE1-vpn-instance-vpn1] tnl-policy policy1
[PE1-vpn-instance-vpn1] quit
[PE1] tunnel-policy policy1
[PE1-tunnel-policy-policy1] tunnel select-seq cr-lsp load-balance-number 1
[PE1-tunnel-policy-policy1] quit
[PE1] interface gigabitethernet 1/0
[PE1-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/0] ip address 192.168.1.1 255.255.255.0
[PE1-GigabitEthernet1/0] quit
```

配置 CE 2。

```
[CE2] interface gigabitethernet 1/0
[CE2-GigabitEthernet1/0] ip address 192.168.2.2 255.255.255.0
```

```
[CE2-GigabitEthernet1/0] quit
```

在 PE 2 上配置 VPN 实例，并将此实例绑定到连接 CE 2 的接口。

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 100:2
[PE2-vpn-instance-vpn1] vpn-target 100:1 both
[PE2-vpn-instance-vpn1] quit
[PE2] interface gigabitethernet 1/0
[PE2-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/0] ip address 192.168.2.1 255.255.255.0
[PE2-GigabitEthernet1/0] quit
```

上述配置完成后，在 PE 设备上执行 **display ip vpn-instance** 命令可以看到 VPN 实例的配置情况。

以 PE 1 为例：

```
[PE1] display ip vpn-instance instance-name vpn1
VPN-Instance Name and ID : vpn1, 1
  Create time : 2006/09/27 15:10:29
  Up time : 0 days, 00 hours, 03 minutes and 09 seconds
  Route Distinguisher : 100:1
  Export VPN Targets :  100:1
  Import VPN Targets :  100:1
  Tunnel Policy : policy1
  Interfaces : GigabitEthernet1/0
```

各 PE 能 ping 通自己接入的 CE。

以 PE 1 和 CE 1 为例：

```
[PE1] ping -vpn-instance vpn1 192.168.1.2
  PING 192.168.1.2: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.2: bytes=56 Sequence=1 ttl=255 time=47 ms
    Reply from 192.168.1.2: bytes=56 Sequence=2 ttl=255 time=26 ms
    Reply from 192.168.1.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 192.168.1.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 192.168.1.2: bytes=56 Sequence=5 ttl=255 time=26 ms
  --- 192.168.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 26/30/47 ms
```

(6) 配置 BGP

配置 CE 1。

```
[CE1] bgp 65001
[CE1-bgp] peer 192.168.1.1 as-number 100
[CE1-bgp] quit
```

配置 PE 1：与 CE 1 建立 EBGP 对等体，与 PE 2 建立 IBGP 对等体。

```
[PE1] bgp 100
[PE1 -bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 192.168.1.2 as-number 65001
```

```
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
[PE1-bgp] peer 3.3.3.3 as-number 100
[PE1-bgp] peer 3.3.3.3 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.3 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

配置 CE 2。

```
[CE2] bgp 65002
[CE2-bgp] peer 192.168.2.1 as-number 100
[CE2-bgp] quit
```

配置 PE 2：与 CE 2 建立 EBGP 对等体，与 PE-A 建立 IBGP 对等体。

```
[PE2] bgp 100
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] peer 192.168.2.2 as-number 65002
[PE2-bgp-vpn1] import-route direct
[PE2-bgp-vpn1] quit
[PE2-bgp] peer 2.2.2.2 as-number 100
[PE2-bgp] peer 2.2.2.2 connect-interface loopback 0
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 2.2.2.2 enable
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] quit
```

上述配置完成后，在 PE 设备上执行 **display bgp peer** 命令和 **display bgp vpn-instance peer** 命令，可以看到 PE 之间、PE 与 CE 之间的 BGP 对等体关系都已建立，并达到 Established 状态。

以 PE 1 为例：

```
[PE1-bgp] display bgp peer
 BGP local router ID : 2.2.2.2
 Local AS number : 100
 Total number of peers : 1          Peers in established state : 1
  Peer     V    AS  MsgRcvd  MsgSent  OutQ  Up/Down  State        PrefRcv
  3.3.3.3  4    100       3        3     0  00:00:11 Established  0
[PE1-bgp] display bgp vpn-instance vpn1 peer
 BGP local router ID : 2.2.2.2
 Local AS number : 100
 Total number of peers : 1          Peers in established state : 1
  Peer        V AS    MsgRcvd MsgSent  OutQ  Up/Down  State        PrefRcv
  192.168.1.2 4 65001 4       5        0     00:02:13 Established  0
```

CE 1 和 CE 2 之间能够互相 ping 通。

```
[CE1] ping 192.168.2.2
  PING 192.168.2.2: 56  data bytes, press CTRL_C to break
    Reply from 192.168.2.2: bytes=56 Sequence=1 ttl=253 time=61 ms
    Reply from 192.168.2.2: bytes=56 Sequence=2 ttl=253 time=54 ms
    Reply from 192.168.2.2: bytes=56 Sequence=3 ttl=253 time=53 ms
```

```
    Reply from 192.168.2.2: bytes=56 Sequence=4 ttl=253 time=57 ms
    Reply from 192.168.2.2: bytes=56 Sequence=5 ttl=253 time=36 ms
  --- 192.168.2.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 36/52/61 ms
[CE2] ping 192.168.1.2
  PING 192.168.1.2: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.2: bytes=56 Sequence=1 ttl=253 time=38 ms
    Reply from 192.168.1.2: bytes=56 Sequence=2 ttl=253 time=61 ms
    Reply from 192.168.1.2: bytes=56 Sequence=3 ttl=253 time=74 ms
    Reply from 192.168.1.2: bytes=56 Sequence=4 ttl=253 time=36 ms
    Reply from 192.168.1.2: bytes=56 Sequence=5 ttl=253 time=35 ms
  --- 192.168.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 35/48/74 ms
```

(7) 配置完成后的检验

在 PE 1 上执行 **display mpls lsp verbose** 命令，可以确认 LspIndex 为 2050 的是使用 CR-LDP 建立的 LSP，即 MPLS TE 隧道。

```
[PE1] display mpls lsp verbose
------------------------------------------------------------------------
                 LSP Information: CRLDP LSP
------------------------------------------------------------------------
  No                  :  1
  IngressLsrID        :  2.2.2.2
  LocalLspID          :  1
  Tunnel-Interface    :  Tunnel1
  Fec                 :  3.3.3.3/32
  Nexthop             :  10.0.0.2
  In-Label            :  NULL
  Out-Label           :  1024
  In-Interface        :  ----------
  Out-Interface       :  POS5/1
  LspIndex            :  2050
  Tunnel ID           :  0x22004
  LsrType             :  Ingress
  Bypass In Use       :  Not Exists
  BypassTunnel        :  Tunnel Index[---]
  Mpls-Mtu            :  1500
------------------------------------------------------------------------
                 LSP Information: BGP  LSP
------------------------------------------------------------------------
  No                  :  2
  VrfIndex            :  vpn1
```

```
  Fec                 :  192.168.1.0/24
  Nexthop             :  192.168.1.1
  In-Label            :  1024
  Out-Label           :  NULL
  In-Interface        :  ----------
  Out-Interface       :  ----------
  LspIndex            :  8193
  Tunnel ID           :  0x0
  LsrType             :  Egress
  Outgoing Tunnel ID  :  0x0
  Label Operation     :  POP
------------------------------------------------------------------------
                 LSP Information: LDP LSP
------------------------------------------------------------------------
  No                  :  3
  VrfIndex            :
  Fec                 :  2.2.2.2/32
  Nexthop             :  127.0.0.1
  In-Label            :  3
  Out-Label           :  NULL
  In-Interface        :  POS5/1
  Out-Interface       :  ----------
  LspIndex            :  10241
  Tunnel ID           :  0x0
  LsrType             :  Egress
  Outgoing Tunnel ID  :  0x0
  Label Operation     :  POP

  No                  :  4
  VrfIndex            :
  Fec                 :  3.3.3.3/32
  Nexthop             :  10.0.0.2
  In-Label            :  NULL
  Out-Label           :  3
  In-Interface        :  ----------
  Out-Interface       :  POS5/1
  LspIndex            :  10242
  Tunnel ID           :  0x22000
  LsrType             :  Ingress
  Outgoing Tunnel ID  :  0x0
  Label Operation     :  PUSH
```

在 PE 1 上执行 **display interface tunnel** 命令可以看到流量通过 TE 隧道 CR-LSP 转发。

```
[PE1] display interface tunnel 1
Tunnel1 current state: UP
Line protocol current state: UP
Description: Tunnel1 Interface
The Maximum Transmit Unit is 1500
Internet Address is 12.1.1.1/24 Primary
```

```
Encapsulation is TUNNEL, service-loopback-group ID not set
Tunnel source unknown, destination 3.3.3.3
Tunnel protocol/transport CR_LSP
Output queue : (Urgent queuing : Size/Length/Discards)  0/100/0
Output queue : (Protocol queuing : Size/Length/Discards)  0/500/0
Output queue : (FIFO queuing : Size/Length/Discards)  0/75/0
    Last 300 seconds input:  5 bytes/sec, 0 packets/sec
    Last 300 seconds output:  5 bytes/sec, 0 packets/sec
    34 packets input,  2856 bytes
    0 input error
    34 packets output,  2856 bytes
    0 output error
```

1.16 MPLS TE 常见配置错误举例

1.16.1 不能产生 TE LSA

1. 故障现象

配置 OSPF TE，无法产生描述 MPLS TE 信息的 TE LSA。

2. 分析

至少有一个 OSPF 邻居达到 FULL 状态时，才可能产生 TE LSA。

3. 处理过程

- 执行 **display current-configuration** 命令，检查是否在相关接口上配置了 MPLS TE；
- 执行 **debugging ospf mpls-te** 命令打开 OSPF TE 的调试开关，检查 OSPF 是否收到建立 TE LINK 的消息；
- 执行 **display ospf peer** 命令，检查 OSPF 邻居是否正常建立。

目 录

1 MPLS L2VPN配置 1-1
1.1 MPLS L2VPN简介 1-1
1.1.1 MPLS L2VPN概述 1-1
1.1.2 CCC方式MPLS L2VPN 1-3
1.1.3 SVC方式MPLS L2VPN 1-3
1.1.4 Martini方式MPLS L2VPN 1-3
1.1.5 Kompella方式MPLS L2VPN 1-4
1.2 MPLS L2VPN配置任务简介 1-4
1.3 配置MPLS L2VPN 1-5
1.4 配置CE侧接口 1-5
1.4.1 配置CE侧接口的封装类型为PPP、HDLC和FR 1-5
1.4.2 配置CE侧接口的封装类型为Ethernet 1-6
1.4.3 配置CE侧接口的封装类型为VLAN 1-6
1.5 配置CCC方式MPLS L2VPN 1-7
1.5.1 配置准备 1-7
1.5.2 配置CCC方式MPLS L2VPN 1-7
1.6 配置SVC方式MPLS L2VPN 1-8
1.6.1 配置准备 1-8
1.6.2 配置SVC方式MPLS L2VPN 1-9
1.7 配置Martini方式MPLS L2VPN 1-9
1.7.1 配置准备 1-9
1.7.2 配置Martini方式MPLS L2VPN 1-10
1.8 配置Kompella方式MPLS L2VPN 1-10
1.8.1 配置准备 1-10
1.8.2 配置Kompella方式MPLS L2VPN 1-11
1.9 MPLS L2VPN显示和维护 1-12
1.9.1 显示L2VPN的运行状态 1-12
1.9.2 复位BGP的L2VPN连接 1-13
1.10 MPLS L2VPN典型配置举例 1-13
1.10.1 配置CCC本地连接示例 1-13
1.10.2 配置CCC远程连接示例 1-15
1.10.3 配置SVC方式MPLS L2VPN示例 1-19
1.10.4 配置Martini方式MPLS L2VPN示例 1-23
1.10.5 配置Kompella方式MPLS L2VPN示例 1-27
1.10.6 配置Kompella方式本地连接示例 1-29
1.11 常见配置错误举例 1-31
1.11.1 配置后不能Ping通 1-31
1.11.2 不同链路协议的串口不能建立LDP方式链接 1-31

1 MPLS L2VPN 配置

1.1 MPLS L2VPN 简介

1.1.1 MPLS L2VPN 概述

1. 传统的 VPN

传统的基于 ATM 或 Frame Relay（帧中继，以下简称 FR）的 VPN（Virtual Private Network，虚拟专用网）应用非常广泛，它们能在不同 VPN 间共享运营商的网络结构。这种 VPN 的不足在于：

- 依赖于专用的介质（如 ATM 或 FR）：为提供基于 ATM 的 VPN 服务，运营商必须建立覆盖全部服务范围的 ATM 网络；为提供基于 FR 的 VPN 服务，又需要建立覆盖全部服务范围的 FR 网络，在网络建设上造成浪费。
- 部署复杂：尤其是向已有的 VPN 加入新的 Site（站点）时，需要同时修改所有接入此 VPN 站点的边缘节点的配置。

由于以上缺点，新的 VPN 替代方案应运而生，MPLS L2VPN 就是其中的一种。

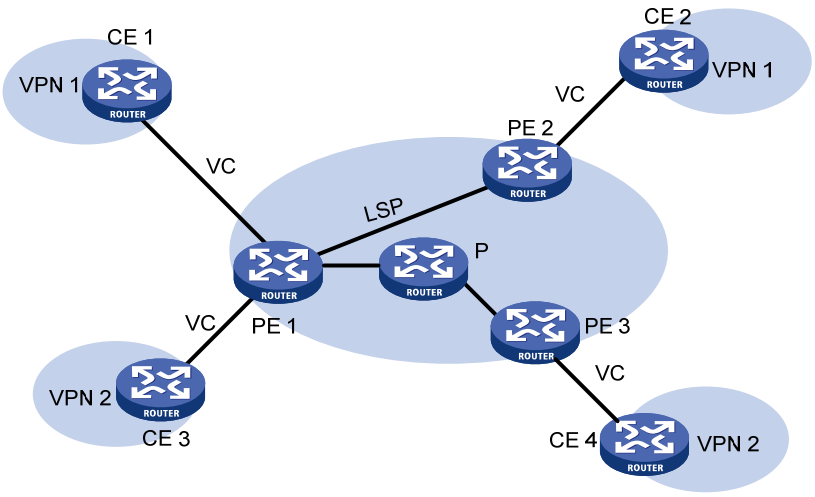
2. MPLS L2VPN

MPLS L2VPN 提供基于 MPLS（Multiprotocol Label Switching，多协议标签交换）网络的二层 VPN 服务，使运营商可以在统一的 MPLS 网络上提供基于不同数据链路层的二层 VPN，包括 FR、Ethernet、PPP 等。

简单来说，MPLS L2VPN 就是在 MPLS 网络上透明传输用户二层数据。从用户的角度来看，MPLS 网络是一个二层交换网络，可以在不同节点间建立二层连接。

以以太网为例，每个用户边缘设备通过 MPLS 网络与远端 CE 相连，这与通过交换网络实现互联类似。

图1-1 MPLS L2VPN 组网示意图



3. 与 MPLS L3VPN 相比

相对于 MPLS L3VPN，MPLS L2VPN 具有以下优点：

- 可扩展性强：MPLS L2VPN 只建立二层连接关系，不引入和管理用户的路由信息。这大大减轻了 PE（Provider Edge，服务提供商边缘设备）甚至整个 SP（Service Provider，服务提供商）网络的负担，使服务提供商能支持更多的 VPN 和接入更多的用户。
- 可靠性和私网路由的安全性得到保证：由于不引入用户的路由信息，MPLS L2VPN 不能获得和处理用户路由，保证了用户 VPN 路由的安全。
- 支持多种网络层协议：包括 IP、IPX、SNA 等。

4. MPLS L2VPN 的基本概念

在 MPLS L2VPN 中，CE、PE、P 的概念与 MPLS L3VPN 一样，原理也相似。

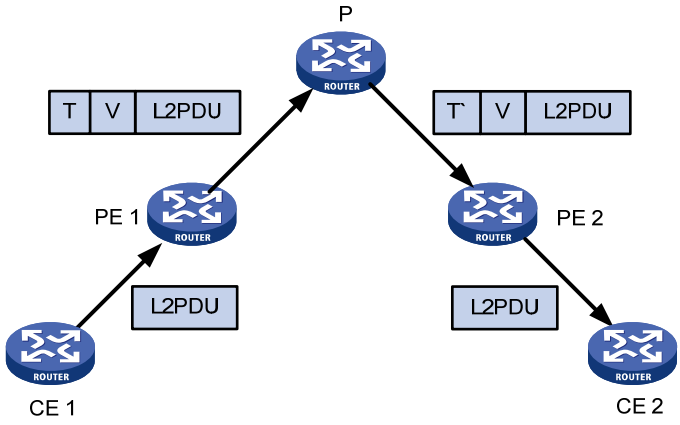
- CE（Customer Edge）设备：用户网络边缘设备，有接口直接与 SP 相连。CE 可以是路由器或交换机，也可以是一台主机。CE“感知”不到 VPN 的存在，也不需要必须支持 MPLS。
- PE（Provider Edge）路由器：服务提供商边缘路由器，是服务提供商网络的边缘设备，与用户的 CE 直接相连。在 MPLS 网络中，对 VPN 的所有处理都发生在 PE 上。
- P（Provider）路由器：服务提供商网络中的骨干路由器，不与 CE 直接相连。P 设备只需要具备基本 MPLS 转发能力。

MPLS L2VPN 通过标签栈实现用户报文在 MPLS 网络中的透明传送：

- 外层标签（称为 Tunnel 标签）用于将报文从一个 PE 传递到另一个 PE；
- 内层标签（称为 VC 标签）用于区分不同 VPN 中的不同连接；
- 接收方 PE 根据 VC 标签决定将报文转发给哪个 CE。

图 1-2 是MPLS L2VPN转发过程中报文标签栈变化的示意图。

图1-2 MPLS L2VPN 标签栈处理



(1) L2PDU 是链路层报文，PDU 即 Protocol Data Unit，协议数据单元

(2) T 是 Tunnel 标签；V 是 VC 标签；T'表示转发过程中外层标签被替换

5. MPLS L2VPN 的实现方式

当前 MPLS L2VPN 还没有形成正式的标准。IETF 的 PPVPN（Provider-provisioned Virtual Private Network）工作组制订了多个框架草案，其中最主要的两种称为 Martini 草案和 Kompella 草案：

- draft-martini-l2circuit-trans-mpls
- draft-kompella-ppvpn-l2vpn

Martini 草案定义了通过建立点到点的链路来实现 MPLS L2VPN 的方法。它以 LDP（Label Distribution Protocol，标记分发协议）为信令协议来传递双方的 VC 标签，称为 Martini 方式 MPLS L2VPN。

Kompella 草案则定义了在 MPLS 网络上以端到端（CE 到 CE）的方式建立 MPLS L2VPN。目前它采用扩展了的 BGP（Border Gateway Protocol，边界网关协议）为信令协议来发布二层可达信息和 VC 标签，称为 Kompella 方式 MPLS L2VPN。

另外，可以采用静态配置 VC 标签的方式来实现 MPLS L2VPN 服务。CCC（Circuit Cross Connect，电路交叉连接）和 SVC（Static Virtual Circuit，静态虚拟电路）就是两种静态配置 MPLS L2VPN 的实现方式。

下面分别介绍这几种 MPLS L2VPN 的特点和实现方式。

1.1.2 CCC 方式 MPLS L2VPN

与普通 MPLS L2VPN 不同，CCC 采用一层标签传送用户数据，因此，CCC 对 LSP（Label Switched Path，标签交换路径）的使用是独占性的。CCC 的 LSP 只用于传递这个 CCC 连接的数据，不能用于其他 MPLS L2VPN 连接，也不能用于 MPLS L3VPN 或承载普通的 IP 报文。

这种方式的最大优点是：不需要任何标签信令传递二层 VPN 信息，只要能支持 MPLS 转发即可，保证在任何情况下，运营商之间可以进行互连。此外，由于 LSP 是专用的，可以提供 QoS 保证。

CCC 连接有两种：本地连接和远程连接。

- 本地连接：在两个本地 CE 之间建立的连接，即两个 CE 连在同一个 PE 上。PE 的作用类似二层交换机，可以直接完成交换，不需要配置静态 LSP。
- 远程连接：在本地 CE 和远程 CE 之间建立的连接，即两个 CE 连在不同的 PE 上，需要配置静态 LSP 来把报文从一个 PE 传递到另一个 PE。



说明

远程连接中的 P 设备上必须单独为每一个 CCC 连接手工配置两条不同方向的 LSP。

1.1.3 SVC 方式 MPLS L2VPN

SVC（Static Virtual Circuit，静态虚拟电路）也是一种静态的 MPLS L2VPN，在 L2VPN 信息传递中不使用信令协议。

SVC 方式与 Martini 方式的 MPLS L2PVN 非常类似，但它不使用 LDP 传递二层 VC 和链路信息，手工配置 VC 标签信息即可，其实是 Martini 方式的一种静态实现。



说明

CCC 和 SVC 使用的标签范围是 16～1023，即保留给静态 LSP 使用的标签。

1.1.4 Martini 方式 MPLS L2VPN

Martini 方式 MPLS L2VPN 着重于在两个 CE 之间建立 VC（Virtual Circuit，虚电路）。

Martini 方式采用 VC-TYPE 加上 VC ID 来标识一个 VC。VC-TYPE 表明 VC 的封装类型：VLAN 或 PPP；VC ID 则用于唯一标识一个 VC。同一个 VC-TYPE 的所有 VC 中，其 VC ID 必须在整个 PE 中唯一。

连接两个 CE 的 PE 通过 LDP 交换 VC 标签，并通过 VC ID 绑定对应的 CE。当连接两个 PE 的 LSP 建立成功，双方的标签交换和绑定完成后，一个 VC 就建立起来了，CE 之间可以通过此 VC 传递二层数据。

为了在 PE 之间交换 VC 标签，Martini 草案对 LDP 进行了扩展，增加了 VC FEC（Forwarding Equivalence Class，转发等价类）的 FEC 类型。此外，由于交换 VC 标签的两个 PE 可能不是直接相连的，所以 LDP 必须使用 remote peer 来建立会话（Session），并在这个会话上传递 VC FEC 和 VC 标签。

在 Martini 方式中，由于在运营商网络中，只有 PE 设备需要保存少量的 VC label 与 LSP 的映射等信息，P 设备不包含任何二层 VPN 信息，所以扩展性很好。此外，当需要新增加一条 VC 时，只在相关的两端 PE 设备上各配置一个单方向 VC 连接即可，不影响网络的运行。
Martini 方式适合稀疏的二层连接，例如星型连接。

1.1.5 Kompella 方式 MPLS L2VPN

与 Martini 方式不同，Kompella 方式的 MPLS L2VPN 不直接对 CE 与 CE 之间的连接进行操作，而是在整个运营商网络中划分不同的 VPN，在 VPN 内部对 CE 进行编号。要建立两个 CE 之间的连接，只需在 PE 上设置本地 CE 和远程 CE 的 CE ID，并指定本地 CE 为这个连接分配的 Circuit ID。

Kompella 方式 MPLS L2VPN 以 BGP 扩展为信令协议来分发 VC 标签。在分配标签时，Kompella 方式采用标签块（Label block）的方式，一次为多个连接分配标签。

用户可以指定一个本地 CE 的范围（CE range），表明这个 CE 能与多少个 CE 建立连接。系统一次为这个 CE 分配一个标签块，标签块的大小等于 CE range。这种方式允许用户为 VPN 分配一些额外的标签，留待以后使用。这样短期来看会造成标签资源的浪费，但是却带来一个很大的好处，即可以减少 VPN 部署和扩容时的配置工作量。

假设一个企业的 VPN 包括 10 个 CE，但是考虑到企业会扩展业务，将来可能会有 20 个 CE。这样可以把每个 CE 的 CE range 设置为 20，系统会预先为未来的 10 个 CE 分配标签。以后 VPN 添加 CE 节点时，配置的修改仅限于与新 CE 直接相连的 PE，其他 PE 不需要作任何修改。这使得 VPN 的扩容变得非常简单。

类似于 MPLS L3VPN，Kompella 方式 MPLS L2VPN 也使用 VPN Target 来区分不同的 VPN，这使得 VPN 组网具有极大的灵活性。

1.2 MPLS L2VPN 配置任务简介

表1-1 MPLS L2VPN 配置任务简介

操作	说明	详细配置
配置 MPLS L2VPN	必选	1.3
配置 CE 侧接口	必选	1.4

操作	说明	详细配置
配置 CCC 方式 MPLS L2VPN	选择其一 根据 MPLS L2VPN 的实现方式，选择相应的配置方法	1.5
配置 SVC 方式 MPLS L2VPN		1.6
配置 Martini 方式 MPLS L2VPN		1.7
配置 Kompella 方式 MPLS L2VPN		1.8

1.3 配置 MPLS L2VPN

MPLS L2VPN 有多种实现方式，用户可以根据实际需求选择合适的组网方式。但不论采用哪种方式，以下两项配置都是必须的：

- 配置 MPLS 基本能力
- 使能 MPLS L2VPN

具体配置参见 表 1-2。

表1-2 配置 MPLS L2VPN

操作	命令	说明
进入系统视图	**system-view**	-
配置 LSR-ID	**mpls lsr-id**	必选
配置 MPLS 基本能力，进入到 MPLS 视图	**mpls**	必选
退回系统视图	**quit**	-
使能 MPLS L2VPN	**mpls l2vpn**	必选

1.4 配置 CE 侧接口

CE 侧接口为 PE 设备上与 CE 相连的接口。

CE 侧接口的封装类型有如下几种：

- PPP
- HDLC
- FR
- Ethernet

1.4.1 配置 CE 侧接口的封装类型为 PPP、HDLC 和 FR

同步串口 Serial 和 POS 接口可以封装为 PPP、HDLC 和 FR 的链路类型。

有关 Serial 和 POS 的配置请参考“接入分册”中的“WAN 接口配置”和“POS 接口配置”。

1. 配置 CE 侧接口的封装类型为 PPP

表1-3 配置 CE 侧接口的封装类型为 PPP

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** { **serial** | **pos** } *number*	-
配置接口的链路协议类型	**link-protocol ppp**	可选 缺省情况下，接口的链路协议类型为 PPP

2. 配置 CE 侧接口的封装类型为 HDLC

表1-4 配置 CE 侧接口的封装类型为 HDLC

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** { **serial** | **pos** } *number*	-
配置接口的链路协议类型	**link-protocol hdlc**	必选 缺省情况下，接口的链路协议类型为 PPP

3. 配置 CE 侧接口的封装类型为 FR

表1-5 配置 CE 侧接口的封装类型为 FR

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** { **serial** | **pos** } *interface-number*	-
配置接口的链路协议类型	**link-protocol fr** [**nonstandard** | **ietf** | **mfr** *interface-number*]	必选
配置虚电路	**fr dlci** *dlci-number*	必选 缺省情况下，接口上没有配置虚电路



说明

CE 侧接口的 DLCI 号与 CE 设备上的 DLCI 号必须一致。

1.4.2 配置 CE 侧接口的封装类型为 Ethernet

当接口类型为以太网接口时，封装类型为 Ethernet。有关以太网接口的配置请参见“接入分册”中“以太网接口配置”。

1.4.3 配置 CE 侧接口的封装类型为 VLAN

当接口类型为以太网子接口，封装类型为 VLAN。有关以太网子接口的配置请参见“接入分册”中“逻辑接口配置”。

1.5 配置 CCC 方式 MPLS L2VPN

1.5.1 配置准备

在配置 CCC L2VPN 之前，需完成以下任务：

- 对 MPLS 骨干网（PE、P）配置 MPLS 基本能力
- 对 MPLS 骨干网 PE 路由器上使能 MPLS L2VPN，P 路由器上不需要使能 MPLS L2VPN

说明

当 CE 侧接口类型是链路协议为 FR 的 Serial 时，必须使用 **reset fr inarp** 命令刷新 PE 与 CE 之间的信息。有关 **reset fr inarp** 命令的解释请参见“接入分册”中的“帧中继命令”。

在配置 CCC L2VPN 之前，需准备以下数据：

- CCC 连接的名称
- 确定需要创建的连接类型：本地连接还是远程连接
- 本地连接：CCC 连接的入接口和出接口类型和编号
- 远程连接：CCC 连接的入接口类型和编号、下一跳地址或出接口类型和编号、CCC 连接沿途 LSR 的入标签值和出标签值

说明

配置远程连接时，如果出接口连接的链路不是点到点链路，如出接口类型为三层以太网接口、VLAN 接口时，请必须使用 **nexthop** 参数指定下一跳 IP 地址；否则，只需指定出接口即可。

1.5.2 配置 CCC 方式 MPLS L2VPN

1. 配置 CCC 本地连接

在 PE 上进行下述配置后即完成 CCC 本地连接的创建。

表1-6 配置 CCC 本地连接

操作	命令	说明
进入系统视图	**system-view**	-
在连接相同 PE 的两个 CE 间创建一条本地 CCC 连接	**ccc** *ccc-connection-name* **interface** *interface-type interface-number* **out-interface** *interface-type interface-number*	必选

2. 配置 CCC 远程连接

(1) PE 的配置

表1-7 PE 的配置

操作	命令	说明
进入系统视图	**system-view**	-
在连接不同 PE 的 CE 间创建一条远程 CCC 连接	**ccc** *ccc-connection-name* **interface** *interface-type interface-number* **in-label** *in-label-value* **out-label** *out-label-value* { **nexthop** *ip-address* | **out-interface** *interface-type interface-number* } [**control-word** | **no-control-word**]	必选

注意

每个 CCC 远程连接不需要配置两条静态 LSP，仅需要为它配置入标签和出标签，入标签为此 CCC 连接独占，相当于配置了静态 LSP。

(2) P 的配置

表1-8 P 的配置

操作	命令	说明
进入系统视图	**system-view**	-
为 Transit 节点配置静态 LSP	**static-lsp transit** *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label* { **nexthop** *next-hop-addr* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*	必选

注意

- 使用 CCC 时，PE 路由器上不需要配置静态 LSP。但 PE 之间的所有 P 路由器上需要配置双向的转发静态 LSP，用于专门传输 CCC 连接的数据。
- 有关配置静态 LSP 的命令请参见“MPLS 分册”中的“MPLS 基本配置命令”。

1.6 配置 SVC 方式 MPLS L2VPN

SVC 方式 MPLS L2VPN 不使用信令协议传输 L2VPN 信息，数据是通过隧道在 PE 之间传递。

SVC 支持的隧道类型包括 LDP LSP、CR-LSP、GRE，缺省情况下使用 LDP LSP 隧道。

1.6.1 配置准备

在配置 SVC 方式 MPLS L2VPN 之前，需完成以下任务：

- 对 MPLS 骨干网（PE、P）配置 IGP，实现骨干网的 IP 连通性
- 对 MPLS 骨干网（PE、P）配置 MPLS 基本能力
- 对 PE 配置 MPLS L2VPN
- 对 MPLS 骨干网（PE、P）配置 MPLS LDP，建立 LDP LSP
- 当 CE 接入类型是 VLAN 时配置子接口

- 根据使用的隧道策略，在 PE 之间建立相应类型的隧道

在配置 SVC 方式 MPLS L2VPN 之前，需准备以下数据：

- 接入 CE 的接口类型和编号
- SVC 的目的 LSR ID
- L2VPN 连接的收、发标签值
- SVC 的隧道策略

1.6.2 配置 SVC 方式 MPLS L2VPN

下述配置需要在 PE 上进行。

表1-9 配置 SVC 方式 MPLS L2VPN

操作	命令	说明
进入系统视图	**system-view**	-
进入连接 CE 的接口视图	**interface** *interface-type interface-number*	-
创建 SVC 方式 MPLS L2VPN 连接	**mpls static-l2vc destination** *destination-router-id* **transmit-vpn-label** *transmit-label-value* **receive-vpn-label** *receive-label-value* [**tunnel-policy** *tunnel-policy-name*] [**control-word** | **no-control-word**]	必选

1.7 配置 Martini 方式 MPLS L2VPN

Martini 方式使用扩展的 LDP 传递二层信息和 VC 标签。配置 Martini 方式的 MPLS L2VPN 时需要：

- 创建 Martini 方式 MPLS L2VPN 连接。

在三层接口上创建 Martini 方式 MPLS L2VPN 连接后，从接口接收到的报文将通过该 MPLS L2VPN 连接转发。

- 配置远端对等体。

Martini 方式的 MPLS L2VPN 中，需要在 PE 之间交换 VC 标签。由于交换 VC 标签的两个 PE 可能不是直接相连的，所以需要在 PE 之间建立 LDP 远端会话（Session），以便在这个会话上传递 VC FEC 和 VC 标签。

1.7.1 配置准备

在配置 Martini 方式 MPLS L2VPN 之前，需完成以下任务：

- 对 MPLS 骨干网（PE、P）配置 IGP，实现骨干网的 IP 连通性
- 对 MPLS 骨干网（PE、P）配置 MPLS 基本能力
- 对 PE 配置 MPLS L2VPN
- 对 MPLS 骨干网（PE、P）配置 MPLS LDP 能力，建立 LDP LSP
- 在 PE 之间建立远端 LDP 会话
- 当 CE 接入类型是 VLAN 时配置子接口

在配置 Martini 方式 MPLS L2VPN 之前，需准备以下数据：

- 接入 CE 的接口类型和接口编号

- L2VC 的目的地址和 VC ID
- 隧道策略

1.7.2 配置 Martini 方式 MPLS L2VPN

下述配置需要在 PE 上进行。

表1-10 配置 Martini 方式 MPLS L2VPN

操作	命令	说明
进入系统视图	**system-view**	-
进入连接 CE 的接口视图	**interface** *interface-type interface-number*	-
创建Martini方式MPLS L2VPN连接	**mpls l2vc** *destination vcid* [**tunnel-policy** *tunnel-policy-name*] [**control-word** | **no-control-word**]	必选
退回系统视图	**quit**	-
配置远端对等体	**mpls ldp remote-peer** *remote-peer-name*	必选
	remote-ip *ip-address*	必选



说明

有关远端对等体的配置请参见“MPLS 分册”中的“MPLS 基本配置”。



注意

Martini 方式连接命令主要参数有两个：一个是对端 PE 的 IP 地址，一个是 VC ID。其中，VC ID 与封装类型的组合必须在 PE 上唯一，修改封装有可能会造成 VC ID 的冲突。

1.8 配置 Kompella 方式 MPLS L2VPN

Kompella 方式 MPLS L2VPN 使用 BGP 扩展作为信令协议在 PE 间传递 L2VPN 信息。

配置Kompella方式本地连接时，不需要配置IGP，也不需要 配置BGP的L2VPN能力，只在PE设备上配置VPN和CE连接即可使CE设备之间建立连接。

1.8.1 配置准备

在配置 Kompella 方式 MPLS L2VPN 之前，需完成以下任务：

- 对 MPLS 骨干网（PE、P）配置 IGP，实现骨干网的 IP 连通性
- 对 MPLS 骨干网（PE、P）配置 MPLS 基本能力
- 对 PE 配置 MPLS L2VPN
- 对 MPLS 骨干网（PE、P）配置 MPLS LDP 能力，建立 LDP LSP

在配置 Kompella 方式 MPLS L2VPN 之前，需准备以下数据：

- 本 PE 和对端 PE 的 AS 号
- L2VPN 连接的名称、RD 和 VPN Target

- CE 的名称、CE ID 和 CE range
- CE offset

1.8.2 配置 Kompella 方式 MPLS L2VPN

1. 配置 BGP 的 L2VPN 能力

表1-11 配置 BGP 的 L2VPN 能力

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
与对端 PE 建立对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
指定建立 TCP 连接的接口	**peer** { *group-name* | *ip-address* } **connect-interface** *interface-type interface-number*	必选
进入 BGP-L2VPN 地址族视图	**l2vpn-family**	必选
对接收到的 VPNv4 路由使能 VPN-Target 过滤功能	**policy vpn-target**	可选 缺省情况下，对接收的路由信息进行 VPN-target 扩展团体属性的过滤
使能对等体，并使能交换 BGP-L2VPN 地址族的 BGP 路由信息的能力	**peer** { *group-name* | *ip-address* } **enable**	必选

说明

有关 BGP-L2VPN 地址族的配置请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

2. 配置 VPN

表1-12 配置 VPN

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN，并进入 MPLS-L2VPN 视图	**mpls l2vpn** *vpn-name* [**encapsulation** { **ethernet** | **fr** | **hdlc** | **ppp** | **vlan** } [**control-word** | **no-control-word**]]	必选
为 VPN 配置 RD	**route-distinguisher** *route-distinguisher*	必选
将一个指定 VPN 和一个或多个 VPN Target 相关联	**vpn-target** *vpn-target*&<1-16> [**both** | **export-extcommunity** | **import-extcommunity**]	必选
配置 VPN 的二层 MTU	**mtu** *mtu*	可选

注意

- **mtu** 命令只进行有可能存在的协议上的参数协商，并不指导转发，因此不建议使用此命令。
- Kompella 方式 MPLS L2VPN 必须在 PE 上为每个直接相连的 CE 所在的 VPN 创建 L2VPN 实例。创建 L2VPN 时指定的封装类型应与 CE 侧接口的封装类型对应。
- 上述配置中 VPN-Target 和 RD（Route Distinguisher，路由标识）的用法与 MPLS L3VPN 相同。对于 Kompella L2VPN，必须配置 RD。RD 配置后不能修改，除非先删除创建的 VPN，然后重新创建。

3. 配置 CE 连接

CE ID 用于在一个 VPN 中唯一确定一个 CE。为了方便配置，建议 CE ID 从 1 开始，采用连续自然数编号。

CE range 表明这个 CE 最多能与多少个 CE 建立连接。在标签资源足够丰富的情况下（一般来说，标签资源总是足够丰富的），可以根据对此 VPN 规模发展的预计，把 CE range 设置得比实际需要大一些。这样当以后对 VPN 进行扩容，增加 VPN 中的 CE 数目时，就可以尽量少的修改配置。

为 CE 创建连接时，如果用户没有指定 CE offset：

- 对于此 CE 的第一个连接，CE offset 为执行 **ce** 命令时，通过 **default-offset** 参数指定的值；
- 对于其他连接，CE offset 是上一个连接的 CE offset+1；
- 在规划 VPN 时，建议 CE ID 编号从 1 顺序递增；然后在配置连接时按 CE ID 顺序配置，这样，大多数连接都可以省略 **ce-offset** 参数，使用缺省值，从而简化配置。

修改 CE range 只能把 CE range 变大，不能变小。例如：原来的 CE range 为 10，则可以把它改为 20，但如果想改为 5，则会失败。把 CE range 改小的唯一方法是：删除这个 CE，重新创建。

对 VPN 扩容时，如果原来设置的 CE range 比所需要的小，修改 CE range 不会导致原来业务的中断。例如：扩容后需要连接的 CE 数目为 20，但 CE range 为 10。此时可以把 CE range 修改为 20。修改 CE range 时，为了保证原来的 10 个连接不中断，系统并不释放原来的标签块，重新申请大小为 20 的标签块，而是在原来的标签块之外，重新申请一个新的标签块，大小为 10。

表1-13 配置 CE 连接

操作	命令	说明
进入系统视图	**system-view**	-
并进入 MPLS-L2VPN 视图	**mpls l2vpn** *vpn-name*	-
创建 CE，并进入 MPLS-L2VPN-CE 视图	**ce** *ce-name* [**id** *ce-id* [**range** *ce-range*] [**default-offset** *ce-offset*]]	必选
创建 Kompella 方式连接	**connection** [**ce-offset** *id*] **interface** *interface-type interface-number* [**tunnel-policy** *tunnel-policy-name*]	必选

1.9 MPLS L2VPN 显示和维护

1.9.1 显示 L2VPN 的运行状态

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MPLS L2VPN 的运行情况，通过查看显示信息验证配置的效果。

表1-14 MPLS L2VPN 显示和维护

操作	命令
显示 CCC 连接信息	**display ccc** [**ccc-name** *ccc-name* | **type** { **local** | **remote** }]
显示 L2VPN 的 VC 使用的接口信息	**display l2vpn ccc-interface vc-type** { **all** | **bgp-vc** | **ccc** | **ldp-vc** | **static-vc** } [**up** | **down**]
显示路由器上配置的静态 VC 信息	**display mpls static-l2vc** [**interface** *interface-type interface-number*]
显示路由器上 Martini 方式的虚电路 VC	**display mpls l2vc**[**interface** *interface-type interface-number* | **remote-info**]
显示 Kompella 方式的 L2VPN 连接信息	**display mpls l2vpn connection** [**vpn-name** *vpn-name* [**remote-ce** *ce-id* | **down** | **up** | **verbose**] | **summary** | **interface** *interface-type interface-number*]
显示 BGP 路由表中的 L2VPN 信息	**display bgp l2vpn** { **all** | **group** [*group-name*] | **peer** [[*ip-address*] **verbose**] | **route-distinguisher** *rd* [**ce-id** *ce-id* [**label-offset** *label-offset*]] }
显示 PE 上的 L2VPN 信息	**display mpls l2vpn** [**export-route-target-list** | **import-route-target-list** | **vpn-name** *vpn-name* [**local-ce** | **remote-ce**]]
显示 MPLS L2VPN 的转发信息	**display mpls l2vpn forwarding-info** [*vc-label*] **interface** *interface-type interface-number* [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示 MPLS L2VPN 的接口报文统计信息	**display interface** *interface-type interface-number*

1.9.2 复位 BGP 的 L2VPN 连接

在完成上述配置后，在用户视图下执行复位 BGP 的 L2VPN 连接命令。

表1-15 复位 L2VPN 的 BGP 连接

操作	命令
复位 BGP 的 L2VPN 连接	**reset bgp l2vpn** { *as-number* | *ip-address* | **all** | **external** | **internal** }

1.10 MPLS L2VPN 典型配置举例

1.10.1 配置 CCC 本地连接示例

1. 组网需求

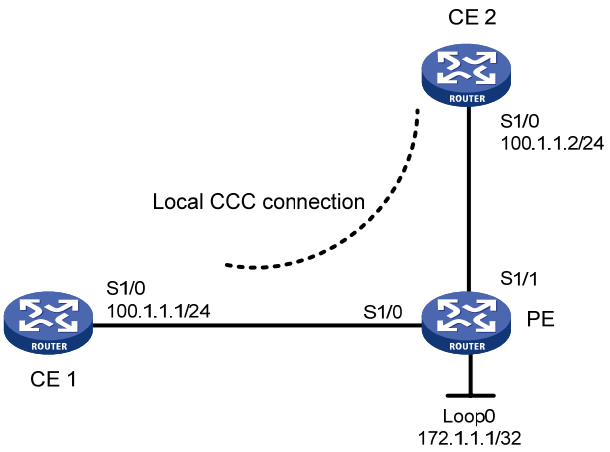
- CE 与 PE 之间通过 Serial 口相连，链路层封装 PPP。
- CE 1 和 CE 2 之间建立 CCC 本地连接。


说明

- 由于 CCC 的本地连接是双向的，因此只需要创建一条连接即可。
- PE 设备上连接 CE 的接口不必配置 IP 地址。

2. 组网图

图1-3 配置 CCC 本地连接组网图



3. 配置步骤

(1) 配置 CE 1

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] interface serial 1/0
[CE1-Serial1/0] link-protocol ppp
[CE1-Serial1/0] ip address 100.1.1.1 24
```

(2) 配置 PE

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE
[PE] interface loopback 0
[PE-LoopBack0] ip address 172.1.1.1 32
[PE-LoopBack0] quit
[PE] mpls lsr-id 172.1.1.1
[PE] mpls
[PE-mpls] quit
```

全局使能 MPLS L2VPN。

```
[PE] mpls l2vpn
```

配置接口 Serial1/0。

```
[PE] interface serial 1/0
[PE-Serial1/0] link-protocol ppp
[PE-Serial1/0] quit
```

配置接口 Serial1/1。

```
[PE] interface serial 1/1
[PE-Serial1/1] link-protocol ppp
[PE-Serial1/1] quit
```

创建 CE 1 到 CE 2 的本地连接。

```
[PE] ccc ce1-ce2 interface serial 1/0 out-interface serial 1/1
```

(3) 配置 CE 2

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] interface serial 1/0
[CE2-Serial1/0] link-protocol ppp
[CE2-Serial1/0] ip address 100.1.1.2 24
```

(4) 配置完成后的检验

配置完成后，在 PE 上查看 CCC 连接信息，可以看到建立了一条 CCC 本地连接。

```
[PE] display ccc
    Total  ccc vc          : 1
    Local  ccc vc          : 1,  1 up
    Remote ccc vc          : 0,  0 up
 ***Name                   : ce1-ce2
    Type                   : local
    State                  : up
    Intf1                   : Serial1/0 (up)
    Intf2                   : Serial1/1 (up)
```

在 CE 1 与 CE 2 之间能够 ping 通。

```
[CE1] ping 100.1.1.2
  PING 100.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 100.1.1.2: bytes=56 Sequence=1 ttl=255 time=180 ms
    Reply from 100.1.1.2: bytes=56 Sequence=2 ttl=255 time=60 ms
    Reply from 100.1.1.2: bytes=56 Sequence=3 ttl=255 time=10 ms
    Reply from 100.1.1.2: bytes=56 Sequence=4 ttl=255 time=70 ms
    Reply from 100.1.1.2: bytes=56 Sequence=5 ttl=255 time=60 ms
  --- 100.1.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 10/76/180 ms
```

在 CE2 上 ping CE1 的接口地址 100.1.1.1，也能够 ping 通。

1.10.2 配置 CCC 远程连接示例

1. 组网需求

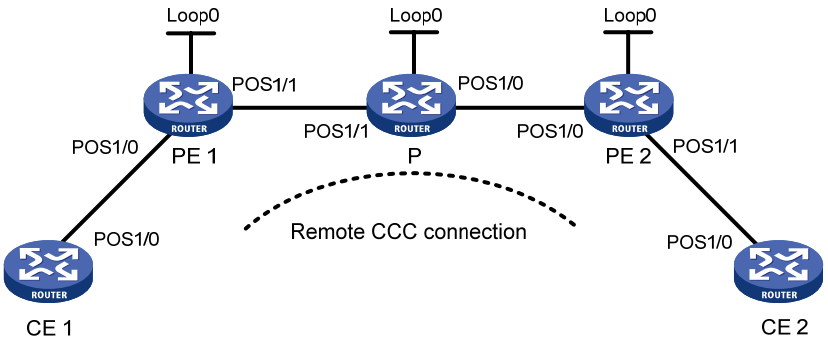
- CE 与 PE 之间通过 POS 口相连，链路层封装 PPP。
- CE 1 和 CE 2 之间建立 CCC 远程连接。

配置 CCC 远程连接的关键步骤包括：

- 在 PE 上创建 CCC 远程连接（不需要配置静态 LSP）。
- PE 上需要使能 MPLS L2VPN，P 设备上是不需要使能 MPLS L2VPN。
- 在 P 上配置两条静态 LSP，用于双向传递报文。

2. 组网图

图1-4 配置 CCC 远程连接组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	POS1/0	100.1.1.1/24	P	Loop0	10.0.0.2/32
PE 1	Loop0	10.0.0.1/32		POS1/0	10.2.2.2/24
	POS1/1	10.1.1.1/24		POS1/1	10.1.1.2/24
CE 2	POS1/0	100.1.1.2/24	PE 2	Loop0	10.0.0.3/32
				POS1/0	10.2.2.1/24

3. 配置步骤

(1) 配置 CE 1

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] interface pos 1/0
[CE1-POS1/0] link-protocol ppp
[CE1-POS1/0] ip address 100.1.1.1 24
```

(2) 配置 PE 1

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 10.0.0.1 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 10.0.0.1
[PE1] mpls
[PE1-mpls] quit
```

全局使能 MPLS L2VPN。

```
[PE1] mpls l2vpn
```

配置接口 POS1/0。

```
[PE1] interface pos 1/0
[PE1-POS1/0] link-protocol ppp
[PE1-POS1/0] quit
```

配置接口 POS1/1，使能 MPLS。

```
[PE1] interface pos 1/1
[PE1-POS1/1] link-protocol ppp
[PE1-POS1/1] ip address 10.1.1.1 24
```

```
[PE1-POS1/1] mpls
[PE1-POS1/1] quit
```

创建 CE 1 到 CE 2 的远程连接：入接口为连接 CE 1 的接口，出接口为连接 P 的接口；入标签为 100，出标签为 200。

```
[PE1] ccc ce1-ce2 interface pos 1/0 in-label 100 out-label 200 out-interface pos 1/1
```

(3) 配置 P

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname P
[P] interface loopback 0
[P-LoopBack0] ip address 10.0.0.2 32
[P-LoopBack0] quit
[P] mpls lsr-id 10.0.0.2
[P] mpls
[P-mpls] quit
```

配置接口 POS1/1，使能 MPLS。

```
[P] interface pos 1/1
[P-POS1/1] link-protocol ppp
[P-POS1/1] ip address 10.1.1.2 24
[P-POS1/1] mpls
[P-POS1/1] quit
```

配置接口 POS1/0，使能 MPLS。

```
[P] interface pos 1/0
[P-POS1/0] link-protocol ppp
[P-POS1/0] ip address 10.2.2.2 24
[P-POS1/0] mpls
[P-POS1/0] quit
```

配置一条静态 LSP 用于转发由 PE1 去往 PE2 的报文。

```
[P] static-lsp transit pe1_pe2 incoming-interface pos 1/1 in-label 200 outgoing-interface pos 1/0 out-label 201
```

配置另一条静态 LSP 用于转发由 PE2 去往 PE1 的报文。

```
[P] static-lsp transit pe2_pe1 incoming-interface pos 1/0 in-label 101 outgoing-interface pos 1/1 out-label 100
```

(4) 配置 PE 2

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 10.0.0.3 32
[PE2-LoopBack0] quit
[PE2] mpls lsr-id 10.0.0.3
[PE2] mpls
[PE2-mpls] quit
```

全局使能 MPLS L2VPN。

```
[PE2] mpls l2vpn
```

配置接口 POS1/1。

```
[PE2] interface pos 1/1
[PE2-POS1/1] link-protocol ppp
[PE2-POS1/1] quit
```

配置接口 POS1/0，使能 MPLS。

```
[PE2] interface pos 1/0
[PE2-POS1/0] link-protocol ppp
[PE2-POS1/0] ip address 10.2.2.1 24
[PE2-POS1/0] mpls
[PE2-POS1/0] quit
```

创建 CE2 到 CE1 的远程连接：入接口为连接 CE2 的接口，出接口为连接 P 的接口；入标签为 201，出标签为 101。

```
[PE2] ccc ce2-ce1 interface pos 1/1 in-label 201 out-label 101 out-interface pos 1/0
```

(5) 配置 CE 2

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] interface pos 1/0
[CE2-POS1/0] link-protocol ppp
[CE2-POS1/0] ip address 100.1.1.2 24
```

(6) 配置完成后的检验

配置完成后，在 PE 上查看 CCC 连接信息，可以看到建立了一条 CCC 本地连接。

```
[PE] display ccc
    Total  ccc vc          : 1
    Local  ccc vc          : 0,  0 up
    Remote ccc vc          : 1,  1 up
 ***Name                   : ce1-ce2
    Type                   : remote
    State                  : up
    Intf                   : POS1/1 (up)
    In-label               : 100
    Out-label              : 200
    Out-interface          : POS1/0 (up)
```

CE 1 与 CE 2 之间能够 ping 通。

```
[CE1] ping 100.1.1.2
  PING 100.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 100.1.1.2: bytes=56 Sequence=1 ttl=255 time=180 ms
    Reply from 100.1.1.2: bytes=56 Sequence=2 ttl=255 time=60 ms
    Reply from 100.1.1.2: bytes=56 Sequence=3 ttl=255 time=10 ms
    Reply from 100.1.1.2: bytes=56 Sequence=4 ttl=255 time=70 ms
    Reply from 100.1.1.2: bytes=56 Sequence=5 ttl=255 time=60 ms
  --- 100.1.1.2 ping statistics ---
    5 packet(s) transmitted
```

```
5 packet(s) received
0.00% packet loss
round-trip min/avg/max = 10/76/180 ms
```

在 CE2 上 ping CE1 的接口地址 100.1.1.1，也能够 ping 通。

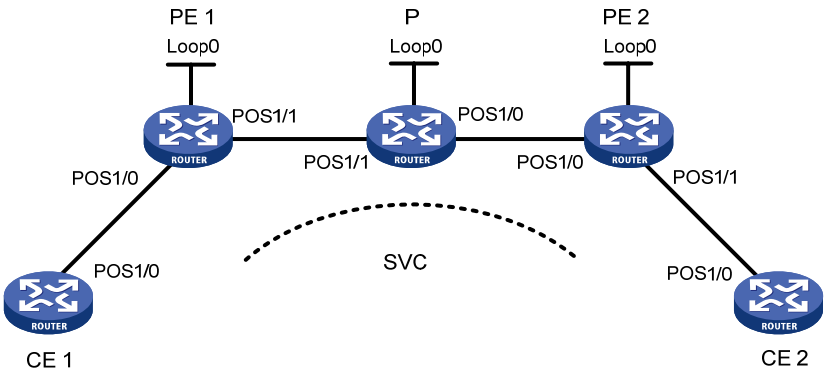
1.10.3 配置 SVC 方式 MPLS L2VPN 示例

1. 组网需求

- 与 PE 之间通过 POS 口相连，链路层封装 PPP。
- CE 1 和 CE 2 之间建立 SVC 方式的 MPLS L2VPN。

2. 组网图

图1-5 配置 SVC 方式 MPLS L2VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	POS1/0	100.1.1.1/24	P	Loop0	192.4.4.4/32
PE 1	Loop0	192.2.2.2/32		POS1/0	10.2.2.2/24
	POS1/1	10.1.1.1/24		POS1/1	10.1.1.2/24
CE 2	POS1/0	100.1.1.2/24	PE 2	Loop0	192.3.3.3/32
				POS1/0	10.2.2.1/24

3. 配置步骤

主要的配置步骤可分为两部分：

- 在 PE 和 P 上配置 MPLS 基本转发能力：包括配置 LSR ID、使能 MPLS 和 LDP、在 PE 1－P－PE 2 之间运行 IGP（本配置例中使用 OSPF）以建立 LSP。
- 建立 SVC 方式的 MPLS L2VPN 连接：包括在 PE 1 和 PE 2 上使能 MPLS L2VPN、创建 SVC 连接并指定 VC 标签。

下面介绍具体的配置过程：

(1) 配置 CE 1

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] interface pos 1/0
[CE1-POS1/0] link-protocol ppp
[CE1-POS1/0] ip address 100.1.1.1 24
```

(2) 配置 PE 1

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 192.2.2.2 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 192.2.2.2
[PE1] mpls
```

配置触发 LSP 的建立。

```
[PE1-mpls] lsp-trigger all
[PE1-mpls] quit
```

全局使能 MPLS L2VPN 和 LDP。

```
[PE1] mpls l2vpn
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

配置连接 P 的接口 POS1/1，在此接口上使能 LDP。

```
[PE1] interface pos 1/1
[PE1-POS1/1] link-protocol ppp
[PE1-POS1/1] ip address 10.1.1.1 24
[PE1-POS1/1] mpls
[PE1-POS1/1] mpls ldp
[PE1-POS1/1] quit
```

在 PE 1 上运行 OSPF，用于建立 LSP。

```
[PE1] ospf
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 10.1.1.1 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 192.2.2.2 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

在接入 CE 1 的接口 POS1/0 上创建一条 SVC 方式 MPLS L2VPN 连接。此接口不需配置 IP 地址。

```
[PE1] interface pos 1/0
[PE1-POS1/0] link-protocol ppp
[PE1-POS1/0] mpls static-l2vc destination 192.3.3.3 transmit-vpn-label 100
receive-vpn-label 200
[PE1-POS1/0] quit
```

(3) 配置 P

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname P
[P] interface loopback 0
[P-LoopBack0] ip address 192.4.4.4 32
[P-LoopBack0] quit
[P] mpls lsr-id 192.4.4.4
[P] mpls
```

配置触发 LSP 的建立。

```
[P-mpls] lsp-trigger all
[P-mpls] quit
```

全局使能 LDP。

```
[P] mpls ldp
[P-mpls-ldp] quit
```

配置连接 PE 1 的接口 POS1/1，在此接口上使能 LDP。

```
[P] interface pos 1/1
[P-POS1/1] link-protocol ppp
[P-POS1/1] ip address 10.1.1.2 24
[P-POS1/1] mpls
[P-POS1/1] mpls ldp
[P-POS1/1] quit
```

配置连接 PE 2 的接口 POS1/0，在此接口上使能 LDP。

```
[P] interface pos 1/0
[P-POS1/0] link-protocol ppp
[P-POS1/0] ip address 10.2.2.2 24
[P-POS1/0] mpls
[P-POS1/0] mpls ldp
[P-POS1/0] quit
```

在 P 上运行 OSPF，用于建立 LSP。

```
[P] ospf
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 10.1.1.2 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 10.2.2.2 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 192.4.4.4 0.0.0.0
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

(4) 配置 PE 2

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 192.3.3.3 32
[PE2-LoopBack0] quit
[PE2] mpls lsr-id 192.3.3.3
[PE2] mpls
```

配置触发 LSP 的建立。

```
[PE2-mpls] lsp-trigger all
[PE2-mpls] quit
```

全局使能 MPLS L2VPN 和 LDP。

```
[PE2] mpls l2vpn
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

配置连接 P 的接口 POS1/0，在此接口上使能 LDP。

```
[PE2] interface pos 1/0
[PE2-POS1/0] link-protocol ppp
[PE2-POS1/0] ip address 10.2.2.1 24
[PE2-POS1/0] mpls
[PE2-POS1/0] mpls ldp
[PE2-POS1/0] quit
```

在 PE 2 上运行 OSPF，用于建立 LSP。

```
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 10.2.2.1 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 192.3.3.3 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

在接入 CE 2 的接口 POS1/1 上创建一条 SVC 方式 MPLS L2VPN 连接。此接口不需配置 IP 地址。

```
[PE2] interface pos 1/1
[PE2-POS1/1] link-protocol ppp
[PE2-POS1/1] mpls static-l2vc destination 192.2.2.2 transmit-vpn-label 200
receive-vpn-label 100
[PE2-POS1/1] quit
```

(5) 配置 CE 2

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] interface pos 1/0
[CE2-POS1/0] link-protocol ppp
[CE2-POS1/0] ip address 100.1.1.2 24
```

(6) 配置完成后的检验

在 PE 1 上查看 SVC 的 L2VPN 连接信息，可以看到建立了一条 L2VPN 连接。

```
[PE1] display mpls static-l2vc
Total connections:  1,  1 up,  0 down
ce-intf           state destination     tr-label  rcv-label tnl-policy
POS1/0            up    192.3.3.3       100       200       default
```

在 PE2 上也可以看到 SVC 的 L2VPN 连接。

```
[PE2] display mpls static-l2vc
Total connections:  1,  1 up,  0 down
ce-intf           state destination     tr-label  rcv-label tnl-policy
POS1/1            up    192.2.2.2       200       100       default
```

在 CE 1 与 CE 2 之间能够 ping 通。

```
[CE1] ping 100.1.1.2
  PING 100.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 100.1.1.2: bytes=56 Sequence=1 ttl=255 time=150 ms
    Reply from 100.1.1.2: bytes=56 Sequence=2 ttl=255 time=130 ms
    Reply from 100.1.1.2: bytes=56 Sequence=3 ttl=255 time=130 ms
    Reply from 100.1.1.2: bytes=56 Sequence=4 ttl=255 time=140 ms
```

```
  Reply from 100.1.1.2: bytes=56 Sequence=5 ttl=255 time=80 ms
--- 100.1.1.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 80/126/150 ms
```

在 CE 2 上 ping CE 1 的接口地址 100.1.1.1，也能够 ping 通。

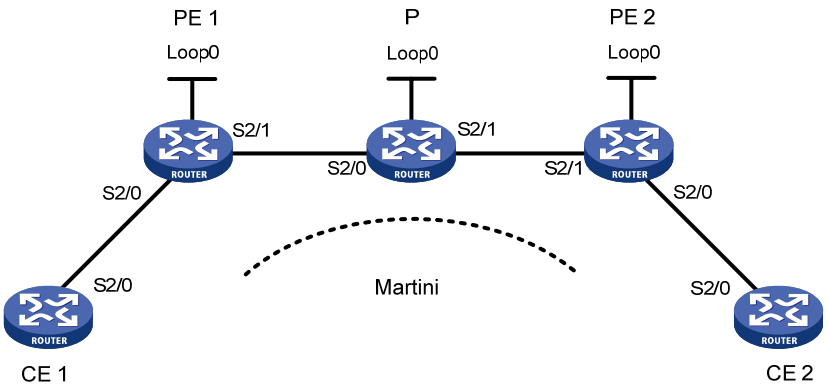
1.10.4 配置 Martini 方式 MPLS L2VPN 示例

1. 组网需求

- CE 与 PE 之间通过 Serial 口相连，链路层封装 PPP。
- CE 1 和 CE 2 之间建立 Martini 方式的 MPLS L2VPN。

2. 组网图

图1-6 配置 Martini 方式 MPLS L2VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	S2/0	100.1.1.1/24	P	Loop0	192.4.4.4/32
PE 1	Loop0	192.2.2.2/32		S2/0	10.1.1.2/24
	S2/1	10.1.1.1/24		S2/1	10.2.2.2/24
CE 2	S2/0	100.1.1.2/24	PE 2	Loop0	192.3.3.3/32
				S2/1	10.2.2.1/24

3. 配置步骤

(1) 配置 CE 1

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] interface serial 2/0
[CE1-Serial2/0] link-protocol ppp
[CE1-Serial2/0] ip address 100.1.1.1 24
```

(2) 配置 PE 1

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 192.2.2.2 32
[PE1-LoopBack0] quit
```

```
[PE1] mpls lsr-id 192.2.2.2
[PE1] mpls
```

配置触发 LSP 的建立。

```
[PE1-mpls] lsp-trigger all
[PE1-mpls] quit
```

全局使能 MPLS L2VPN 和 LDP。

```
[PE1] mpls l2vpn
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

配置 PE 1 与 PE 2 建立 LDP 远程会话。

```
[PE1] mpls ldp remote-peer 1
[PE1-mpls-ldp-remote-1] remote-ip 192.3.3.3
[PE1-mpls-ldp-remote-1] quit
```

配置连接 P 的接口 Serial2/1，在此接口上使能 LDP。

```
[PE1] interface serial 2/1
[PE1-Serial2/1] link-protocol ppp
[PE1-Serial2/1] ip address 10.1.1.1 24
[PE1-Serial2/1] mpls
[PE1-Serial2/1] mpls ldp
[PE1-Serial2/1] quit
```

在 PE 1 上运行 OSPF，用于建立 LSP。

```
[PE1] ospf
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 10.1.1.1 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 192.2.2.2 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

在接入 CE 1 的接口 Serial2/0 上创建 L2VPN 连接。此接口不需配置 IP 地址。

```
[PE1] interface serial 2/0
[PE1-Serial2/0] mpls l2vc 192.3.3.3 101
[PE1-Serial2/0] quit
```

(3) 配置 P

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname P
[P] interface loopback 0
[P-LoopBack0] ip address 192.4.4.4 32
[P-LoopBack0] quit
[P] mpls lsr-id 192.4.4.4
[P] mpls
```

配置触发 LSP 的建立。

```
[P-mpls] lsp-trigger all
[P-mpls] quit
```

全局使能 LDP。

```
[P] mpls ldp
[P-mpls-ldp] quit
```

配置连接 PE 1 的接口 Serial2/0，在此接口上使能 LDP。

```
[P] interface serial 2/0
[P-Serial2/0] link-protocol ppp
[P-Serial2/0] ip address 10.1.1.2 24
[P-Serial2/0] mpls
[P-Serial2/0] mpls ldp
[P-Serial2/0] quit
```

配置连接 PE 2 的接口 Serial2/1，在此接口上使能 LDP。

```
[P] interface serial 2/1
[P-Serial2/1] link-protocol ppp
[P-Serial2/1] ip address 10.2.2.2 24
[P-Serial2/1] mpls
[P-Serial2/1] mpls ldp
[P-Serial2/1] quit
```

在 P 上运行 OSPF，用于建立 LSP。

```
[P] ospf
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 10.1.1.2 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 10.2.2.2 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 192.4.4.4 0.0.0.0
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

(4) 配置 PE 2

配置 LSR ID，全局使能 MPLS。

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 192.3.3.3 32
[PE2-LoopBack0] quit
[PE2] mpls lsr-id 192.3.3.3
[PE2] mpls
```

配置触发 LSP 的建立。

```
[PE2-mpls] lsp-trigger all
[PE2-mpls] quit
```

全局使能 MPLS L2VPN 和 LDP。

```
[PE2] mpls l2vpn
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

配置 PE 2 与 PE 1 建立 LDP 远程会话。

```
[PE2] mpls ldp remote-peer 2
```

```
[PE2-mpls-ldp-remote-2] remote-ip 192.2.2.2
[PE2-mpls-ldp-remote-2] quit
```

配置连接 P 的接口 Serial2/1，在此接口上使能 LDP。

```
[PE2] interface serial 2/1
[PE2-Serial2/1] link-protocol ppp
[PE2-Serial2/1] ip address 10.2.2.1 24
[PE2-Serial2/1] mpls
[PE2-Serial2/1] mpls ldp
[PE2-Serial2/1] quit
```

在 PE 2 上运行 OSPF，用于建立 LSP。

```
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 192.3.3.3 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.2.2.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

在接入 CE 2 的接口 Serial2/0 上创建 L2VPN 连接。此接口不需配置 IP 地址。

```
[PE2] interface serial 2/0
[PE2-Serial2/0] mpls l2vc 192.2.2.2 101
[PE2-Serial2/0] quit
```

(5) 配置 CE 2

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] interface serial 2/0
[CE2-Serial2/0] ip address 100.1.1.2 24
```

(6) 配置完成后的检验

在 PE 1 上查看 L2VPN 连接信息，可以看到建立了一条 L2VC。

```
[PE1] display mpls l2vc
total  ldp vc : 1     1 up       0 down

Transport  Client       VC      Local      Remote     Tunnel
VC ID      Intf         State   VC Label   VC Label   Policy
101        S2/0    up      1024       1025       default
```

在 PE 2 上也可以看到 L2VC 连接。

```
[PE2] display mpls l2vc
total  ldp vc : 1     1 up       0 down

Transport  Client       VC      Local      Remote     Tunnel
VC ID      Intf         State   VC Label   VC Label   Policy
101        S2/0    up      1025       1024       default
```

在 CE 1 上 ping CE 2 的接口地址 100.1.1.2，能够 ping 通。

```
[CE1] ping 100.1.1.2
  PING 100.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 100.1.1.2: bytes=56 Sequence=1 ttl=255 time=30 ms
```

```
  Reply from 100.1.1.2: bytes=56 Sequence=2 ttl=255 time=60 ms
  Reply from 100.1.1.2: bytes=56 Sequence=3 ttl=255 time=50 ms
  Reply from 100.1.1.2: bytes=56 Sequence=4 ttl=255 time=40 ms
  Reply from 100.1.1.2: bytes=56 Sequence=5 ttl=255 time=70 ms
--- 100.1.1.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 30/50/70 ms
```

在 CE 1 与 CE 2 之间能够 ping 通。

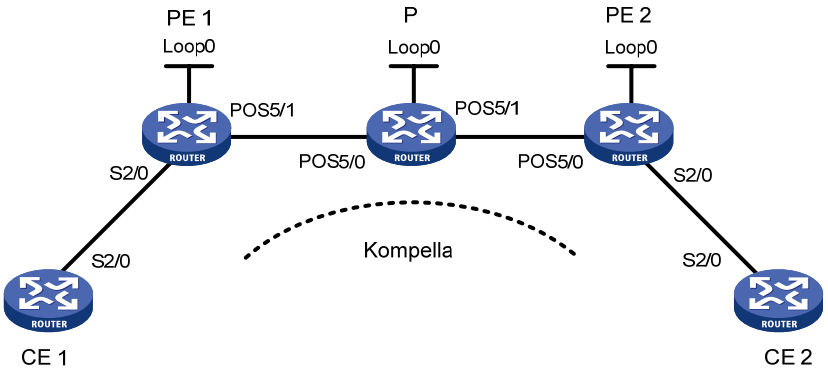
1.10.5 配置 Kompella 方式 MPLS L2VPN 示例

1. 组网需求

- CE 与 PE 之间通过 Serial 口相连，链路层封装 PPP。
- CE 1 和 CE 2 之间建立 Kompella 方式的 MPLS L2VPN。

2. 组网图

图1-7 配置 Kompella 方式 MPLS L2VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	S2/0	30.1.1.1/24	P	Loop0	2.2.2.9/32
PE 1	Loop0	1.1.1.9/32		POS5/0	168.1.1.2/24
	POS5/1	168.1.1.1/24		POS5/1	169.1.1.1/24
CE 2	S2/0	30.1.1.2/24	PE 2	Loop0	3.3.3.9/32
				POS5/0	169.1.1.2/24

3. 配置步骤

(1) 在 MPLS 骨干网上配置 IGP

本例中使用 OSPF，具体配置步骤略。

配置完成后，在各 LSR 上执行 **display ip routing-table** 命令可以看到都已学到彼此 LSR ID 的路由；执行 **display ospf peer** 命令可以看到建立了 OSPF 邻居关系，状态为 FULL。

(2) 配置 MPLS 基本能力和 LDP，建立 LDP LSP

具体配置步骤略。

配置完成后，在各 LSR 上执行 **display mpls ldp session** 和 **display mpls ldp peer** 命令可以看到 LDP 会话和对等体的建立情况，执行 **display mpls lsp** 命令可以看到 LSP 的建立情况。

(3) 配置 BGP 的 L2VPN 能力

配置 PE 1。

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] mpls l2vpn
[PE1] bgp 100
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp] l2vpn-family
[PE1-bgp-af-l2vpn] policy vpn-target
[PE1-bgp-af-l2vpn] peer 3.3.3.9 enable
[PE1-bgp-af-l2vpn] quit
[PE1-bgp] quit
```

配置 PE 2。

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] mpls l2vpn
[PE2] bgp 100
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 0
[PE2-bgp] l2vpn-family
[PE2-bgp-af-l2vpn] policy vpn-target
[PE2-bgp-af-l2vpn] peer 1.1.1.9 enable
[PE2-bgp-af-l2vpn] quit
[PE2-bgp] quit
```

配置完成后，在 PE 1 和 PE 2 上执行 **display bgp l2vpn peer** 命令可以看到 PE 之间建立了对等体关系，状态为 Established。

以 PE 1 为例：

```
[PE1] display bgp l2vpn peer
BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1                 Peers in established state : 1
  Peer      V   AS   MsgRcvd   MsgSent  OutQ PrefRcv Up/Down  State
  3.3.3.9   4   100        2        5     0       0  00:01:07 Established
```

(4) 配置 L2VPN 和 CE 连接

配置 PE 1。

```
[PE1] mpls l2vpn vpn1 encapsulation ppp
[PE1-mpls-l2vpn-vpn1] route-distinguisher 100:1
[PE1-mpls-l2vpn-vpn1] vpn-target 1:1
[PE1-mpls-l2vpn-vpn1] ce ce1 id 1 range 10
[PE1-mpls-l2vpn-ce-vpn1-ce1] connection ce-offset 2 interface serial 2/0
[PE1-mpls-l2vpn-ce-vpn1-ce1] quit
[PE1-mpls-l2vpn-vpn1] quit
```

配置 PE 2。

```
[PE2] mpls l2vpn vpn1 encapsulation ppp
[PE2-mpls-l2vpn-vpn1] route-distinguisher 100:1
```

```
[PE2-mpls-l2vpn-vpn1] vpn-target 1:1
[PE2-mpls-l2vpn-vpn1] ce ce2 id 2 range 10
[PE2-mpls-l2vpn-ce-vpn1-ce2] connection ce-offset 1 interface serial 2/0
[PE2-mpls-l2vpn-ce-vpn1-ce2] quit
[PE2-mpls-l2vpn-vpn1] quit
```

(5) 配置完成后的检验

完成上述配置后，在 PE 上执行 **display mpls l2vpn connection** 命令，可以看到建立了一条 L2VPN 连接，状态为 up。

以 PE 1 为例：

```
[PE1] display mpls l2vpn connection
1 total connections,
connections: 1 up, 0 down, 0 local, 1 remote, 0 unknown
VPN name: vpn1,
1 total connections,
connections: 1 up, 0 down, 0 local, 1 remote, 0 unknown
  CE name: ce1, id: 1,
  Rid type status peer-id         route-distinguisher   intf
  2   rmt  up     3.3.3.9         100:1                 S2/0
```

CE 1 与 CE 2 之间能够 ping 通。

```
[CE1] ping 30.1.1.2
  PING 30.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 30.1.1.2: bytes=56 Sequence=1 ttl=255 time=90 ms
    Reply from 30.1.1.2: bytes=56 Sequence=2 ttl=255 time=77 ms
    Reply from 30.1.1.2: bytes=56 Sequence=3 ttl=255 time=34 ms
    Reply from 30.1.1.2: bytes=56 Sequence=4 ttl=255 time=46 ms
    Reply from 30.1.1.2: bytes=56 Sequence=5 ttl=255 time=94 ms
  --- 30.1.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 34/68/94 ms
```

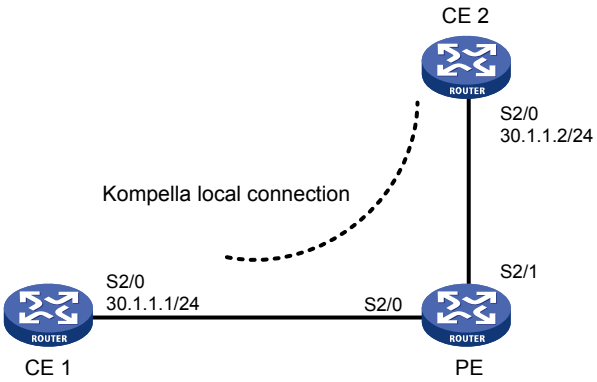
1.10.6 配置 Kompella 方式本地连接示例

1. 组网需求

CE 1 和 CE 2 之间建立 Kompella 方式本地连接。

2. 组网图

图1-8 配置 Kompella 方式本地连接组网图



3. 配置步骤

(1) 配置 PE

配置 MPLS 基本能力。

具体配置步骤略。

配置 L2VPN 和 CE 连接。

```
<Sysname> system-view
[Sysname] sysname PE
[PE] mpls l2vpn
[PE] mpls l2vpn vpn1 encapsulation ppp
[PE-mpls-l2vpn-vpn1] route-distinguisher 100:1
[PE-mpls-l2vpn-vpn1] vpn-target 111:1
[PE-mpls-l2vpn-vpn1] ce ce1 id 1
[PE-mpls-l2vpn-ce-vpn1-ce1] connection ce-offset 2 interface serial 2/0
[PE-mpls-l2vpn-ce-vpn1-ce1] quit
[PE-mpls-l2vpn-vpn1] ce ce2 id 2
[PE-mpls-l2vpn-ce-vpn1-ce2] connection ce-offset 1 interface serial 2/1
[PE-mpls-l2vpn-vpn1] quit
```

(2) 配置完成后的检验

完成上述配置后，在 PE 上执行 **display mpls l2vpn connection** 命令，可以看到建立了两条 L2VPN 的本地连接，状态为 up。

```
[PE] display mpls l2vpn connection
2 total connections,
connections: 2 up, 0 down, 2 local, 0 remote, 0 unknown

VPN name: vpn1,
2 total connections,
connections: 2 up, 0 down, 2 local, 0 remote, 0 unknown
  CE name: ce1, id: 1,
  Rid type status peer-id         route-distinguisher   intf
  2   loc  up     ---             ---                   S2/0
```

```
  CE name: ce2, id: 2,
  Rid type status peer-id         route-distinguisher   intf
  1   loc  up     ---             ---                   S2/1
```

查看 VPN 本地连接的配置。

```
[PE] display mpls l2vpn vpn-name vpn1 local-ce
ce-name              ce-id   range   conn-num  LB
ce1                  1       10      1         8192/0/10
ce2                  2       10      1         8202/0/10
```

CE 1 与 CE 2 之间能够 ping 通。

```
[CE1] ping 30.1.1.2
  PING 30.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 30.1.1.2: bytes=56 Sequence=1 ttl=255 time=90 ms
    Reply from 30.1.1.2: bytes=56 Sequence=2 ttl=255 time=77 ms
    Reply from 30.1.1.2: bytes=56 Sequence=3 ttl=255 time=34 ms
    Reply from 30.1.1.2: bytes=56 Sequence=4 ttl=255 time=46 ms
    Reply from 30.1.1.2: bytes=56 Sequence=5 ttl=255 time=94 ms
  --- 30.1.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 34/68/94 ms
```

1.11 常见配置错误举例

1.11.1 配置后不能 Ping 通

1. 错误现象

L2VPN 配置后，Ping 对端失败，查看 VC 状态，发现 VC 状态为 down，Remote 值为无效值。

2. 分析

VC 状态为 down 可能是因为两端封装类型不一致。

3. 处理过程

- 查看本端和对端 PE 设备配置的封装类型是否一致，如果配置的封装类型不一样，连接将失败。
- 检查两端是否已配置了 Remote 参数，并正确设置了对端的地址。

1.11.2 不同链路协议的串口不能建立 LDP 方式链接

1. 错误现象

Serial1/0 和 Serial1/1 分别封装 HDLC 和 PPP，各创建一个 LDP 方式连接，VC ID 都是 1。将 Serial1/1 的封装类型改为 HDLC，则不能建立 LDP 方式链接。

2. 分析

将 Serial1/1 的封装类型改为 HDLC，则会造成两个 HDLC 封装的 LDP 连接，其 VC ID 都是 1。为了避免 VC ID 冲突，此时 Serial1/1 上的 LDP 连接会被自动删除。

3. 处理过程

重新建立链接，基于接口 Serial1/0 和 Serial1/1 创建 LDP 方式链接，VC ID 不同。

目 录

1 MPLS L3VPN配置 1-1
1.1 MPLS L3VPN简介 1-1
1.1.1 MPLS L3VPN概述 1-1
1.1.2 MPLS L3VPN的基本概念 1-2
1.1.3 MPLS L3VPN的报文转发 1-4
1.1.4 MPLS L3VPN的网络架构 1-5
1.1.5 MPLS L3VPN的路由信息发布 1-8
1.1.6 跨域VPN 1-9
1.1.7 运营商的运营商 1-12
1.1.8 嵌套VPN 1-13
1.1.9 多角色主机 1-15
1.1.10 HoVPN 1-15
1.1.11 OSPF VPN扩展 1-17
1.1.12 BGP的AS号替换 1-20
1.2 MPLS L3VPN配置任务简介 1-21
1.3 配置VPN实例 1-22
1.3.1 创建VPN实例 1-22
1.3.2 配置VPN实例与接口关联 1-22
1.3.3 配置VPN实例的路由相关属性 1-23
1.3.4 配置VPN实例的隧道策略 1-24
1.4 配置基本MPLS L3VPN 1-25
1.4.1 配置准备 1-25
1.4.2 配置VPN实例 1-25
1.4.3 配置PE-CE间路由交换 1-25
1.4.4 配置PE-PE间的路由交换 1-29
1.4.5 配置BGP-VPNv4 子地址族下的路由特性 1-29
1.5 配置跨域VPN 1-32
1.5.1 配置准备 1-32
1.5.2 配置跨域VPN-OptionA 1-32
1.5.3 配置跨域VPN-OptionB 1-32
1.5.4 配置跨域VPN-OptionC 1-33
1.6 配置嵌套VPN 1-35
1.6.1 配置准备 1-35
1.6.2 配置嵌套VPN 1-35
1.7 配置多角色主机 1-36
1.7.1 配置准备 1-36
1.7.2 配置策略路由及应用 1-37
1.7.3 配置静态路由 1-37
1.8 配置HoVPN 1-37
1.8.1 配置准备 1-37
1.8.2 配置HoVPN 1-37

1.9 配置OSPF伪连接 1-38
1.9.1 配置准备 1-38
1.9.2 配置Loopback接口 1-38
1.9.3 发布Loopback接口的路由 1-38
1.9.4 配置伪连接 1-39
1.10 配置Multi-VPN-Instance CE 1-39
1.10.1 配置准备 1-39
1.10.2 配置Multi-VPN-Instance CE与各路由协议进行路由交换 1-39
1.11 Egress PE上私网路由标签操作方式简介 1-43
1.12 配置BGP的AS号替换 1-44
1.12.1 配置准备 1-44
1.12.2 配置BGP的AS号替换 1-44
1.13 MPLS L3VPN显示和维护 1-45
1.13.1 复位BGP连接 1-45
1.13.2 显示MPLS L3VPN的运行状态 1-45
1.14 MPLS L3VPN典型配置举例 1-46
1.14.1 配置MPLS L3VPN示例 1-46
1.14.2 配置MPLS L3VPN采用GRE隧道示例 1-54
1.14.3 配置跨域VPN-OptionA方式示例 1-59
1.14.4 配置跨域VPN-OptionB方式示例 1-64
1.14.5 配置跨域VPN-OptionC方式示例 1-69
1.14.6 配置运营商的运营商示例 1-75
1.14.7 配置嵌套VPN示例 1-82
1.14.8 配置多角色主机示例 1-92
1.14.9 配置HoVPN示例 1-93
1.14.10 配置OSPF伪连接 1-100
1.14.11 配置BGP的AS号替换 1-106

1 MPLS L3VPN 配置

说明

本章只涉及 MPLS L3VPN 的介绍和配置，有关 MPLS 基本配置、BGP 协议和 VPN 等的介绍和配置请参见相关分册。

1.1 MPLS L3VPN 简介

1.1.1 MPLS L3VPN 概述

MPLS L3VPN 是服务提供商 VPN 解决方案中一种基于 PE 的 L3VPN 技术，它使用 BGP 在服务提供商骨干网上发布 VPN 路由，使用 MPLS 在服务提供商骨干网上转发 VPN 报文。

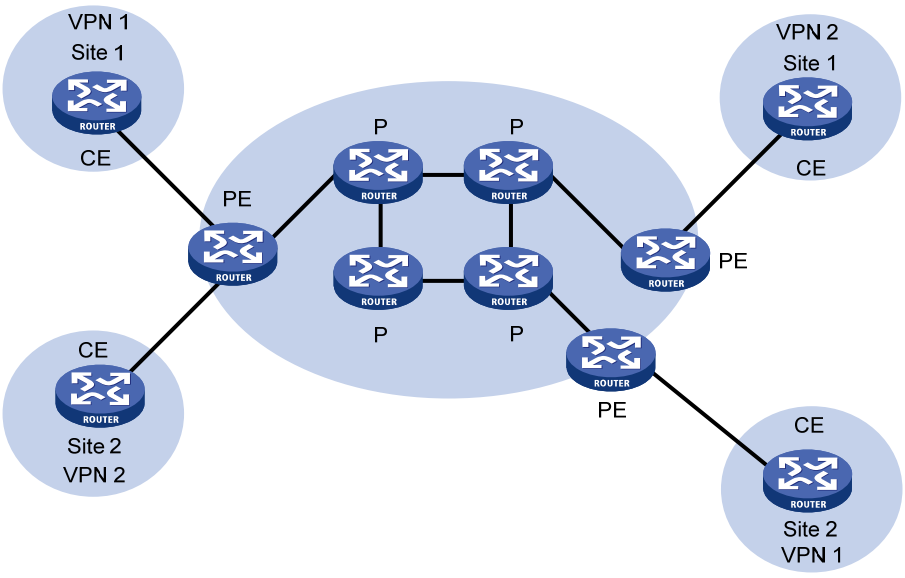
MPLS L3VPN 组网方式灵活、可扩展性好，并能够方便地支持 MPLS QoS 和 MPLS TE，因此得到越来越多的应用。

MPLS L3VPN 模型由三部分组成：CE、PE 和 P。

- CE（Customer Edge）设备：用户网络边缘设备，有接口直接与 SP（Service Provider，服务提供商）相连。CE 可以是路由器或交换机，也可以是一台主机。CE“感知”不到 VPN 的存在，也不需要必须支持 MPLS。
- PE（Provider Edge）路由器：服务提供商边缘路由器，是服务提供商网络的边缘设备，与用户的 CE 直接相连。在 MPLS 网络中，对 VPN 的所有处理都发生在 PE 上。
- P（Provider）路由器：服务提供商网络中的骨干路由器，不与 CE 直接相连。P 设备只需要具备基本 MPLS 转发能力。

图 1-1是一个MPLS L3VPN组网方案的示意图。

图1-1 MPLS L3VPN 组网



CE 和 PE 的划分主要是根据 SP 与用户的管理范围，CE 和 PE 是两者管理范围的边界。

CE 设备通常是一台路由器，当 CE 与直接相连的 PE 建立邻接关系后，CE 把本站点的 VPN 路由发布给 PE，并从 PE 学到远端 VPN 的路由。CE 与 PE 之间使用 BGP/IGP 交换路由信息，也可以使用静态路由。

PE 从 CE 学到 CE 本地的 VPN 路由信息后，通过 BGP 与其它 PE 交换 VPN 路由信息。PE 路由器只维护与它直接相连的 VPN 的路由信息，不维护服务提供商网络中的所有 VPN 路由。

P 路由器只维护到 PE 的路由，不需要了解任何 VPN 路由信息。

当在 MPLS 骨干网上传输 VPN 流量时，入口 PE 做为 Ingress LSR（Label Switch Router，标签交换路由器），出口 PE 做为 Egress LSR，P 路由器则做为 Transit LSR。

1.1.2 MPLS L3VPN 的基本概念

1. Site

在介绍 VPN 时经常会提到“Site”，Site（站点）的含义可以从下述几个方面理解：

- Site 是指相互之间具备 IP 连通性的一组 IP 系统，并且，这组 IP 系统的 IP 连通性不需通过服务提供商网络实现；
- Site 的划分是根据设备的拓扑关系，而不是地理位置，尽管在大多数情况下一个 Site 中的设备地理位置相邻；
- 一个 Site 中的设备可以属于多个 VPN，换言之，一个 Site 可以属于多个 VPN；
- Site 通过 CE 连接到服务提供商网络，一个 Site 可以包含多个 CE，但一个 CE 只属于一个 Site。

对于多个连接到同一服务提供商网络的 Sites，通过制定策略，可以将它们划分为不同的集合（set），只有属于相同集合的 Sites 之间才能通过服务提供商网络互访，这种集合就是 VPN。

2. 地址空间重叠

VPN 是一种私有网络，不同的 VPN 独立管理自己使用的地址范围，也称为地址空间（Address Space）。

不同 VPN 的地址空间可能会在一定范围内重合，比如，VPN1 和 VPN2 都使用了 10.110.10.0/24 网段的地址，这就发生了地址空间重叠（Overlapping Address Spaces）。

3. VPN 实例

在 MPLS VPN 中，不同 VPN 之间的路由隔离通过 VPN 实例（VPN-instance）实现。

PE 为每个直接相连的 Site 建立并维护专门的 VPN 实例。VPN 实例中包含对应 Site 的 VPN 成员关系和路由规则。如果一个 Site 中的用户同时属于多个 VPN，则该 Site 的 VPN 实例中将包括所有这些 VPN 的信息。

为保证 VPN 数据的独立性和安全性，PE 上每个 VPN 实例都有相对独立的路由表和 LFIB（Label Forwarding Information Base，标签转发表）。

具体来说，VPN 实例中的信息包括：标签转发表、IP 路由表、与 VPN 实例绑定的接口以及 VPN 实例的管理信息。VPN 实例的管理信息包括 RD（Route Distinguisher，路由标识符）、路由过滤策略、成员接口列表等。

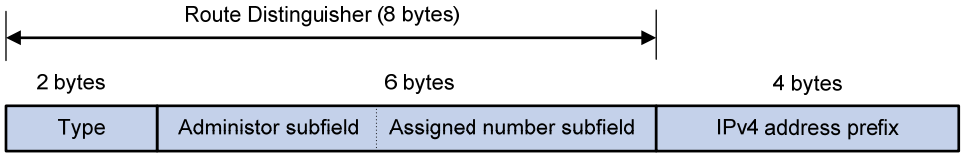
4. VPN-IPv4 地址

传统 BGP 无法正确处理地址空间重叠的 VPN 的路由。假设 VPN1 和 VPN2 都使用了 10.110.10.0/24 网段的地址，并各自发布了一条去往此网段的路由，BGP 将只会选择其中一条路由，从而导致去往另一个 VPN 的路由丢失。

PE 路由器之间使用 MP-BGP 来发布 VPN 路由，并使用 VPN-IPv4 地址族来解决上述问题。

VPN-IPv4 地址共有 12 个字节，包括 8 字节的RD和 4 字节的IPv4 地址前缀，如 图 1-2所示。

图1-2 VPN-IPv4 地址结构



PE 从 CE 接收到普通 IPv4 路由后，需要将这些私网 VPN 路由发布给对端 PE。私网路由的独立性是通过为这些路由附加 RD 实现的。

SP 可以独立地分配 RD，但必须保证 RD 的全局唯一性。这样，即使来自不同服务提供商的 VPN 使用了同样的 IPv4 地址空间，PE 路由器也可以向各 VPN 发布不同的路由。

建议为 PE 上每个 VPN 实例配置专门的 RD，以保证到达同一 CE 的路由都使用相同的 RD。RD 为 0 的 VPN-IPv4 地址相当于全局唯一的 IPv4 地址。

RD 的作用是添加到一个特定的 IPv4 前缀，使之成为全局唯一的 VPN IPv4 前缀。

RD 或者是与自治系统号（ASN）相关的，在这种情况下，RD 是由一个自治系统号和一个任意的数组成；或者是与 IP 地址相关的，在这种情况下，RD 是由一个 IP 地址和一个任意的数组成。

RD 有三种格式，通过 2 字节的 Type 字段区分：

- Type 为 0 时，Administrator 子字段占 2 字节，Assigned number 子字段占 4 字节，格式为：16bits 自治系统号:32bits 用户自定义数字。例如：100:1
- Type 为 1 时，Administrator 子字段占 4 字节，Assigned number 子字段占 2 字节，格式为：32bitsIPv4 地址:16bits 用户自定义数字。例如：172.1.1.1:1
- Type 为 2 时，Administrator 子字段占 4 字节，Assigned number 子字段占 2 字节，格式为：32bits 自治系统号：16bits 用户自定义数字，其中的自治系统号最小值为 65536。例如：65536:1

为保证 RD 的全局唯一性，建议不要将 Administrator 子字段的值设置为私有 AS 号或私有 IP 地址。

5. VPN Target 属性

MPLS L3VPN 使用 BGP 扩展团体属性——VPN Target（也称为 Route Target）来控制 VPN 路由信息的发布。

PE 路由器上的 VPN 实例有两类 VPN Target 属性：

- Export Target 属性：在本地 PE 将从与自己直接相连的 Site 学到的 VPN-IPv4 路由发布给其它 PE 之前，为这些路由设置 Export Target 属性；
- Import Target 属性：PE 在接收到其它 PE 路由器发布的 VPN-IPv4 路由时，检查其 Export Target 属性，只有当此属性与 PE 上 VPN 实例的 Import Target 属性匹配时，才把路由加入到相应的 VPN 路由表中。

也就是说，VPN Target 属性定义了一条 VPN-IPv4 路由可以为哪些 Site 所接收，PE 路由器可以接收哪些 Site 发送来的路由。

与 RD 类似，VPN Target 也有三种格式：

- 16bits 自治系统号:32bits 用户自定义数字，例如：100:1。
- 32bits IPv4 地址:16bits 用户自定义数字，例如：172.1.1.1:1。
- 32bits 自治系统号：16bits 用户自定义数字，其中的自治系统号最小值为 65536。例如：65536:1。

6. MP-BGP

MP-BGP（Multiprotocol extensions for BGP-4）在 PE 路由器之间传播 VPN 组成信息和路由。MP-BGP 向下兼容，既可以支持传统的 IPv4 地址族，又可以支持其他地址族（比如 VPN-IPv4 地址族）。使用 MP-BGP 既确保 VPN 的私网路由只在 VPN 内发布，又实现了 MPLS VPN 成员间的通信。

7. 路由策略（Routing Policy）

在通过入口、出口扩展团体来控制 VPN 路由发布的基础上，如果需要更精确地控制 VPN 路由的引入和发布，可以使用入方向或出方向路由策略。

入方向路由策略根据路由的 VPN Target 属性进一步过滤可引入到 VPN 实例的路由，它可以拒绝接收引入列表中的团体选定的路由，而出方向路由策略则可以拒绝发布输出列表中的团体选定的路由。

VPN 实例创建完成后，可以选择是否需要配置入方向或出方向路由策略。

8. 隧道策略（Tunneling Policy）

隧道策略用于选择给特定 VPN 实例的报文使用的隧道。

隧道策略是可选配的，VPN 实例创建完成后，就可以配置隧道策略。缺省情况下，选择 LSP 做为隧道，不进行负载分担（负载分担条数为 1）。另外，隧道策略只在同一 AS 域内生效。

1.1.3 MPLS L3VPN 的报文转发

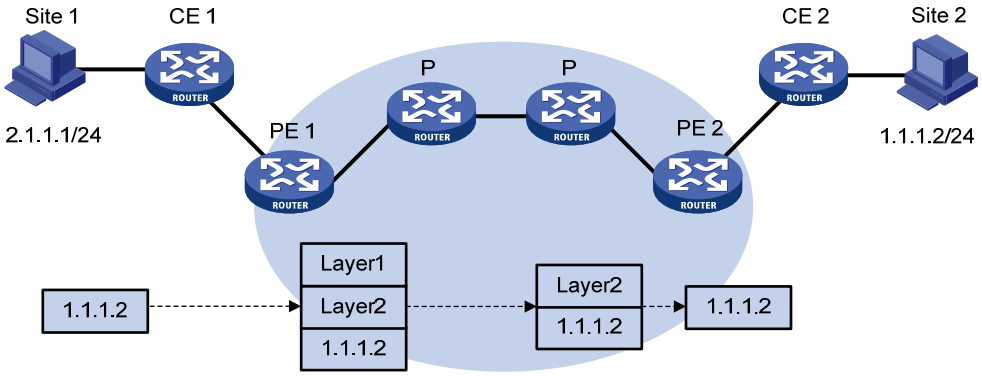
在基本 MPLS L3VPN 应用中（不包括跨域的情况），VPN 报文转发采用两层标签方式：

- 第一层（外层）标签在骨干网内部进行交换，指示从 PE 到对端 PE 的一条 LSP。VPN 报文利用这层标签，可以沿 LSP 到达对端 PE；
- 第二层（内层）标签在从对端 PE 到达 CE 时使用，指示报文应被送到哪个 Site，或者更具体一些，到达哪一个 CE。这样，对端 PE 根据内层标签可以找到转发报文的接口。

特殊情况下，属于同一个 VPN 的两个 Site 连接到同一个 PE，这种情况下只需要知道如何到达对端 CE。

以 图 1-3为例，说明VPN报文的转发：

图1-3 VPN 报文转发示意图



(1) Site 1 发出一个目的地址为 1.1.1.2 的 IP 报文，由 CE 1 将报文发送至 PE 1。

(2) PE 1 根据报文到达的接口及目的地址查找 VPN 实例表项，匹配后将报文转发出去，同时打上内层和外层两个标签。

(3) MPLS 网络利用报文的外层标签，将报文传送到 PE 2（报文在到达 PE 2 前一跳时已经被剥离外层标签，仅含内层标签）。

(4) PE 2 根据内层标签和目的地址查找 VPN 实例表项，确定报文的出接口，将报文转发至 CE 2。

(5) CE 2 根据正常的 IP 转发过程将报文传送到目的地。

1.1.4 MPLS L3VPN 的网络架构

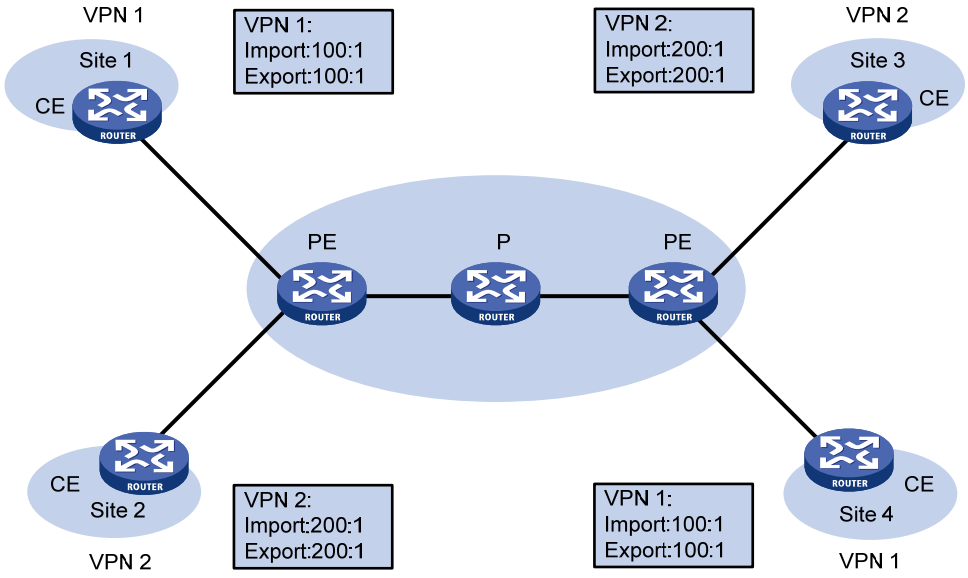
在 MPLS L3VPN 网络中，通过 VPN Target 属性来控制 VPN 路由信息在各 Site 之间的发布和接收。VPN Export Target 和 Import Target 的设置相互独立，并且都可以设置多个值，能够实现灵活的 VPN 访问控制，从而实现多种 VPN 组网方案。

1. 基本的 VPN 组网方案

最简单的情况下，一个 VPN 中的所有用户形成闭合用户群，相互之间能够进行流量转发，VPN 中的用户不能与任何本 VPN 以外的用户通信。

对于这种组网，需要为每个 VPN 分配一个 VPN Target，作为该 VPN 的 Export Target 和 Import Target，并且，此 VPN Target 不能被其他 VPN 使用。

图1-4 基本的 VPN 组网方案



在 图 1-4中，PE上为VPN 1 分配的VPN Target值为 100:1，为VPN 2 分配的VPN Target值为 200:1。VPN 1 的两个Site之间可以互访，VPN 2 的两个Site之间也可以互访，但VPN 1 和VPN 2 的Site之间不能互访。

2. Hub&Spoke 组网方案

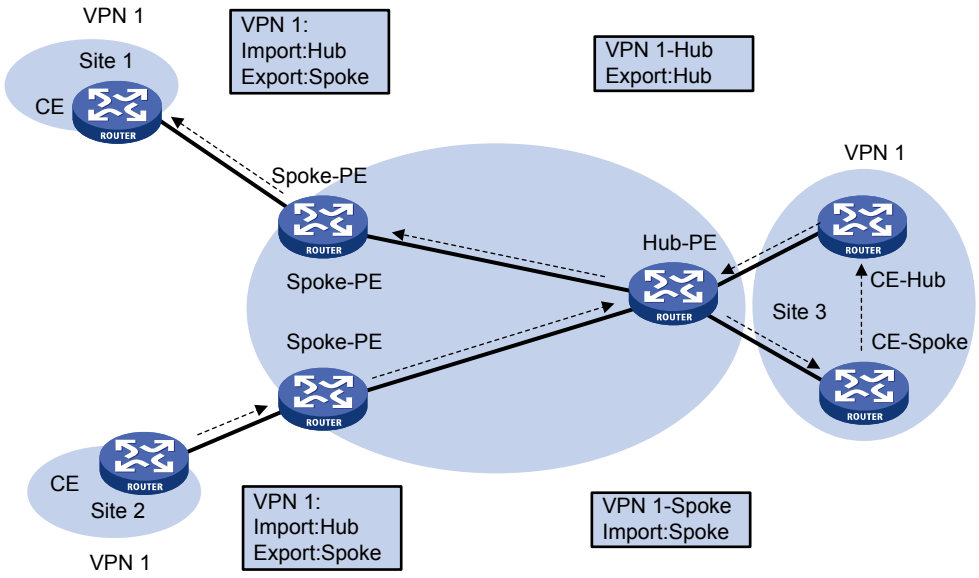
如果希望在 VPN 中设置中心访问控制设备，其它用户的互访都通过中心访问控制设备进行，可以使用 Hub&Spoke 组网方案，从而实现中心设备对两端设备之间的互访进行监控和过滤等功能。

对于这种组网，需要设置两个 VPN Target，一个表示“Hub”，另一个表示“Spoke”。

各 Site 在 PE 上的 VPN 实例的 VPN Target 设置规则为：

- 连接 Spoke 站点（Site 1 和 Site 2）的 Spoke-PE：Export Target 为“Spoke”，Import Target 为“Hub”；
- 连接 Hub 站点（Site 3）的 Hub-PE：Hub-PE 上需要使用两个接口或子接口，一个用于接收 Spoke-PE 发来的路由，其 VPN 实例的 Import Target 为“Spoke”；另一个用于向 Spoke-PE 发布路由，其 VPN 实例的 Export Target 为“Hub”。

图1-5 Hub&Spoke 组网方案



在 图 1-5中，Spoke站点之间的通信通过Hub站点进行（图中箭头所示为Site 2 的路由向Site 1 的发布过程）：

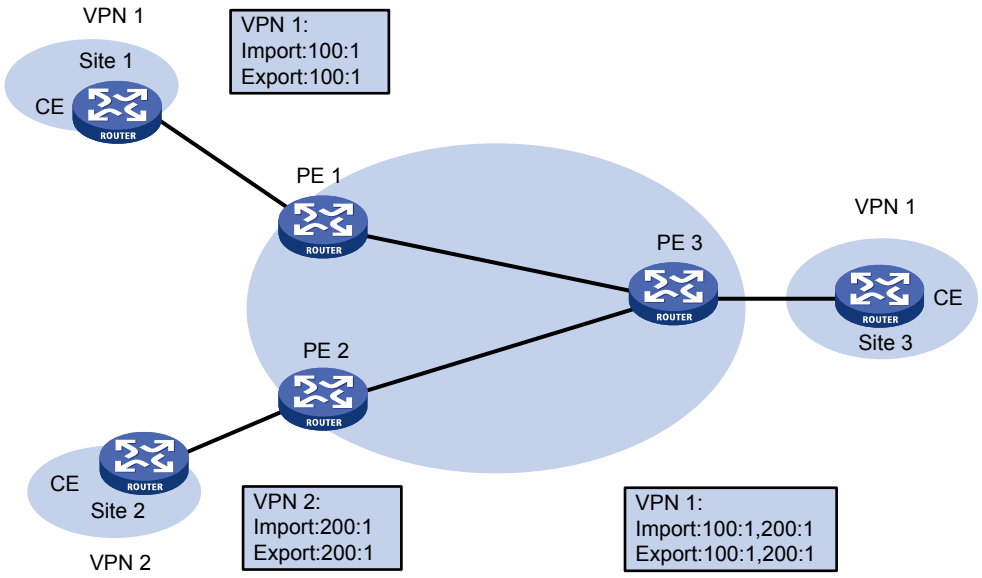
- Hub-PE 能够接收所有 Spoke-PE 发布的 VPN-IPv4 路由；
- Hub-PE 发布的 VPN-IPv4 路由能够为所有 Spoke-PE 接收；
- Hub-PE 将从 Spoke-PE 学到的路由发布给其他 Spoke-PE，因此，Spoke 站点之间可以通过 Hub 站点互访。
- 任意 Spoke-PE 的 Import Target 属性不与其它 Spoke-PE 的 Export Target 属性相同。因此，任意两个 Spoke-PE 之间不直接发布 VPN-IPv4 路由，Spoke 站点之间不能直接互访。

3. Extranet 组网方案

如果一个 VPN 用户希望提供部分本 VPN 的站点资源给非本 VPN 的用户访问，可以使用 Extranet 组网方案。

对于这种组网，如果某个 VPN 需要访问共享站点，则该 VPN 的 Export Target 必须包含在共享站点的 VPN 实例的 Import Target 中，而其 Import Target 必须包含在共享站点 VPN 实例的 Export Target 中。

图1-6 Extranet 组网方案



在 图 1-6中，VPN 1 的Site 3 能够被VPN 1 和VPN 2 访问：

- PE 3 能够接受 PE 1 和 PE 2 发布的 VPN-IPv4 路由；
- PE 3 发布的 VPN-IPv4 路由能够为 PE 1 和 PE 2 接受；
- 基于以上两点，VPN 1 的 Site 1 和 Site 3 之间能够互访，VPN 2 的 Site 2 和 VPN 1 的 Site 3 之间能够互访。

PE 3 不把从 PE 1 接收的 VPN-IPv4 路由发布给 PE 2，也不把从 PE 2 接收的 VPN-IPv4 路由发布给 PE 1（IBGP 邻居学来的条目是不会再发送给别的 IBGP 邻居），因此，VPN 1 的 Site 1 和 VPN 2 的 Site 2 之间不能互访。

1.1.5 MPLS L3VPN 的路由信息发布

在基本 MPLS L3VPN 组网中，VPN 路由信息的发布涉及 CE 和 PE，P 路由器只维护骨干网的路由，不需要了解任何 VPN 路由信息。PE 路由器也只维护与它直接相连的 VPN 的路由信息，不维护所有 VPN 路由。因此，MPLS L3VPN 网络具有良好的可扩展性。

VPN 路由信息的发布过程包括三部分：本地 CE 到入口 PE、入口 PE 到出口 PE、出口 PE 到远端 CE。完成这三部分后，本地 CE 与远端 CE 之间将建立可达路由，VPN 私网路由信息能够在骨干网上发布。

下面分别对这三部分进行介绍。

1. 本地 CE 到入口 PE 的路由信息交换

CE 与直接相连的 PE 建立邻接关系后，把本站点的 VPN 路由发布给 PE。

CE 与 PE 之间可以使用静态路由、RIP、OSPF、IS-IS 或 EBGP。无论使用哪种路由协议，CE 发布给 PE 的都是标准的 IPv4 路由。

2. 入口 PE 到出口 PE 的路由信息交换

PE 从 CE 学到 VPN 路由信息后，为这些标准 IPv4 路由增加 RD 和 VPN Target 属性，形成 VPN-IPv4 路由，存放到为 CE 创建的 VPN 实例中。

入口 PE 通过 MP-BGP 把 VPN-IPv4 路由发布给出口 PE。出口 PE 根据 VPN-IPv4 路由的 Export Target 属性与自己维护的 VPN 实例的 Import Target 属性，决定是否将该路由加入到 VPN 实例的路由表。

PE 之间通过 IGP 来保证内部的连通性。

3. 出口 PE 到远端 CE 的路由信息交换

远端 CE 有多种方式可以从出口 PE 学习 VPN 路由，包括静态路由、RIP、OSPF、IS-IS 和 EBGP，与本地 CE 到入口 PE 的路由信息交换相同。

1.1.6 跨域 VPN

实际组网应用中，某用户一个 VPN 的多个 Site 可能会连接到使用不同 AS 号的多个服务提供商，或者连接到一个服务提供商的多个 AS。这种 VPN 跨越多个自治系统的应用方式被称为跨域 VPN（Multi-AS VPN）。

RFC 2547bis 中提出了三种跨域 VPN 解决方案，分别是：

- VRF-to-VRF：ASBR 间使用子接口管理 VPN 路由，也称为 Inter-Provider Option A；
- EBGP Redistribution of labeled VPN-IPv4 routes：ASBR 间通过 MP-EBGP 发布标签 VPN-IPv4 路由，也称为 Inter-Provider Option B；
- Multihop EBGP redistribution of labeled VPN-IPv4 routes：PE 间通过 MP-EBGP 发布标签 VPN-IPv4 路由，也称为 Inter-Provider Option C。

下面逐一对这些方案进行介绍。

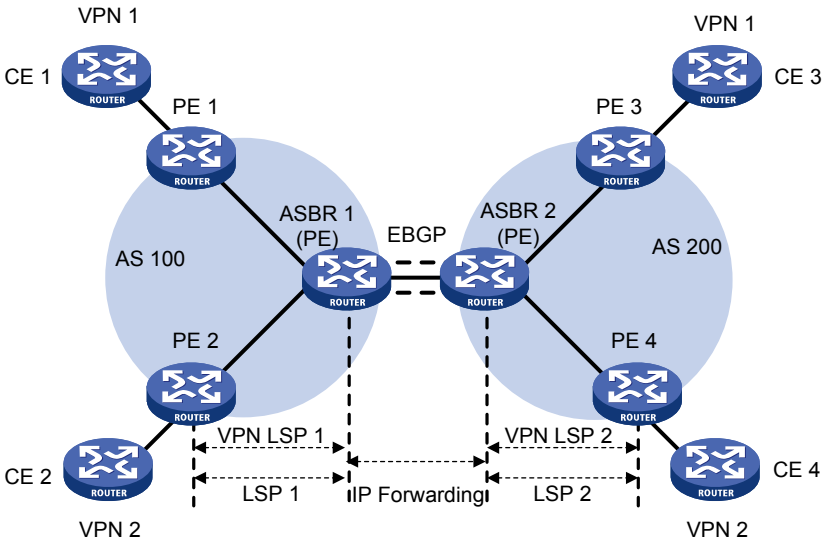
1. ASBR 间使用子接口管理 VPN 路由

这种方式下，两个 AS 的 PE 路由器直接相连，PE 路由器同时也是各自所在自治系统的边界路由器 ASBR。

作为 ASBR 的 PE 之间通过多个子接口相连，两个 PE 都把对方作为自己的 CE 设备对待，使用传统的 EBGP 方式向对端发布 IPv4 路由。报文在 AS 内部作为 VPN 报文，采用两层标签转发方式；在 ASBR 之间则采用普通 IP 转发方式。

理想情况下，每个跨域的 VPN 都有一对子接口与之对应，用来交换 VPN 路由信息。

图1-7 ASBR 间使用子接口管理 VPN 路由组网图



使用子接口实现跨域 VPN 的优点是实现简单：两个作为 ASBR 的 PE 之间不需要为跨域进行特殊配置。

缺点是可扩展性差：作为 ASBR 的 PE 需要管理所有 VPN 路由，为每个 VPN 创建 VPN 实例。这将导致 PE 上的 VPN-IPv4 路由数量过于庞大。并且，为每个 VPN 单独创建子接口也提高了对 PE 设备的要求。

2. ASBR 间通过 MP-EBGP 发布标签 VPN-IPv4 路由

这种方式下，两个 ASBR 通过 MP-EBGP 交换它们从各自 AS 的 PE 路由器接收的标签 VPN-IPv4 路由。

路由发布过程可分为以下步骤：

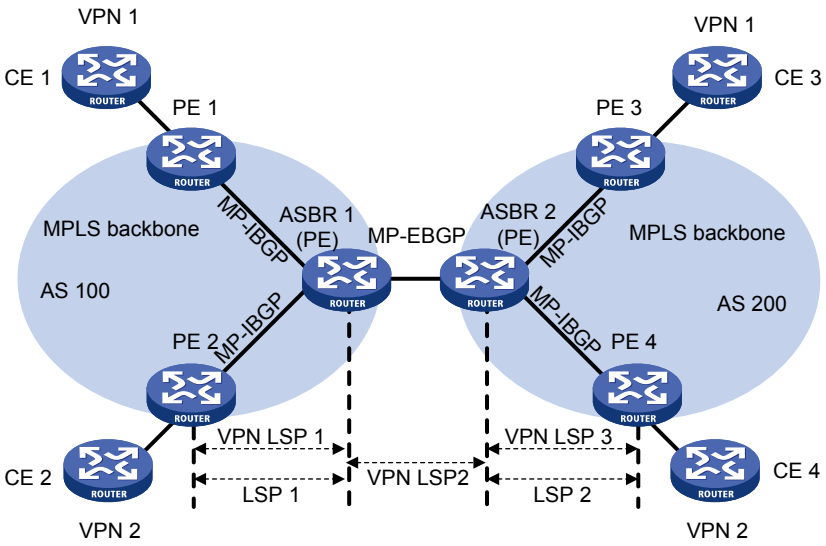
(1) AS 100 内的 PE 先通过 MP-IBGP 方式把标签 VPN-IPv4 路由发布给 AS 100 的边界路由器 PE，或发布给为 ASBR PE 反射路由的路由反射器；

(2) 作为 ASBR 的 PE 通过 MP-EBGP 方式把标签 VPN-IPv4 路由发布给 AS 200 的 PE（也是 AS 200 的边界路由器）；

(3) AS 200 的 ASBR PE 再通过 MP-IBGP 方式把标签 VPN-IPv4 路由发布给 AS 200 内的 PE，或发布给为 PE 反射路由的路由反射器。

这种方式的 ASBR 需要对标签 VPN-IPv4 路由进行特殊处理，因此也称为 ASBR 扩展方式。

图1-8 ASBR 间通过 MP-EBGP 发布标签 VPN-IPv4 路由组网图



在可扩展性方面，通过 MP-EBGP 发布标签 VPN-IPv4 路由优于 ASBR 间通过子接口管理 VPN。

采用 MP-EBGP 方式时，需要注意：

- ASBR 之间不对接收的 VPN-IPv4 路由进行 VPN Target 过滤，因此，交换 VPN-IPv4 路由的各 AS 服务提供商之间需要就这种路由交换达成信任协议；
- VPN-IPv4 路由交换仅发生在私网对等点之间，不能与公网交换 VPN-IPv4 路由，也不能与没有达成信任协议的 MP-EBGP 对等体交换 VPN-IPv4 路由。

3. PE 间通过 MP-EBGP 发布标签 VPN-IPv4 路由

前面介绍的两种方式都能够满足跨域 VPN 的组网需求，但这两种方式也都需要 ASBR 参与

VPN-IPv4 路由的维护和发布。当每个 AS 都有大量的 VPN 路由需要交换时，ASBR 就很可能成为阻碍网络进一步扩展的瓶颈。

解决上述可扩展性问题的方案是：ASBR 不维护或发布 VPN-IPv4 路由，PE 之间直接交换 VPN-IPv4 路由。

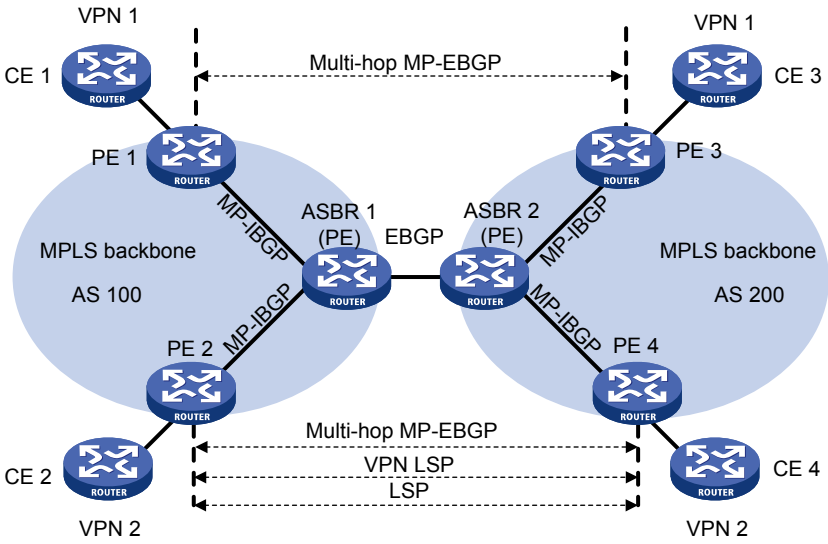
两个 ASBR 通过 MP-IBGP 向各自 AS 内的 PE 路由器发布标签 IPv4 路由。

ASBR 上不保存 VPN-IPv4 路由，相互之间也不通告 VPN-IPv4 路由。

ASBR 保存 AS 内 PE 的带标签的 IPv4 路由，并通告给其它 AS 的对等体。另一个自治系统中的 ASBR 也通告带标签的 IPv4 路由。这样，在入口 PE 和出口 PE 之间建立起一条 LSP。

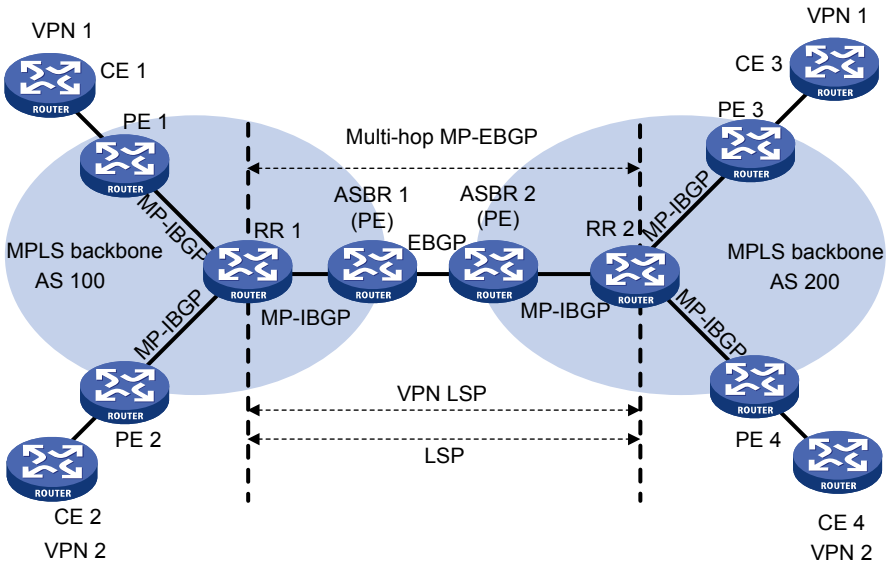
不同 AS 的 PE 之间建立 Multihop 方式的 EBGP 连接，交换 VPN-IPv4 路由。

图1-9 PE 间通过 Multi-hop MP-EBGP 发布标签 VPN-IPv4 路由组网图



为提高可扩展性，可以在每个AS中指定一个路由反射器RR（Route Reflector），由RR保存所有VPN-IPv4 路由，与AS的PE交换VPN-IPv4 路由信息。两个AS的RR之间建立跨域VPNv4 连接，通告VPN-IPv4 路由。如 图 1-10所示。

图1-10 采用 RR 的跨域 VPN OptionC 方式组网图



1.1.7 运营商的运营商

1. 简介

MPLS L3VPN 服务提供商的用户可能也是一个服务提供商。这种情况下，前者称为提供商运营商（Provider Carrier）或一级运营商（First Carrier），后者称为客户运营商（Customer Carrier）或二级运营商（Second Carrier）。这种组网模型称为运营商的运营商（Carriers' carriers），低级别的 SP 作为更高级别 SP 的 CE。

为保持良好的可扩展性，一级运营商并不引入二级运营商的外部路由，只引入为不同站点的二级运营商交换报文的路由。二级运营商维护的外部路由通过在相关二级运营商路由器间建立 BGP 会话交换。这可以大大减少一级运营商网络中需要维护的路由数量。

2. 实现

与普通 MPLS L3VPN 相比，配置运营商的运营商的主要区别在于一级运营商 PE 接入二级运营商 CE 这一部分：

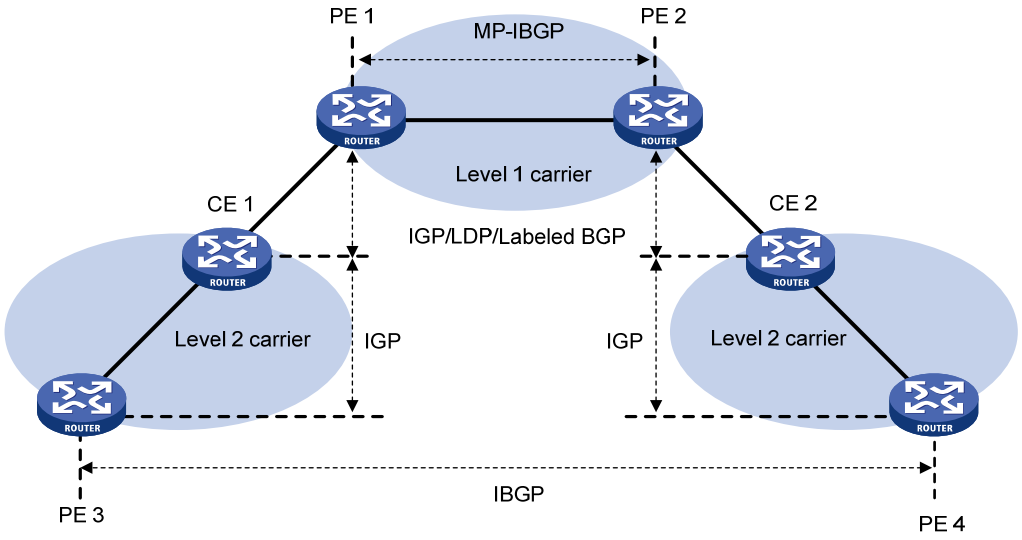
- 如果一级运营商 PE 与二级运营商 CE 位于同一个 AS 内，则它们之间配置 IGP 和 LDP；
- 如果一级运营商 PE 与二级运营商 CE 位于不同的 AS，则它们之间配置 MP-EBGP。MP-EBGP 为 PE 与 CE 之间交换的路由分配标签。

无论哪种情况，CE 上都需要使能 MPLS。并且，CE 上虽然有二级运营商的 VPN 路由，但它们并不把这些路由发布给一级运营商的 PE，只在二级运营商 PE 之间交换。

二级运营商可能只是普通 ISP，也可能是 MPLS L3VPN 服务提供商。

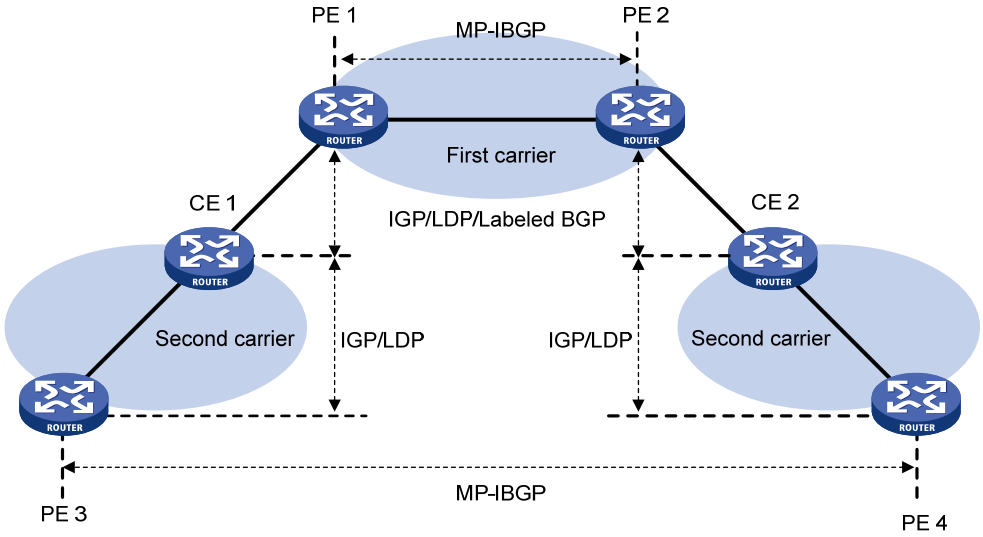
二级运营商是普通ISP时，其PE不需要运行MPLS，与CE之间运行IGP。PE 3 和PE 4 之间通过IBGP会话交换二级运营商的VPN路由。如 图 1-11所示。

图1-11 二级运营商是 ISP



二级运营商是MPLS L3VPN服务提供商时，其PE也需要运行MPLS，与CE之间运行IGP和LDP。PE 3 和PE 4 之间通过MP-IBGP会话交换二级运营商的VPN路由。如 图 1-12所示。

图1-12 二级运营商是 MPLS L3VPN 服务提供商



注意

如果一级运营商和二级运营商之间存在等价路由，则建议在一级运营商和二级运营商之间建立对应的等价 LSP。

1.1.8 嵌套 VPN

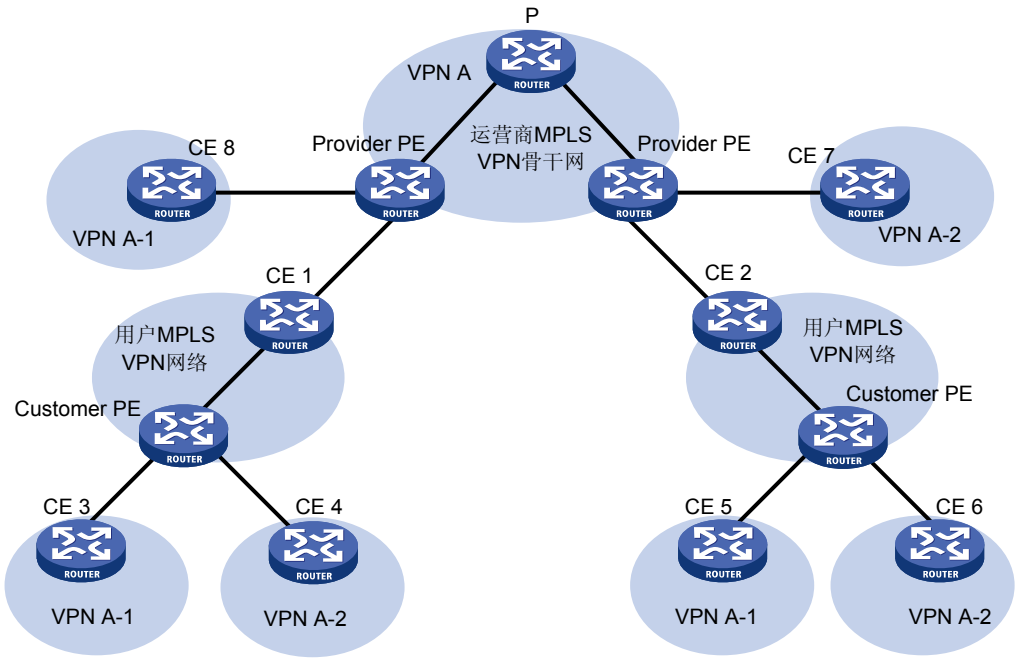
1. 产生背景

在基本的 MPLS L3VPN 模型中，由运营商经营 MPLS VPN 骨干网，通过 PE 设备提供 VPN 服务。VPN 用户通过 CE 设备与运营商的 PE 设备互连，接入 MPLS VPN 网络，实现属于用户 VPN 的不同 Site 之间的通信。在这种情况下，用户网络是普通的 IP 网络，VPN 用户不能再划分子 VPN。

实际组网中，VPN 用户网络复杂多样，繁简不同。当 VPN 用户内部需要进一步划分多个 VPN 时，传统的解决方案是直接在运营商的 PE 设备上配置用户内部的 VPN。这种方案实现简单，缺点是将导致 PE 上承载的 VPN 数量急剧增加，并且，如果用户内部需要调整 VPN 关系，也必须通过运营商操作。这不仅增加了网络的运营成本，也带来了管理和安全方面的问题。

嵌套VPN是一种更为完善的解决方案，它的主要思想是在普通MPLS L3VPN的PE和CE之间传递VPNv4 路由，由用户管理自己内部的VPN划分，运营商不参与用户内部VPN的管理。如 图 1-13所示，用户在运营商MPLS VPN网络上所属的VPN为VPN A，在用户网络内部划分子VPN：VPN A-1和VPN A-2。运营商PE设备把用户网络当作普通VPN用户对待，不参与子VPN的划分。VPN用户的CE设备（CE 1、CE 2、CE 7 和CE 8）和运营商PE设备之间传递包括子VPN路由信息的VPNv4 路由，从而实现用户网络内部子VPN路由信息的传递。

图1-13 嵌套 VPN 组网应用



2. 路由信息的传播过程

嵌套 VPN 中，路由信息的传播过程为：

(1) 运营商的 PE 设备与 VPN 用户的 CE 设备交互 VPNv4 路由，VPNv4 路由携带用户内部的 VPN 信息。

(2) 运营商的 PE 设备收到 VPNv4 路由后，保留用户内部的 VPN 信息，并附加用户在运营商网络上的 MPLS VPN 属性，即将该 VPNv4 路由的 RD 更换为用户所处运营商网络 VPN 的 RD，同时将用户所处运营商网络 VPN 的 ERT（Export Route-target，引出路由标识）添加到路由的扩展团体属性列表中。运营商的 PE 设备维护用户内部的 VPN 信息。

(3) 运营商的 PE 设备向其他运营商 PE 设备发布这些携带综合 VPN 信息的 VPNv4 路由。

(4) 其他的运营商PE设备收到VPNv4 路由后，与本地的VPN进行匹配，每个VPN接收属于自己的路由，然后将路由扩散到自己连接的子VPNCE设备（如 图 1-13中的CE 3、CE 4、CE 5 和 CE 6）：如果运营商PE和CE设备之间是IPv4 连接，直接扩散IPv4 路由；如果和CE设备之间是VPNv4 连接，表示通过私网连接的是一个用户MPLS VPN网络，向CE扩散VPNv4 路由。

3. 技术优点

嵌套 VPN 技术的主要优点是：

- 实现了 VPN 聚合功能，可以把用户的多个内部 VPN 聚合成一个用户 VPN，接入运营商的 MPLS VPN 网络；
- 支持对称组网方式和非对称组网方式，即属于同一 VPN 用户的不同 Site 包括的用户内部 VPN 数目可以相同，也可以不同；
- 支持用户内部 VPN 的多层嵌套。

嵌套 VPN 技术降低了用户接入 VPN 网络的复杂度和成本，为用户提供多样化的 VPN 组网方式，实现了用户对内部 VPN 以及多层 VPN 之间的互访权限控制管理。

1.1.9 多角色主机

从 CE 进入 PE 的报文的 VPN 属性由入接口绑定的 VPN 决定，这就决定了由同一入接口经 PE 转发的所有 CE 设备都必须属于同一 VPN。但在实际组网中，一个 CE 设备可能需要经过一个物理接口访问多个 VPN，这可以通过设置不同的逻辑接口来实现，但会增加额外的配置负担，使用起来也有局限性。

多角色主机，是通过在 PE 上配置策略路由，使来自 CE 的报文可以访问多个 VPN。

为使其它 VPN 的信息能够从 PE 发送到 CE，需要在其它 VPN 上配置静态路由，并指定连接 CE 的接口作为下一跳。



说明

配置多角色主机时需要注意：应保证与 PE 路由器关联的所有 IP 地址都是唯一的，以实现多角色主机特性。

在实际应用中，应合理进行地址规划，使每个 VPN 的地址段相对集中，以提高 PE 的转发效率。

1.1.10 HoVPN

1. 产生背景

(1) 分层模型与平面模型

在 MPLS L3VPN 解决方案中，PE 设备最为关键，它完成两方面的功能：首先是为用户提供接入功能，这需要 PE 具有大量接口；然后是管理和发布 VPN 路由，处理用户报文，这需要 PE 设备具有大容量存储和高转发能力。

目前的网络设计大多采用经典的分层结构，例如，城域网的典型结构是三层模型：核心层、汇聚层、接入层。从核心层到接入层，对设备的性能要求依次下降，网络的规模则依次扩大。

而 MPLS L3VPN 是一种平面模型，对网络中所有 PE 设备的性能要求相同，当网络中某些 PE 在性能和可扩展性方面存在问题时，整个网络的性能和可扩展性将受到影响。

由于 MPLS L3VPN 的平面模型与典型的分层网络模型不相符，在每一个层次上部署 PE 都会遇到扩展性问题，不利于大规模部署 VPN。

(2) HoVPN

为解决可扩展性问题，MPLS L3VPN 必然要从平面模型转变为分层模型。

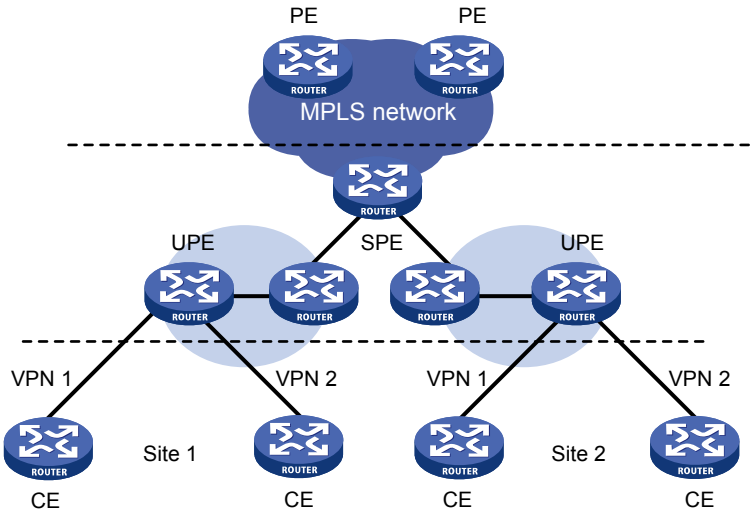
在 MPLS L3VPN 领域，分层 VPN（Hierarchy of VPN，简称 HoVPN）解决方案的提出，实现了将 PE 的功能分布到多个 PE 设备上，多个 PE 承担不同的角色，并形成层次结构，共同完成一个 PE 的功能。

HoVPN 对处于较高层次的设备的路由能力和转发性能要求较高，而对处于较低层次的设备的相应要求也较低，符合典型的分层网络模型。

2. HoVPN 的实现

(1) HoVPN 的基本结构

图1-14 HoVPN 的基本结构



在 图 1-14中，直接连结用户的设备称为下层PE（Underlayer PE）或用户侧PE（User-end PE），简写为UPE；连结UPE并位于网络内部的设备称为上层PE（Superstratum PE）或运营商侧PE（Sevice Provider-end PE），简写为SPE。

多个 UPE 与 SPE 构成分层式 PE，共同完成传统上一个 PE 的功能。

说明

HoVPN 把 PE 的功能分层实现，因此，这种解决方案有时也被称为分层 PE（Hiberarchy of PE，HoPE）。

SPE 与 UPE 的分工是：

- UPE 主要完成用户接入功能。UPE 维护其直接相连的 VPN Site 的路由，但不维护 VPN 中其它远程 Site 的路由或仅维护它们的聚合路由；UPE 为其直接相连的 Site 的路由分配内层标签，并通过 MP-BGP 随 VPN 路由发布此标签给 SPE；
- SPE 主要完成 VPN 路由的管理和发布。SPE 维护其通过 UPE 连接的 VPN 所有路由，包括本地和远程 Site 的路由，SPE 将路由信息发布给 UPE，并携带标签。SPE 发布的路由信息可以是 VPN 实例的缺省路由（或聚合路由），也可以是通过路由策略的路由信息。通过后者可以实现对同一 VPN 下不同站点之间互访的控制。

由于分工的不同，对 SPE 和 UPE 的要求也不同：SPE 的路由表容量大，转发性能强，但接口资源较少；UPE 的路由容量和转发性能较低，但接入能力强。HoVPN 充分利用了 SPE 的性能和 UPE 的接入能力。

需要说明的是，SPE 和 UPE 是相对的概念。在多个层次的 PE 结构中，上层 PE 相对于下层就是 SPE，下层 PE 相对于上层就是 UPE。

分层式 PE 从外部来看同传统上的 PE 没有区别，可以同普通 PE 共存于一个 MPLS 网络。

(2) SPE-UPE

SPE 和 UPE 之间运行 MP-BGP，可以是 MP-IBGP，也可以是 MP-EBGP，这取决于 UPE 和 SPE 是否属于同一个 AS。

采用 MP-IBGP 时，为了在 IBGP 对等体之间通告路由，SPE 将作为路由反射器，把来自 IBGP 对等体 UPE 的 VPN 路由发布给 IBGP 对等体 SPE，但 SPE 不作为其它 PE 的路由反射器。

(3) HoVPN 的嵌套与扩展

HoVPN 支持分层式 PE 的嵌套：

- 一个分层式 PE 可以作为 UPE，同另一个 SPE 组成新的分层式 PE；
- 一个分层式 PE 可以作为 SPE，同多个 UPE 组成新的分层式 PE；
- 以上这种嵌套可以多次进行。

通过分层式 PE 的嵌套，理论上可以将 VPN 无限扩展与延伸。

图1-15 分层式 PE 的嵌套

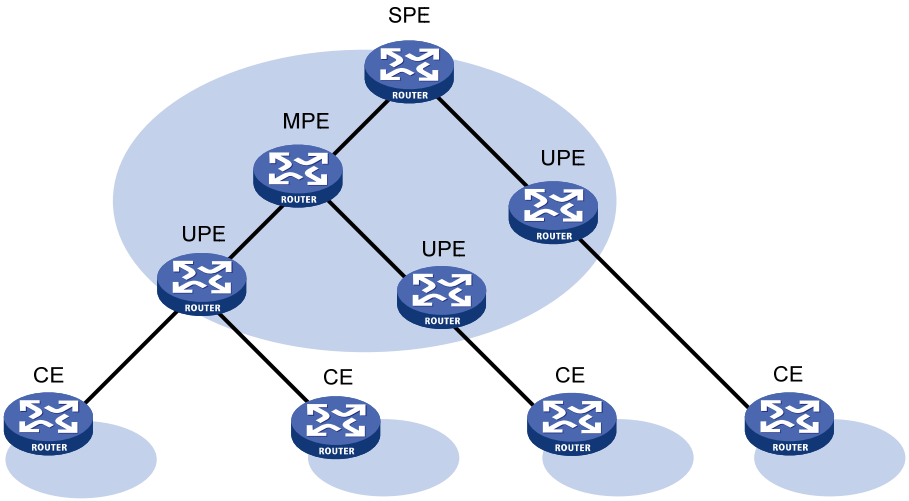


图 1-15是一个三层的分层式PE，称中间的PE为MPE（Middle-level PE）。SPE和MPE之间，以及MPE和UPE之间，均运行MP-BGP。

说明

“MPE”这种说法只是为了表述方便，在 HoVPN 模型中并没有 MPE。

MP-BGP 为上层 PE 发布下层 PE 上的所有 VPN 路由，为下层 PE 发布上层 PE 的 VPN 实例缺省路由或通过路由策略的 VPN 路由。

SPE 维护了这个分层式 PE 接入的所有 Site 的 VPN 路由，路由数目最多；UPE 只维护它所直接连接的 Site 的 VPN 路由，路由数目最少；MPE 的路由数目介于 SPE 和 UPE 之间。

1.1.11 OSPF VPN 扩展

说明

本节重点介绍 OSPF 对 VPN 的扩展，如果需要了解 OSPF 的基本知识，请参见“IP 路由分册”中的“OSPF 配置”。

1. PE 上的 OSPF 多实例

OSPF 是应用广泛的一种 IGP 协议，很多情况下，VPN 客户端通过 BGP 对等体连接，客户端内部则经常以 OSPF 作为内部路由协议。如果能够在 PE-CE 之间使用 OSPF，则 CE 上就不需要再为到 PE 的连接支持其它路由协议，从而简化 CE 的管理和配置。并且，如果客户需要通过传统的 OSPF 骨干区域提供 MPLS L3VPN 服务，在 PE 和 CE 之间使用 OSPF 可以简化这种转换。

为了在 PE-CE 间运行 OSPF，PE 必须支持 OSPF 多实例，每个 OSPF 实例与一个 VPN 实例对应，使用自己的接口、路由表。

下面具体介绍在 PE-CE 间配置 OSPF 需要了解的知识。

(1) PE 和 CE 间的 OSPF 区域配置

PE 与 CE 之间的 OSPF 区域可以是非骨干区域，也可以是骨干区域。

在 OSPF VPN 扩展应用中，MPLS VPN 骨干网被看作是骨干区域 area 0。由于 OSPF 要求骨干区域连续，因此，所有 VPN 站点的 area 0 必须与 MPLS VPN 骨干网相连。

即：如果 VPN 站点存在 OSPF area 0，则 CE 接入的 PE 必须通过 area 0 与这个 VPN 站点的骨干区域相连（可以通过 Virtual-link 实现逻辑上的连通）。

(2) BGP/OSPF 交互

在 PE-CE 间运行 OSPF 后，PE 与 PE 通过 BGP 发布 VPN 路由，PE 通过 OSPF 向 CE 发布 VPN 路由。

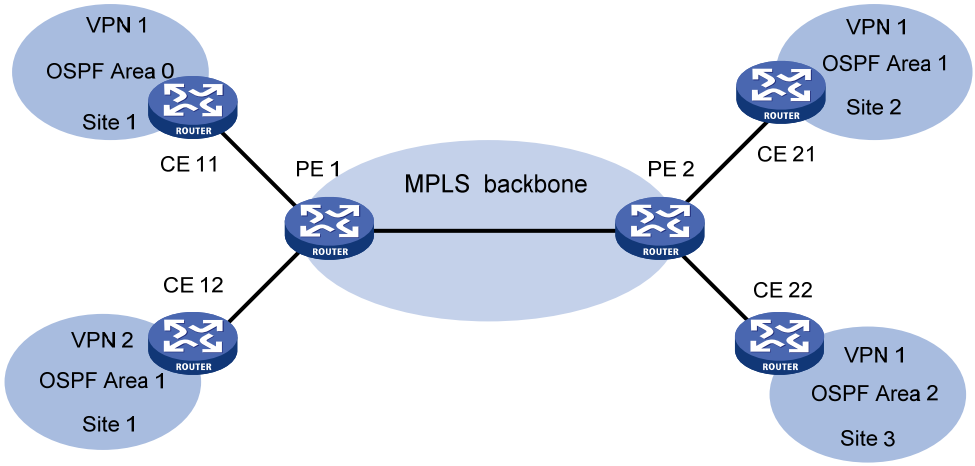
对于普通 OSPF，即使两个不同站点属于同一 VPN，它们也会被看作属于不同的自治系统。这样，在一个站点学到的路由，将被作为外部路由传送给另一站点。这种处理方式导致了比较高的 OSPF 路由协议流量，并带来了一些原本可以避免的网络管理问题。

目前的 OSPF 可以解决上述问题。通过适当配置，运行 OSPF 的不同站点之间彼此看作是直接相连的。这样，PE 路由器交换 OSPF 路由信息时就好像是通过一条专线相连。改善了网络管理并使 OSPF 的应用更为有效。

以 图 1-16为例，PE 1 和PE 2 通过MPLS骨干网相连，CE 11、CE 21 和CE 22 都属于VPN 1。假设图中所有路由器属于同一个域，即，CE 11、CE 21、CE 22 属于同一个OSPF域（OSPF domain）。

VPN 1 路由的发布过程可以描述为：首先在 PE 1 上将 CE 11 的 OSPF 路由引入 BGP；然后通过 BGP 将这些 VPN 路由发布给 PE 2；在 PE 2 上将 BGP 的 VPN 路由引入到 OSPF，再发布给 CE 21 和 CE 22。

图1-16 OSPF 在 VPN 中的使用



如果使用标准的 BGP/OSPF 交互过程，PE 2 将把 BGP VPN 路由通过 Type5 LSAs（即 ASE LSAs）发布给 CE 21 和 CE 22。但 CE 11 与 CE 21、CE 22 是同一个 OSPF 域，它们之间的路由发布应该使用 Type3 LSAs，即区域间路由。

为了解决上述问题，PE 使用一种经过修改的 BGP/OSPF 交互过程（简称为 BGP/OSPF 互操作功能），发布从一个 Site 到另一个 Site 的路由，将这种路由与真正的 AS-External 路由进行区分。这一过程需要 BGP 使用扩展团体属性，携带可以标识 OSPF 属性的信息。

在实现中，要求每个 OSPF 域有一个可配置的域 ID（Domain ID）。一般建议：与每个 VPN 实例相关的网络中的所有 OSPF 实例要么配置一个相同的域 ID，要么都使用缺省的域 ID。这样在收到 BGP 的 VPN 路由时，域 ID 相同的是来自同一 VPN 实例的路由。

(3) 路由环的检测

假设 PE 与 CE 之间通过 OSPF 骨干区域相连，且同一个 VPN 站点（Site）连接到多个不同 PE。这种情况下，当一个 PE 通过 LSA 向 VPN 站点发布从 MPLS/BGP 学的 BGP VPN 路由时，LSA 可能被另一个 PE 接收到，造成路由环。

为了防止产生路由环，对于从 MPLS/BGP 学到的 BGP VPN 路由，无论 PE 与 CE 间是否通过 OSPF 骨干区域相连，PE 在生成 Type3 LSA 时，都会设置标志位 DN。PE 路由器的 OSPF 进程在进行路由计算时，忽略 DN 置位的 Type3 LSAs。

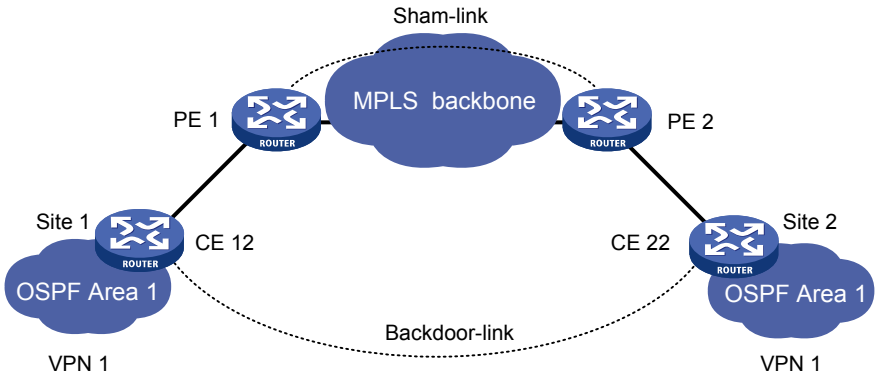
如果 PE 需要向 CE 发布一条来自其它 OSPF 域的路由，则 PE 应表明自己是 ASBR，并将该路由作为 Type5 LSA 发布。

2. OSPF 伪连接

通常情况下，BGP 对等体之间通过 BGP 扩展团体属性在 MPLS VPN 骨干网上承载路由信息。另一端 PE 上运行的 OSPF 可利用这些信息来生成 PE 到 CE 的 Type3 summary LSA，这些路由是区域间路由。

如 图 1-17所示：在OSPF的PE-CE连接中，同一个OSPF区域中有两个Site连接到不同的PE，两个Site属于同一个VPN，它们之间存在一条区域内OSPF链路（backdoor link）。这种情况下，通过PE连接两个Site的路由将作为区域间路由（Inter-Area Route），由于其优先级低于经过backdoor链路的区域内路由（Intra-Area Route），不会被OSPF优选。

图1-17 Sham-link 应用示意图



上述情况导致 VPN 流量总是通过后门路由转发，而不走骨干网。为了避免这一问题，可以在 PE 路由器之间建立 OSPF 伪连接（Sham-link），使经过 MPLS VPN 骨干网的路由也成为 OSPF 区域内路由。

Sham-link 作为区域内的一条点到点链路，包含在 Type1 LSA 中发布。用户可以通过调整度量值在 Sham-link 和 backdoor 之间进行选路。

Sham-link 被看成是两个 VPN 实例之间的链路，每个 VPN 实例中必须有一个 Sham-link 的端点地址，它是 PE 路由器上 VPN 地址空间中的一个有 32 位掩码的 Loopback 接口地址。同一个 OSPF 进程的 Sham-link 可以共用端点地址，但不同 OSPF 进程不能拥有两条端点地址完全相同的 Sham-link。

Sham-link 的端点地址被 BGP 作为 VPN-IPv4 地址发布。如果路由经过了 Sham-link，它就不能再以 VPN-IPv4 路由的形式被引入到 BGP。

Sham-link 可以在任何区域配置。Sham-link 需要手工配置，并且，本端 VPN 实例中必须有到 Sham-link 目的地址的路由。

3. Multi-VPN-Instance CE

OSPF 多实例通常运行在 PE 路由器上，在局域网内部运行 OSPF 多实例的路由器称为 Multi-VPN-Instance CE，即多实例 CE。与 PE 上的 OSPF 多实例相比，Multi-VPN-Instance CE 不需要支持 BGP/OSPF 互操作功能。

Multi-VPN-Instance CE 主要用于以较低的成本解决局域网的安全问题。

单纯使用传统路由器很难实现局域网中不同业务的完全隔离。目前，在一台路由器上可以运行多个 OSPF 进程，这些 OSPF 进程可以是公网进程，也可以是某个 VPN 实例的进程。因此，在一台路由器上可以运行多个 OSPF 进程，将不同的进程绑定到不同的 VPN 实例。

在实际应用中，通过为不同的业务建立各自的 OSPF 实例，可以实现不同业务的隔离，保证各自的安全性。

1.1.12 BGP 的 AS 号替换

在 MPLS L3VPN 中，如果 PE 和 CE 之间运行 EBGP，由于 BGP 使用 AS 号检测路由环路，为保证路由信息的正确发送，需要为物理位置不同的站点分配不同的 AS 号。

如果物理分散的 CE 复用相同的 AS 号，就应该在 PE 上配置 BGP 的 AS 号替换功能。此功能是 BGP 的出口策略，在发布路由时有效。

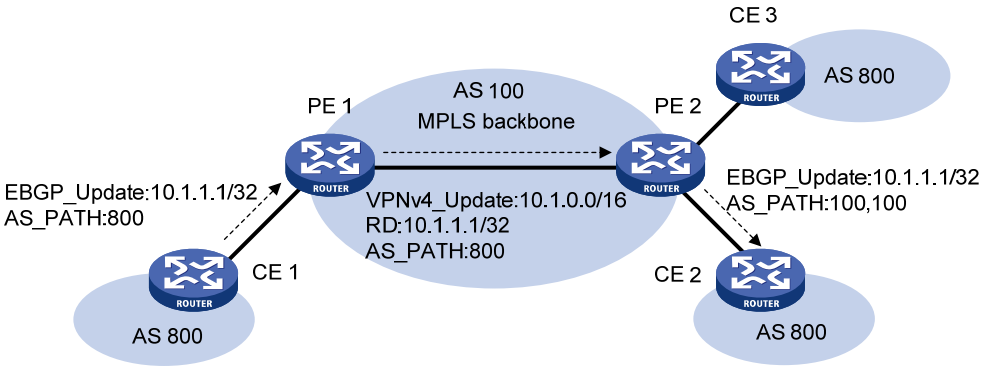
使能了 BGP 的 AS 号替换功能后，当 PE 向指定对等体中的 CE 发布路由时，如果路由的 AS_PATH 中有与 CE 相同的 AS 号，将被替换成 PE 的 AS 号后再发布。



说明

使能 BGP 的 AS 号替换功能后，PE 向对等体组中所有已建立连接的 CE 重新发送所有路由，并对发送路由中的 AS_PATH 属性按上述规则替换。

图1-18 BGP AS 号替换应用示意图



在 图 1-18中，CE 1 和CE 2 都使用AS号 800，在PE 2 上使能针对CE 2 的AS号替换功能。当CE 1 发来的Update信息从PE 2 发布给CE 2 时，PE 2 发现AS_PATH中存在与CE 2 相同的AS号 800，就把它替换为自己的AS号 100，这样，CE 2 就可以正确接收CE 1 的路由信息。

对于 PE 使用不同接口连接多个 CE 的情况，如上图中的 CE 2 和 CE 3，也可以使用 BGP 的 AS 号替换功能。

说明

对于 CE 连接到多个 PE 的情况（Multi-homed CE），单独使用 BGP 的 AS 号替换功能将可能导致路由环路，需要配合 SOO（Site-of-Origin）特性一起使用。

1.2 MPLS L3VPN 配置任务简介

表1-1 MPLS L3VPN 配置任务简介

配置任务	说明	详细配置
配置 VPN 实例	必选	1.3
配置基本 MPLS L3VPN	必选	1.4
配置跨域 VPN	可选 根据组网的需要选择是否配置	1.5
配置嵌套 VPN	可选 根据组网的需要选择是否配置	1.6
配置多角色主机	可选 根据组网的需要选择是否配置	1.7
配置 HoVPN	可选 根据组网的需要选择是否配置	1.8
配置 OSPF 伪连接	可选 根据组网的需要选择是否配置	1.9
配置 Multi-VPN-Instance CE	可选 根据组网的需要选择是否配置	1.10
配置 Egress PE 上私网路由标签操作方式	可选 根据组网的需要选择是否配置	1.11

配置任务	说明	详细配置
配置 BGP 的 AS 号替换	可选 根据组网的需要选择是否配置	1.12

1.3 配置 VPN 实例

VPN 实例用于将 VPN 私网路由与公网路由隔离，在所有 MPLS L3VPN 组网方案中，都需要配置 VPN 实例。

VPN 实例不仅可以将 VPN 私网路由与公网路由隔离，不同 VPN 实例的路由之间也是相互隔离的，这一特点使得 VPN 实例的使用不限于 MPLS L3VPN。

配置 VPN 实例的操作是在 PE 设备上。

1.3.1 创建 VPN 实例

VPN 实例在实现中与 Site 关联。VPN 实例不是直接对应于 VPN，一个 VPN 实例综合了和它所对应 Site 的 VPN 成员关系和路由规则。

一个 VPN 实例只有配置了 RD 后才生效。在配置 RD 之前，除了描述信息外，不能配置 VPN 实例的其他任何参数。

描述信息用于描述 VPN 实例，可以用来记录 VPN 实例与某个 VPN 的关系等信息。

表1-2 创建 VPN 实例

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	必选
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选
配置 VPN 实例的描述信息	**description** *text*	可选

1.3.2 配置 VPN 实例与接口关联

VPN 实例配置完成后，还需要与连接 CE 的接口进行关联。支持 MPLS LDP 能力的接口类型都可以与 VPN 实例相关联。具体接口类型请参见“MPLS 分册”中的“MPLS 基本配置”。

表1-3 配置 VPN 实例与接口关联

操作	命令	说明
进入系统视图	**system-view**	-
进入要关联接口的接口视图	**interface** *interface-type interface-number*	-
将当前接口与 VPN 实例关联	**ip binding vpn-instance** *vpn-instance-name*	必选 缺省情况下，接口不关联任何 VPN 实例

说明

执行 **ip binding vpn-instance** 命令将删除接口上已经配置的 IP 地址，因此需要重新配置接口的 IP 地址。

1.3.3 配置 VPN 实例的路由相关属性

VPN 路由的发布控制过程如下：

- 当从 CE 学习到的一条 VPN 路由引入 BGP 时，BGP 为它关联一个 VPN Target 扩展团体属性列表，通常这个列表是与 CE 相关联的 VPN 实例的输出路由属性列表。
- VPN 实例根据 VPN Target 中 **import-extcommunity** 决定可被接受并引入此 VPN 实例的路由。
- VPN 实例根据 VPN Target 中的 **export-extcommunity** 对向外发布的路由进行 VPN Target 属性的修改。

表1-4 配置 VPN 实例的路由相关属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	必选
将当前 VPN 实例与一个或多个 VPN Target 相关联	**vpn-target** *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]	必选
配置 VPN 实例支持的最大路由数	**routing-table limit** *number* { *warn-threshold* | **simply-alert** }	可选
对当前 VPN 实例应用入方向路由策略	**import route-policy** *route-policy*	可选 缺省情况下，允许所有 VPN Target 属性匹配的路由通过
对当前 VPN 实例应用出方向路由策略	**export route-policy** *route-policy*	可选 缺省情况下，允许所有 VPN Target 属性匹配的路由通过

说明

- 命令 **vpn-target** 最多可以配置 8 个 VPN Target; 一个 VPN 实例最多可以配置 20 个 VPN Target。
- 可以配置一个 VPN 实例可以支持的最大路由数，以防止 PE 路由器的入接口有过多的路由。一个 PE 路由器最多可以支持的路由条数与具体产品有关。
- 为 VPN 实例配置路由策略之前必须已经创建了路由策略，否则将采用缺省路由策略。

1.3.4 配置 VPN 实例的隧道策略

1. 配置隧道策略

表1-5 配置隧道策略

操作	命令	说明
进入系统视图	**system-view**	-
创建隧道策略，并进入隧道策略视图	**tunnel-policy** *tunnel-policy-name*	必选
配置隧道的优先级顺序和负载分担条数	**tunnel select-seq** { **cr-lsp** | **gre** | **lsp** } * **load-balance-number** *number*	必选 缺省情况下，选择的策略是 LSP 隧道，负载分担条数为 1

说明

- 配置的隧道类型不止一种，且负载分担的条数多于单个类型包含的隧道个数时，会使用不同类型的隧道一起进行负载分担。
- 在配置隧道优先级顺序时，隧道类型越靠近关键字 **select-seq** 的，其优先级越高。例如：如果配置了命令 **tunnel select-seq lsp gre load-balance-number** 1，在没有 LSP 的情况下，VPN 将使用 GRE 隧道，当 LSP 建立后，VPN 将选用 LSP。

2. 配置 VPN 实例的隧道策略

表1-6 配置 VPN 实例的隧道策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	必选
配置 VPN 实例的隧道策略	**tnl-policy** *tunnel-policy-name*	必选 缺省情况下，选择的策略是 LSP 隧道，负载分担条数为 1

说明

为 VPN 实例配置隧道策略之前必须先创建隧道策略，否则将采用缺省策略。

1.4 配置基本 MPLS L3VPN

本节介绍比较简单的 MPLS L3VPN 组网配置：只包括一个运营商，MPLS 骨干网不跨域，PE、P、CE 设备都不兼具其它功能（没有一台设备既是 PE，又是 CE）。

如果需要部署一些特殊的 MPLS L3VPN 组网方案，例如 HoVPN、多角色主机、跨域 VPN 等，则还需要另外进行相关的配置。请参考本章中相应小节的介绍。

配置 MPLS L3VPN 的关键在于管理 VPN 私网路由在 MPLS 骨干网上的发布，包括 PE-CE 间的路由发布管理以及 PE-PE 间的路由发布管理。

PE-CE 间的路由交换可以采用静态路由、RIP、OSPF、IS-IS，也可以采用 EBGP。根据实际组网情况选择一种进行配置即可。

1.4.1 配置准备

在配置基本 MPLS L3VPN 之前，需完成以下任务：

- 对 MPLS 骨干网（PE、P）配置 IGP，实现骨干网的 IP 连通性
- 对 MPLS 骨干网（PE、P）配置 MPLS 基本能力
- 对 MPLS 骨干网（PE、P）配置 MPLS LDP，建立 LDP LSP
- 在 CE 上配置接入 PE 的接口的 IP 地址

1.4.2 配置 VPN 实例

表1-7 配置 VPN 实例

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	必选 缺省情况下，无 VPN 实例
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选
将当前 VPN 实例与一个或多个 VPN Target 相关联	**vpn-target** *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]	必选
退回到系统视图	**quit**	-
进入要关联接口的接口视图	**interface** *interface-type interface-number*	-
将当前接口与 VPN 实例关联	**ip binding vpn-instance** *vpn-instance-name*	必选 缺省情况下，无接口关联任何 VPN 实例

1.4.3 配置 PE-CE 间路由交换

配置 PE-CE 间路由交换可以使用静态路由、RIP、OSPF、IS-IS 和 EBGP 路由协议。具体采用哪种协议请根据配置的实际需要。

1. 配置 PE-CE 间使用静态路由

表1-8 配置 PE-CE 间使用静态路由

操作	命令	说明
进入系统视图	**system-view**	-
为指定 VPN 实例配置静态路由	**ip route-static** *dest-address* { *mask* | *mask-length* } { *gateway-address* | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	必选 该配置在 PE 上进行，CE 上的配置方法与普通静态路由相同
	ip route-static vpn-instance *s-vpn-instance-name*&<1-5> *dest-address* { *mask* | *mask-length* } { *gateway-address* [**public**] | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

说明

有关静态路由的配置请参见“IP 路由分册”中的“静态路由配置”。

2. 配置 PE-CE 间使用 RIP

一个 RIP 进程只能属于一个 VPN 实例。如果在启动 RIP 进程时不绑定到 VPN 实例，则该进程属于公网进程。

表1-9 配置 PE-CE 间使用 RIP

操作	命令	说明
进入系统视图	**system-view**	-
创建 PE-CE 间的 RIP 实例，并进入 RIP 视图	**rip** [*process-id*] **vpn-instance** *vpn-instance-name*	必选 该配置在 PE 上进行，CE 上配置普通 RIP 即可

说明

- 配置 RIP 实例后，需要启动 RIP，具体配置与普通的 RIP 相同。
- 有关 RIP 的介绍和详细配置，请参见“IP 路由分册”中的“RIP 配置”。

3. 配置 PE-CE 间使用 OSPF

VPN 实例绑定的 OSPF 进程中不使用系统视图下配置的公网 Router ID，因此用户需要在启动进程时手工配置 Router ID，或者所要绑定的 VPN 实例中至少有一个接口配置了 IP 地址。

一个 OSPF 进程只能属于一个 VPN 实例。如果在启动 OSPF 进程时不绑定到 VPN 实例，则该进程属于公网进程。

表1-10 配置 PE-CE 间使用 OSPF

操作	命令	说明
进入系统视图	**system-view**	-
创建PE-CE间的OSPF实例，并进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *	必选 该配置在 PE 上进行，CE 上配置普通 OSPF 即可
配置 OSPF 域标识符	**domain-id** *domain-id* [**secondary**]	可选 缺省情况下，OSPF 域标识符为 0 该配置在 PE 上进行，CE 上配置普通 OSPF 即可
配置 OSPF 扩展团体属性的类型编码	**ext-community-type** { **domain-id** *type-code1* | **router-id** *type-code2* | **route-type** *type-code3* }	可选 缺省情况下，OSPF 扩展团体属性 Domain ID 的类型编码是 0x0005，Router ID 的类型编码是 0x0107，Route Type 的类型编码是 0x0306 该配置在 PE 上进行

说明

删除 VPN 实例后，相关的所有 OSPF 进程也将全部被删除。

每个 OSPF 进程只能配置一个域 ID，不同进程的域 ID 相互没有影响。不同 VPN 的 PE 的 OSPF 进程域 ID 配置没有限制。但同一 VPN 的所有 OSPF 进程应配置相同的域 ID，以保证路由发布的正确性。

OSPF 进程的域 ID 包含在此进程生成的路由中，在将 OSPF 路由引入 BGP 中时， 域 ID 被附加到 BGP VPN 路由上，作为 BGP 的扩展团体属性传递。

说明

- 配置 OSPF 实例后，需要启动 OSPF，具体配置与普通的 OSPF 相同。
- 有关 OSPF 的介绍和详细配置，请参见“IP 路由分册”中的“OSPF 配置”。

4. 配置 PE-CE 间使用 IS-IS

一个 IS-IS 进程只能属于一个 VPN 实例。如果在启动 IS-IS 进程时不绑定到 VPN 实例，则该进程属于公网进程。

表1-11 配置 PE-CE 间使用 IS-IS

操作	命令	说明
进入系统视图	**system-view**	-
创建 PE-CE 间的 IS-IS 实例，并进入 IS-IS 视图	**isis** [*process-id*] **vpn-instance** *vpn-instance-name*	必选 该配置在 PE 上进行，CE 上配置普通 IS-IS 即可

说明

- 配置 IS-IS 实例后，需要启动 IS-IS，具体配置与普通的 IS-IS 相同。
- 有关 IS-IS 的介绍和详细配置，请参见“IP 路由分册”中的“IS-IS 配置”。

5. 配置 PE-CE 间使用 EBGP

(1) PE 上的配置

表1-12 PE 上的配置

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-VPN 实例视图	**ipv4-family vpn-instance** *vpn-instance-name*	必选
将 CE 配置为 VPN 私网对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
引入本端 CE 路由	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 PE 需要将到本端 CE 的路由引入 VPN 路由表中，以发布给对端 PE
配置对发布的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	可选 缺省情况下，不对发布的路由信息进行过滤
配置对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	可选 缺省情况下，不对接收的路由信息进行过滤
配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	可选 用于 Hub&Spoke 组网方案

说明

通常情况下，BGP 通过 AS 号检测路由环路。但在 Hub&Spoke 组网方式下，如果在 PE 和 CE 之间运行 EBGP，当 PE 将路由信息通告给 CE 时带上本自治系统的 AS 号，再从 CE 接收路由更新时，路由更新消息中会带有本自治系统的 AS 号，这样 PE 就不能接收这条路由更新信息，此时需要配置允许路由环路。

(2) CE 上的配置

表1-13 CE 上的配置

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
将 PE 配置为对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选

操作	命令	说明
配置路由引入	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	可选 CE 需要将自己所能到达的 VPN 网段地址发布给接入的 PE，通过 PE 发布给对端 CE

说明

- 配置 BGP VPN 实例后，VPN 实例中的 BGP 路由交换与普通的 BGP 相同。
- BGP 实例视图下配置任务与 BGP 视图下的相同，有关介绍和详细配置，请参见“IP 路由分册”中的“BGP 配置”和“BGP 命令”。
- BGP 的配置中有对等体和对等体组的配置，有关介绍和详细说明请参见“IP 路由分册”中的“BGP 配置”和“BGP 命令”。本章中不将对等体和对等体组加以区分。

1.4.4 配置 PE-PE 间的路由交换

表1-14 配置 PE-PE 间的路由交换

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
将对端 PE 配置为对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
指定路由更新报文的源接口	**peer** { *group-name* | *ip-address* } **connect-interface** *interface-type interface-number*	必选 缺省情况下，BGP 使用最佳路由更新报文的源接口
进入 BGP-VPNv4 子地址族视图	**ipv4-family vpnv4**	必选
使能对等体交换 BGP-VPNv4 路由信息	**peer** { *group-name* | *ip-address* } **enable**	必选 缺省情况下，BGP 对等体之间只能交换 IPv4 路由信息

1.4.5 配置 BGP-VPNv4 子地址族下的路由特性

BGP-VPNv4 子地址族下路由的很多特性都与 BGP IPv4 单播路由的特性相同。这些特性根据组网的需求来决定是否选择。

1. 配置各种地址族通用路由特性

对于 VPN 的应用，BGP 的地址族包括：BGP-VPNv4 子地址族和 BGP-L2VPN 地址族。对于下面的命令在各地址族中的作用都是相同的，相应地址族中的 BGP 路由起作用。

表1-15 配置各种地址族通用路由特性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
将对端 PE 配置为对等体	**peer** *ip-address* **as-number** *as-number*	必选

操作	命令	说明
指定建立 TCP 连接的接口	**peer** *ip-address* **connect-interface** *interface-type interface-number*	必选
进入地址族视图	**ipv4-family vpnv4**	必选 根据不同的组网情况选择地址族
	l2vpn-family	
配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	可选
在指定的地址族下使能对等体，并使能交换对应地址族的 BGP 路由信息的能力	**peer** { *group-name* | *ip-address* } **enable**	必选 缺省情况下，BGP 对等体之间只能交换 IPv4 路由信息
将对等体加入已存在的对等体组	**peer** *ip-address* **group** *group-name*	可选
向对等体或对等体组发布路由时，将自身地址作为下一跳	**peer** { *group-name* | *ip-address* } **next-hop-local**	可选 缺省情况下，向 EBGP 对等体发布路由时会将下一跳改为自己的地址。在跨域 VPN OptionC 组网中反射器上对多跳 EBGP 邻居和反射客户需要配置 next-hop-invariable 以保证私网路由下一跳不会被改变。
配置将本机作为路由反射器，并将对等体或对等体组作为路由反射器的客户	**peer** { *group-name* | *ip-address* } **reflect-client**	可选 缺省情况下，没有配置路由反射器及其客户
对接收到的相应地址族下 BGP 路由使能 VPN-Target 过滤功能	**policy vpn-target**	可选 缺省情况下，对接收到的相应地址族的 BGP 路由使能 VPN-Target 过滤功能
允许客户到客户的路由反射	**reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
为路由反射器配置群 ID	**reflector cluster-id** { *cluster-id* | *ip-address* }	可选 缺省情况下，群 ID 是群中某个路由反射器的 Router ID
创建路由反射器的反射策略	**rr-filter** *extended-community-list-number*	可选

说明

有关 BGP-L2VPN 地址族的介绍请参见“MPLS 分册”中的“MPLS L2VPN 命令”。

2. 配置 BGP-VPNv4 子地址族的特有路由特性

表1-16 配置 BGP-VPNv4 子地址族的特有路由特性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-

操作	命令	说明
将对端 PE 配置为对等体	**peer** *ip-address* **as-number** *as-number*	必选
指定建立 TCP 连接的接口	**peer** *ip-address* **connect-interface** *interface-type interface-number*	必选
进入 BGP-VPNv4 子地址族视图	**ipv4-family vpnv4**	-
配置本地优先级的缺省值	**default local-preference** *value*	可选 缺省情况下，本地优先级的值为 100
配置缺省的系统度量	**default med** *med-value*	可选 缺省情况下，缺省的系统度量为 0
配置对发布的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	可选 缺省情况下，不对发布的路由信息进行过滤
配置对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	可选 缺省情况下，不对接收的路由信息进行过滤
配置将团体属性通告给对等体或对等体组	**peer** { *group-name* | *ip-address* } **advertise-community**	可选 缺省情况下，不将团体属性传给任何对等体或对等体组
为指定对等体或对等体组配置基于 AS 路径列表的路由过滤	**peer** { *group-name* | *ip-address* } **as-path-acl** *aspath-filter-number* { **import** | **export** }	可选 缺省情况下，对等体及对等体组没有应用 AS 过滤列表
配置向对等体或对等体组发布与指定 VPN 实例相关的所有缺省路由	**peer** { *group-name* | *ip-address* } **default-route-advertise vpn-instance** *vpn-instance-name*	可选 缺省情况下，不向对等体或对等体组发布缺省路由
配置为指定对等体或对等体组应用过滤策略	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* { **export** | **import** }	可选 缺省情况下，没有为对等体或对等体组应用过滤策略
为指定对等体或对等体组应用基于 IP 前缀列表的路由过滤策略	**peer** { *group-name* | *ip-address* } **ip-prefix** *prefix-name* { **export** | **import** }	可选 缺省情况下，对等体或对等体组上不应用基于 IP 前缀列表的路由过滤策略
配置向 EBGP 对等体发布路由时不改变下一跳	**peer** { *group-name* | *ip-address* } **next-hop-invariable**	可选 缺省情况下，向 EBGP 对等体发布路由时会将下一跳改为自己的地址
为从对等体/对等体组接收的路由分配首选值	**peer** { *group-name* | *ip-address* } **preferred-value** *value*	可选 缺省情况下，从对等体/对等体组接收的路由的首选值为 0
配置发送 BGP 更新报文时不携带私有自治系统号	**peer** { *group-name* | *ip-address* } **public-as-only**	可选 缺省情况下，发送 BGP 更新报文时，携带私有自治系统号
为指定对等体或对等体组配置路由策略	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* { **export** | **import** }	可选 缺省情况下，对等体及对等体组没有应用路由策略

说明

有关 BGP 的路由特性请参见“IP 路由分册”中的“BGP 配置”。

1.5 配置跨域 VPN

如果承载 VPN 路由的 MPLS 骨干网跨越多个 AS，就需要配置跨域 VPN。

跨域VPN有三种解决方案，详细描述请参考“1.1.6 跨域VPN”。

1.5.1 配置准备

在配置跨域 VPN 之前，需完成以下任务：

- 为各 AS 的 MPLS 骨干网分别配置 IGP，实现同一 AS 内骨干网的 IP 连通性
- 为各 AS 的 MPLS 骨干网分别配置 MPLS 基本能力
- 为各 AS 的 MPLS 骨干网分别配置 MPLS LDP，建立 LDP LSP
- 为各 AS 分别进行基本 MPLS L3VPN 配置

在对各 AS 分别进行基本 MPLS L3VPN 配置时，根据所选择的跨域 VPN 方式，PE 和 ASBR-PE 上还需要进行专门的配置。

说明

请根据实际应用的跨域 VPN 方式，选择下述配置任务中的对应项。

1.5.2 配置跨域 VPN-OptionA

跨域 VPN-OptionA 的实现比较简单，当 PE 上的 VPN 数量及 VPN 路由数量都比较少时可以采用这种方案。跨域 VPN-OptionA 的配置可以描述为：对各 AS 分别进行基本 MPLS L3VPN 配置，对于 ASBR-PE，将对端 ASBR-PE 看作自己的 CE 配置即可。即：跨域 VPN-OptionA 方式需要在 PE 和 ASBR-PE 上分别配置 VPN 实例，前者用于接入 CE，后者用于接入对端 ASBR-PE。

请参考“1.4 配置基本MPLS L3VPN”。

说明

在跨域 VPN-OptionA 方式中，对于同一个 VPN，同一 AS 内的 ASBR-PE 与 PE 的 VPN 实例的 VPN Target 应能匹配；不同 AS 的 PE 之间的 VPN 实例的 VPN Target 则不需要匹配。

1.5.3 配置跨域 VPN-OptionB

以下配置是在作为 ASBR-PE 的路由器上进行配置的。

表1-17 配置跨域 VPN-OptionB 的 ASBR-PE

操作	命令	说明
进入系统视图	**system-view**	-
进入连接对端 ASBR-PE 接口的接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置接口 IP 地址	**ip address** *ip-address* { *mask* | *mask-length* }	必选
退回系统视图	**quit**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-VPNv4 子地址族视图	**ipv4-family vpnv4**	-
不对 VPNv4 路由进行 VPN Target 过滤	**undo policy vpn-target**	必选 缺省情况下，PE 对收到的 VPNv4 路由进行 VPN Target 过滤 通过过滤的路由会被加入到路由表中，没有通过过滤的路由将被丢弃

在跨域 VPN-OptionB 方式中，ASBR-PE 需要保存所有 VPNv4 路由信息，以通告给对端 ASBR-PE。这种情况下，ASBR-PE 应接收所有的 VPNv4 路由信息，不对它们进行 VPN Target 过滤。

说明

在跨域 VPN-OptionB 方式中，对于同一个 VPN，同一 AS 内的 ASBR-PE 与 PE 的 VPN 实例的 VPN Target 应能匹配；不同 AS 的 PE 之间的 VPN 实例的 VPN Target 也需要匹配。

注意

对于跨域 VPN-OptionB，存在两种配置方式：

- 一种是在 ASBR 不改下一跳的方式，在这种方式下 ASBR 间仍然需要配置 MPLS LDP；
- 另一种在 ASBR 上需要修改下一跳，这种组网下 ASBR 间不需要配置 MPLS LDP。

目前实现只支持在 ASBR 上修改下一跳的方式，因此在实现上，默认情况下对于 MP-EBGP 路由，在发往 MP-IBGP 时是修改下一跳的。而在普通的 BGP 上，对于发往 IBGP 的 EBGP 路由，默认情况下是不改下一跳的，如果需要将下一跳改为本地地址，需要配置 **peer** { *ip-address* | *group-name* } **next-hop-local** 命令。具体说明请参见“IP 路由分册”中的“BGP 配置”。

1.5.4 配置跨域 VPN-OptionC

1. 配置 PE

同一AS内的PE和ASBR-PE之间建立普通的IBGP对等体关系，不同AS的PE之间建立MP-EBGP对等体关系。

同一 AS 内的 PE 和 ASBR-PE 之间应能够交换标签 IPv4 路由。

表1-18 配置 PE

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置本 AS 的 ASBR-PE 为 IBGP 对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选

操作	命令	说明
配置与本 AS 的 ASBR-PE 之间能够交换带标签的 IPv4 路由	**peer** { *group-name* | *ip-address* } **label-route-capability**	必选 缺省情况下，不向 IPv4 对等体/对等体组发送标签路由
将另一 AS 的 PE 配置为 EBGP 对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
进入BGP-VPNv4子地址族视图	**ipv4-family vpnv4**	-
使能对等体交换 BGP-VPNv4 路由信息	**peer** { *group-name* | *ip-address* } **enable**	必选
配置向 EBGP 对等体发送路由时不改变下一跳	**peer** { *group-name* | *ip-address* } **next-hop-invariable**	可选 该步骤用于使用路由反射器 RR（Route Reflector）通告 VPNv4 路由的情况：RR 之间通告 VPNv4 路由时，路由的下一跳不能被改变

2. 配置 ASBR-PE

在跨域 VPN-OptionC 方式中，需要建立一条跨域的 VPN LSP，相关 PE、ASBR 之间发布公网路由时携带 MPLS 标签信息。

ASBR-PE 与本 AS 的 PE 建立普通 IBGP 对等体关系，与对端 ASBR-PE 建立普通 EBGP 对等体关系。都能够交换标签 IPv4 路由。

携带 MPLS 标签的公网路由通过 MP-BGP 发布。根据 RFC 3107（Carrying Label Information in BGP-4）中的描述，一条路由的标签映射信息可以通过发布这条路由的 BGP Update 消息捎带（piggyback）。这种能力使用 BGP 的扩展属性实现，要求 BGP 对等体能够处理标签 IPv4 路由。

表1-19 配置 ASBR-PE

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
配置本 AS 的 PE 为 IBGP 对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
配置与本 AS 的 PE 之间能够交换带标签的 IPv4 路由	**peer** { *group-name* | *ip-address* } **label-route-capability**	必选 缺省情况下，不向 IPv4 对等体/对等体组发送标签路由
向本 AS 的 PE 发布路由时将下一跳改为自己	**peer** { *group-name* | *ip-address* } **next-hop-local**	必选 缺省情况下，BGP 发言者在向 IBGP 对等体/对等体组发布路由时不会将下一跳改为自己的地址
将对端 ASBR-PE 配置为 EBGP 对等体	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选
配置与对端 ASBR-PE 之间能够交换带标签的 IPv4 路由	**peer** { *group-name* | *ip-address* } **label-route-capability**	必选 缺省情况下，不向 IPv4 对等体发送标签路由

操作	命令	说明
对对端 ASBR-PE 发布来的路由应用路由策略	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，对等体及对等体组没有应用路由策略

3. 配置路由策略

配置路由策略应用在 ASBR-PE 上：

- 对于从本 AS 的 PE 接收的路由，在向对端 ASBR-PE 发布时，分配 MPLS 标签；
- 对于向本 AS 的 PE 发布的路由，如果是带标签的 IPv4 路由，为其分配新的 MPLS 标签。

对 IPv4 路由分配 MPLS 标签是通过路由策略控制的，只对满足某些条件的路由分配标签，其它路由还是普通 IPv4 路由。

表1-20 配置路由策略

操作	命令	说明
进入系统视图	**system-view**	-
进入路由策略视图	**route-policy** *policy-name* **permit node** *seq-number*	必选
匹配带标签的 IPv4 路由	**if-match mpls-label**	必选
为 IPv4 路由分配标签	**apply mpls-label**	必选 缺省情况下，IPv4 路由不带标签

说明

有关路由策略的配置请参见“IP 路由分册”中的“路由策略配置”。

1.6 配置嵌套 VPN

如果网络中 VPN 接入数量比较多，管理者想要管理权限层次化（通过 VPN 划分访问权限），且不想让外部知道用户内部 VPN 的部署情况，可以使用嵌套 VPN 的组网方式。在嵌套 VPN 组网方式中，用户通过较低的管理复杂度和组网成本，就可以实现内部 VPN 网络的层次化管理。

1.6.1 配置准备

配置嵌套VPN之前，需完成MPLS L3VPN的基本配置，请参见“1.4 配置基本MPLS L3VPN”。

1.6.2 配置嵌套 VPN

表1-21 配置嵌套 VPN

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-VPN 实例视图	**ipv4-family vpn-instance** *vpn-instance-name*	-

操作	命令	说明
配置 CE 对等体或对等体组	**peer** { *group-name* | *peer-address* } **as-number** *number*	必选
退回 BGP 视图	**quit**	-
进入 BGP-VPNv4 子地址族视图	**ipv4-family vpnv4**	-
使能嵌套 VPN 功能	**nesting-vpn**	必选 缺省情况下，禁止嵌套 VPN 功能
激活嵌套 VPN 对等体或对等体组，并使能其交换 BGP-VPNv4 路由信息的能力	**peer** { *group-name* | *peer-address* } **vpn-instance** *vpn-instance-name* **enable**	必选 缺省情况下，嵌套 VPN 对等体/对等体组之间只能交换 IPv4 路由信息，不能交换 BGP-VPNv4 路由信息
将对等体加入到嵌套 VPN 对等体组中	**peer** *peer-address* **vpn-instance** *vpn-instance-name* **group** *group-name*	可选 缺省情况下，对等体没有加入任何嵌套 VPN 对等体组
配置从嵌套 VPN 对等体或对等体组接收的 VPNv4 路由应用的路由策略	**peer** { *group-name* | *peer-address* } **vpn-instance** *vpn-instance-name* **route-policy** *route-policy-name* **import**	可选 缺省情况下，对于从嵌套 VPN 对等体或对等体组接收的 VPNv4 路由不应用路由策略

说明

- 同一用户 VPN 的不同子 VPN 之间地址空间不能重叠。
- 建议嵌套 VPN 对等体的地址不要与公网中对等体的地址重叠。
- 配置嵌套 VPN 对等体或对等体组之前，需要先配置 BGP-VPN 实例视图下的 CE 对等体或对等体组。
- 目前嵌套 VPN 不支持多跳 EBGP 组网方式，因此运营商 PE 和 CE 之间必须使用直连接口地址建立邻居关系。
- 在某些产品上，如果子 VPN 的 CE 设备直接连接运营商 PE 设备，则需要在该 PE 设备上配置策略路由，使得主 VPN 和子 VPN 可以互访。

1.7 配置多角色主机

多角色主机特性是通过在 PE 上配置策略路由，使来自同一 CE 的报文可以访问不同的 VPN。

多角色主机特性本身的配置都在多角色主机所属 CE 所接入的 PE 上进行。

说明

有关策略路由的配置和命令请参见“IP 业务分册”中的“IP 单播策略路由配置”和“IP 单播策略路由命令”。

1.7.1 配置准备

在配置多角色主机之前，需完成以下任务：

- 为不同的 VPN 分别创建 VPN 实例

- 配置基本 MPLS L3VPN

1.7.2 配置策略路由及应用

表1-22 配置策略路由及应用

操作	命令	说明
进入系统视图	**system-view**	-
创建策略节点，并进入策略路由视图	**policy-based-route** *policy-name* { **deny** | **permit** } **node** *node-number*	-
在策略节点中设置指定 VPN 实例进行转发	**apply access-vpn vpn-instance** *vpn-instance-name*&<1-6>	必选
进入系统视图	**system-view**	-
进入接入 CE 的接口视图	**interface** *interface-type interface-number*	-
在接口上应用策略路由	**ip policy-based-route** *policy-name*	必选

1.7.3 配置静态路由

配置PE-CE间使用静态路由，具体配置参见 1.4.3 1. 配置PE-CE间使用静态路由

在 PE 上配置私网静态路由，指定另外一个私网或者公网出口作为该静态路由的出接口，使多角色主机访问某个 VPN 的报文能够使用非该 VPN 的路由表返回。

1.8 配置 HoVPN

对于层次化比较明显的 VPN 网络，可以采用 HoVPN 方案，以降低对 PE 设备的性能要求。

1.8.1 配置准备

在配置 HoVPN 之前，需完成以下任务：

配置基本 MPLS L3VPN

1.8.2 配置 HoVPN

表1-23 配置 HoVPN

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-VPNv4 子地址族视图	**ipv4-family vpnv4**	必选
使能对等体交换 BGP-VPNv4 路由信息	**peer** { *group-name* | *ip-address* } **enable**	必选
配置 BGP 对等体或对等体组为 UPE	**peer** { *group-name* | *ip-address* } **upe**	必选
向 UPE 发送指定 VPN 实例的缺省路由	**peer** { *group-name* | *ip-address* } **default-route-advertise vpn-instance** *vpn-instance-name*	二者必选其一 缺省情况下，不向对等体发布缺省路由
向 UPE 发送通过路由策略的路由	**peer** { *group-name* | *ip-address* } **upe route-policy** *route-policy-name* **export**	

执行 **peer default-route-advertise vpn-instance** 命令后，不论本地路由表中是否存在缺省路由，SPE 都会向 UPE 发布一条下一跳地址为本地地址的缺省路由。

说明

- 向 BGP 对等体或对等体组发布 VPN 实例缺省路由的前提是：此 BGP 对等体或对等体组必须是 UPE。
- 建议不要同时配置 **peer default-route-advertise vpn-instance** 命令和 **peer upe route-policy** 命令。
- 建议 SPE 不直接同 CE 连接，如必须连接，则 SPE 上 VPN 实例和 UPE 上 VPN 实例必须配置不同的 RD。

1.9 配置 OSPF 伪连接

在 VPN PE 间配置伪连接 Sham-link 后，Sham-link 将被视为 OSPF 区域内路由。这一特性用于使经过 MPLS VPN 骨干网的路由成为 OSPF 区域内路由，避免 VPN 流量经后门路由转发。

Sham-link 的源地址和目的地址应使用 32 位掩码的 Loopback 接口地址，且该 loopback 接口需要绑定到 VPN 实例中，并通过 BGP 发布。

1.9.1 配置准备

在配置 OSPF 伪连接之前，需完成以下任务：

- 配置基本 MPLS L3VPN（PE-CE 间使用 OSPF）
- 配置用户 CE 所在局域网的 OSPF

1.9.2 配置 Loopback 接口

表1-24 配置 Loopback 接口

操作	命令	说明
进入系统视图	**system-view**	-
创建 Loopback 接口并进入 Loopback 接口视图	**interface loopback** *interface-number*	必选
将 Loopback 接口绑定到 VPN 实例	**ip binding vpn-instance** *vpn-instance-name*	必选 缺省情况下，接口不关联到任何 VPN 实例，属于公网接口
配置 Loopback 接口的地址	**ip address** *ip-address* { *mask* | *mask-length* }	必选

1.9.3 发布 Loopback 接口的路由

表1-25 发布 Loopback 接口的路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选

操作	命令	说明
进入 BGP-VPN 实例视图	**ipv4-family vpn-instance** *vpn-instance-name*	必选
引入直连路由（将 Loopback 主机路由引入 BGP）	**import-route direct**	必选

1.9.4 配置伪连接

表1-26 创建伪连接

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *	-
配置 route tag	**route-tag** *tag-value*	必选
进入 OSPF 区域视图	**area** *area-id*	必选
配置伪连接	**sham-link** *source-ip-address destination-ip-address* [**cost** *cost* | **dead** *dead-interval* | **hello** *hello-interval* | **retransmit** *retrans-interval* | **trans-delay** *delay* | **simple** [**cipher** | **plain**] *password* | { **md5** | **hmac-md5** } *key-id* [**cipher** | **plain**] *password*] *	必选 缺省情况下，未配置伪连接

说明

- 启动 OSPF 而不手工配置路由器 ID 时，系统会自动选取出一个路由器 ID。选举规则请参见“IP 路由分册”中的“OSPF 配置”。需要注意的是，由于选举规则相同会得到相同的路由器 ID。建议用户启动 OSPF 进程时手工配置路由器 ID。
- 配置 OSPF 的 VPN 多实例而不配置 route tag 时，系统会根据配置的 BGP 自治系统号生成一个 tag 值，如果没有配置 BGP 则 tag 为 0。需要注意的是，在 BGP 的自治系统号相同的 PE 或同一台 PE 上，因为计算规则相同会得到相同的 tag 值。建议用户配置 OSPF 的 VPN 多实例时手工配置 tag 值。

1.10 配置 Multi-VPN-Instance CE

Multi-VPN-Instance CE 在局域网内使用，通过在 CE 设备上使用 OSPF 多实例实现不同业务的隔离。

1.10.1 配置准备

在配置 Multi-VPN-Instance CE 之前，需完成以下任务：

- 配置 VPN 实例
- 配置相关接口的链路层协议和网络层协议，保证路由可达

1.10.2 配置 Multi-VPN-Instance CE 与各路由协议进行路由交换

Multi-VPN-Instance CE 可以看作一种通过路由隔离实现业务隔离的组网方案，配置上，与传统的 CE 接入方式相比并没有特殊之处，但需要使能多 VPN 实例 CE 功能。

配置该功能后，路由计算时关闭 PE 路由上的路由环路检测功能，防止路由丢失；同时禁止各路由协议互操作功能，节省系统资源。

1. 配置 MCE 与静态路由进行路由交换

CE 可以通过静态路由与 Site 连接。传统 CE 配置的静态路由对全局生效，无法解决多 VPN 间的地址重叠问题。以太网交换机提供的 MCE 功能可以将静态路由与 VPN 实例相绑定，将各 VPN 之间的静态路由进行隔离。

表1-27 配置 MCE 与静态路由进行路由交换

操作	命令	说明
进入系统视图	**system-view**	-
为指定 VPN 实例配置静态路由	**ip route-static** *dest-address* { *mask* | *mask-length* } { *gateway-address* | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	必选 缺省情况下，静态路由的优先级 **preference** 为 60，静态路由 **tag** 值为 0，未配置描述信息
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* | *mask-length* } { *gateway-address* [**public**] | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	
配置静态路由的缺省优先级	**ip route-static default-preference** *default-preference-value*	可选 缺省情况下，静态路由的缺省优先级为 60

2. 配置 MCE 与 RIP 进行路由交换

一个 RIP 进程只能属于一个 VPN 实例。如果在启动 RIP 进程时不绑定到 VPN 实例，则该进程属于公网进程。通过在 CE 和 Site 间配置相同的绑定关系，使不同 VPN 内的私网路由可以通过不同的 RIP 进程在 Site 和 CE 间进行交互，保证了私网路由的隔离和安全。

表1-28 配置 MCE 与 RIP 进行路由交换

操作	命令	说明
进入系统视图	**system-view**	-
进入 RIP 视图	**rip** [*process-id*] **vpn-instance** *vpn-instance-name*	必选
分别引入链接 PE 和 Site 端其它协议的路由	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	必选 缺省情况下，RIP 未引入其它路由
配置引入路由的缺省度量值	**default cost** *value*	可选 缺省情况下，引入路由的缺省度量值为 0



说明

配置 RIP 实例后，需要启动 RIP，具体配置与普通的 RIP 相同，有关 RIP 协议的配置，请参见“IP 路由分册”中的“RIP 配置”。

3. 配置 MCE 与 OSPF 进行路由交换

一个 OSPF 进程只能属于一个 VPN 实例。如果在启动 OSPF 进程时不绑定到 VPN 实例，则该进程属于公网进程。

在 MCE 与 PE 之间，除了配置与 VPN 实例绑定的 OSPF 进程和 Router ID 外，还需要手工将 MCE 中维护的 Site 内的 VPN 路由引入到 PE 的路由表中。

表1-29 配置 MCE 与 OSPF 进行路由交换

操作	命令	说明
进入系统视图	**system-view**	-
进入 OSPF 视图	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *	必选
开启 OSPF 的 VPN 多实例功能	**vpn-instance-capability simple**	必选 缺省情况下，没有开启 OSPF 的 VPN 多实例功能
配置 OSPF 域标识符	**domain-id** *domain-id* [**secondary**]	可选 缺省情况下，OSPF 域标识符为 0 该配置在 MCE 上进行，Site 上配置普通 OSPF 即可
分别引入链接 PE 和 Site 端其它协议的路由	**import-route** *protocol* [*process-id* | **allow-ibgp**] [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] *	必选 缺省情况下，没有引入其他协议的路由信息
配置对引入的路由进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*]]	可选 缺省情况下，没有对引入的路由信息进行过滤
配置引入外部路由时的参数缺省值（开销、路由数量、标记、类型）（该配置仅连接 PE 端使用）	**default** { **cost** *cost* | **limit** *limit* | **tag** *tag* | **type** *type* } *	可选 缺省情况下，OSPF 引入外部路由的度量值为 1，单位时间内引入外部路由数量的上限为 1000，外部路由标记值为 1，引入的外部路由类型为 Type2

说明

- VPN 实例绑定的 OSPF 进程中不使用系统视图下配置的公网 Router ID，因此用户需要在启动进程时手工配置 Router ID。同一 VPN 的所有 OSPF 进程应配置相同的域 ID，以保证路由发布的正确性。
- 一个 OSPF 进程只能属于一个 VPN 实例，但一个 VPN 实例可以使用多个 OSPF 进程为其传播私网路由。同一 VPN 实例内的 OSPF 进程应配置有相同的域 ID，以保证路由发布的正确性。
- 配置 OSPF 实例后，需要启动 OSPF，具体配置与普通的 OSPF 相同。有关 OSPF 协议的配置，请参见“IP 路由分册”中的“OSPF 配置”。

4. 配置 MCE 与 IS-IS 进行路由交换

一个 IS-IS 进程只能属于一个 VPN 实例。如果在启动 IS-IS 进程时不绑定到 VPN 实例，则该进程属于公网进程。

表1-30 配置 MCE 与 IS-IS 进行路由交换

操作	命令	说明
进入系统视图	**system-view**	-
进入 IS-IS 视图	**isis** [*process-id*] **vpn-instance** *vpn-instance-name*	必选
分别引入链接 PE 和 Site 端其它协议的路由	**import-route** { **isis** [*process-id*] | **ospf** [*process-id*] | **rip** [*process-id*] | **bgp** [**allow-ibgp**] | **direct** | **static** } [**cost** *cost* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	可选 缺省情况下，IS-IS 不引入其它协议的路由信息 如果 **import-route** 命令中不指定引入的级别，则默认为引入路由到 Level-2 路由表中
对引入的路由进行过滤（该配置仅连接 PE 端使用）	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export** [**isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **bgp** | **direct** | **static**]	可选 缺省情况下，IS-IS 不对引入的路由信息进行过滤



说明

配置 IS-IS 实例后，需要启动 IS-IS，具体配置与普通的 IS-IS 相同。有关 IS-IS 协议的配置，请参见“IP 路由分册”中的“IS-IS 配置”。

5. 配置 MCE 与 EBGP 进行路由交换

在 MCE 与 PE 间使用 EBGP 交换路由信息时，需要在两端的 BGP-VPN 实例中将对端设置为对等体，并在 MCE 上引入 Site 内的 VPN 路由发布到 PE 上。

在 MCE 与 Site 间使用 EBGP 交换路由信息时，需要在 CE 上为每个 VPN 实例配置 BGP 对等体，并引入相应 VPN 内的 IGP 路由信息。由于各个 Site 间正常情况下是处在不同的 AS 内，因此使用 EBGP 进行路由的传播。

(1) 配置 EBGP 引入各 Site 内的 IGP 路由

为正确的将私网路由发布到 PE，需要 MCE 首先将与其直连 Site 内的 IGP 路由引入到自己的 BGP 路由表中。

(2) 为每个 VPN 实例配置对等体组

为准确地与各 Site 交换路由信息，可以在 BGP 的 IPv4 地址族视图下，为每个 VPN 实例配置对等体组，并指定对等体组的 AS 编号。

(3) 通过 Filter-policy 对路由进行过滤

为正确地将路由信息传播到 Site 和 PE 设备，在指定对等体后，还需要使用 Filter-policy 对接收/发布的路由进行过滤。

表1-31 配置 MCE 与 EBGP 进行路由交换

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-

操作	命令	说明
进入 BGP-VPN 实例视图	**ipv4-family vpn-instance** *vpn-instance-name*	必选
配置对等体组中加入对等体	**peer** { *group-name* | *ip-address* } [**as-number** *as-number*]	二者必选其一
配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	
分别引入链接 PE 和 Site 端其它协议的路由	**import-route** *protocol* [*process-id* | **all-processes**] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 缺省情况下，BGP 不引入且不通告其它协议的路由
配置对发布的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	可选 缺省情况下，不对发布的路由信息进行过滤
配置对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	可选 缺省情况下，不对接收的路由信息进行过滤

说明

- 通常情况下，BGP 通过 AS 号检测路由环路。如果在 MCE 和 Site 之间运行 EBGP，当 MCE 将路由信息通告给 Site 时带上本自治系统的 AS 号，再从 Site 接收路由更新时，路由更新消息中会带有本自治系统的 AS 号，这样 MCE 就不能接收这条路由更新信息，此时需要配置允许路由环路。
- 对于标准的 BGP/OSPF 互相引用功能，当在 MCE 设备上配置引入 BGP 路由到 OSPF 中时，该路由的原 OSPF 属性将无法恢复，造成该路由与从其他域引入的路由无法区分。为了区分原属于不同 OSPF 域的路由，需要在远端 PE 将 OSPF 路由引入到 BGP 时携带标识域的属性，即 OSPF 的域 ID（Domain ID）。OSPF 进程的域 ID 包含在此进程生成的路由中，在将 OSPF 路由引入 BGP 中时，域 ID 被附加到 BGP VPN 路由上，作为 BGP 的扩展团体属性传递。
- 配置 BGP VPN 实例后，VPN 实例中的 BGP 路由交换与普通的 BGP 相同。有关 BGP 协议的配置，请参见“IP 路由分册”中的“BGP 配置”。

1.11 Egress PE 上私网路由标签操作方式简介

Egress PE 上私网路由的标签操作方式根据标签查找 FIB 进行转发与根据标签查找出接口进行转发两种：

- 私网标签的 POPGO 转发方式：标签分配时，都明确了标签跟出接口和下一跳的对应关系，弹出标签后，直接从出接口发送。

私网标签的 POP 转发方式：忽略标签分配时明确的标签跟出接口和下一跳对应关系，弹出标签后，再查 FIB 表转。

表1-32 配置 Egress PE 上私网路由标签操作方式

操作	命令	说明
进入系统视图	**system-view**	-
配置 Egress PE 上私网路由的标签操作方式为 POPGO 转发	**vpn popgo**	必选 缺省情况下，Egress PE 上私网路由的标签操作方式为 POP 转发



说明

在设备上执行该命令前需保存当前配置并根据提示信息重启设备。配置成功后此命令不会在设备上显示。可以通过 **display vpn label operation** 命令查看当前生效的转发方式。

1.12 配置 BGP 的 AS 号替换

1.12.1 配置准备

在配置 BGP 的 AS 号替换之前，需完成以下任务：

- 配置基本 MPLS L3VPN
- 不同 Site 的 CE 具有相同的 AS 号

1.12.2 配置 BGP 的 AS 号替换

不同 Site 的 CE 具有相同的 AS 号时，PE 上需要开启 BGP 的 AS 号替换功能，从而避免路由被丢弃。

使能了 BGP 的 AS 号替换功能后，当 PE 向指定对等体中的 CE 发布路由时，如果路由的 AS_PATH 中有与 CE 相同的 AS 号，将被替换成 PE 的 AS 号后再发布。

表1-33 配置 BGP 的 AS 号替换

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	必选
进入 BGP-VPN 实例视图	**ipv4-family vpn-instance** *vpn-instance-name*	必选
使能 BGP 的 AS 号替换功能	**peer** { *ip-address* | *group-name* } **substitute-as**	必选 缺省情况下，BGP 的 AS 号替换功能是被禁止的



说明

有关命令 **peer** { *ip-address* | *group-name* } **substitute-as** 介绍请参见“IP 路由分册”中的“BGP 命令”。

1.13 MPLS L3VPN 显示和维护

1.13.1 复位 BGP 连接

当 BGP 配置变化后，可以通过软复位或复位 BGP 连接使新的配置生效。软复位需要 BGP 对等体具备路由刷新能力（支持 ROUTE-REFRESH 消息）。

请在用户视图下进行下列操作。

表1-34 复位 BGP 连接

操作	命令
软复位指定 VPN 实例的 BGP 连接	**refresh bgp vpn-instance** *vpn-instance-name* { *ip-address* | **all** | **external** | **group** *group-name* } { **export** | **import** }
软复位 BGP 的 VPNv4 连接	**refresh bgp vpnv4** { *ip-address* | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** }
复位指定 VPN 实例的 BGP 连接	**reset bgp vpn-instance** *vpn-instance-name* { *as-number* | *ip-address* | **all** | **external** | **group** *group-name* }
复位 BGP 的 VPNv4 连接	**reset bgp vpnv4** { *as-number* | *ip-address* | **all** | **external** | **internal** | **group** *group-name* }

1.13.2 显示 MPLS L3VPN 的运行状态

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MPLS L3VPN 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除路由振荡衰减信息。

表1-35 显示 MPLS L3VPN 的运行状态

操作	命令
显示与 VPN 实例相关联的 IP 路由表	**display ip routing-table vpn-instance** *vpn-instance-name* [**verbose**]
显示指定 VPN 实例信息	**display ip vpn-instance** [**instance-name** *vpn-instance-name*]
显示指定 VPN 实例的 FIB 信息	**display fib vpn-instance** *vpn-instance-name* [**|** { **begin** | **exclude** | **include** } *regular-expression* | **acl** *acl-number* | **ip-prefix** *ip-prefix-name*]
显示指定 VPN 实例中与指定目的 IP 地址匹配的 FIB 信息	**display fib vpn-instance** *vpn-instance-name ip-address* [*mask* | *mask-length*]
显示 BGP 路由表中的标签路由信息	**display bgp vpnv4** { **all** | **vpn-instance** *vpn-instance-name* } **routing-table label**
显示 BGP VPNv4 对等体组信息	**display bgp vpnv4** { **all** | **vpn-instance** *vpn-instance-name* } **group** [*group-name*]
显示引入的 BGP VPNv4 路由信息	**display bgp vpnv4** { **all** | **vpn-instance** *vpn-instance-name* } **network**
显示 BGP VPNv4 的 AS 路径信息	**display bgp vpnv4** { **all** | **vpn-instance** *vpn-instance-name* } **paths** [*as-regular-expression*]
显示 BGP VPNv4 对等体信息	**display bgp vpnv4** { **all** | **vpn-instance** *vpn-instance-name* } **peer** [*group-name* **log-info** | *ip-address* { **log-info** | **verbose** } | **verbose**]

操作	命令
显示所有 BGP VPNv4 路由信息	**display bgp vpnv4 all routing-table** [*network-address* [{ *mask* | *mask-length* } [**longer-prefixes**]] | **as-path-acl** *as-path-acl-number* | **cidr** | **community** [*aa:nn*]&<1-13> [**no-export-subconfed** | **no-advertise** | **no-export**] * [**whole-match**] | **community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16> | **different-origin-as** | **peer** *ip-address* { **advertised-routes** | **received-routes** } [**statistic**] | **regular-expression** *as-regular-expression* | **statistic**]
显示指定路由标识符的 BGP VPNv4 路由信息	**display bgp vpnv4 route-distinguisher** *route-distinguisher* **routing-table** [*network-address* [{ *mask* | *mask-length* } [**longer-prefixes**]] | **as-path-acl** *as-path-acl-number* | **cidr** | **community** [*aa:nn*]&<1-13> [**no-export-subconfed** | **no-advertise** | **no-export**]* [**whole-match**] | **community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16> | **different-origin-as** | **regular-expression** *as-regular-expression*]
显示指定 VPN 实例的 BGP VPNv4 路由信息	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **routing-table** [*network-address* [{ *mask* | *mask-length* } [**longer-prefixes**]] | **as-path-acl** *as-path-acl-number* | **cidr** | **community** [*aa:nn*]&<1-13>[**no-export-subconfed** | **no-advertise** | **no-export**]* [**whole-match**] | **community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16> | **dampened** | **dampening parameter** | **different-origin-as** | **flap-info** [**as-path-acl** *as-path-acl-number* | *network-address* [*mask* [**longer-match**]] | *mask-length* [**longer-match**]] | **regular-expression** *as-regular-expression*] | **peer** *ip-address* { **advertised-routes** | **received-routes** } | **regular-expression** *as-regular-expression* | **statistic**]
显示 OSPF 伪连接信息	**display ospf** [*process-id*] **sham-link** [**area** *area-id*]
显示所有或指定隧道策略的信息	**display tunnel-policy** { **all** | **policy-name** *tunnel-policy-name* }
显示 Egress PE 上私网路由的标签操作方式信息	**display vpn label operation**
清除指定 VPN 实例的路由振荡衰减信息	**reset bgp vpn-instance** *vpn-instance-name* **dampening** [*network-address* [*mask* | *mask-length*]
清除指定 VPN 实例的 BGP 对等体路由振荡历史信息	**reset bgp vpn-instance** *vpn-instance-name ip-address* **flap-info** **reset bgp vpn-instance** *vpn-instance-name* **flap-info** [*ip-address* [*mask* | *mask-length*] | **as-path-acl** *as-path-acl-number* | **regexp** *as-path-regexp*]

说明

有关查看路由表的命令介绍请参见“IP 路由分册”中的“IP 路由表显示命令”。

1.14 MPLS L3VPN 典型配置举例

1.14.1 配置 MPLS L3VPN 示例

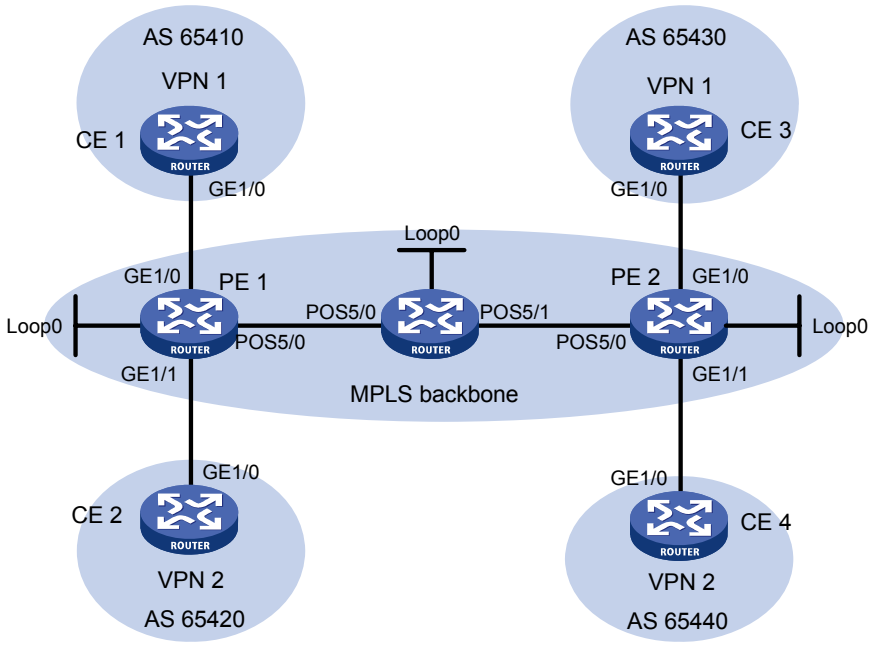
1. 组网需求

- CE 1、CE 3 属于 VPN 1，CE 2、CE 4 属于 VPN 2；
- VPN 1 使用的 VPN Target 属性为 111:1，VPN 2 使用的 VPN Target 属性为 222:2。不同 VPN 用户之间不能互相访问；

- CE 与 PE 之间配置 EBGP 交换 VPN 路由信息；
- PE 与 PE 之间配置 OSPF 实现 PE 内部的互通、配置 MP-IBGP 交换 VPN 路由信息。

2. 组网图

图1-19 配置 MPLS L3VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.1.1.1/24	P	Loop0	2.2.2.9/32
PE 1	Loop0	1.1.1.9/32		POS5/0	172.1.1.2/24
	GE1/0	10.1.1.2/24		POS5/1	172.2.1.1/24
	GE1/1	10.2.1.2/24	PE 2	Loop0	3.3.3.9/32
	POS5/0	172.1.1.1/24		GE1/0	10.3.1.2/24
CE 2	GE1/0	10.2.1.1/24		GE1/1	10.4.1.2/24
CE 3	GE1/0	10.3.1.1/24		POS5/0	172.2.1.2/24
CE 4	GE1/0	10.4.1.1/24			

3. 配置步骤

(1) 在 MPLS 骨干网上配置 IGP 协议，实现骨干网 PE 和 P 的互通

配置 PE 1。

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 1.1.1.9 32
[PE1-LoopBack0] quit
[PE1] interface pos 5/0
[PE1-POS5/0] ip address 172.1.1.1 24
[PE1-POS5/0] quit
[PE1] ospf
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

配置 P。

```
<P> system-view
[P] interface loopback 0
[P-LoopBack0] ip address 2.2.2.9 32
[P-LoopBack0] quit
[P] interface pos 5/0
[P-POS5/0] ip address 172.1.1.2 24
[P-POS5/0] quit
[P] interface pos 5/1
[P-POS5/1] ip address 172.2.1.1 24
[P-POS5/1] quit
[P] ospf
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

配置 PE 2。

```
<PE2> system-view
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 3.3.3.9 32
[PE2-LoopBack0] quit
[PE2] interface pos 5/0
[PE2-POS5/0] ip address 172.2.1.2 24
[PE2-POS5/0] quit
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

配置完成后，PE 1、P、PE 2 之间应能建立 OSPF 邻居，执行 **display ospf peer** 命令可以看到邻居达到 FULL 状态。执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback 路由。

以 PE 1 为例：

```
[PE1] display ip routing-table
Routing Tables: Public
         Destinations : 9        Routes : 9
Destination/Mask  Proto  Pre  Cost     NextHop         Interface
1.1.1.9/32        Direct 0    0        127.0.0.1       InLoop0
2.2.2.9/32        OSPF   10   1        172.1.1.2       POS5/0
3.3.3.9/32        OSPF   10   2        172.1.1.2       POS5/0
127.0.0.0/8       Direct 0    0        127.0.0.1       InLoop0
127.0.0.1/32      Direct 0    0        127.0.0.1       InLoop0
172.1.1.0/24      Direct 0    0        172.1.1.1       POS5/0
```

```
172.1.1.1/32      Direct 0    0        127.0.0.1       InLoop0
172.1.1.2/32      Direct 0    0        172.1.1.2       POS5/0
172.2.1.0/24      OSPF   10   1        172.1.1.2       POS5/0
[PE1] display ospf peer verbose
          OSPF Process 1 with Router ID 1.1.1.9
                  Neighbors
 Area 0.0.0.0 interface 172.1.1.1(POS5/0)'s neighbors
 Router ID: 172.1.1.2        Address: 172.1.1.2        GR State: Normal
   State: Full  Mode:Nbr is  Master  Priority: 1
   DR: None   BDR: None   MTU: 1500
   Dead timer due in 38  sec
   Neighbor is up for 00:02:44
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 5
```

(2) 在 MPLS 骨干网上配置 MPLS 基本能力和 MPLS LDP，建立 LDP LSP

配置 PE 1。

```
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface pos 5/0
[PE1-POS5/0] mpls
[PE1-POS5/0] mpls ldp
[PE1-POS5/0] quit
```

配置 P。

```
[P] mpls lsr-id 2.2.2.9
[P] mpls
[P-mpls] quit
[P] mpls ldp
[P-mpls-ldp] quit
[P] interface pos 5/0
[P-POS5/0] mpls
[P-POS5/0] mpls ldp
[P-POS5/0] quit
[P] interface pos 5/1
[P-POS5/1] mpls
[P-POS5/1] mpls ldp
[P-POS5/1] quit
```

配置 PE 2。

```
[PE2] mpls lsr-id 3.3.3.9
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface pos 5/0
```

```
[PE2-POS5/0] mpls
[PE2-POS5/0] mpls ldp
[PE2-POS5/0] quit
```

上述配置完成后，PE 1、P、PE 2 之间应能建立 LDP 会话，执行 **display mpls ldp session** 命令可以看到显示结果中 Session State 项为 Operational。执行 **display mpls ldp lsp** 命令，可以看到 LDP LSP 的建立情况。

以 PE 1 为例：

```
[PE1] display mpls ldp session
               LDP Session(s) in Public Network
------------------------------------------------------------------
 Peer-ID         Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 2.2.2.9:0       Operational   DU   Passive  Off  Off  5/5
 ------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
[PE1] display mpls ldp lsp
                             LDP LSP Information
 ------------------------------------------------------------------------
 SN  DestAddress/Mask   In/OutLabel  Next-Hop     In/Out-Interface
 ------------------------------------------------------------------------
 1   1.1.1.9/32         3/NULL       127.0.0.1     POS5/0/InLoop0
 2   2.2.2.9/32         NULL/3       172.1.1.2     -------/POS5/0
 3   3.3.3.9/32         NULL/1024    172.1.1.2     -------/POS5/0
------------------------------------------------------------------------
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
```

(3) 在 PE 设备上配置 VPN 实例，将 CE 接入 PE

配置 PE 1。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 111:1
[PE1-vpn-instance-vpn1] quit
[PE1] ip vpn-instance vpn2
[PE1-vpn-instance-vpn2] route-distinguisher 100:2
[PE1-vpn-instance-vpn2] vpn-target 222:2
[PE1-vpn-instance-vpn2] quit
[PE1] interface gigabitethernet 1/0
[PE1-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/0] ip address 10.1.1.2 24
[PE1-GigabitEthernet1/0] quit
[PE1] interface gigabitethernet 1/1
[PE1-GigabitEthernet1/1] ip binding vpn-instance vpn2
[PE1-GigabitEthernet1/1] ip address 10.2.1.2 24
[PE1-GigabitEthernet1/1] quit
```

配置 PE 2。

```
[PE2] ip vpn-instance vpn1
```

```
[PE2-vpn-instance-vpn1] route-distinguisher 200:1
[PE2-vpn-instance-vpn1] vpn-target 111:1
[PE2-vpn-instance-vpn1] quit
[PE2] ip vpn-instance vpn2
[PE2-vpn-instance-vpn2] route-distinguisher 200:2
[PE2-vpn-instance-vpn2] vpn-target 222:2
[PE2-vpn-instance-vpn2] quit
[PE2] interface gigabitethernet 1/0
[PE2-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/0] ip address 10.3.1.2 24
[PE2-GigabitEthernet1/0] quit
[PE2] interface gitabitethernet 1/1
[PE2-GigabitEthernet1/1] ip binding vpn-instance vpn2
[PE2-GigabitEthernet1/1] ip address 10.4.1.2 24
[PE2-GigabitEthernet1/1] quit
```

按 图 1-19配置各CE的接口IP地址，配置过程略。

配置完成后，在 PE 设备上执行 **display ip vpn-instance** 命令可以看到 VPN 实例的配置情况。各 PE 能 ping 通自己接入的 CE。

以 PE 1 和 CE 1 为例：

```
[PE1] display ip vpn-instance
  Total VPN-Instances configured : 2
  VPN-Instance Name      RD          Create Time
  vpn1                  100:1        2006/08/13 09:32:45
  vpn2                  100:2        2006/08/13 09:42:59
[PE1] ping -vpn-instance vpn1 10.1.1.1
  PING 10.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 10.1.1.1: bytes=56 Sequence=1 ttl=255 time=56 ms
    Reply from 10.1.1.1: bytes=56 Sequence=2 ttl=255 time=4 ms
    Reply from 10.1.1.1: bytes=56 Sequence=3 ttl=255 time=4 ms
    Reply from 10.1.1.1: bytes=56 Sequence=4 ttl=255 time=52 ms
    Reply from 10.1.1.1: bytes=56 Sequence=5 ttl=255 time=3 ms
  --- 10.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 3/23/56 ms
```

(4) 在 PE 与 CE 之间建立 EBGP 对等体，引入 VPN 路由

配置 CE 1。

```
<CE1> system-view
[CE1] bgp 65410
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct
[CE1-bgp] quit
```

说明

另外 3 个 CE 设备（CE 2～CE 4）配置与 CE 1 设备配置类似，配置过程省略。

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 10.1.1.1 as-number 65410
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
[PE1-bgp] ipv4-family vpn-instance vpn2
[PE1-bgp-vpn2] peer 10.2.1.1 as-number 65420
[PE1-bgp-vpn2] import-route direct
[PE1-bgp-vpn2] quit
[PE1-bgp] quit
```

说明

PE 2 的配置与 PE 1 类似，配置过程省略。

配置完成后，在 PE 设备上执行 **display bgp vpnv4 vpn-instance peer** 命令，可以看到 PE 与 CE 之间的 BGP 对等体关系已建立，并达到 Established 状态。

以 PE 1 与 CE 1 的对等体关系为例：

```
[PE1] display bgp vpnv4 vpn-instance vpn1 peer
 BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1            Peers in established state : 1

  Peer        AS    MsgRcvd MsgSent OutQ  PrefRcv   Up/Down    State
  10.1.1.1  65410 11      9       0     1         00:06:37   Established
```

(5) 在 PE 之间建立 MP-IBGP 对等体

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

配置 PE 2。

```
[PE2] bgp 100
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 0
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 1.1.1.9 enable
```

```
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] quit
```

配置完成后，在 PE 设备上执行 **display bgp peer** 或 **display bgp vpnv4 all peer** 命令，可以看到 PE 之间的 BGP 对等体关系已建立，并达到 Established 状态。

```
[PE1] display bgp peer
 BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1          Peers in established state : 1
  Peer        AS  MsgRcvd  MsgSent  OutQ   PrefRcv   Up/Down  State
  3.3.3.9    100        2        6     0         0   00:00:12 Established
```

(6) 配置完成后的检验

在 PE 设备上执行 **display ip routing-table vpn-instance** 命令，可以看到去往对端 CE 的路由。

以 PE 1 为例：

```
[PE1] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 3        Routes : 3
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
10.1.1.0/24       Direct 0    0         10.1.1.2        GE1/0
10.1.1.2/32       Direct 0    0         127.0.0.1       InLoop0
10.3.1.0/24       BGP    255  0         3.3.3.9         NULL0
[PE1] display ip routing-table vpn-instance vpn2
Routing Tables: vpn2
         Destinations : 3        Routes : 3
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
10.2.1.0/24       Direct 0    0         10.2.1.2        GE1/1
10.2.1.2/32       Direct 0    0         127.0.0.1       InLoop0
10.4.1.0/24       BGP    255  0         3.3.3.9         NULL0
```

同一 VPN 的 CE 能够相互 Ping 通，不同 VPN 的 CE 不能相互 Ping 通。

例如：CE 1 能够 Ping 通 CE 3（10.3.1.1），但不能 Ping 通 CE 4（10.4.1.1）。

```
[CE1] ping 10.3.1.1
  PING 10.3.1.1: 56  data bytes, press CTRL_C to break
    Reply from 10.3.1.1: bytes=56 Sequence=1 ttl=253 time=72 ms
    Reply from 10.3.1.1: bytes=56 Sequence=2 ttl=253 time=34 ms
    Reply from 10.3.1.1: bytes=56 Sequence=3 ttl=253 time=50 ms
    Reply from 10.3.1.1: bytes=56 Sequence=4 ttl=253 time=50 ms
    Reply from 10.3.1.1: bytes=56 Sequence=5 ttl=253 time=34 ms
  --- 10.3.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 34/48/72 ms
[CE1] ping 10.4.1.1
  PING 10.4.1.1: 56  data bytes, press CTRL_C to break
    Request time out
    Request time out
```

```
  Request time out
  Request time out
  Request time out
--- 10.4.1.1 ping statistics ---
  5 packet(s) transmitted
  0 packet(s) received
  100.00% packet loss
```

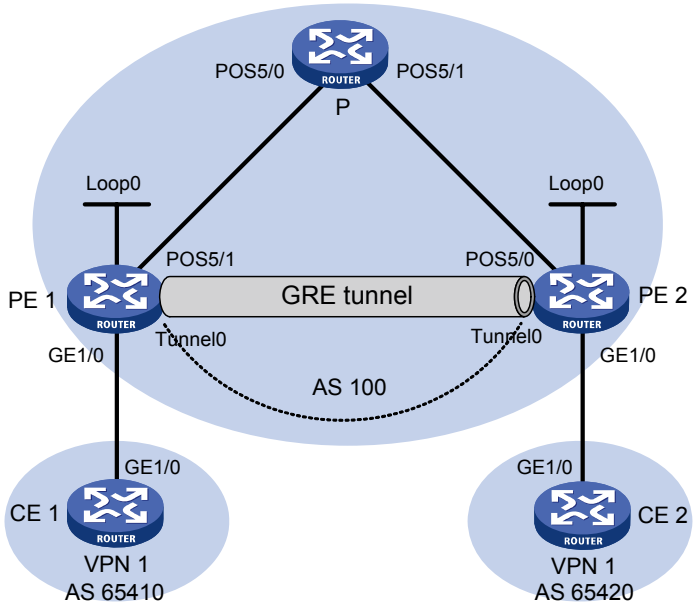
1.14.2 配置 MPLS L3VPN 采用 GRE 隧道示例

1. 组网需求

- CE 1 和 CE 2 属于 VPN 1。
- 在运营商骨干网上，PE 设备具备 MPLS 能力，P 设备只提供纯 IP 功能，不具备 MPLS 能力。
- 在骨干网上使用 GRE 隧道封装并转发 VPN 报文，实现 MPLS L3VPN。
- 这种组网下，需要在 PE 上配置隧道策略，指定 VPN 流量使用的隧道类型为 GRE。

2. 组网图

图1-20 配置采用 GRE 隧道的 MPLS L3VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.1.1.1/24	P	POS5/0	172.1.1.2/24
PE 1	Loop0	1.1.1.9/32		POS5/1	172.2.1.1/24
	GE1/0	10.1.1.2/24	PE 2	Loop0	2.2.2.9/32
	POS5/1	172.1.1.1/24		GE1/0	10.2.1.2/24
	Tunnel0	20.1.1.1/24		POS5/0	172.2.1.2/24
CE 2	GE1/0	10.2.1.1/24		Tunnel0	20.1.1.2/24

3. 配置步骤

(1) 在 MPLS 骨干网上配置 IGP 协议，实现骨干网 PE 和 P 的互通

本例中采用 OSPF，具体配置过程略。

配置完成后，PE 1、P、PE 2 之间应能建立 OSPF 邻居，执行 **display ospf peer** 命令可以看到邻居达到 FULL 状态。执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback 路由。

(2) 在 PE 设备上使能 MPLS 基本能力

配置 PE 1。

```
<PE1> system-view
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] quit
```

配置 PE 2。

```
<PE2> system-view
[PE2] mpls lsr-id 2.2.2.9
[PE2] mpls
[PE2-mpls] quit
```

(3) 在 PE 设备上配置 VPN 实例，将 CE 接入 PE，并在 PE 上应用隧道策略，指定使用 GRE 隧道转发 VPN 报文

配置 PE 1。

```
[PE1] tunnel-policy gre1
[PE1-tunnel-policy-gre1] tunnel select-seq gre load-balance-number 1
[PE1-tunnel-policy-gre1] quit
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 100:1 both
[PE1-vpn-instance-vpn1] tnl-policy gre1
[PE1-vpn-instance-vpn1] quit
[PE1] interface gigabitethernet 1/0
[PE1-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/0] ip address 10.1.1.2 24
[PE1-GigabitEthernet1/0] quit
```

配置 PE 2。

```
[PE2] tunnel-policy gre1
[PE2-tunnel-policy-gre1] tunnel select-seq gre load-balance-number 1
[PE2-tunnel-policy-gre1] quit
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 100:2
[PE2-vpn-instance-vpn1] vpn-target 100:1 both
[PE2-vpn-instance-vpn1] tnl-policy gre1
[PE2-vpn-instance-vpn1] quit
[PE2] interface gigabitethernet 1/0
[PE2-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/0] ip address 10.2.1.2 24
[PE2-GigabitEthernet1/0] quit
```

配置 CE 1。

```
<CE1> system-view
[CE1] interface gigabitethernet 1/0
[CE1-GigabitEthernet1/0] ip address 10.1.1.1 24
[CE1-GigabitEthernet1/0] quit
```

配置 CE2。

```
<CE2> system-view
[CE2] interface gigabitethernet 1/0
[CE2-GigabitEthernet1/0] ip address 10.2.1.1 24
[CE2-GigabitEthernet1/0] quit
```

配置完成后，在 PE 设备上执行 **display ip vpn-instance** 命令可以看到 VPN 实例的配置情况。各 PE 能 ping 通自己接入的 CE。

以 PE 1 为例：

```
[PE1] display ip vpn-instance
  Total VPN-Instances configured : 1
  VPN-Instance Name      RD          Create Time
  vpn1                  100:1        2006/08/13 09:32:45
[PE1] ping -vpn-instance vpn1 10.1.1.1
  PING 10.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 10.1.1.1: bytes=56 Sequence=1 ttl=255 time=27 ms
    Reply from 10.1.1.1: bytes=56 Sequence=2 ttl=255 time=33 ms
    Reply from 10.1.1.1: bytes=56 Sequence=3 ttl=255 time=7 ms
    Reply from 10.1.1.1: bytes=56 Sequence=4 ttl=255 time=29 ms
    Reply from 10.1.1.1: bytes=56 Sequence=5 ttl=255 time=9 ms
  --- 10.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 7/21/33 ms
```

(4) 在 PE 与 CE 之间建立 EBGP 对等体，引入 VPN 路由

配置 CE 1。

```
[CE1] bgp 65410
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct
[CE1-bgp] quit
```

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 10.1.1.1 as-number 65410
[PE1-bgp-vpn1] peer 10.1.1.1 next-hop-local
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
[PE1-bgp] quit
```



说明

CE 2 的配置与 CE 1 类似，PE 2 的配置与 PE 1 类似，配置过程省略。

配置完成后，在 PE 设备上执行 **display bgp vpnv4 vpn-instance peer** 命令，可以看到 PE 与 CE 之间的 BGP 对等体关系已建立，并达到 Established 状态。

以 PE 1 为例：

```
[PE1] display bgp vpnv4 vpn-instance vpn1 peer
 BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1             Peers in established state : 1
  Peer        AS  MsgRcvd  MsgSent  OutQ  PrefRcv   Up/Down  State
  10.1.1.1   65410      5        5     0        1   00:02:03 Established
```

(5) 在 PE 之间建立 MP-IBGP 对等体

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] peer 2.2.2.9 as-number 100
[PE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 2.2.2.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```



说明

PE 2 的配置与 PE 1 类似，配置过程省略。

配置完成后，在 PE 设备上执行 **display bgp peer** 或 **display bgp vpnv4 all peer** 命令，可以看到 PE 之间的 BGP 对等体关系已建立，并达到 Established 状态。

```
[PE1] display bgp vpnv4 all peer
 BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1           Peers in established state : 1
  Peer         AS  MsgRcvd  MsgSent  OutQ  PrefRcv   Up/Down  State
  2.2.2.9     100        3        3     0        1   00:00:34 Established
```

(6) 配置 GRE 隧道

配置 PE 1。

```
[PE1] interface tunnel 0
[PE1-Tunnel0] tunnel-protocol gre
[PE1-Tunnel0] source loopback 0
[PE1-Tunnel0] destination 2.2.2.9
[PE1-Tunnel0] ip address 20.1.1.1 24
[PE1-Tunnel0] mpls
[PE1-Tunnel0] quit
```

配置 PE 2。

```
[PE2] interface tunnel 0
[PE2-Tunnel0] tunnel-protocol gre
[PE2-Tunnel0] source loopback 0
[PE2-Tunnel0] destination 1.1.1.9
[PE2-Tunnel0] ip address 20.1.1.2 24
[PE2-Tunnel0] mpls
```

```
[PE2-Tunnel0] quit
```

(7) 配置完成后的检验

配置完成后，CE 能学到对端 CE 的接口路由。

以 CE 1 为例：

```
[CE1] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
10.1.1.0/24         Direct 0    0         10.1.1.1        GE1/0
10.1.1.1/32         Direct 0    0         127.0.0.1       InLoop0
10.2.1.0/24         BGP    255  0         10.1.1.2        GE1/0
127.0.0.0/8         Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0         127.0.0.1       InLoop0
```

以 PE 1 为例：

```
[PE1] display ip routing-table
Routing Tables: Public
         Destinations : 11       Routes : 11
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
1.1.1.9/32          Direct 0    0         127.0.0.1       InLoop0
2.2.2.9/32          OSPF   10   3125      172.1.1.2       POS5/1
10.2.1.0/24         Static 60   0         20.1.1.1        Tunnel0
20.1.1.0/24         Direct 0    0         20.1.1.1        Tunnel0
20.1.1.1/32         Direct 0    0         127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0         127.0.0.1       InLoop0
172.1.1.0/24        Direct 0    0         172.1.1.1       POS5/1
172.1.1.1/32        Direct 0    0         127.0.0.1       InLoop0
172.1.1.2/32        Direct 0    0         172.1.1.2       POS5/1
172.2.1.0/24        OSPF   10   3124      172.1.1.2       POS5/1
[PE1] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 3        Routes : 3
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
10.1.1.0/24         Direct 0    0         10.1.1.2        GE1/0
10.1.1.2/32         Direct 0    0         127.0.0.1       InLoop0
10.2.1.0/24         BGP    255  0         2.2.2.9         NULL0
```

CE 之间能够 ping 通。

```
[CE1] ping 10.2.1.1
  PING 10.2.1.1: 56  data bytes, press CTRL_C to break
    Reply from 10.2.1.1: bytes=56 Sequence=1 ttl=253 time=41 ms
    Reply from 10.2.1.1: bytes=56 Sequence=2 ttl=253 time=69 ms
    Reply from 10.2.1.1: bytes=56 Sequence=3 ttl=253 time=68 ms
    Reply from 10.2.1.1: bytes=56 Sequence=4 ttl=253 time=68 ms
    Reply from 10.2.1.1: bytes=56 Sequence=5 ttl=253 time=67 ms
  --- 10.2.1.1 ping statistics ---
    5 packet(s) transmitted
```

```
5 packet(s) received
0.00% packet loss
round-trip min/avg/max = 41/62/69 ms
```

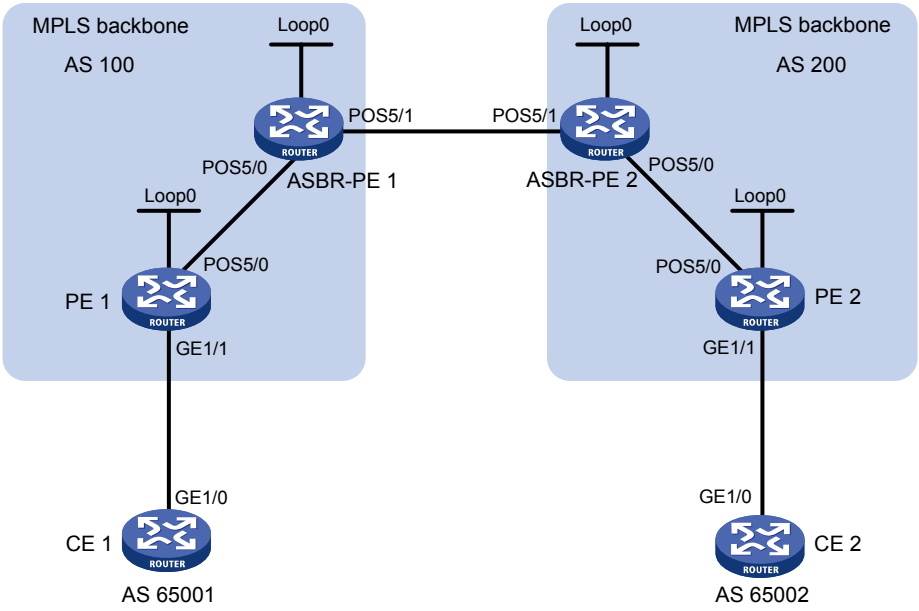
1.14.3 配置跨域 VPN-OptionA 方式示例

1. 组网需求

- CE 1 和 CE 2 属于同一个 VPN。
- CE 1 通过 AS100 的 PE 1 接入，CE2 通过 AS200 的 PE 2 接入。
- 采用 OptionA 方式实现跨域的 MPLS L3VPN，即，采用 VRF-to-VRF 方式管理 VPN 路由。
- 同一个 AS 内部的 MPLS 骨干网使用 OSPF 作为 IGP。

2. 组网图

图1-21 配置跨域 VPN-OptionA 方式组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.1.1.1/24	CE 2	GE1/0	10.2.1.1/24
PE 1	Loop0	1.1.1.9/32	PE 2	Loop0	4.4.4.9/32
	GE1/1	10.1.1.2/24		GE1/1	10.2.1.2/24
	POS5/0	172.1.1.2/24		POS5/0	162.1.1.2/24
ASBR-PE1	Loop0	2.2.2.9/32	ASBR-PE2	Loop0	3.3.3.9/32
	POS5/0	172.1.1.1/24		POS5/0	162.1.1.1/24
	POS5/1	192.1.1.1/24		POS5/1	192.1.1.2/24

3. 配置步骤

(1) 在 MPLS 骨干网上配置 IGP 协议，实现骨干网内互通

本例中采用 OSPF，具体配置步骤略。



说明

需要将作为 LSR ID 的 Loopback 接口的 32 位地址通过 OSPF 发布出去。

配置完成后，ASBR-PE 与本 AS 的 PE 之间应能建立 OSPF 邻居，执行 **display ospf peer** 命令可以看到邻居达到 FULL 状态，PE 之间能学习到对方的 Loopback 地址。

ASBR-PE 与本 AS 的 PE 之间能够互相 ping 通。

(2) 在 MPLS 骨干网上配置 MPLS 基本能力和 MPLS LDP，建立 LDP LSP

配置 PE 1 的 MPLS 基本能力，并在与 ASBR-PE 1 相连的接口上使能 LDP。

```
<PE1> system-view
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface pos 5/0
[PE1-POS5/0] mpls
[PE1-POS5/0] mpls ldp
[PE1-POS5/0] quit
```

配置 ASBR-PE 1 的 MPLS 基本能力，并在与 PE 1 相连的接口上使能 LDP。

```
<ASBR-PE1> system-view
[ASBR-PE1] mpls lsr-id 2.2.2.9
[ASBR-PE1] mpls
[ASBR-PE1-mpls] quit
[ASBR-PE1] mpls ldp
[ASBR-PE1-mpls-ldp] quit
[ASBR-PE1] interface pos 5/0
[ASBR-PE1-POS5/0] mpls
[ASBR-PE1-POS5/0] mpls ldp
[ASBR-PE1-POS5/0] quit
```

配置 ASBR-PE 2 的 MPLS 基本能力，并在与 PE 2 相连的接口上使能 LDP。

```
<ASBR-PE2> system-view
[ASBR-PE2] mpls lsr-id 3.3.3.9
[ASBR-PE2] mpls
[ASBR-PE2-mpls] quit
[ASBR-PE2] mpls ldp
[ASBR-PE2-mpls-ldp] quit
[ASBR-PE2] interface pos 5/0
[ASBR-PE2-POS5/0] mpls
[ASBR-PE2-POS5/0] mpls ldp
[ASBR-PE2-POS5/0] quit
```

配置 PE 2 的 MPLS 基本能力，并在与 ASBR-PE 2 相连的接口上使能 LDP。

```
<PE2> system-view
[PE2] mpls lsr-id 4.4.4.9
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface pos 5/0
[PE2-POS5/0] mpls
[PE2-POS5/0] mpls ldp
```

```
[PE2-POS5/0] quit
```

上述配置完成后，同一 AS 的 PE 和 ASBR-PE 之间应该建立起 LDP 邻居，在各设备上执行 **display mpls ldp session** 命令可以看到显示结果中 Session State 项为“Operational”。

(3) 在 PE 设备上配置 VPN 实例，将 CE 接入 PE



说明

同一 AS 内的 ASBR-PE 与 PE 的 VPN 实例的 VPN Target 应能匹配，不同 AS 的 PE 的 VPN 实例的 VPN Target 则不需要匹配。

配置 CE 1。

```
<CE1> system-view
[CE1] interface gigabitethernet 1/0
[CE1-GigabitEthernet1/0] ip address 10.1.1.1 24
[CE1-GigabitEthernet1/0] quit
```

配置 PE 1。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 100:1 both
[PE1-vpn-instance-vpn1] quit
[PE1] interface gigabitethernet 1/1
[PE1-GigabitEthernet1/1] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/1] ip address 10.1.1.2 24
[PE1-GigabitEthernet1/1] quit
```

配置 CE 2。

```
<CE2> system-view
[CE2] interface gigabitethernet 1/0
[CE2-GigabitEthernet1/0] ip address 10.2.1.1 24
[CE2-GigabitEthernet1/0] quit
```

配置 PE 2。

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 200:2
[PE2-vpn-instance-vpn1] vpn-target 100:1 both
[PE2-vpn-instance-vpn1] quit
[PE2] interface gigabitethernet 1/1
[PE2-GigabitEthernet1/1] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/1] ip address 10.2.1.2 24
[PE2-GigabitEthernet1/1] quit
```

配置 ASBR-PE 1：创建 VPN 实例，并将此实例绑定到连接 ASBR-PE 2 的接口（ASBR-PE 1 认为 ASBR-PE 2 是自己的 CE）。

```
[ASBR-PE1] ip vpn-instance vpn1
[ASBR-PE1-vpn-vpn1] route-distinguisher 100:1
[ASBR-PE1-vpn-vpn1] vpn-target 100:1 both
[ASBR-PE1-vpn-vpn1] quit
```

```
[ASBR-PE1] interface pos 5/1
[ASBR-PE1-POS5/1] ip binding vpn-instance vpn1
[ASBR-PE1-POS5/1] ip address 192.1.1.1 24
[ASBR-PE1-POS5/1] quit
```

配置 ASBR-PE 2：创建 VPN 实例，并将此实例绑定到连接 ASBR-PE 1 的接口（ASBR-PE 2 认为 ASBR-PE 1 是自己的 CE）。

```
[ASBR-PE2] ip vpn-instance vpn1
[ASBR-PE2-vpn-vpn1] route-distinguisher 200:1
[ASBR-PE2-vpn-vpn1] vpn-target 100:1 both
[ASBR-PE2-vpn-vpn1] quit
[ASBR-PE2] interface pos 5/1
[ASBR-PE2-POS5/1] ip binding vpn-instance vpn1
[ASBR-PE2-POS5/1] ip address 192.1.1.2 24
[ASBR-PE2-POS5/1] quit
```

上述配置完成后，在各 PE 设备上执行 **display ip vpn-instance** 命令能正确显示 VPN 实例配置。各 PE 能 ping 通各自的 CE。ASBR-PE 之间也能互相 ping 通。

(4) 在 PE 与 CE 之间建立 EBGP 对等体，引入 VPN 路由

配置 CE 1。

```
[CE1] bgp 65001
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct
[CE1-bgp] quit
```

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 10.1.1.1 as-number 65001
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
[PE1-bgp] quit
```

配置 CE 2。

```
[CE2] bgp 65002
[CE2-bgp] peer 10.2.1.2 as-number 200
[CE2-bgp] import-route direct
[CE2-bgp] quit
```

配置 PE 2。

```
[PE2] bgp 200
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] peer 10.2.1.1 as-number 65002
[PE2-bgp-vpn1] import-route direct
[PE2-bgp-vpn1] quit
[PE2-bgp] quit
```

(5) PE 与本 AS 的 ASBR-PE 之间建立 IBGP 对等体，ASBR-PE 之间建立 EBGP 对等体

配置 PE 1。

```
[PE1] bgp 100
[PE1-bgp] peer 2.2.2.9 as-number 100
[PE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 2.2.2.9 enable
[PE1-bgp-af-vpnv4] peer 2.2.2.9 next-hop-local
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

配置 ASBR-PE 1。

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] ipv4-family vpn-instance vpn1
[ASBR-PE1-bgp-vpn1] peer 192.1.1.2 as-number 200
[ASBR-PE1-bgp-vpn1] quit
[ASBR-PE1-bgp] peer 1.1.1.9 as-number 100
[ASBR-PE1-bgp] peer 1.1.1.9 connect-interface loopback 0
[ASBR-PE1-bgp] ipv4-family vpnv4
[ASBR-PE1-bgp-af-vpnv4] peer 1.1.1.9 enable
[ASBR-PE1-bgp-af-vpnv4] peer 1.1.1.9 next-hop-local
[ASBR-PE1-bgp-af-vpnv4] quit
[ASBR-PE1-bgp] quit
```

配置 ASBR-PE 2。

```
[ASBR-PE2] bgp 200
[ASBR-PE2-bgp] ipv4-family vpn-instance vpn1
[ASBR-PE2-bgp-vpn1] peer 192.1.1.1 as-number 100
[ASBR-PE2-bgp-vpn1] quit
[ASBR-PE2-bgp] peer 4.4.4.9 as-number 200
[ASBR-PE2-bgp] peer 4.4.4.9 connect-interface loopback 0
[ASBR-PE2-bgp] ipv4-family vpnv4
[ASBR-PE2-bgp-af-vpnv4] peer 4.4.4.9 enable
[ASBR-PE2-bgp-af-vpnv4] peer 4.4.4.9 next-hop-local
[ASBR-PE2-bgp-af-vpnv4] quit
[ASBR-PE2-bgp] quit
```

配置 PE 2。

```
[PE2] bgp 200
[PE2-bgp] peer 3.3.3.9 as-number 200
[PE2-bgp] peer 3.3.3.9 connect-interface loopback 0
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 3.3.3.9 enable
[PE2-bgp-af-vpnv4] peer 3.3.3.9 next-hop-local
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] quit
```

(6) 配置完成后的检验

上述配置完成后，CE 之间能学习到对方的接口路由，CE 1 和 CE 2 能够相互 ping 通。

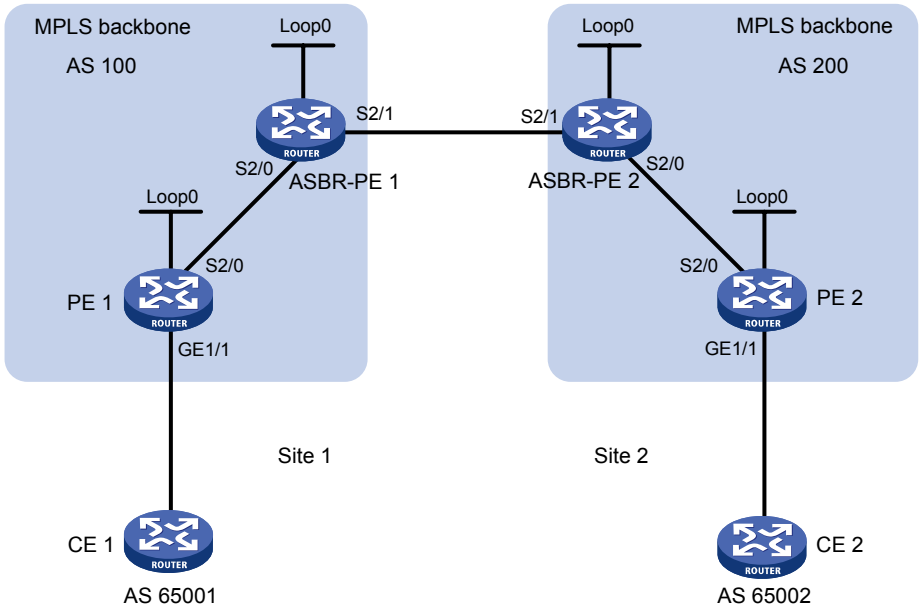
1.14.4 配置跨域 VPN-OptionB 方式示例

1. 组网需求

- Site 1 和 Site 2 属于同一个 VPN，Site 1 的 CE 1 通过 AS 100 的 PE 1 接入，Site 2 的 CE 2 通过 AS 600 的 PE 2 接入；
- 同一自治系统内的 PE 设备之间运行 IS-IS 作为 IGP；
- PE 1 与 ASBR-PE 1 间通过 MP-IBGP 交换标签 IPv4 路由；
- PE 2 与 ASBR-PE 2 间通过 MP-IBGP 交换标签 IPv4 路由；
- ASBR-PE 1 与 ASBR-PE 2 间通过 MP-EBGP 交换标签 IPv4 路由；
- ASBR 上不对接收的 VPN-IPv4 路由进行 VPN Target 过滤。

2. 组网图

图1-22 配置跨域 VPN-OptionB 方式组网图



设备	接口	IP 地址	设备	接口	IP 地址
PE 1	Loop0	2.2.2.9/32	PE 2	Loop0	5.5.5.9/32
	GE1/1	30.0.0.1/8		GE1/1	20.0.0.1/8
	S2/0	1.1.1.2/8		S2/0	9.1.1.2/8
ASBR-PE 1	Loop0	3.3.3.9/32	ASBR-PE 2	Loop0	4.4.4.9/32
	S2/0	1.1.1.1/8		S2/0	9.1.1.1/8
	S2/1	11.0.0.2/8		S2/1	11.0.0.1/8

3. 配置步骤

(1) 配置 PE 1

在 PE 1 上运行 IS-IS。

```
<PE1> system-view
[PE1] isis 1
[PE1-isis-1] network-entity 10.111.111.111.111.00
[PE1-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[PE1] mpls lsr-id 2.2.2.9
```

```
[PE1] mpls
[PE1-mpls] label advertise non-null
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

配置接口 Serial2/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[PE1] interface serial 2/0
[PE1-Serial2/0] ip address 1.1.1.2 255.0.0.0
[PE1-Serial2/0] isis enable 1
[PE1-Serial2/0] mpls
[PE1-Serial2/0] mpls ldp
[PE1-Serial2/0] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.9 32
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
```

创建一个 VPN 实例，名为 vpn1，配置 RD 和 VPN Target 属性。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE1-vpn-instance-vpn1] quit
```

将连接 CE 1 的接口绑定到创建的 VPN 实例。

```
[PE1] interface gigabitethernet 1/1
[PE1-GigabitEthernet1/1] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/1] ip address 30.0.0.1 8
[PE1-GigabitEthernet1/1] quit
```

在 PE 1 上运行 BGP。

```
[PE1] bgp 100
```

配置 IBGP 对等体 3.3.3.9 为 VPNv4 对等体。

```
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-af-vpnv4] quit
```

将直连路由引入 vpn1 的 VPN 路由表。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
```

(2) 配置 ASBR-PE 1

在 ASBR-PE 1 上运行 IS-IS。

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 10.222.222.222.222.00
[ASBR-PE1-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls
[ASBR-PE1-mpls] label advertise non-null
[ASBR-PE1-mpls] quit
[ASBR-PE1] mpls ldp
[ASBR-PE1-mpls-ldp] quit
```

配置接口 Serial2/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[ASBR-PE1] interface serial 2/0
[ASBR-PE1-Serial2/0] ip address 1.1.1.1 255.0.0.0
[ASBR-PE1-Serial2/0] isis enable 1
[ASBR-PE1-Serial2/0] mpls
[ASBR-PE1-Serial2/0] mpls ldp
[ASBR-PE1-Serial2/0] quit
```

配置接口 Serial2/1，使能 MPLS。

```
[ASBR-PE1] interface serial 2/1
[ASBR-PE1-Serial2/1] ip address 11.0.0.2 255.0.0.0
[ASBR-PE1-Serial2/1] mpls
[ASBR-PE1-Serial2/1] quit
```

创建 Loopback0 接口，并运行 IS-IS。

```
[ASBR-PE1] interface loopback 0
[ASBR-PE1-LoopBack0] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack0] isis enable 1
[ASBR-PE1-LoopBack0] quit
```

在 ASBR-PE 1 上运行 BGP。

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 2.2.2.9 as-number 100
[ASBR-PE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[ASBR-PE1-bgp] peer 11.0.0.1 as-number 200
[ASBR-PE1-bgp] peer 11.0.0.1 connect-interface serial 2/1
```

不对接收的 VPNv4 路由进行 Import VPN-target 过滤。

```
[ASBR-PE1-bgp] ipv4-family vpnv4
[ASBR-PE1-bgp-af-vpnv4] undo policy vpn-target
```

将 IBGP 对等体 2.2.2.9 和 EBGP 对等体 11.0.0.1 都配置为 VPNv4 对等体。

```
[ASBR-PE1-bgp-af-vpnv4] peer 11.0.0.1 enable
[ASBR-PE1-bgp-af-vpnv4] peer 2.2.2.9 enable
[ASBR-PE1-bgp-af-vpnv4] quit
```

(3) 配置 ASBR-PE 2

在 ASBR-PE 2 上运行 IS-IS。

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 10.222.222.222.222.00
[ASBR-PE2-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls
[ASBR-PE2-mpls] label advertise non-null
[ASBR-PE2-mpls] quit
[ASBR-PE2] mpls ldp
[ASBR-PE2-mpls-ldp] quit
```

配置接口 Serial2/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[ASBR-PE2] interface serial 2/0
[ASBR-PE2-Serial2/0] ip address 9.1.1.1 255.0.0.0
[ASBR-PE2-Serial2/0] isis enable 1
[ASBR-PE2-Serial2/0] mpls
[ASBR-PE2-Serial2/0] mpls ldp
[ASBR-PE2-Serial2/0] quit
```

配置接口 Serial2/1，使能 MPLS。

```
[ASBR-PE2] interface serial 2/1
[ASBR-PE2-Serial2/1] ip address 11.0.0.1 255.0.0.0
[ASBR-PE2-Serial2/1] mpls
[ASBR-PE2-Serial2/1] quit
```

创建 Loopback0 接口，并运行 IS-IS。

```
[ASBR-PE2] interface loopback 0
[ASBR-PE2-LoopBack0] ip address 4.4.4.9 32
[ASBR-PE2-LoopBack0] isis enable 1
[ASBR-PE2-LoopBack0] quit
```

在 ASBR-PE 2 上运行 BGP。

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp] peer 11.0.0.2 connect-interface serial 2/1
[ASBR-PE2-bgp] peer 5.5.5.9 as-number 200
[ASBR-PE2-bgp] peer 5.5.5.9 connect-interface loopback 0
```

不对接收的 VPNv4 路由进行 Import VPN-target 过滤。

```
[ASBR-PE2-bgp] ipv4-family vpnv4
[ASBR-PE2-bgp-af-vpnv4] undo policy vpn-target
```

将 IBGP 对等体 5.5.5.9 和 EBGP 对等体 11.0.0.2 都配置为 VPNv4 对等体。

```
[ASBR-PE2-bgp-af-vpnv4] peer 11.0.0.2 enable
[ASBR-PE2-bgp-af-vpnv4] peer 5.5.5.9 enable
[ASBR-PE2-bgp-af-vpnv4] quit
[ASBR-PE2-bgp] quit
```

(4) 配置 PE 2

在 PE 2 上运行 IS-IS。

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 10.111.111.111.111.00
[PE2-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls
[PE2-mpls] label advertise non-null
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

配置接口 Serial2/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[PE2] interface serial 2/0
[PE2-Serial2/0] ip address 9.1.1.2 255.0.0.0
[PE2-Serial2/0] isis enable 1
[PE2-Serial2/0] mpls
[PE2-Serial2/0] mpls ldp
[PE2-Serial2/0] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 5.5.5.9 32
[PE2-LoopBack0] isis enable 1
[PE2-LoopBack0] quit
```

创建一个 VPN 实例，名为 vpn1，配置 RD 和 VPN Target 属性。

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 12:12
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

将连接 CE 1 的接口绑定到创建的 VPN 实例。

```
[PE2] interface gigabitethernet 1/1
[PE2-GigabitEthernet1/1] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/1] ip address 20.0.0.1 8
[PE2-GigabitEthernet1/1] quit
```

在 PE 2 上运行 BGP。

```
[PE2] bgp 600
```

配置 IBGP 对等体 4.4.4.9 为 VPNv4 对等体。

```
[PE2-bgp] peer 4.4.4.9 as-number 200
[PE2-bgp] peer 4.4.4.9 connect-interface loopback 0
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 4.4.4.9 enable
[PE2-bgp-af-vpnv4] quit
```

将直连路由引入 vpn1 的 VPN 路由表。

```
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] import-route direct
[PE2-bgp-vpn1] quit
[PE2-bgp] quit
```

(5) 配置完成后的检验

配置完成后，在 PE 2 上对 PE 1 进行 Ping 操作，可以 Ping 通。

在 PE 1 上对 PE 2 进行 Ping 操作，也可以 Ping 通。

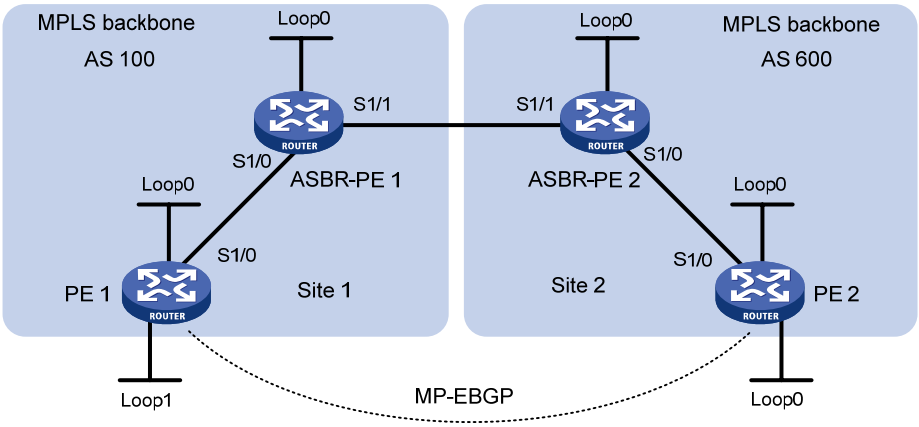
1.14.5 配置跨域 VPN-OptionC 方式示例

1. 组网需求

- Site 1 和 Site 2 属于同一个 VPN，Site 1 通过 AS 100 的 PE 1 接入，Site 2 通过 AS 600 的 PE 2 接入；
- 同一自治系统内的 PE 设备之间运行 IS-IS 作为 IGP；
- PE 1 与 ASBR-PE 1 间通过 MP-IBGP 交换标签 IPv4 路由；
- PE 2 与 ASBR-PE 2 间通过 MP-IBGP 交换标签 IPv4 路由；
- PE 1 与 PE 2 建立 MP-EBGP 对等体；
- ASBR-PE 1 和 ASBR-PE 2 上分别配置路由策略，对从对方接收的路由压入标签；
- ASBR-PE 1 与 ASBR-PE 2 间通过 MP-EBGP 交换标签 IPv4 路由。

2. 组网图

图1-23 配置跨域 VPN-OptionC 方式组网图



设备	接口	IP 地址	设备	接口	IP 地址
PE 1	Loop0	2.2.2.9/32	PE 2	Loop0	5.5.5.9/32
	Loop1	30.0.0.1/32		Loop1	20.0.0.1/32
	S1/0	1.1.1.2/8		S1/0	9.1.1.2/8
ASBR-PE 1	Loop0	3.3.3.9/32	ASBR-PE 2	Loop0	4.4.4.9/32
	S1/0	1.1.1.1/8		S1/0	9.1.1.1/8
	S1/1	11.0.0.2/8		S1/1	11.0.0.1/8

3. 配置步骤

(1) 配置 PE 1

在 PE 1 上运行 IS-IS。

```
<PE1> system-view
[PE1] isis 1
[PE1-isis-1] network-entity 10.111.111.111.111.00
[PE1-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[PE1] mpls lsr-id 2.2.2.9
[PE1] mpls
[PE1-mpls] label advertise non-null
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

配置接口 Serial1/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[PE1] interface serial 1/0
[PE1-Serial1/0] ip address 1.1.1.2 255.0.0.0
[PE1-Serial1/0] isis enable 1
[PE1-Serial1/0] mpls
[PE1-Serial1/0] mpls ldp
[PE1-Serial1/0] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.9 32
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
```

创建 VPN 实例，名称为 vpn1，为其配置 RD 和 VPN Target 属性。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE1-vpn-instance-vpn1] quit
```

创建 Loopback1 接口，并将此接口绑定到 vpn1。

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip binding vpn-instance vpn1
[PE1-LoopBack1] ip address 30.0.0.1 32
[PE1-LoopBack1] quit
```

在 PE 1 上运行 BGP。

```
[PE1] bgp 100
```

配置 PE 1 向 IBGP 对等体 3.3.3.9 发布标签路由及从 3.3.3.9 接收标签路由的能力。

```
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp] peer 3.3.3.9 label-route-capability
```

配置 PE 1 到 EBGP 对等体 5.5.5.9 的最大跳数为 10。

```
[PE1-bgp] peer 5.5.5.9 as-number 600
[PE1-bgp] peer 5.5.5.9 connect-interface loopback 0
```

```
[PE1-bgp] peer 5.5.5.9 ebgp-max-hop 10
```

配置对等体 5.5.5.9 作为 VPNv4 对等体。

```
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 5.5.5.9 enable
[PE1-bgp-af-vpnv4] quit
```

将直连路由引入到 vpn1 的路由表。

```
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] import-route direct
[PE1-bgp-vpn1] quit
[PE1-bgp] quit
```

(2) 配置 ASBR-PE1

在 ASBR-PE1 上运行 IS-IS。

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 10.222.222.222.222.00
[ASBR-PE1-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls
[ASBR-PE1-mpls] label advertise non-null
[ASBR-PE1-mpls] quit
[ASBR-PE1] mpls ldp
[ASBR-PE1-mpls-ldp] quit
```

配置接口 Serial1/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[ASBR-PE1] interface serial 1/0
[ASBR-PE1-Serial1/0] ip address 1.1.1.1 255.0.0.0
[ASBR-PE1-Serial1/0] isis enable 1
[ASBR-PE1-Serial1/0] mpls
[ASBR-PE1-Serial1/0] mpls ldp
[ASBR-PE1-Serial1/0] quit
```

配置接口 Serial1/1，并在接口上使能 MPLS。

```
[ASBR-PE1] interface serial 1/1
[ASBR-PE1-Serial1/1] ip address 11.0.0.2 255.0.0.0
[ASBR-PE1-Serial1/1] mpls
[ASBR-PE1-Serial1/1] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[ASBR-PE1] interface loopback 0
[ASBR-PE1-LoopBack0] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack0] isis enable 1
[ASBR-PE1-LoopBack0] quit
```

创建路由策略。

```
[ASBR-PE1] route-policy policy1 permit node 1
[ASBR-PE1-route-policy1] apply mpls-label
```

```
[ASBR-PE1-route-policy1] quit
[ASBR-PE1] route-policy policy2 permit node 1
[ASBR-PE1-route-policy2] if-match mpls-label
[ASBR-PE1-route-policy2] apply mpls-label
[ASBR-PE1-route-policy2] quit
```

在 ASBR-PE 1 上运行 BGP，引入 IS-IS 进程 1 的路由。

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] import-route isis 1
```

对 IBGP 对等体 2.2.2.9 发布来的路由应用已配置的路由策略 policy2。

```
[ASBR-PE1-bgp] peer 2.2.2.9 as-number 100
[ASBR-PE1-bgp] peer 2.2.2.9 route-policy policy2 export
```

向 IBGP 对等体 2.2.2.9 发布标签路由及从 2.2.2.9 接收标签路由的能力。

```
[ASBR-PE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[ASBR-PE1-bgp] peer 2.2.2.9 label-route-capability
```

对 EBGP 对等体 11.0.0.1 发布来的路由应用已配置的路由策略 policy1。

```
[ASBR-PE1-bgp] peer 11.0.0.1 as-number 600
[ASBR-PE1-bgp] peer 11.0.0.1 route-policy policy1 export
```

向 EBGP 对等体 11.0.0.1 发布标签路由及从 11.0.0.1 接收标签路由的能力。

```
[ASBR-PE1-bgp] peer 11.0.0.1 label-route-capability
[ASBR-PE1-bgp] quit
```

(3) 配置 ASBR-PE 2

在 ASBR-PE 2 上运行 IS-IS。

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 10.222.222.222.222.00
[ASBR-PE2-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls
[ASBR-PE2-mpls] label advertise non-null
[ASBR-PE2-mpls] quit
[ASBR-PE2] mpls ldp
[ASBR-PE2-mpls-ldp] quit
```

配置接口 Serial1/0，在接口上运行 IS-IS，并在接口上使能 MPLS 和 LDP。

```
[ASBR-PE2] interface serial 1/0
[ASBR-PE2-Serial1/0] ip address 9.1.1.1 255.0.0.0
[ASBR-PE2-Serial1/0] isis enable 1
[ASBR-PE2-Serial1/0] mpls
[ASBR-PE2-Serial1/0] mpls ldp
[ASBR-PE2-Serial1/0] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[ASBR-PE2] interface loopback 0
[ASBR-PE2-LoopBack0] ip address 4.4.4.9 32
```

```
[ASBR-PE2-LoopBack0] isis enable 1
[ASBR-PE2-LoopBack0] quit
```

配置接口 Serial1/1，在接口上使能 MPLS。

```
[ASBR-PE2] interface serial 1/1
[ASBR-PE2-Serial1/1] ip address 11.0.0.1 255.0.0.0
[ASBR-PE2-Serial1/1] mpls
[ASBR-PE2-Serial1/1] quit
```

创建路由策略。

```
[ASBR-PE2] route-policy policy1 permit node 1
[ASBR-PE2-route-policy1] apply mpls-label
[ASBR-PE2-route-policy1] quit
[ASBR-PE2] route-policy policy2 permit node 1
[ASBR-PE2-route-policy2] if-match mpls-label
[ASBR-PE2-route-policy2] apply mpls-label
[ASBR-PE2-route-policy2] quit
```

在 ASBR-PE 2 上运行 BGP，引入 IS-IS 进程 1 的路由。

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp] import-route isis 1
```

向 IBGP 对等体 5.5.5.9 发布标签路由及从 5.5.5.9 接收标签路由的能力。

```
[ASBR-PE2-bgp] peer 5.5.5.9 as-number 600
[ASBR-PE2-bgp] peer 5.5.5.9 connect-interface loopback 0
[ASBR-PE2-bgp] peer 5.5.5.9 label-route-capability
```

对 IBGP 对等体 5.5.5.9 发布来的路由应用已配置的路由策略 policy2。

```
[ASBR-PE2-bgp] peer 5.5.5.9 route-policy policy2 export
```

对 EBGP 对等体 11.0.0.2 发布来的路由应用已配置的路由策略 policy1。

```
[ASBR-PE2-bgp] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp] peer 11.0.0.2 route-policy policy1 export
```

向 EBGP 对等体 11.0.0.2 发布标签路由及从 11.0.0.2 接收标签路由的能力。

```
[ASBR-PE2-bgp] peer 11.0.0.2 label-route-capability
[ASBR-PE2-bgp] quit
```

(4)　配置 PE 2

在 PE 2 上运行 IS-IS。

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 10.111.111.111.111.00
[PE2-isis-1] quit
```

配置 LSR ID，使能 MPLS 和 LDP。

```
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls
[PE2-mpls] label advertise non-null
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

配置接口 Serial1/0，在接口上运行 IS-IS，并使能 MPLS 和 LDP。

```
[PE2] interface serial 1/0
[PE2-Serial1/0] ip address 9.1.1.2 255.0.0.0
[PE2-Serial1/0] isis enable 1
[PE2-Serial1/0] mpls
[PE2-Serial1/0] mpls ldp
[PE2-Serial1/0] quit
```

创建 Loopback0 接口，在接口上运行 IS-IS。

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 5.5.5.9 32
[PE2-LoopBack0] isis enable 1
[PE2-LoopBack0] quit
```

创建 VPN 实例，名称为 vpn1，为其配置 RD 和 VPN Target 属性。

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 11:11
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

创建 Loopback1 接口，并将此接口绑定到 vpn1。

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip binding vpn-instance vpn1
[PE2-LoopBack1] ip address 20.0.0.1 32
[PE2-LoopBack1] quit
```

在 PE 2 上运行 BGP。

```
[PE2] bgp 600
```

配置 PE 2 向 IBGP 对等体 4.4.4.9 发布标签路由及从 4.4.4.9 接收标签路由的能力。

```
[PE2-bgp] peer 4.4.4.9 as-number 600
[PE2-bgp] peer 4.4.4.9 connect-interface loopback 0
[PE2-bgp] peer 4.4.4.9 label-route-capability
```

配置 PE 2 到 EBGP 对等体 2.2.2.9 的最大跳数为 10。

```
[PE2-bgp] peer 2.2.2.9 as-number 100
[PE2-bgp] peer 2.2.2.9 connect-interface loopback 0
[PE2-bgp] peer 2.2.2.9 ebgp-max-hop 10
```

配置对等体 2.2.2.9 作为 VPNv4 对等体。

```
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 2.2.2.9 enable
[PE2-bgp-af-vpnv4] quit
```

将直连路由引入 vpn1 的路由表。

```
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] import-route direct
[PE2-bgp-vpn1] quit
[PE2-bgp] quit
```

配置完成后，在 PE2 上对 PE1 进行 Ping 操作，可以 Ping 通：

```
[PE2] ping -vpn-instance vpn1 30.0.0.1
```

在 PE1 上对 PE2 进行 Ping 操作，也可以 Ping 通：

```
[PE1] ping -vpn-instance vpn1 20.0.0.1
```

1.14.6 配置运营商的运营商示例

1. 组网需求

二级运营商向自己的客户提供 MPLS L3VPN 服务。

在 图 1-24中：

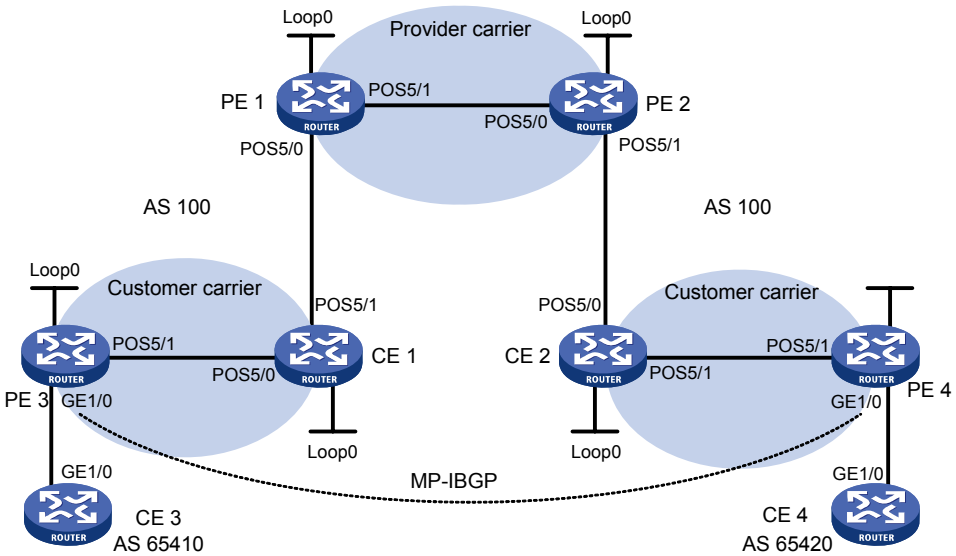
- PE 1 和 PE 2 是一级运营商骨干网的 PE 设备；
- CE 1 和 CE 2 是二级运营商的设备，作为 CE 接入一级运营商的骨干网；
- PE 3 和 PE 4 是二级运营商的设备，为二级运营商的客户提供接入；
- CE 3 和 CE 4 是二级运营商的客户。

运营商的运营商的配置关键在于理解两类路由的交换过程，即：

- 二级运营商 VPN 内部路由在一级运营商骨干网上的交换：一级运营商将二级运营商作为自己的 CE 接入；
- 二级运营商本身客户的 VPN 路由在二级运营商 PE 设备间的交换：需要在二级运营商 PE 设备（PE 3 和 PE 4）间建立 MP-IBGP 对等体关系。

2. 组网图

图1-24 配置 Carriers' carriers 配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 3	GE1/0	100.1.1.1/24	CE 4	GE1/0	120.1.1.1/24
PE 3	Loop0	1.1.1.9/32	PE 4	Loop0	6.6.6.9/32
	GE1/0	100.1.1.2/24		GE1/0	120.1.1.2/24
	POS5/1	10.1.1.1/24		POS5/1	20.1.1.2/24
CE 1	Loop0	2.2.2.9/32	CE 2	Loop0	5.5.5.9/32
	POS5/0	10.1.1.2/24		POS5/0	21.1.1.2/24
	POS5/1	11.1.1.1/24		POS5/1	20.1.1.1/24
PE 1	Loop0	3.3.3.9/32	PE 2	Loop0	4.4.4.9/32
	POS5/0	11.1.1.2/24		POS5/0	30.1.1.2/24
	POS5/1	30.1.1.1/24		POS5/1	21.1.1.1/24

3. 配置步骤

(1) 配置一级运营商骨干网的 MPLS L3VPN，使用 IS-IS 作为骨干网的 IGP 协议，PE 1 和 PE 2 之间使能 LDP，并建立 MP-IBGP 对等体关系

配置 PE 1。

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 3.3.3.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 3.3.3.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE1-isis-1] quit
[PE1] interface loopback 0
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
[PE1] interface pos 5/1
[PE1-POS5/1] ip address 30.1.1.1 24
[PE1-POS5/1] isis enable 1
[PE1-POS5/1] mpls
[PE1-POS5/1] mpls ldp
[PE1-POS5/1] mpls ldp transport-address interface
[PE1-POS5/1] quit
[PE1] bgp 100
[PE1-bgp] peer 4.4.4.9 as-number 100
[PE1-bgp] peer 4.4.4.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 4.4.4.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```



说明

PE 2 的配置与 PE 1 类似，配置过程省略。

配置完成后，在 PE 1 或 PE 2 上执行 **display mpls ldp session** 命令可以看到 LDP 会话建立成功；执行 **display bgp peer** 命令可以看到 BGP 对等体关系已建立，并达到 Established 状态；执行 **display isis peer** 命令可以看到 IS-IS 邻居关系已建立，状态为 up。

以 PE 1 为例：

```
[PE1] display mpls ldp session
               LDP Session(s) in Public Network
 ------------------------------------------------------------------
 Peer-ID        Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
```

```
 --------------------------------------------------------------------
 4.4.4.9:0      Operational   DU   Active   Off  Off  378/378
 --------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
[PE1] display bgp peer
 BGP local router ID : 3.3.3.9
 Local AS number : 100
 Total number of peers : 1          Peers in established state : 1
  Peer          AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
  4.4.4.9      100      162      145     0        0  02:12:47 Established
[PE1] display isis peer
                         Peer information for ISIS(1)
                         ---------------------------
  System Id    Interface   Circuit Id    State HoldTime Type   PRI
0000.0000.0005 POS5/1           001       Up   29s      L1L2   --
```

(2) 配置二级运营商网络：使用 IS-IS 作为 IGP 协议，PE 3 和 CE 1、PE 4 和 CE 2 之间分别使能 LDP

配置 PE 3。

```
<PE3> system-view
[PE3] interface loopback 0
[PE3-LoopBack0] ip address 1.1.1.9 32
[PE3-LoopBack0] quit
[PE3] mpls lsr-id 1.1.1.9
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
[PE3] isis 2
[PE3-isis-2] network-entity 10.0000.0000.0000.0001.00
[PE3-isis-2] quit
[PE3] interface loopback 0
[PE3-LoopBack0] isis enable 2
[PE3-LoopBack0] quit
[PE3] interface pos 5/1
[PE3-POS5/1] ip address 10.1.1.1 24
[PE3-POS5/1] isis enable 2
[PE3-POS5/1] mpls
[PE3-POS5/1] mpls ldp
[PE3-POS5/1] mpls ldp transport-address interface
[PE3-POS5/1] quit
```

配置 CE 1。

```
<CE1> system-view
[CE1] interface loopback 0
[CE1-LoopBack0] ip address 2.2.2.9 32
[CE1-LoopBack0] quit
[CE1] mpls lsr-id 2.2.2.9
```

```
[CE1] mpls
[CE1-mpls] quit
[CE1] mpls ldp
[CE1-mpls-ldp] quit
[CE1] isis 2
[CE1-isis-2] network-entity 10.0000.0000.0000.0002.00
[CE1-isis-2] quit
[CE1] interface loopback 0
[CE1-LoopBack0] isis enable 2
[CE1-LoopBack0] quit
[CE1] interface POS 5/0
[CE1-POS5/0] ip address 10.1.1.2 24
[CE1-POS5/0] isis enable 2
[CE1-POS5/0] mpls
[CE1-POS5/0] mpls ldp
[CE1-POS5/0] mpls ldp transport-address interface
[CE1-POS5/0] quit
```

配置完成后，PE 3 和 CE 1 之间应能建立 LDP 和 IS-IS 邻居关系。



说明

PE 4 和 CE 2 之间的配置与 PE 3 和 CE 1 之间的配置类似，配置过程省略。

---

(3) 配置二级运营商 CE 接入到一级运营商的 PE

配置 PE1，并引入 IS-IS 路由。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 200:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] mpls ldp vpn-instance vpn1
[PE1-mpls-ldp-vpn-instance-vpn1] quit
[PE1] isis 2 vpn-instance vpn1
[PE1-isis-2] network-entity 10.0000.0000.0000.0003.00
[PE1-isis-2] import-route bgp
[PE1-isis-2] quit
[PE1] interface pos 5/0
[PE1-POS5/0] ip binding vpn-instance vpn1
[PE1-POS5/0] ip address 11.1.1.2 24
[PE1-POS5/0] isis enable 2
[PE1-POS5/0] mpls
[PE1-POS5/0] mpls ldp
[PE1-POS5/0] mpls ldp transport-address interface
[PE1-POS5/0] quit
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] import isis 2
[PE1-bgp-vpn1] quit
```

```
[PE1-bgp] quit
```

配置 CE1。

```
[CE1] interface pos 5/1
[CE1-POS5/1] ip address 11.1.1.1 24
[CE1-POS5/1] isis enable 2
[CE1-POS5/1] mpls
[CE1-POS5/1] mpls ldp
[CE1-POS5/1] mpls ldp transport-address interface
[CE1-POS5/1] quit
```

配置完成后，PE 1 和 CE 1 之间应能建立 LDP 和 IS-IS 邻居关系。



说明

PE 2 和 CE 2 之间的配置与 PE 1 和 CE 1 之间的配置类似，配置过程省略。

(4) 配置二级运营商的客户接入 PE

配置 CE 3。

```
<CE3> system-view
[CE3] interface gigabitethernet 1/0
[CE3-GigabitEthernet1/0] ip address 100.1.1.1 24
[CE3-GigabitEthernet1/0] quit
[CE3] bgp 65410
[CE3-bgp] peer 100.1.1.2 as-number 100
[CE3-bgp] import-route direct
[CE3-bgp] quit
```

配置 PE3。

```
[PE3] ip vpn-instance vpn1
[PE3-vpn-instance-vpn1] route-distinguisher 100:1
[PE3-vpn-instance-vpn1] vpn-target 1:1
[PE3-vpn-instance-vpn1] quit
[PE3] interface gigabitethernet 1/0
[PE3-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE3-GigabitEthernet1/0] ip address 100.1.1.2 24
[PE3-GigabitEthernet1/0] quit
[PE3] bgp 100
[PE3-bgp] ipv4-family vpn-instance vpn1
[PE3-bgp-vpn1] peer 100.1.1.1 as-number 65410
[PE3-bgp-vpn1] import-route direct
[PE3-bgp-vpn1] quit
[PE3-bgp] quit
```



说明

PE 4 和 CE 4 之间的配置与 PE 3 和 CE 3 之间的配置类似，配置过程省略。

(5) 在二级运营商的 PE 之间建立 MP-IBGP 对等体关系，交换二级运营商的客户的 VPN 路由

配置 PE 3。

```
[PE3] bgp 100
[PE3-bgp] peer 6.6.6.9 as-number 100
[PE3-bgp] peer 6.6.6.9 connect-interface loopback 0
[PE3-bgp] ipv4-family vpnv4
[PE3-bgp-af-vpnv4] peer 6.6.6.9 enable
[PE3-bgp-af-vpnv4] quit
[PE3-bgp] quit
```



说明

PE 4 的配置与 PE 3 类似，配置过程省略。

(6) 配置完成后的检验

所有配置完成后，在 PE 1 和 PE 2 上执行 **display ip routing-table** 命令，可以看到 PE 1 和 PE 2 的公网路由表中只有一级运营商网络的路由。以 PE 1 为例：

```
[PE1] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
3.3.3.9/32          Direct 0    0         127.0.0.1       InLoop0
4.4.4.9/32          ISIS   15   10        30.1.1.2        POS5/1
30.1.1.0/24         Direct 0    0         30.1.1.1        POS5/1
30.1.1.1/32         Direct 0    0         127.0.0.1       InLoop0
30.1.1.2/32         Direct 0    0         30.1.1.2        POS5/1
127.0.0.0/8         Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0         127.0.0.1       InLoop0
```

在 PE 1 和 PE 2 上执行 **display ip routing-table vpn-instance** 命令，可以看到 VPN 路由表中有二级运营商网络的内部路由，但没有二级运营商维护的 VPN 路由。以 PE 1 为例：

```
[PE1] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 11        Routes : 11
Destination/Mask    Proto  Pre  Cost      NextHop         Interface
1.1.1.9/32          ISIS   15   20        11.1.1.1        POS5/0
2.2.2.9/32          ISIS   15   10        11.1.1.1        POS5/0
5.5.5.9/32          BGP    255  0         4.4.4.9         NULL0
6.6.6.9/32          BGP    255  0         4.4.4.9         NULL0
10.1.1.0/24         ISIS   15   20        11.1.1.1        POS5/0
11.1.1.0/24         Direct 0    0         11.1.1.1        POS5/0
11.1.1.1/32         Direct 0    0         127.0.0.1       InLoop0
11.1.1.2/32         Direct 0    0         11.1.1.2        POS5/0
20.1.1.0/24         BGP    255  0         4.4.4.9         NULL0
21.1.1.0/24         BGP    255  0         4.4.4.9         NULL0
21.1.1.2/32         BGP    255  0         4.4.4.9         NULL0
```

在 CE 1 和 CE 2 上执行 **display ip routing-table** 命令，可以看到公网路由表中有二级运营商网络的内部路由，但没有二级运营商维护的 VPN 路由。以 CE 1 为例：

```
[CE1] display ip routing-table
Routing Tables: Public
         Destinations : 16       Routes : 16
Destination/Mask    Proto  Pre  Cost   NextHop         Interface
1.1.1.9/32          ISIS   15   10     10.1.1.2        POS5/0
2.2.2.9/32          Direct 0    0      127.0.0.1       InLoop0
5.5.5.9/32          ISIS   15   74     11.1.1.2        POS5/1
6.6.6.9/32          ISIS   15   74     11.1.1.2        POS5/1
10.1.1.0/24         Direct 0    0      10.1.1.2        POS5/0
10.1.1.1/32         Direct 0    0      10.1.1.1        POS5/0
10.1.1.2/32         Direct 0    0      127.0.0.1       InLoop0
11.1.1.0/24         Direct 0    0      11.1.1.1        POS5/1
11.1.1.1/32         Direct 0    0      127.0.0.1       InLoop0
11.1.1.2/32         Direct 0    0      11.1.1.2        POS5/1
20.1.1.0/24         ISIS   15   74     11.1.1.2        POS5/1
21.1.1.0/24         ISIS   15   74     11.1.1.2        POS5/1
21.1.1.2/32         ISIS   15   74     11.1.1.2        POS5/1
127.0.0.0/8         Direct 0    0      127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0      127.0.0.1       InLoop0
```

在 PE 3 和 PE 4 上执行 **display ip routing-table** 命令，可以看到公网路由表中有二级运营商网络的内部路由。以 PE 3 为例：

```
[PE3] display ip routing-table
Routing Tables: Public
         Destinations : 11       Routes : 11
Destination/Mask    Proto  Pre  Cost   NextHop         Interface
1.1.1.9/32          Direct 0    0      127.0.0.1       InLoop0
2.2.2.9/32          ISIS   15   10     10.1.1.2        POS5/1
5.5.5.9/32          ISIS   15   84     10.1.1.2        POS5/1
6.6.6.9/32          ISIS   15   84     10.1.1.2        POS5/1
10.1.1.0/24         Direct 0    0      10.1.1.1        POS5/1
10.1.1.1/32         Direct 0    0      127.0.0.1       InLoop0
10.1.1.2/32         Direct 0    0      10.1.1.2        POS5/1
11.1.1.0/24         ISIS   15   20     10.1.1.2        POS5/1
20.1.1.0/24         ISIS   15   84     10.1.1.2        POS5/1
21.1.1.0/24         ISIS   15   84     10.1.1.2        POS5/1
21.1.1.2/32         ISIS   15   84     10.1.1.2        POS5/1
127.0.0.0/8         Direct 0    0      127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0      127.0.0.1       InLoop0
```

在 PE 3 和 PE 4 上执行 **display ip routing-table vpn-instance** 命令，可以看到 VPN 路由表中有远端 VPN 客户的路由。以 PE 3 为例：

```
[PE3] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 3        Routes : 3
Destination/Mask    Proto  Pre  Cost    NextHop         Interface
100.1.1.0/24        Direct 0    0       100.1.1.2       GE1/0
100.1.1.2/32        Direct 0    0       127.0.0.1       InLoop0
120.1.1.0/24        BGP    255  0       6.6.6.9         NULL0
```

PE 3 和 PE 4 可以相互 Ping 通：

```
[PE3] ping 20.1.1.2
  PING 20.1.1.2: 56  data bytes, press CTRL_C to break
    Reply from 20.1.1.2: bytes=56 Sequence=1 ttl=252 time=127 ms
    Reply from 20.1.1.2: bytes=56 Sequence=2 ttl=252 time=97 ms
    Reply from 20.1.1.2: bytes=56 Sequence=3 ttl=252 time=83 ms
    Reply from 20.1.1.2: bytes=56 Sequence=4 ttl=252 time=70 ms
    Reply from 20.1.1.2: bytes=56 Sequence=5 ttl=252 time=60 ms

  --- 20.1.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 60/87/127 ms
```

CE 3 和 CE 4 可以互相 Ping 通。

```
[CE3] ping 120.1.1.1
  PING 120.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 120.1.1.1: bytes=56 Sequence=1 ttl=252 time=102 ms
    Reply from 120.1.1.1: bytes=56 Sequence=2 ttl=252 time=69 ms
    Reply from 120.1.1.1: bytes=56 Sequence=3 ttl=252 time=105 ms
    Reply from 120.1.1.1: bytes=56 Sequence=4 ttl=252 time=88 ms
    Reply from 120.1.1.1: bytes=56 Sequence=5 ttl=252 time=87 ms

  --- 120.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 69/90/105 ms
```

1.14.7 配置嵌套 VPN 示例

1. 组网需求

运营商向客户提供嵌套VPN服务。如 图 1-25所示：

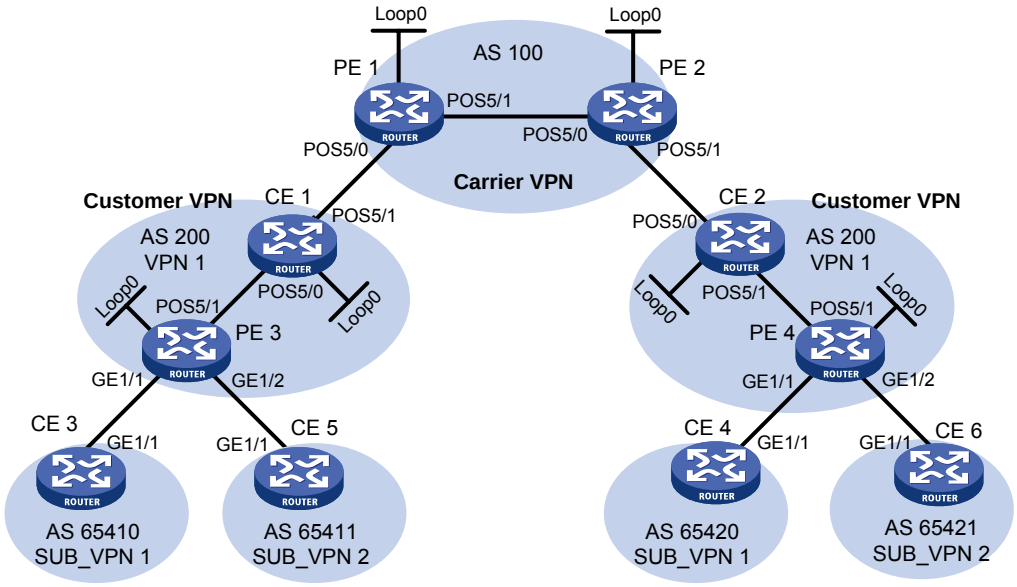
- PE 1 和 PE 2 是运营商骨干网的 PE 设备，支持嵌套 VPN 功能；
- CE 1 和 CE 2 是 VPN 用户的 CE 设备，接入运营商的骨干网，该 CE 设备支持发送 VPNv4 路由；
- PE 3 和 PE 4 是用户网络内部的 PE 设备，支持 MPLS L3VPN；
- CE 3、CE 4、CE 5 和 CE 6 是用户网络内部的子 VPN 的 CE 设备；

配置嵌套 VPN 的关键在于理解子 VPN 路由在运营商 PE 设备上的处理过程：

- 运营商 PE（PE 1 和 PE 2）收到用户 CE（CE 1 和 CE 2）发送来的 VPNv4 路由时，需要将该 VPNv4 路由的 RD 更换为用户 CE 所处 VPN 的 RD，同时将用户 CE 所处 VPN 的 ERT 添加到路由的扩展团体属性列表中，然后再按照一般的 VPNv4 路由发送出去；
- 为了实现用户网络内部子 VPN 的路由在用户 PE 和运营商 PE 间交换，需要在运营商 PE 和用户 CE 间建立 MP-EBGP 对等体关系。

2. 组网图

图1-25 嵌套 VPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	Loop0	2.2.2.9/32	CE 2	Loop0	5.5.5.9/32
	POS5/0	10.1.1.2/24		POS5/0	21.1.1.2/24
	POS5/1	11.1.1.1/24		POS5/1	20.1.1.1/24
CE 3	GE1/1	100.1.1.1/24	CE 4	GE1/1	120.1.1.1/24
CE 5	GE1/1	110.1.1.1/24	CE 6	GE1/1	130.1.1.1/24
PE 1	Loop0	3.3.3.9/32	PE 2	Loop0	4.4.4.9/32
	POS5/0	11.1.1.2/24		POS5/0	30.1.1.2/24
	POS5/1	30.1.1.1/24		POS5/1	21.1.1.1/24
PE 3	Loop0	1.1.1.9/32	PE 4	Loop0	6.6.6.9/32
	GE1/1	100.1.1.2/24		GE1/1	120.1.1.2/24
	GE1/2	110.1.1.2/24		GE1/2	130.1.1.2/24
	POS5/1	10.1.1.1/24		POS5/1	20.1.1.2/24

3. 配置步骤

(1) 配置运营商骨干网的 MPLS L3VPN，使用 IS-IS 作为骨干网的 IGP 协议，PE 1 和 PE 2 之间使能 LDP，并建立 MP-IBGP 对等体关系

配置 PE 1。

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 3.3.3.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 3.3.3.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE1-isis-1] quit
[PE1] interface loopback 0
[PE1-LoopBack0] isis enable 1
```

```
[PE1-LoopBack0] quit
[PE1] interface pos 5/1
[PE1-POS5/1] ip address 30.1.1.1 24
[PE1-POS5/1] isis enable 1
[PE1-POS5/1] mpls
[PE1-POS5/1] mpls ldp
[PE1-POS5/1] mpls ldp transport-address interface
[PE1-POS5/1] quit
[PE1] bgp 100
[PE1-bgp] peer 4.4.4.9 as-number 100
[PE1-bgp] peer 4.4.4.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 4.4.4.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```



说明

PE 2 的配置与 PE 1 类似，配置过程略。

配置完成后，在 PE 1 或 PE 2 上执行 **display mpls ldp session** 命令可以看到 LDP 会话建立成功；执行 **display bgp peer** 命令可以看到 BGP 对等体关系已建立，并达到 Established 状态；执行 **display isis peer** 命令可以看到 IS-IS 邻居关系已建立，状态为 up。

以 PE 1 为例：

```
[PE1] display mpls ldp session
               LDP Session(s) in Public Network
 ------------------------------------------------------------------
 Peer-ID        Status        LAM  SsnRole  FT   MD5  KA-Sent/Rcv
 ------------------------------------------------------------------
 4.4.4.9:0      Operational   DU   Active   Off  Off  378/378
 ------------------------------------------------------------------
 LAM : Label Advertisement Mode         FT  : Fault Tolerance
[PE1] display bgp peer
 BGP local router ID : 3.3.3.9
 Local AS number : 100
 Total number of peers : 1          Peers in established state : 1
  Peer          AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
  4.4.4.9      100      162      145     0        0  02:12:47 Established
[PE1] display isis peer
                         Peer information for ISIS(1)
                         ---------------------------
  System Id    Interface   Circuit Id    State HoldTime Type   PRI
0000.0000.0005 POS5/1           001       Up   29s      L1L2   --
```

(2) 配置用户网络：使用 IS-IS 作为 IGP 协议，PE 3 和 CE 1、PE 4 和 CE 2 之间分别使能 LDP

配置 PE 3。

```
<PE3> system-view
```

```
[PE3] interface loopback 0
[PE3-LoopBack0] ip address 1.1.1.9 32
[PE3-LoopBack0] quit
[PE3] mpls lsr-id 1.1.1.9
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
[PE3] isis 2
[PE3-isis-2] network-entity 10.0000.0000.0000.0001.00
[PE3-isis-2] quit
[PE3] interface loopback 0
[PE3-LoopBack0] isis enable 2
[PE3-LoopBack0] quit
[PE3] interface pos 5/1
[PE3-POS5/1] ip address 10.1.1.1 24
[PE3-POS5/1] isis enable 2
[PE3-POS5/1] mpls
[PE3-POS5/1] mpls ldp
[PE3-POS5/1] quit
```

配置 CE 1。

```
<CE1> system-view
[CE1] interface loopback 0
[CE1-LoopBack0] ip address 2.2.2.9 32
[CE1-LoopBack0] quit
[CE1] mpls lsr-id 2.2.2.9
[CE1] mpls
[CE1-mpls] quit
[CE1] mpls ldp
[CE1-mpls-ldp] quit
[CE1] isis 2
[CE1-isis-2] network-entity 10.0000.0000.0000.0002.00
[CE1-isis-2] quit
[CE1] interface loopback 0
[CE1-LoopBack0] isis enable 2
[CE1-LoopBack0] quit
[CE1] interface pos 5/0
[CE1-POS5/0] ip address 10.1.1.2 24
[CE1-POS5/0] isis enable 2
[CE1-POS5/0] mpls
[CE1-POS5/0] mpls ldp
[CE1-POS5/0] quit
```

配置完成后，PE 3 和 CE 1 之间可以建立 LDP 和 IS-IS 邻居关系。

 说明

PE 4 和 CE 2 之间的配置与 PE 3 和 CE 1 之间的配置类似，配置过程略。

---

(3) 配置用户 CE 接入到运营商的 PE

配置 PE 1。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 200:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] interface pos 5/0
[PE1-POS5/0] ip binding vpn-instance vpn1
[PE1-POS5/0] ip address 11.1.1.2 24
[PE1-POS5/0] mpls
[PE1-POS5/0] quit
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 11.1.1.1 as-number 200
[PE1-bgp-vpn1] quit
[PE1-bgp] quit
```

配置 CE 1。

```
[CE1] interface pos 5/1
[CE1-POS5/1] ip address 11.1.1.1 24
[CE1-POS5/1] mpls
[CE1-POS5/1] quit
[CE1] bgp 200
[CE1-bgp] peer 11.1.1.2 as-number 100
[CE1-bgp] quit
```



说明

PE 2 和 CE 2 之间的配置与 PE 1 和 CE 1 之间的配置类似，配置过程省略。

(4) 配置子 VPN 的 CE 接入用户网络的 PE

配置 CE 3。

```
<CE3> system-view
[CE3] interface gigabitethernet 1/1
[CE3-GigabitEthernet1/1] ip address 100.1.1.1 24
[CE3-GigabitEthernet1/1] quit
[CE3] bgp 65410
[CE3-bgp] peer 100.1.1.2 as-number 200
[CE3-bgp] import-route direct
[CE3-bgp] quit
```

配置 CE 5。

```
<CE5> system-view
[CE5] interface gigabitethernet 1/1
[CE5-GigabitEthernet1/1] ip address 110.1.1.1 24
[CE5-GigabitEthernet1/1] quit
[CE5] bgp 65411
```

```
[CE5-bgp] peer 110.1.1.2 as-number 200
[CE5-bgp] import-route direct
[CE5-bgp] quit
```

配置 PE 3。

```
[PE3] ip vpn-instance SUB_VPN1
[PE3-vpn-instance-SUB_VPN1] route-distinguisher 100:1
[PE3-vpn-instance-SUB_VPN1] vpn-target 2:1
[PE3-vpn-instance-SUB_VPN1] quit
[PE3] interface gigabitethernet 1/1
[PE3-GigabitEthernet1/1] ip binding vpn-instance SUB_VPN1
[PE3-GigabitEthernet1/1] ip address 100.1.1.2 24
[PE3-GigabitEthernet1/1] quit
[PE3] ip vpn-instance SUB_VPN2
[PE3-vpn-instance-SUB_VPN2] route-distinguisher 101:1
[PE3-vpn-instance-SUB_VPN2] vpn-target 2:2
[PE3-vpn-instance-SUB_VPN2] quit
[PE3] interface gigabitethernet 1/2
[PE3-GigabitEthernet1/2] ip binding vpn-instance SUB_VPN2
[PE3-GigabitEthernet1/2] ip address 110.1.1.2 24
[PE3-GigabitEthernet1/2] quit
[PE3] bgp 200
[PE3-bgp] ipv4-family vpn-instance SUB_VPN1
[PE3-bgp-SUB_VPN1] peer 100.1.1.1 as-number 65410
[PE3-bgp-SUB_VPN1] import-route direct
[PE3-bgp-SUB_VPN1] quit
[PE3-bgp] ipv4-family vpn-instance SUB_VPN2
[PE3-bgp-SUB_VPN2] peer 110.1.1.1 as-number 65411
[PE3-bgp-SUB_VPN2] import-route direct
[PE3-bgp-SUB_VPN2] quit
[PE3-bgp] quit
```



说明

PE 4 和 CE 4，CE 6 之间的配置与 PE 3 和 CE 3，CE 5 之间的配置类似，配置过程省略。

(5) 在运营商的 PE 和用户 CE 之间建立 MP-EBGP 对等体关系，交换用户的 VPNv4 路由

配置 PE 1，使能嵌套 VPN 特性。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] nesting-vpn
[PE1-bgp-af-vpnv4] peer 11.1.1.1 vpn-instance vpn1 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

配置 CE 1，使能 VPNv4 能力，使之与 PE 1 建立 VPNv4 邻居关系。

```
[CE1] bgp 200
[CE1-bgp] ipv4-family vpnv4
```

```
[CE1-bgp-af-vpnv4] peer 11.1.1.2 enable
```

配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现。

```
[CE1-bgp-af-vpnv4] peer 11.1.1.2 allow-as-loop 2
```

配置接收所有 VPNv4 路由。

```
[CE1-bgp-af-vpnv4] undo policy vpn-target
[CE1-bgp-af-vpnv4] quit
[CE1-bgp] quit
```



说明

PE 2 和 CE 2 之间的配置与 PE 1 和 CE 1 之间的配置类似，配置过程省略。

(6) 在用户网络内部的 PE 和 CE 之间建立 MP-IBGP 对等体关系，交换用户内部子 VPN 的 VPNv4 路由

配置 PE 3。

```
[PE3] bgp 200
[PE3-bgp] peer 2.2.2.9 as-number 200
[PE3-bgp] peer 2.2.2.9 connect-interface loopback 0
[PE3-bgp] ipv4-family vpnv4
[PE3-bgp-af-vpnv4] peer 2.2.2.9 enable
```

配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现。

```
[PE3-bgp-af-vpnv4] peer 2.2.2.9 allow-as-loop 2
[PE3-bgp-af-vpnv4] quit
[PE3-bgp] quit
```

配置 CE 1。

```
[CE1] bgp 200
[CE1-bgp] peer 1.1.1.9 as-number 200
[CE1-bgp] peer 1.1.1.9 connect-interface loopback 0
[CE1-bgp] ipv4-family vpnv4
[CE1-bgp-af-vpnv4] peer 1.1.1.9 enable
[CE1-bgp-af-vpnv4] undo policy vpn-target
[CE1-bgp-af-vpnv4] quit
[CE1-bgp] quit
```



说明

PE 4 和 CE 2 之间的配置与 PE 3 和 CE 1 之间的配置类似，配置过程省略。

(7) 配置完成后的检验

所有配置完成后，在 PE 1 和 PE 2 上执行 **display ip routing-table** 命令，可以看到 PE 1 和 PE 2 的公网路由表中只有运营商网络的路由。以 PE 1 为例：

```
[PE1] display ip routing-table
Routing Tables: Public
```

```
         Destinations : 7        Routes : 7

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.3.3.9/32          Direct 0    0            127.0.0.1       InLoop0
4.4.4.9/32          ISIS   15   10           30.1.1.2        POS5/1
30.1.1.0/24         Direct 0    0            30.1.1.1        POS5/1
30.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.2/32         Direct 0    0            30.1.1.2        POS5/1
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

在 PE 1 和 PE 2 上执行 **display ip routing-table vpn-instance** 命令，可以看到 VPN 路由表中有用户子 VPN 网络的路由。以 PE 1 为例：

```
[PE1] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

11.1.1.0/24         Direct 0    0            11.1.1.1        POS5/0
11.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
11.1.1.2/32         Direct 0    0            11.1.1.2        POS5/0
100.1.1.0/24        BGP    255  0            11.1.1.1        NULL0
110.1.1.0/24        BGP    255  0            11.1.1.1        NULL0
120.1.1.0/24        BGP    255  0            4.4.4.9         NULL0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
130.1.1.0/24        BGP    255  0            4.4.4.9         NULL0
```

在 CE 1 和 CE 2 上执行 **display bgp vpnv4 all routing-table** 命令，可以看到用户网络中 VPNv4 路由表中有子 VPN 网络的内部路由。以 CE 1 为例：

```
[CE1] display bgp vpnv4 all routing-table
BGP Local router ID is 11.11.11.11
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete


 Total number of routes from all PE: 4

Route Distinguisher: 100:1

    Network            NextHop         In/Out Label     MED        LocPrf

*>  100.1.1.0/24       1.1.1.9         1024/1024
```

```
Route Distinguisher: 101:1

    Network            NextHop         In/Out Label     MED        LocPrf

*>  110.1.1.0/24       1.1.1.9         1025/1025

 Route Distinguisher: 200:1

    Network            NextHop         In/Out Label     MED        LocPrf

*>  120.1.1.0/24       11.1.1.2        1026/1027

Route Distinguisher: 201:1

    Network            NextHop         In/Out Label     MED        LocPrf

*>  130.1.1.0/24       11.1.1.2        1027/1028
```

在 PE 3 和 PE 4 上执行 **display ip routing-table vpn-instance SUB_VPN1** 命令，可以看到私网路由表中有从运营商 PE 发布到用户网络子 VPN 内部的路由。以 PE 3 为例：

```
[PE3] display ip routing-table vpn-instance SUB_VPN1
Routing Tables: SUB_VPN1
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

100.1.1.0/24        Direct 0    0            100.1.1.2       GE1/1
100.1.1.2/32        Direct 0    0            127.0.0.1       InLoop0
120.1.1.0/24        BGP    255  0            2.2.2.9         NULL0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

在 CE 3 和 CE 4 上执行 **display ip routing-table** 命令，可以看到路由表中有远端子 VPN 的路由。以 CE 3 为例：

```
[CE3] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

100.1.1.0/24        Direct 0    0            100.1.1.1       GE1/1
100.1.1.1/32        Direct 0    0            127.0.0.1       InLoop0
120.1.1.0/24        BGP    255  0            100.1.1.2       GEh1/1
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

在 CE 5 和 CE 6 上执行 **display ip routing-table** 命令，可以看到路由表中有远端子 VPN 的路由。以 CE 5 为例：

```
[CE5] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

110.1.1.0/24        Direct 0    0            110.1.1.1       GE1/1
110.1.1.1/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
130.1.1.0/24        BGP    255  0            110.1.1.2       GE1/1
```

CE 3 和 CE 4 可以互相 Ping 通。

```
[CE3] ping 120.1.1.1
  PING 120.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 120.1.1.1: bytes=56 Sequence=1 ttl=252 time=102 ms
    Reply from 120.1.1.1: bytes=56 Sequence=2 ttl=252 time=69 ms
    Reply from 120.1.1.1: bytes=56 Sequence=3 ttl=252 time=105 ms
    Reply from 120.1.1.1: bytes=56 Sequence=4 ttl=252 time=88 ms
    Reply from 120.1.1.1: bytes=56 Sequence=5 ttl=252 time=87 ms

  --- 120.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 69/90/105 ms
```

CE 5 和 CE 6 可以互相 Ping 通。

```
[CE5] ping 130.1.1.1
  PING 130.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 130.1.1.1: bytes=56 Sequence=1 ttl=252 time=102 ms
    Reply from 130.1.1.1: bytes=56 Sequence=2 ttl=252 time=69 ms
    Reply from 130.1.1.1: bytes=56 Sequence=3 ttl=252 time=105 ms
    Reply from 130.1.1.1: bytes=56 Sequence=4 ttl=252 time=88 ms
    Reply from 130.1.1.1: bytes=56 Sequence=5 ttl=252 time=87 ms

  --- 130.1.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 69/90/105 ms
```

CE 3 和 CE 6 不能互相 Ping 通。

```
[CE3] ping 130.1.1.1
  PING 130.1.1.1: 56  data bytes, press CTRL_C to break
    Request time out
```

```
  Request time out
  Request time out
  Request time out
  Request time out

--- 130.1.1.1 ping statistics ---
  5 packet(s) transmitted
  0 packet(s) received
  100.00% packet loss
```

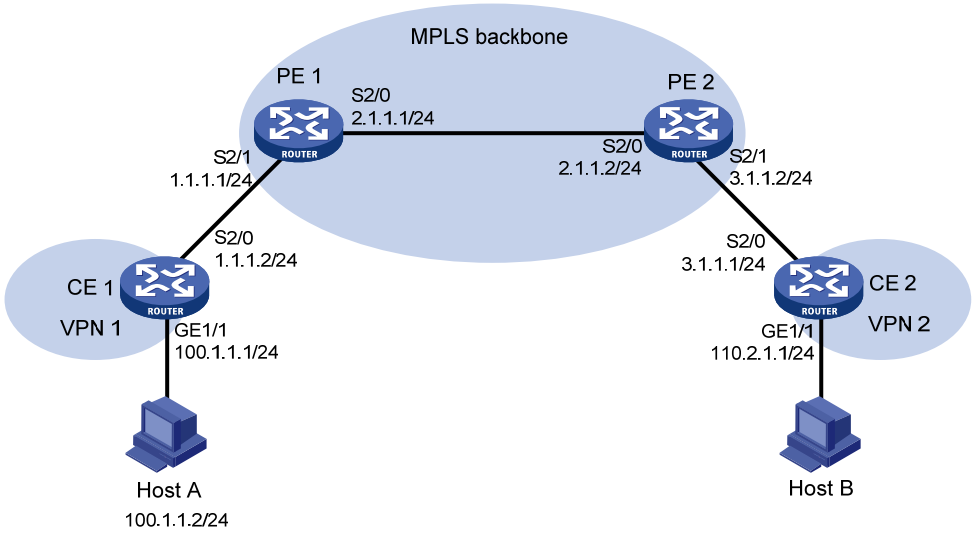
1.14.8 配置多角色主机示例

1. 组网需求

- 主机 Host A 通过 CE 1 接入，其 IP 地址为 100.1.1.2。Host A 可以访问 VPN 1 和 VPN 2。
- 将 PE 1 与 CE 1 相连的 Serial2/1 接口绑定 VPN 1，将 PE2 与 CE2 相连的接口 Serial2/1 绑定到 VPN 2。

2. 组网图

图1-26 配置多角色主机组网图



3. 配置步骤

(1) 配置 CE 1

配置 CE 1 的接口 IP 地址。

```
<CE1> system-view
[CE1] interface gigabitethernet 1/1
[CE1-GigabitEthernet1/1] ip address 100.1.1.1 24
[CE1-GigabitEthernet1/1] quit
[CE1] interface serial 2/0
[CE1-Serial2/0] ip address 1.1.1.2 24
[CE1-Serial2/0] quit
```

在 CE 1 上配置一条指向 PE1 的缺省路由。

```
[CE1] ip route-static 0.0.0.0 0 1.1.1.1
```

(2) 配置 PE 1

在 PE 1 上为 VPN 1 和 VPN 2 分别创建 VPN 实例，并配置 RD 和不同的 VPN Target 属性。

```
<PE1> system-view
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 100:1 both
[PE1-vpn-instance-vpn1] quit
[PE1] ip vpn-instance vpn2
[PE1-vpn-instance-vpn2] route-distinguisher 100:2
[PE1-vpn-instance-vpn2] vpn-target 100:2 both
[PE1-vpn-instance-vpn2] quit
```

将 PE 1 与 CE 1 相连的接口绑定到 VPN 1。

```
[PE1] interface serial 2/1
[PE1-Serial2/1] ip binding vpn-instance vpn1
[PE1-Serial2/1] ip address 1.1.1.1 255.255.255.0
[PE1-Serial2/1] quit
```

配置静态路由，并引入到 BGP 中，使 Host A 访问 VPN 2 的返回报文能够在 PE 1 的 VPN 实例 vpn1 中找到正确的路由，回到 Host A。

```
[PE1] ip route-static vpn-instance vpn2 100.1.1.0 24 vpn-instance vpn1 1.1.1.2
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn2
[PE1-bgp-vpn2] import-route static
[PE1-bgp-vpn2] quit
[PE1-bgp] quit
```

配置策略路由，对于 Host A 发出的报文，如果在本接口所属的 VPN-instance 中没有找到路由，就在名为 vpn2 的 VPN-instance 中查找私网路由并转发。

```
[PE1] acl number 3001
[PE1-acl-adv-3001] rule 0 permit ip vpn-instance vpn1 source 100.1.1.2 0
[PE1-acl-adv-3001] quit
[PE1] policy-based-route policy1 permit node 10
[PE1-policy-based-route] if-match acl 3001
[PE1-policy-based-route] apply access-vpn vpn-instance vpn1 vpn2
[PE1-policy-based-route] quit
```

在接口 Serial2/1 上应用定义的策略路由。

```
[PE1] interface serial2/1
[PE1-Serial2/1] ip policy-based-route policy1
```

其他配置按照基本的 MPLS L3VPN 配置即可。

1.14.9 配置 HoVPN 示例

1. 组网需求

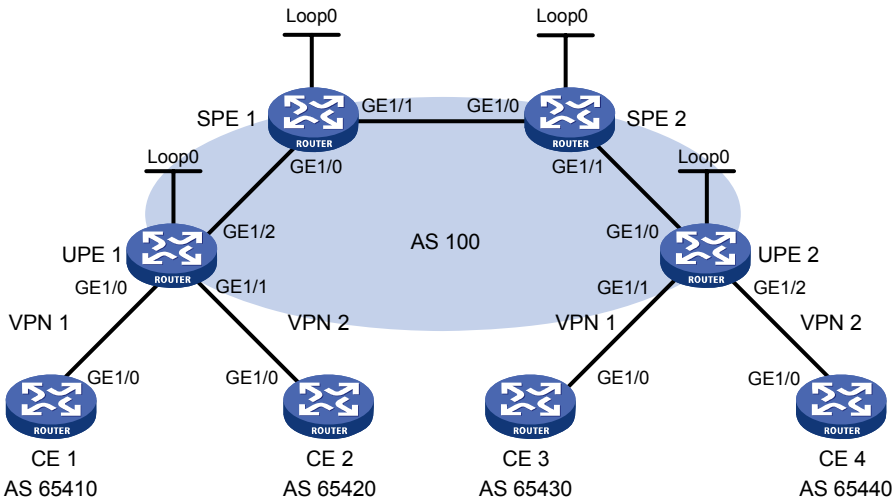
以一个包括省骨干和地市的 MPLS VPN 网络为例：

- SPE 作为省网的 PE 设备，接入地市的 MPLS VPN 网络；

- UPE 作为下层地市网络的 PE 设备，最终接入 VPN 客户。对 UPE 的性能要求低于对 SPE 的性能要求。
- SPE 将通过路由策略的路由发送给 UPE，限制不同 Site 之间的互相访问权限，使得 VPN 1 内的 CE 1 和 CE 3 可以互相访问，VPN 2 内的 CE 2 和 CE 4 不能互相访问。

2. 组网图

图1-27 配置 HoVPN 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.2.1.1/24	CE 3	GE1/0	10.1.1.1/24
CE 2	GE1/0	10.4.1.1/24	CE 4	GE1/0	10.3.1.1/24
UPE 1	Loop0	1.1.1.9/32	UPE 2	Loop0	4.4.4.9/32
	GE1/0	10.2.1.2/24		GE1/0	172.2.1.1/24
	GE1/1	10.4.1.2/24		GE1/1	10.1.1.2/24
	GE1/2	172.1.1.1/24		GE1/2	10.3.1.2/24
SPE 1	Loop0	2.2.2.9/32	SPE 2	Loop0	3.3.3.9/32
	GE1/0	172.1.1.2/24		GE1/0	180.1.1.2/24
	GE1/1	180.1.1.1/24		GE1/1	172.2.1.2/24

3. 配置步骤

(1) 配置 UPE 1

配置 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<UPE1> system-view
[UPE1] interface loopback 0
[UPE1-LoopBack0] ip address 1.1.1.9 32
[UPE1-LoopBack0] quit
[UPE1] mpls lsr-id 1.1.1.9
[UPE1] mpls
[UPE1-mpls] quit
[UPE1] mpls ldp
[UPE1-mpls-ldp] quit
[UPE1] interface gigabitethernet 1/2
[UPE1-GigabitEthernet1/2] ip address 172.1.1.1 24
[UPE1-GigabitEthernet1/2] mpls
[UPE1-GigabitEthernet1/2] mpls ldp
[UPE1-GigabitEthernet1/2] quit
```

配置 IGP 协议，以 OSPF 为例。

```
[UPE1] ospf
[UPE1-ospf-1] area 0
[UPE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[UPE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[UPE1-ospf-1-area-0.0.0.0] quit
[UPE1-ospf-1] quit
```

配置 VPN 实例 vpn1 和 vpn2，将 CE 1 和 CE 2 接入 UPE 1。

```
[UPE1] ip vpn-instance vpn1
[UPE1-vpn-instance-vpn1] route-distinguisher 100:1
[UPE1-vpn-instance-vpn1] vpn-target 100:1 both
[UPE1-vpn-instance-vpn1] quit
[UPE1] ip vpn-instance vpn2
[UPE1-vpn-instance-vpn2] route-distinguisher 100:2
[UPE1-vpn-instance-vpn2] vpn-target 100:2 both
[UPE1-vpn-instance-vpn2] quit
[UPE1] interface gigabitethernet 1/0
[UPE1-GigabitEthernet1/0] ip binding vpn-instance vpn1
[UPE1-GigabitEthernet1/0] ip address 10.2.1.2 24
[UPE1-GigabitEthernet1/0] quit
[UPE1] interface gigabitethernet 1/1
[UPE1-GigabitEthernet1/1] ip binding vpn-instance vpn2
[UPE1-GigabitEthernet1/1] ip address 10.4.1.2 24
[UPE1-GigabitEthernet1/1] quit
```

配置 UPE 1 与 SPE 1 建立 MP-IBGP 对等体，并引入 VPN 路由。

```
[UPE1] bgp 100
[UPE1-bgp] peer 2.2.2.9 as-number 100
[UPE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[UPE1-bgp] ipv4-family vpnv4
[UPE1-bgp-af-vpnv4] peer 2.2.2.9 enable
[UPE1-bgp-af-vpnv4] quit
[UPE1-bgp] ipv4-family vpn-instance vpn1
[UPE1-bgp-vpn1] peer 10.2.1.1 as-number 65410
[UPE1-bgp-vpn1] import-route direct
[UPE1-bgp-vpn1] quit
[UPE1-bgp] ipv4-family vpn-instance vpn2
[UPE1-bgp-vpn1] peer 10.4.1.1 as-number 65420
[UPE1-bgp-vpn1] import-route direct
[UPE1-bgp-vpn1] quit
[UPE1-bgp] quit
```

(2) 配置 CE 1

```
<CE1> system-view
[CE1] interface gigabitethernet 1/0
[CE1-GigabitEthernet1/0] ip address 10.2.1.1 255.255.255.0
[CE1-GigabitEthernet1/0] quit
[CE1] bgp 65410
```

```
[CE1-bgp] peer 10.2.1.2 as-number 100
[CE1-bgp] import-route direct
[CE1] quit
```

(3) 配置 CE 2

```
<CE2> system-view
[CE2] interface gigabitethernet 1/0
[CE2-GigabitEthernet1/0] ip address 10.4.1.1 255.255.255.0
[CE2-GigabitEthernet1/0] quit
[CE2] bgp 65420
[CE2-bgp] peer 10.4.1.2 as-number 100
[CE2-bgp] import-route direct
[CE2] quit
```

(4) 配置 UPE 2

配置 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<UPE2> system-view
[UPE2] interface loopback 0
[UPE2-LoopBack0] ip address 4.4.4.9 32
[UPE2-LoopBack0] quit
[UPE2] mpls lsr-id 4.4.4.9
[UPE2] mpls
[UPE2-mpls] quit
[UPE2] mpls ldp
[UPE2-mpls-ldp] quit
[UPE2] interface gigabitethernet 1/0
[UPE2-GigabitEthernet1/0] ip address 172.2.1.1 24
[UPE2-GigabitEthernet1/0] mpls
[UPE2-GigabitEthernet1/0] mpls ldp
[UPE2-GigabitEthernet1/0] quit
```

配置 IGP 协议，以 OSPF 为例。

```
[UPE2] ospf
[UPE2-ospf-1] area 0
[UPE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[UPE2-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[UPE2-ospf-1-area-0.0.0.0] quit
[UPE2-ospf-1] quit
```

配置 VPN 实例 vpn1 和 vpn2，将 CE 3 和 CE 4 接入 UPE 2。

```
[UPE2] ip vpn-instance vpn1
[UPE2-vpn-instance-vpn1] route-distinguisher 300:1
[UPE2-vpn-instance-vpn1] vpn-target 100:1 both
[UPE2-vpn-instance-vpn1] quit
[UPE2] ip vpn-instance vpn2
[UPE2-vpn-instance-vpn2] route-distinguisher 400:2
[UPE2-vpn-instance-vpn2] vpn-target 100:2 both
[UPE2-vpn-instance-vpn2] quit
[UPE2] interface gigabitethernet 1/1
[UPE2-GigabitEthernet1/1] ip binding vpn-instance vpn1
```

```
[UPE2-GigabitEthernet1/1] ip address 10.1.1.2 24
[UPE2-GigabitEthernet1/1] quit
[UPE2] interface gigabitethernet 1/2
[UPE2-GigabitEthernet1/2] ip binding vpn-instance vpn2
[UPE2-GigabitEthernet1/2] ip address 10.3.1.2 24
[UPE2-GigabitEthernet1/2] quit
```

配置 UPE 2 与 SPE 2 建立 MP-IBGP 对等体，并引入 VPN 路由。

```
[UPE2] bgp 100
[UPE2-bgp] peer 3.3.3.9 as-number 100
[UPE2-bgp] peer 3.3.3.9 connect-interface loopback 0
[UPE2-bgp] ipv4-family vpnv4
[UPE2-bgp-af-vpnv4] peer 3.3.3.9 enable
[UPE2-bgp-af-vpnv4] quit
[UPE2-bgp] ipv4-family vpn-instance vpn1
[UPE2-bgp-vpn1] peer 10.1.1.1 as-number 65430
[UPE2-bgp-vpn1] import-route direct
[UPE2-bgp-vpn1] quit
[UPE2-bgp] ipv4-family vpn-instance vpn2
[UPE2-bgp-vpn1] peer 10.3.1.1 as-number 65440
[UPE2-bgp-vpn1] import-route direct
[UPE2-bgp-vpn1] quit
[UPE2-bgp] quit
```

(5) 配置 CE 3

```
<CE3> system-view
[CE3] interface gigabitethernet 1/0
[CE3-GigabitEthernet1/0] ip address 10.1.1.1 255.255.255.0
[CE3-GigabitEthernet1/0] quit
[CE3] bgp 65430
[CE3-bgp] peer 10.1.1.2 as-number 100
[CE3-bgp] import-route direct
[CE3] quit
```

(6) 配置 CE 4

```
<CE4> system-view
[CE4] interface gigabitethernet 1/0
[CE4-GigabitEthernet1/0] ip address 10.3.1.1 255.255.255.0
[CE4-GigabitEthernet1/0] quit
[CE4] bgp 65440
[CE4-bgp] peer 10.3.1.2 as-number 100
[CE4-bgp] import-route direct
[CE4] quit
```

(7) 配置 SPE 1

配置 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<SPE1> system-view
[SPE1] interface loopback 0
[SPE1-LoopBack0] ip address 2.2.2.9 32
[SPE1-LoopBack0] quit
```

```
[SPE1] mpls lsr-id 2.2.2.9
[SPE1] mpls
[SPE1-mpls] quit
[SPE1] mpls ldp
[SPE1-mpls-ldp] quit
[SPE1] interface gigabitethernet 1/0
[SPE1-GigabitEthernet1/0] ip address 172.1.1.2 24
[SPE1-GigabitEthernet1/0] mpls
[SPE1-GigabitEthernet1/0] mpls ldp
[SPE1-GigabitEthernet1/0] quit
[SPE1] interface gigabitethernet 1/1
[SPE1-GigabitEthernet1/1] ip address 180.1.1.1 24
[SPE1-GigabitEthernet1/1] mpls
[SPE1-GigabitEthernet1/1] mpls ldp
[SPE1-GigabitEthernet1/1] quit
```

配置 IGP 协议，以 OSPF 为例。

```
[SPE1] ospf
[SPE1-ospf-1] area 0
[SPE1-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[SPE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[SPE1-ospf-1-area-0.0.0.0] network 180.1.1.0 0.0.0.255
[SPE1-ospf-1-area-0.0.0.0] quit
[SPE1-ospf-1] quit
```

配置 VPN 实例 vpn1 和 vpn2。

```
[SPE1] ip vpn-instance vpn1
[SPE1-vpn-instance-vpn1] route-distinguisher 500:1
[SPE1-vpn-instance-vpn1 ] vpn-target 100:1 both
[SPE1-vpn-instance-vpn1] quit
[SPE1] ip vpn-instance vpn2
[SPE1-vpn-instance-vpn2] route-distinguisher 700:1
[SPE1-vpn-instance-vpn2] vpn-target 100:2 both
[SPE1-vpn-instance-vpn2] quit
```

配置 SPE 1 与 UPE 1 建立 MP-IBGP 对等体，指定 UPE 1，并引入 VPN 路由。

```
[SPE1] bgp 100
[SPE1-bgp] peer 1.1.1.9 as-number 100
[SPE1-bgp] peer 1.1.1.9 connect-interface loopback 0
[SPE1-bgp] peer 1.1.1.9 next-hop-local
[SPE1-bgp] peer 3.3.3.9 as-number 100
[SPE1-bgp] peer 3.3.3.9 connect-interface loopback 0
[SPE1-bgp] ipv4-family vpnv4
[SPE1-bgp-af-vpnv4] peer 3.3.3.9 enable
[SPE1-bgp-af-vpnv4] peer 1.1.1.9 enable
[SPE1-bgp-af-vpnv4] peer 1.1.1.9 upe
 [SPE1-bgp-af-vpnv4] quit
[SPE1-bgp]ipv4-family vpn-instance vpn1
[SPE1-bgp-vpn1] quit
```

```
[SPE1-bgp]ipv4-family vpn-instance vpn2
[SPE1-bgp-vpn2] quit
[SPE1-bgp] quit
```

配置 SPE 1 向 UPE 1 发送通过策略的路由信息，允许 CE 3 的路由发送给 UPE 1。

```
[SPE1] ip ip-prefix hope index 10 permit 10.1.1.1 24
[SPE1] route-policy hope permit node 0
[SPE1-route-policy] if-match ip-prefix hope
[SPE1-route-policy] quit
[SPE1] bgp 100
[SPE1-bgp] ipv4-family vpnv4
[SPE1-bgp-af-vpnv4] peer 1.1.1.9 upe route-policy hope export
```

(8) 配置 SPE2

配置 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<SPE2> system-view
[SPE2] interface loopback 0
[SPE2-LoopBack0] ip address 3.3.3.9 32
[SPE2-LoopBack0] quit
[SPE2] mpls lsr-id 3.3.3.9
[SPE2] mpls
[SPE2-mpls] quit
[SPE2] mpls ldp
[SPE2-mpls-ldp] quit
[SPE2] interface gigabitethernet 1/0
[SPE2-GigabitEthernet1/0] ip address 180.1.1.2 24
[SPE2-GigabitEthernet1/0] mpls
[SPE2-GigabitEthernet1/0] mpls ldp
[SPE2-GigabitEthernet1/0] quit
[SPE2] interface gigabitethernet 1/1
[SPE2-GigabitEthernet1/1] ip address 172.2.1.2 24
[SPE2-GigabitEthernet1/1] mpls
[SPE2-GigabitEthernet1/1] mpls ldp
[SPE2-GigabitEthernet1/1] quit
```

配置 IGP 协议，以 OSPF 为例。

```
[SPE2] ospf
[SPE2-ospf-1] area 0
[SPE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[SPE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[SPE2-ospf-1-area-0.0.0.0] network 180.1.1.0 0.0.0.255
[SPE2-ospf-1-area-0.0.0.0] quit
[SPE2-ospf-1] quit
```

配置 VPN 实例 vpn1 和 vpn2。

```
[SPE2] ip vpn-instance vpn1
[SPE2-vpn-instance-vpn1] route-distinguisher 600:1
[SPE2-vpn-instance-vpn1 ] vpn-target 100:1 both
[SPE2-vpn-instance-vpn1] quit
[SPE2] ip vpn-instance vpn2
```

```
[SPE2-vpn-instance-vpn2] route-distinguisher 800:1
[SPE2-vpn-instance-vpn2] vpn-target 100:2 both
[SPE2-vpn-instance-vpn2] quit
```

配置 SPE 2 与 UPE 2 建立 MP-IBGP 对等体，指定 UPE 2，并引入 VPN 路由。

```
[SPE2] bgp 100
[SPE2-bgp] peer 4.4.4.9 as-number 100
[SPE2-bgp] peer 4.4.4.9 connect-interface loopback 0
[SPE2-bgp] peer 4.4.4.9 next-hop-local
[SPE2-bgp] peer 2.2.2.9 as-number 100
[SPE2-bgp] peer 2.2.2.9 connect-interface loopback 0
[SPE2-bgp] ipv4-family vpnv4
[SPE2-bgp-af-vpnv4] peer 2.2.2.9 enable
[SPE2-bgp-af-vpnv4] peer 4.4.4.9 enable
[SPE2-bgp-af-vpnv4] peer 4.4.4.9 upe
 [SPE2-bgp-af-vpnv4] quit
[SPE2-bgp]ipv4-family vpn-instance vpn1
[SPE2-bgp-vpn1] quit
[SPE2-bgp]ipv4-family vpn-instance vpn2
[SPE2-bgp-vpn2] quit
[SPE2-bgp] quit
```

配置 SPE 2 向 UPE 2 发送通过策略的路由信息，允许 CE 1 的路由发送给 UPE 2。

```
[SPE2] ip ip-prefix hope index 10 permit 10.2.1.1 24
[SPE2] route-policy hope permit node 0
[SPE2-route-policy] if-match ip-prefix hope
[SPE2-route-policy] quit
[SPE2] bgp 100
[SPE2-bgp] ipv4-family vpnv4
[SPE2-bgp-af-vpnv4] peer 4.4.4.9 upe route-policy hope export
```

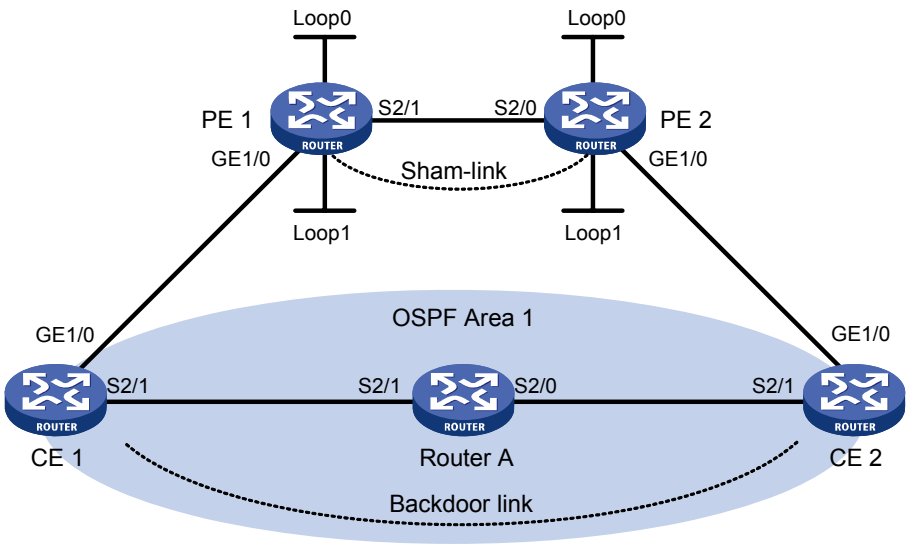
1.14.10 配置 OSPF 伪连接

1. 组网需求

- CE 1 和 CE 2 都属于 VPN 1，它们分别接入 PE 1 和 PE 2；
- CE 1 和 CE 2 在同一个 OSPF 区域中；
- CE 1 与 CE 2 之间的 VPN 流量通过 MPLS 骨干网转发，不使用 OSPF 的区域内路由。

2. 组网图

图1-28 OSPF 伪连接配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	100.1.1.1/24	CE 2	GE1/0	120.1.1.1/24
	S2/1	20.1.1.1/24		S2/1	30.1.1.2/24
PE 1	Loop0	1.1.1.9/32	PE 2	Loop0	2.2.2.9/32
	Loop1	3.3.3.3/32		Loop1	5.5.5.5/32
	GE1/0	100.1.1.2/24		GE1/0	120.1.1.2/24
	S2/1	10.1.1.1/24		S2/0	10.1.1.2/24
Router A	S2/0	30.1.1.1/24			
	S2/1	20.1.1.2/24			

3. 配置步骤

(1) 配置用户网络上的 OSPF

在CE 1、Router A、CE 2 上配置普通OSPF，发布 图 1-28中所示各接口的网段地址。具体配置过程略。

配置完成后，CE 1 和 CE 2 应学到到对端 GigabitEthernet 接口的 OSPF 路由。

以 CE 1 为例：

```
<CE1> display ip routing-table
Routing Tables: Public
         Destinations : 9        Routes : 9
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
20.1.1.0/24       Direct 0    0         20.1.1.1        S2/1
20.1.1.1/32       Direct 0    0         127.0.0.1       InLoop0
20.1.1.2/32       Direct 0    0         20.1.1.2        S2/1
30.1.1.0/24       OSPF   10   3124      20.1.1.2        S2/1
100.1.1.0/24      Direct 0    0         100.1.1.1       GE1/0
100.1.1.1/32      Direct 0    0         127.0.0.1       InLoop0
120.1.1.0/24      OSPF   10   3125      20.1.1.2        S2/1
127.0.0.0/8       Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32      Direct 0    0         127.0.0.1       InLoop0
```

(2) 在骨干网上配置 MPLS L3VPN

配置 PE 1 的 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 1.1.1.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface serial 2/1
[PE1-Serial2/1] ip address 10.1.1.1 24
[PE1-Serial2/1] mpls
[PE1-Serial2/1] mpls ldp
[PE1-Serial2/1] quit
```

配置 PE 1 的 MP-IBGP 对等体为 PE2。

```
[PE1] bgp 100
[PE1-bgp] peer 2.2.2.9 as-number 100
[PE1-bgp] peer 2.2.2.9 connect-interface loopback 0
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 2.2.2.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

配置 PE 1 的 OSPF。

```
[PE1]ospf 1
[PE1-ospf-1]area 0
[PE1-ospf-1-area-0.0.0.0]network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0]network 10.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0]quit
[PE1-ospf-1]quit
```

配置 PE 2 的 MPLS 基本能力和 MPLS LDP 能力，建立 LDP LSP。

```
<PE2> system-view
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 2.2.2.9 32
[PE2-LoopBack0] quit
[PE2] mpls lsr-id 2.2.2.9
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface serial 2/1
[PE2-Serial2/1] ip address 10.1.1.2 24
[PE2-Serial2/1] mpls
[PE2-Serial2/1] mpls ldp
[PE2-Serial2/1] quit
```

配置 PE 2 的 MP-IBGP 对等体为 PE1。

```
[PE2] bgp 100
```

```
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 0
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] quit
```

配置 PE 2 的 OSPF。

```
[PE2]ospf 1
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

(3) 配置 PE 接入 CE

配置 PE 1 接入 CE 1。

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] interface gigabitethernet 1/0
[PE1-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE1-GigabitEthernet1/0] ip address 100.1.1.2 24
[PE1-GigabitEthernet1/0] quit
[PE1] ospf 100 vpn-instance vpn1
[PE1-ospf-100] domain-id 10
[PE1-ospf-100] area 1
[PE1-ospf-100-area-0.0.0.1] network 100.1.1.0 0.0.0.255
[PE1-ospf-100-area-0.0.0.1] quit
[PE1-ospf-100] quit
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] import-route ospf 100
[PE1-bgp-vpn1] quit
[PE1-bgp] quit
```

配置 PE 2 接入 CE 2。

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 100:2
[PE2-vpn-instance-vpn1] vpn-target 1:1
[PE2-vpn-instance-vpn1] quit
[PE2] interface gigabitethernet 1/0
[PE2-GigabitEthernet1/0] ip binding vpn-instance vpn1
[PE2-GigabitEthernet1/0] ip address 120.1.1.2 24
[PE2-GigabitEthernet1/0] quit
[PE2] ospf 100 vpn-instance vpn1
[PE2-ospf-100] domain-id 10
[PE2-ospf-100] area 1
```

```
[PE2-ospf-100-area-0.0.0.1] network 120.1.1.0 0.0.0.255
[PE2-ospf-100-area-0.0.0.1] quit
[PE2-ospf-100] quit
[PE2] bgp 100
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] import-route ospf 100
[PE2-bgp-vpn1] quit
[PE2-bgp] quit
```

完成上述配置后，在 PE 设备上执行 **display ip routing-table vpn-instance** 命令，可以看到去往对端 CE 的路由是通过用户网络的 OSPF 路由，不是通过骨干网的 BGP 路由。

以 PE 1 为例：

```
[PE1] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 5        Routes : 5
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
20.1.1.0/24       OSPF   10   1563      100.1.1.1       GE1/0
30.1.1.0/24       OSPF   10   3125      100.1.1.1       GE1/0
100.1.1.0/24      Direct 0    0         100.1.1.2       GE1/0
100.1.1.2/32      Direct 0    0         127.0.0.1       InLoop0
120.1.1.0/24      OSPF   10   3126      100.1.1.1       GE1/0
```

(4) 配置 Sham-link

配置 PE 1。

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip binding vpn-instance vpn1
[PE1-LoopBack1] ip address 3.3.3.3 32
[PE1-LoopBack1] quit
[PE1] ospf 100
[PE1-ospf-100] area 1
[PE1-ospf-100-area-0.0.0.1] sham-link 3.3.3.3 5.5.5.5 cost 10
[PE1-ospf-100-area-0.0.0.1] quit
[PE1-ospf-100] quit
```

配置 PE 2。

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip binding vpn-instance vpn1
[PE2-LoopBack1] ip address 5.5.5.5 32
[PE2-LoopBack1] quit
[PE2] ospf 100
[PE2-ospf-100] area 1
[PE2-ospf-100-area-0.0.0.1] sham-link 5.5.5.5 3.3.3.3 cost 10
[PE2-ospf-100-area-0.0.0.1] quit
[PE2-ospf-100] quit
```

完成上述配置后，在 PE 设备上再次执行 **display ip routing-table vpn-instance** 命令，可以看到去往对端 CE 的路由变成了通过骨干网的 BGP 路由，并且有去往 Sham-link 目的地址的路由。

以 PE 1 为例：

```
[PE1] display ip routing-table vpn-instance vpn1
```

```
Routing Tables: vpn1
         Destinations : 6        Routes : 6
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
3.3.3.3/32        Direct 0    0         127.0.0.1       InLoop0
5.5.5.5/32        BGP    255  0         2.2.2.9         NULL0
20.1.1.0/24       OSPF   10   1563      100.1.1.1       GE1/0
100.1.1.0/24      Direct 0    0         100.1.1.2       GE1/0
100.1.1.2/32      Direct 0    0         127.0.0.1       InLoop0
120.1.1.0/24      BGP    255  0         2.2.2.9         NULL0
```

在 CE 设备上执行 **display ip routing-table** 命令，可以看到去往对端 CE 的 OSPF 路由开销变为 10（为 Sham-link 配置的开销），下一跳变为接入 PE 的 GigabitEthernet 接口，即去往对端的 VPN 流量将通过骨干网转发。

以 CE 1 为例：

```
[CE1] display ip routing-table
Routing Tables: Public
         Destinations : 9        Routes : 9
Destination/Mask  Proto  Pre  Cost      NextHop         Interface
20.1.1.0/24       Direct 0    0         20.1.1.1        S2/1
20.1.1.1/32       Direct 0    0         127.0.0.1       InLoop0
20.1.1.2/32       Direct 0    0         20.1.1.2        S2/1
30.1.1.0/24       OSPF   10   1574      100.1.1.2       GE1/0
100.1.1.0/24      Direct 0    0         100.1.1.1       GE1/0
100.1.1.1/32      Direct 0    0         127.0.0.1       InLoop0
120.1.1.0/24      OSPF   10   12        100.1.1.2       GE1/0
127.0.0.0/8       Direct 0    0         127.0.0.1       InLoop0
127.0.0.1/32      Direct 0    0         127.0.0.1       InLoop0
```

在 PE 上执行 **display ospf sham-link** 命令可以看到 Sham-link 的建立情况。

以 PE 1 为例：

```
[PE1] display ospf sham-link
           OSPF Process 100 with Router ID 100.1.1.2
 Sham Link:
 Area        RouterId     Source-IP     Destination-IP  State Cost
 0.0.0.1     100.1.1.2    3.3.3.3       5.5.5.5         P-2-P 10
```

执行 **display ospf sham-link area** 命令可以看到对端状态为 Full。

```
[PE1] display ospf sham-link area 1
          OSPF Process 100 with Router ID 100.1.1.2
  Sham-Link: 3.3.3.3 --> 5.5.5.5
  Neighbour State: Full
  Area: 0.0.0.1
  Cost: 10  State: P-2-P, Type: Sham
  Timers: Hello 10 , Dead 40 , Retransmit 5 , Transmit Delay 1
```

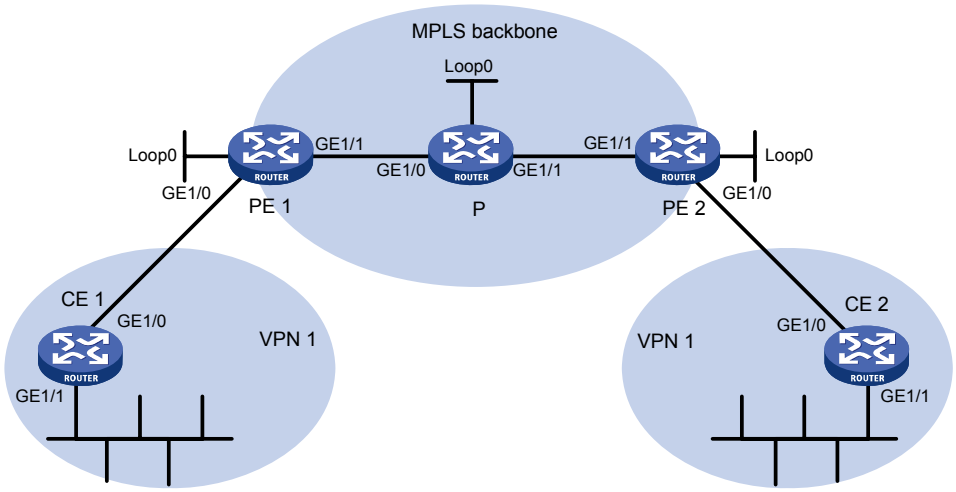
1.14.11 配置 BGP 的 AS 号替换

1. 组网需求

如图所示，CE 1 和 CE 2 同属于 VPN 1，分别接入 PE 1 和 PE 2，并且 CE 1 和 CE 2 复用 AS 号 600。

2. 组网图

图1-29 BGP 的 AS 号替换组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.1.1.1/24	P	Loop0	2.2.2.9/32
	GE1/1	100.1.1.1/24		GE1/0	20.1.1.2/24
PE 1	Loop0	1.1.1.9/32		GE1/1	30.1.1.1/24
	GE1/0	10.1.1.2/24	PE 2	Loop0	3.3.3.9/32
	GE1/1	20.1.1.1/24		GE1/0	10.2.1.2/24
CE 2	GE1/0	10.2.1.1/24		GE1/1	30.1.1.2/24
	GE1/1	200.1.1.1/24			

3. 配置步骤

(1) 配置基本 MPLS L3VPN

- 在 MPLS 骨干网上配置 OSPF，PE 和 P 之间能够学到对方 Loopback 接口的路由；
- 在 MPLS 骨干网上配置 MPLS 基本能力和 MPLS LDP，建立 LDP LSP；
- PE 之间建立 MP-IBGP 对等体关系，发布 VPN-IPv4 路由；
- 在 PE 2 上配置 VPN 1 的 VPN 实例，接入 CE 2；
- 在 PE 1 上配置 VPN 1 的 VPN 实例，接入 CE 1；
- PE 1 和 CE 1、PE 2 和 CE 2 之间配置 BGP，将 CE 的路由引入 PE。

完成上述配置后，在 CE 2 上执行 **display ip routing-table** 命令，可以看到 CE 2 能够学到 CE 1 接入 PE 1 的接口所在网段（10.1.1.0/24）的路由，但没有到达 CE 1 内部 VPN（100.1.1.0/24）的路由。CE 1 上也存在同样的现象。

```
<CE2> display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8
Destination/Mask    Proto   Pre  Cost      NextHop         Interface
10.1.1.0/24         BGP     255  0         10.2.1.2        GE1/0
10.1.1.1/32         BGP     255  0         10.2.1.2        GE1/0
```

```
10.2.1.0/24         Direct 0    0          10.2.1.1        GE1/0
10.2.1.1/32         Direct 0    0          127.0.0.1       InLoop0
10.2.1.2/32         Direct 0    0          10.2.1.2        GE1/0
127.0.0.0/8         Direct 0    0          127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0          127.0.0.1       InLoop0
200.1.1.0/24        Direct 0    0          200.1.1.1       InLoop0
200.1.1.1/32        Direct 0    0          127.0.0.1       InLoop0
```

在 PE 上执行 **display ip routing-table vpn-instance** 命令，可以看到 PE 的 VPN 实例中有到达对端 CE 内部 VPN 的路由。

以 PE 2 为例：

```
<PE2> display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 7        Routes : 7
Destination/Mask    Proto  Pre  Cost       NextHop         Interface
10.1.1.0/24         BGP    255  0          1.1.1.9         NULL0
10.1.1.1/32         BGP    255  0          1.1.1.9         NULL0
10.2.1.0/24         Direct 0    0          10.2.1.2        GE1/0
10.2.1.1/32         Direct 0    0          10.2.1.1        GE1/0
10.2.1.2/32         Direct 0    0          127.0.0.1       InLoop0
100.1.1.1/32        BGP    255  0          1.1.1.9         NULL0
200.1.1.1/32        BGP    255  0          10.2.1.1        GE1/0
```

在 PE 2 上打开 BGP 的 Update 报文调试信息开关，可以看到 PE 2 发布了去往 100.1.1.1/32 的路由，AS 路径信息为“100 600”。

```
<PE2> terminal monitor
<PE2> terminal debugging
<PE2> debugging bgp update vpn-instance vpn1 verbose
<PE2> refresh bgp vpn-instance vpn1 all export
*0.4402392 PE2 RM/7/RMDEBUG:
         BGP.vpn1: Send UPDATE to 10.2.1.1 for following destinations :
         Origin    : Incomplete
         AS Path   : 100 600
         Next Hop  : 10.2.1.2
         100.1.1.1/32,
```

在 CE 2 上执行 **display bgp routing-table peer received-routes** 命令，可以看到 CE 2 没有接收 100.1.1.1/32 的路由。

```
<CE2> display bgp routing-table peer 10.2.1.2 received-routes
 Total Number of Routes: 4
 BGP Local router ID is 10.2.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop        MED     LocPrf    PrefVal Path/Ogn
 *>   10.1.1.0/24      10.2.1.2                           0      100?
 *>   10.1.1.1/32      10.2.1.2                           0      100?
 *    10.2.1.0/24      10.2.1.2        0                  0      100?
 *    10.2.1.1/32      10.2.1.2        0                  0      100?
```

(2) 配置 BGP 的 AS 号替换功能

在 PE 2 上配置 BGP 的 AS 号替换功能。

```
<PE2> system-view
[PE2] bgp 100
[PE2-bgp] ipv4-family vpn-instance vpn1
[PE2-bgp-vpn1] peer 10.2.1.1 substitute-as
[PE2-bgp-vpn1] quit
[PE2-bgp] quit
```

可以看到 PE 2 向 CE 2 发布的路由中，100.1.1.1/32 的 AS 路径信息由“100 600”变为“100 100”：

```
*0.13498737 PE2 RM/7/RMDEBUG:
        BGP.vpn1: Send UPDATE to 10.2.1.1 for following destinations :
        Origin    : Incomplete
        AS Path   : 100 100
        Next Hop  : 10.2.1.2
        100.1.1.1/32
```

再次查看 CE 2 接收的路由信息和路由表：

```
<CE2> display bgp routing-table peer 10.2.1.2 received-routes
 Total Number of Routes: 5
 BGP Local router ID is 10.2.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED      LocPrf    PrefVal Path/Ogn
 *>   10.1.1.0/24      10.2.1.2                             0      100?
 *>   10.1.1.1/32      10.2.1.2                             0      100?
 *    10.2.1.0/24      10.2.1.2       0                     0      100?
 *    10.2.1.1/32      10.2.1.2       0                     0      100?
 *>   100.1.1.1/32     10.2.1.2                             0      100 100?
<CE2> display ip routing-table
Routing Tables: Public
        Destinations : 9        Routes : 9
Destination/Mask    Proto  Pre  Cost       NextHop         Interface
10.1.1.0/24         BGP    255  0          10.2.1.2        GE1/0
10.1.1.1/32         BGP    255  0          10.2.1.2        GE1/0
10.2.1.0/24         Direct 0    0          10.2.1.1        GE1/0
10.2.1.1/32         Direct 0    0          127.0.0.1       InLoop0
10.2.1.2/32         Direct 0    0          10.2.1.2        GE1/0
100.1.1.1/32        BGP    255  0          10.2.1.2        GE1/0
127.0.0.0/8         Direct 0    0          127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0          127.0.0.1       InLoop0
200.1.1.1/32        Direct 0    0          127.0.0.1       InLoop0
```

在 PE 1 上也配置 BGP 的 AS 号替换功能后，CE 1 和 CE 2 的 GigabitEthernet 接口能够相互 Ping 通。

```
<CE1> ping –a 100.1.1.1 200.1.1.1
  PING 200.1.1.1: 56  data bytes, press CTRL_C to break
```

```
  Reply from 200.1.1.1: bytes=56 Sequence=1 ttl=253 time=109 ms
  Reply from 200.1.1.1: bytes=56 Sequence=2 ttl=253 time=67 ms
  Reply from 200.1.1.1: bytes=56 Sequence=3 ttl=253 time=66 ms
  Reply from 200.1.1.1: bytes=56 Sequence=4 ttl=253 time=85 ms
  Reply from 200.1.1.1: bytes=56 Sequence=5 ttl=253 time=70 ms
--- 200.1.1.1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 66/79/109 ms
```

目 录

1 GRE配置 1-1
1.1 GRE简介 1-1
1.1.1 协议简介 1-1
1.1.2 GRE的安全选项 1-2
1.1.3 应用范围 1-2
1.1.4 协议规范 1-4
1.2 配置GRE over IPv4 隧道 1-4
1.2.1 配置准备 1-4
1.2.2 配置GRE over IPv4 隧道 1-4
1.3 配置GRE over IPv6 隧道 1-5
1.3.1 配置准备 1-5
1.3.2 配置GRE over IPv6 隧道 1-6
1.4 GRE显示和维护 1-7
1.5 GRE over IPv4 典型配置举例 1-7
1.5.1 组网需求 1-7
1.5.2 组网图 1-8
1.5.3 配置步骤 1-8
1.6 GRE over IPv6 典型配置举例 1-9
1.6.1 GRE over IPv6 典型配置举例 1-9
1.7 常见配置错误举例 1-11

1 GRE 配置

1.1 GRE 简介

1.1.1 协议简介

GRE（Generic Routing Encapsulation，通用路由封装）协议是对某些网络层协议（如 IP 和 IPX）的数据报文进行封装，使这些被封装的数据报文能够在另一个网络层协议（如 IP）中传输。GRE 采用了 Tunnel（隧道）技术，是 VPN（Virtual Private Network）的第三层隧道协议。

Tunnel 是一个虚拟的点对点的连接，提供了一条通路使封装的数据报文能够在这个通路上传输，并且在一个 Tunnel 的两端分别对数据报进行封装及解封装。

一个X协议的报文要想穿越IP网络在Tunnel中传输，必须要经过加封装与解封装两个过程，下面以 图1-1的网络为例说明这两个过程：

图1-1 X 协议通过 GRE 隧道互连



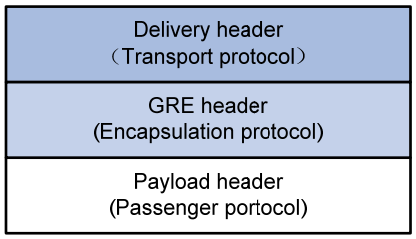
1. 加封装过程

- Router A 连接 Group 1 的接口收到 X 协议报文后，首先交由 X 协议处理；
- X 协议检查报头中的目的地址域来确定如何路由此包；
- 若报文的目的地址要经过 Tunnel 才能到达，则设备将此报文发给相应的 Tunnel 接口；
- Tunnel 口收到此报文后进行 GRE 封装，在封装 IP 报文头后，设备根据此 IP 包的目的地址及路由表对报文进行转发，从相应的网络接口发送出去。

2. GRE 封装后的报文格式

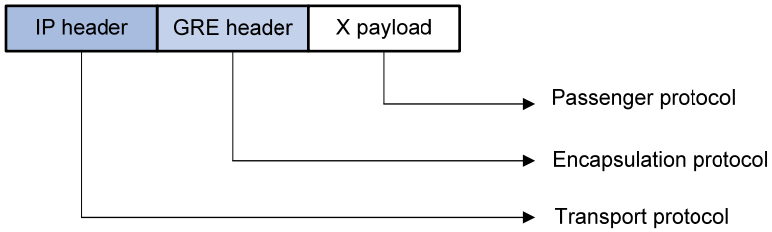
封装好的报文的形式如下图所示：

图1-2 封装好的 Tunnel 报文格式



举例来说，一个封装在 IP Tunnel 中的 X 协议报文的格式如下：

图1-3 Tunnel 中传输报文的格式



需要封装和传输的数据报文，称之为净荷（Payload），净荷的协议类型为乘客协议（Passenger Protocol）。系统收到一个净荷后，首先使用封装协议（Encapsulation Protocol）对这个净荷进行 GRE 封装，即把乘客协议报文进行了“包装”，加上了一个 GRE 头部成为 GRE 报文；然后再把封装好的原始报文和 GRE 头部封装在 IP 报文中，这样就可完全由 IP 层负责此报文的前向转发（Forwarding）。通常把这个负责前向转发的 IP 协议称为传输协议（Delivery Protocol 或者 Transport Protocol）。根据传输协议的不同，可以分为 GRE over IPv4 和 GRE over IPv6 两种隧道模式。

3. 解封装的过程

解封装过程和加封装的过程相反。

- RouterB 从 Tunnel 接口收到 IP 报文，检查目的地址；
- 如果发现目的地是本路由器，则 RouterB 剥掉此报文的 IP 报头，交给 GRE 协议处理（进行检验密钥、检查校验和及报文的序列号等）；
- GRE 协议完成相应的处理后，剥掉 GRE 报头，再交由 X 协议对此数据报进行后续的转发处理。


说明

GRE 收发双方的加封装、解封装处理，以及由于封装造成的数据量增加，会导致使用 GRE 后设备的数据转发效率有一定程度的下降。

1.1.2 GRE 的安全选项

为了提高 GRE 隧道的安全性，GRE 还支持由用户选择设置 Tunnel 接口的识别关键字（或称密钥），和对隧道封装的报文进行端到端校验。

在 RFC1701 中规定：

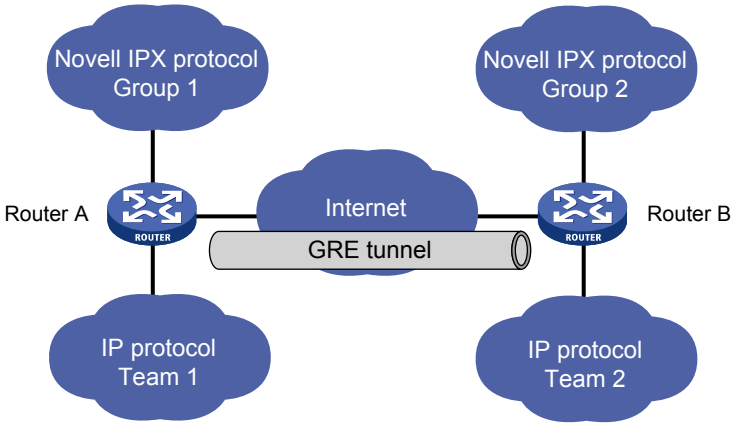
- 若 GRE 报文头中的 Key 标识位置 1，则收发双方将进行通道识别关键字的验证，只有 Tunnel 两端设置的识别关键字完全一致时才能通过验证，否则将报文丢弃。
- 若 GRE 报文头中的 Checksum 标识位置 1，则校验和有效。发送方将根据 GRE 头及 Payload 信息计算校验和，并将包含校验和的报文发送给对端。接收方对接收到的报文计算校验和，并与报文中的校验和比较，如果一致则对报文进一步处理，否则丢弃。

1.1.3 应用范围

GRE 主要能实现以下几种服务类型：

1. 多协议的本地网通过单一协议的骨干网传输

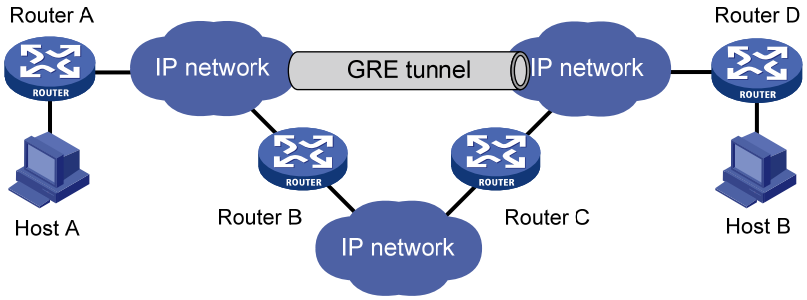
图1-4 多协议本地网通过单一协议骨干网传输



上图中，Group 1 和 Group 2 是运行 Novell IPX 协议的本地网，Team 1 和 Team 2 是运行 IP 协议的本地网。通过在 Router A 和 Router B 之间采用 GRE 协议封装的隧道（Tunnel），Group 1 和 Group 2、Team 1 和 Team 2 可以互不影响地进行通信。

2. 扩大了步跳数受限协议（如 RIP）的网络的工作范围

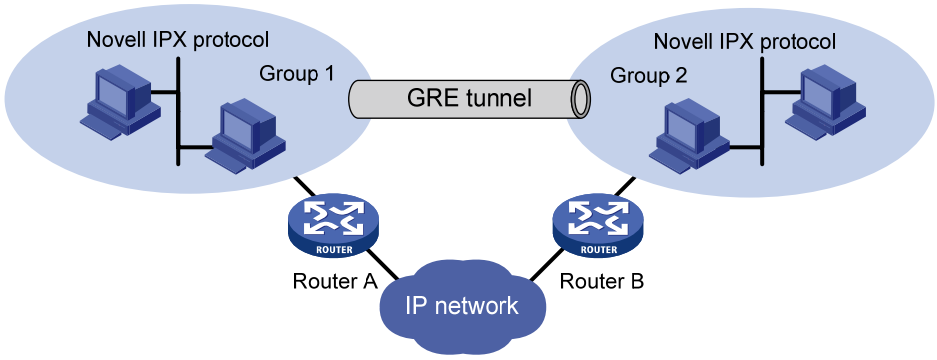
图1-5 扩大网络工作范围



两台终端之间的步跳数超过 15，它们将无法通信。而通过在网络中使用隧道（Tunnel）可以隐藏一部分步跳，从而扩大网络的工作范围。

3. 将一些不能连续的子网连接起来，用于组建 VPN

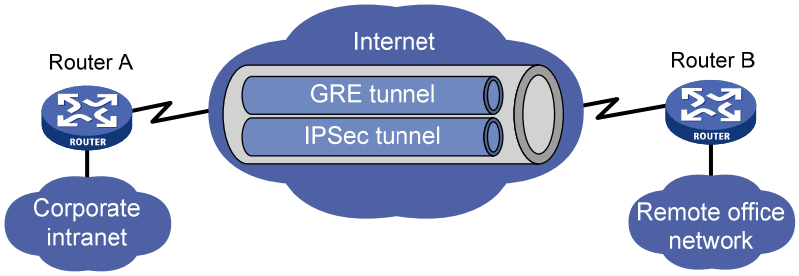
图1-6 Tunnel 连接不连续子网



运行 Novell IPX 协议的两个子网 Group 1 和 Group 2 分别在不同的城市，通过使用隧道可以实现跨越广域网的 VPN。

4. 与 IPsec 结合使用

图1-7 GRE-IPsec 隧道应用



对于诸如路由协议、语音、视频等数据先进行 GRE 封装，然后再对封装后的报文进行 IPsec 的加密处理。

1.1.4 协议规范

与 GRE 相关的协议规范有：

- RFC1701：Generic Routing Encapsulation (GRE)
- RFC1702：Generic Routing Encapsulation over IPv4 networks
- RFC2784：Generic Routing Encapsulation (GRE)

1.2 配置 GRE over IPv4 隧道

1.2.1 配置准备

设备上的接口（如 GigabitEthernet 接口，Loopback 接口等）已经配置 IP 地址，能够进行正常通讯。这些接口将作为 Tunnel 虚接口的源接口，以保证隧道目的地址路由可达。

1.2.2 配置 GRE over IPv4 隧道

表1-1 配置 GRE over IPv4 隧道

操作	命令	说明
进入系统视图	**system-view**	-
创建一个 Tunnel 接口，并进入该 Tunnel 接口视图	**interface tunnel** *interface-number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv4 地址	**ip address** *ip-address* { *mask* | *mask-length* }	必选 缺省情况下，Tunnel 接口上没有设置 IPv4 地址
配置隧道模式为 GRE over IPv4	**tunnel-protocol gre**	可选 缺省情况下，采用 GRE over IPv4 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口	**source** { *ip-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置 Tunnel 接口的目的端地址	**destination** *ip-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址

操作	命令	说明
使能 GRE 的 keepalive 功能，探测 Tunnel 接口状态，并配置 keepalive 报文发送周期及最大发送次数	**keepalive** [*seconds* [*times*]]	可选 缺省情况下，不启用 GRE 的 keepalive 功能
使能 GRE 报文校验和功能	**gre checksum**	可选 缺省情况下，未使能 GRE 报文校验和功能
设置 GRE 类型隧道接口的密钥	**gre key** *key-number*	可选 缺省情况下，没有设置 GRE 类型隧道接口的密钥 隧道两端要么设置相同的密钥，要么都不设置密钥
配置通过 Tunnel 的路由	配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置	可选 在源端路由器和目的端路由器上都必须存在经过 Tunnel 转发的路由，这样需要进行 GRE 封装的报文才能正确转发。可以选择配置静态路由，也可以选择配置动态路由。在 Tunnel 的两端都要进行此项配置
设置隧道接口的 MTU 值	**mtu** *mtu-size*	可选
指定转发当前接口流量的业务处理板	**service slot** *slot-number*	可选 缺省情况下，接口没有指定转发当前接口流量的业务处理板

注意

- Tunnel 的源端地址与目的端地址唯一标识了一个通道。这些配置在 Tunnel 两端必须配置，且两端地址互为源地址和目的地址。
- 两个或两个以上使用同种封装协议的 Tunnel 接口不能配置完全相同的源地址和目的地址。
- 配置 Tunnel 接口的源端地址时，若采用配置源接口形式，则 Tunnel 的源地址取的是源接口的主 IP 地址。
- 隧道两端可以根据实际应用的需要决定配置校验和或禁止校验和。如果本端配置了校验和而对端没有配置，则本端将不会对接收到的报文进行校验和检查，但对发送的报文计算校验和；相反，如果本端没有配置校验和而对端已配置，则本端将对从对端发来的报文进行校验和检查，但对发送的报文不计算校验和。
- 配置通过 Tunnel 转发的路由时，可以手工配置一条静态路由，目的地址是未进行 GRE 封装的报文的目的地址，下一跳是对端 Tunnel 接口的地址。也可以在 Tunnel 接口上和与私网相连的路由器接口上分别使能动态路由协议，由动态路由协议来建立通过 Tunnel 转发的路由表项。
- 在隧道接口配置的静态路由的目的地址不能与隧道接口的地址在同一网段中。

1.3 配置 GRE over IPv6 隧道

1.3.1 配置准备

设备上的接口（如 VLAN 接口，Ethernet 接口，Loopback 接口等）已经配置 IPv6 地址，能够进行正常通讯。这些接口将作为 Tunnel 虚接口的源接口，以保证隧道目的地址路由可达。

1.3.2 配置 GRE over IPv6 隧道

表1-2 配置 GRE over IPv6 隧道

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 报文转发功能	**ipv6**	必选 缺省情况下，关闭 IPv6 报文转发功能
创建一个 Tunnel 接口，并进入该 Tunnel 接口视图	**interface tunnel** *interface-number*	必选 缺省情况下，设备上无 Tunnel 接口
设置 Tunnel 接口的 IPv4 地址	**ip address** *ip-address* { *mask* | *mask-length* }	必选 缺省情况下，Tunnel 接口上没有设置 IPv4 地址
配置隧道模式为 GRE over IPv6	**tunnel-protocol gre ipv6**	必选 缺省情况下，采用 GRE over IPv4 隧道模式 在隧道的两端应配置相同的隧道模式，否则可能造成报文传输失败
设置 Tunnel 接口的源端地址或接口	**source** { *ipv6-address* | *interface-type interface-number* }	必选 缺省情况下，Tunnel 接口上没有设置源端地址和接口
设置 Tunnel 接口的目的端地址	**destination** *ipv6-address*	必选 缺省情况下，Tunnel 接口上没有设置目的端地址
设置隧道允许的最大嵌套封装次数	**encapsulation-limit** [*number*]	可选 缺省情况下，最大嵌套封装次数为 4
使能 GRE 报文校验和功能	**gre checksum**	可选 缺省情况下，未使能 GRE 报文校验和功能
设置 GRE 类型隧道接口的密钥	**gre key** *key-number*	可选 缺省情况下，没有设置 GRE 类型隧道接口的密钥 隧道两端要么设置相同的密钥，要么都不设置密钥
配置通过 Tunnel 的路由	配置的详细情况请参见“IP 路由分册”中的“静态路由配置”或其他路由协议配置	可选 在源端路由器和目的端路由器上都必须存在经过 Tunnel 转发的路由，这样需要进行 GRE 封装的报文才能正确转发。可以选择配置静态路由，也可以选择配置动态路由。在 Tunnel 的两端都要进行此项配置
设置隧道接口发送 IPv6 报文的 MTU 值	**ipv6 mtu** *mtu-size*	可选 缺省情况下，MTU 值的范围为 1280~1500
指定转发当前接口流量的业务处理板	**service slot** *slot-number*	可选 缺省情况下，接口没有指定转发当前接口流量的业务处理板



说明

- **interface tunnel**、**tunnel-protocol**、**mtu** 和 **service slot** 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”；**source** 和 **destination** 命令的详细介绍，请参见“IP 业务分册”中的“隧道技术命令”。
- 有关 **ipv6 mtu** 的相关配置可参考“IP 业务分册”中的“IPv6 基础配置命令”。

注意

- 以上各项 Tunnel 接口下进行的功能特性配置，在删除 Tunnel 接口后，该接口上的所有配置也将被删除。
- Tunnel 的源端地址与目的端地址唯一标识了一个通道。这些配置在 Tunnel 两端必须配置，且两端地址互为源地址和目的地址。
- 两个或两个以上使用同种封装协议的 Tunnel 接口不能配置完全相同的源地址和目的地址。
- 配置 Tunnel 接口的源端地址时，若采用配置源接口形式，则 Tunnel 的源地址取的是源接口的主 IP 地址。
- 隧道两端可以根据实际应用的需要决定配置校验和或禁止校验和。如果本端配置了校验和而对端没有配置，则本端将不会对接收到的报文进行校验和检查，但对发送的报文计算校验和；相反，如果本端没有配置校验和而对端已配置，则本端将对从对端发来的报文进行校验和检查，但对发送的报文不计算校验和。
- 配置通过 Tunnel 转发的路由时，可以手工配置一条静态路由，目的地址是未进行 GRE 封装的报文的目的地址，下一跳是对端 Tunnel 接口的地址。也可以在 Tunnel 接口上和与私网相连的路由器接口上分别使能动态路由协议，由动态路由协议来建立通过 Tunnel 转发的路由表项。
- 在隧道接口配置的静态路由的目的地址不能与隧道接口的地址在同一网段中。

1.4 GRE 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 GRE 的运行情况，通过查看显示信息验证配置的效果。

表1-3 GRE 的显示和维护

操作	命令
显示指定 Tunnel 接口的相关信息	**display interface tunnel** [*number*]
显示指定 Tunnel 接口的 IPv6 相关信息	**display ipv6 interface tunnel** [*number*] [**verbose**]

说明

display interface tunnel 和 **display ipv6 interface tunnel** 命令的详细介绍，请参见“接入分册”中的“逻辑接口命令”。

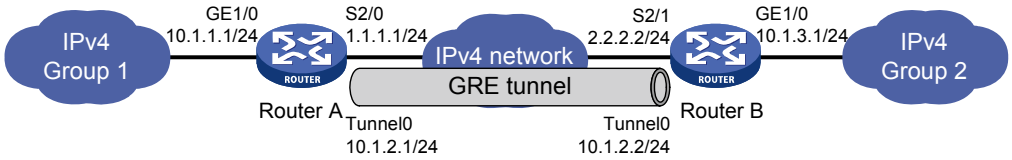
1.5 GRE over IPv4 典型配置举例

1.5.1 组网需求

路由器 Router A 和路由器 Router B 之间通过 Internet 相连。运行 IP 协议的私有网络的两个子网 Group 1 和 Group 2，通过在两台路由器之间使用 GRE 建立隧道实现互联。

1.5.2 组网图

图1-8 GRE over IPv4 应用组网图



1.5.3 配置步骤

(1) 配置路由器 Router A

配置接口 GigabitEthernet1/0。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ip address 10.1.1.1 255.255.255.0
[RouterA-GigabitEthernet1/0] quit
```

配置接口 Serial2/0（隧道的实际物理接口）。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 1.1.1.1 255.255.255.0
[RouterA-Serial2/0] quit
```

创建 Tunnel0 接口。

```
[RouterA] interface tunnel 0
```

配置 Tunnel0 接口的 IP 地址。

```
[RouterA-Tunnel0] ip address 10.1.2.1 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterA-Tunnel0] tunnel-protocol gre
```

配置 Tunnel0 接口的源地址（Serial2/0 的 IP 地址）。

```
[RouterA-Tunnel0] source 1.1.1.1
```

配置 Tunnel0 接口的目的地址（Router B 的 Serial2/1 的 IP 地址）。

```
[RouterA-Tunnel0] destination 2.2.2.2
[RouterA-Tunnel0] quit
```

配置从 Router A 经过 Tunnel0 接口到 Group 2 的静态路由。

```
[RouterA] ip route-static 10.1.3.0 255.255.255.0 tunnel 0
```

(2) 配置路由器 Router B

配置接口 GigabitEthernet1/0。

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] ip address 10.1.3.1 255.255.255.0
[RouterB-GigabitEthernet1/0] quit
```

配置接口 Serial2/1（隧道的实际物理接口）。

```
[RouterB] interface serial 2/1
```

```
[RouterB-Serial2/1] ip address 2.2.2.2 255.255.255.0
[RouterB-Serial2/1] quit
```

创建 Tunnel0 接口。

```
[RouterB] interface tunnel 0
```

配置 Tunnel0 接口的 IP 地址。

```
[RouterB-Tunnel0] ip address 10.1.2.2 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterB-Tunnel0] tunnel-protocol gre
```

配置 Tunnel0 接口的源地址（Serial2/1 的 IP 地址）。

```
[RouterB-Tunnel0] source 2.2.2.2
```

配置 Tunnel0 接口的目的地址（Router A 的 Serial2/0 的 IP 地址）。

```
[RouterB-Tunnel0] destination 1.1.1.1
[RouterB-Tunnel0] quit
```

配置从 Router B 经过 Tunnel0 接口到 Group 1 的静态路由。

```
[RouterB] ip route-static 10.1.1.0 255.255.255.0 tunnel 0
```

1.6 GRE over IPv6 典型配置举例

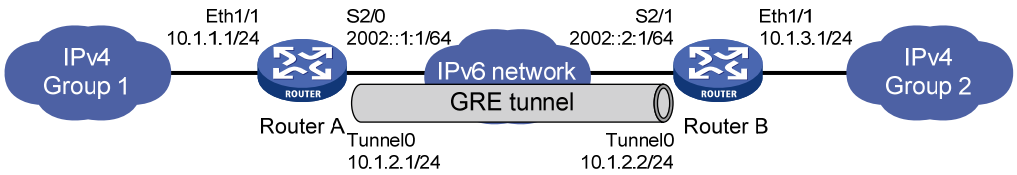
1.6.1 GRE over IPv6 典型配置举例

1. 组网需求

运行 IP 协议的两个子网 Group 1 和 Group 2，通过在路由器 Router A 和路由器 Router B 之间使用三层隧道协议 GRE，穿越 IPv6 网络实现互联。

2. 组网图

图1-9 GRE over IPv6 应用组网图



3. 配置步骤

(1) 配置路由器 Router A

```
<RouterA> system-view
```

使能 IPv6。

```
[RouterA] ipv6
```

配置接口 Ethernet1/1。

```
[RouterA] interface ethernet 1/1
[RouterA-Ethernet1/1] ip address 10.1.1.1 255.255.255.0
[RouterA-Ethernet1/1] quit
```

配置接口 Serial2/0（隧道的实际物理接口）。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ipv6 address 2002::1:1 64
[RouterA-Serial2/0] quit
```

创建 Tunnel0 接口。

```
[RouterA] interface tunnel 0
```

配置 Tunnel0 接口的 IP 地址。

```
[RouterA-Tunnel0] ip address 10.1.2.1 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterA-Tunnel0] tunnel-protocol gre ipv6
```

配置 Tunnel0 接口的源地址（Serial2/0 的 IP 地址）。

```
[RouterA-Tunnel0] source 2002::1:1
```

配置 Tunnel0 接口的目的地址（Router B 的 Serial2/1 的 IP 地址）。

```
[RouterA-Tunnel0] destination 2002::2:1
[RouterA-Tunnel0] quit
```

配置从 Router A 经过 Tunnel0 接口到 Group 2 的静态路由。

```
[RouterA] ip route-static 10.1.3.0 255.255.255.0 tunnel 0
```

(2) 配置路由器 Router B

```
<RouterB> system-view
```

使能 IPv6。

```
[RouterB] ipv6
```

配置接口 Ethernet1/1。

```
[RouterB] interface ethernet 1/1
[RouterB-Ethernet1/1] ip address 10.1.3.1 255.255.255.0
[RouterB-Ethernet1/1] quit
```

配置接口 Serial2/1（隧道的实际物理接口）。

```
[RouterB] interface serial 2/1
[RouterB-Serial2/1] ipv6 address 2002::2:1 64
[RouterB-Serial2/1] quit
```

创建 Tunnel0 接口。

```
[RouterB] interface tunnel 0
```

配置 Tunnel0 接口的 IP 地址。

```
[RouterB-Tunnel0] ip address 10.1.2.2 255.255.255.0
```

配置 Tunnel 封装模式。

```
[RouterB-Tunnel0] tunnel-protocol gre ipv6
```

配置 Tunnel0 接口的源地址（Serial2/1 的 IP 地址）。

```
[RouterB-Tunnel0] source 2002::2:1
```

配置 Tunnel0 接口的目的地址（Router A 的 Serial2/0 的 IP 地址）。

```
[RouterB-Tunnel0] destination 2002::1:1
```

```
[RouterB-Tunnel0] quit
```

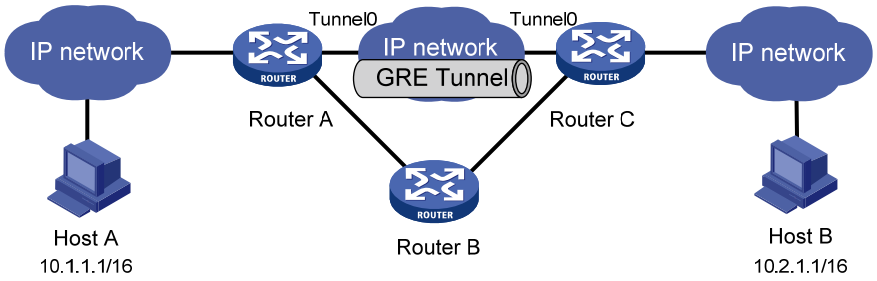
配置从 Router B 经过 Tunnel0 接口到 Group 1 的静态路由。

```
[RouterB] ip route-static 10.1.1.0 255.255.255.0 tunnel 0
```

1.7 常见配置错误举例

GRE的配置相对比较简单，但要注意配置的一致性，大部分的错误都可以通过使用调试命令**debugging gre**和**debugging tunnel**定位。这里仅就一种错误进行分析，如 图 1-10所示。

图1-10 GRE 排错示例



故障之一：Tunnel 两端接口配置正确且 Tunnel 两端可以 ping 通，但 Host A 和 Host B 之间却无法 ping 通。

故障排除：可以按照如下步骤进行。

- 在任意视图下，在 Router A 和 Router C 分别执行 **display ip routing-table** 命令，观察在 Router A 是否有经过 Tunnel0 接口到 10.2.0.0/16 的路由；在 Router C 是否有经过 Tunnel0 接口到 10.1.0.0/16 的路由。
- 如果在上一步的输出中发现缺少相应的静态路由，在系统视图下使用 **ip route-static** 命令添加。以 Router A 为例，配置如下：

```
[RouterA] ip route-static 10.2.0.0 255.255.0.0 tunnel 0
```

目 录

1 L2TP配置 1-1
1.1 L2TP简介 1-1
1.1.1 VPDN概述 1-1
1.1.2 L2TP典型组网应用 1-1
1.1.3 L2TP基本概念 1-2
1.1.4 L2TP隧道模式及隧道建立过程 1-4
1.1.5 L2TP协议的特点 1-6
1.1.6 协议规范 1-7
1.2 L2TP配置任务简介 1-7
1.3 配置L2TP基本功能 1-7
1.4 配置LAC端 1-8
1.4.1 配置LAC发起L2TP连接请求 1-8
1.4.2 配置AVP数据的隐藏传输 1-9
1.4.3 配置LAC侧VPN用户认证 1-9
1.5 配置LNS端 1-10
1.5.1 配置虚接口模板 1-10
1.5.2 配置本端地址及分配的地址池 1-10
1.5.3 配置LNS接收L2TP连接请求 1-11
1.5.4 配置LNS侧的用户验证 1-11
1.5.5 配置LNS侧VPN用户认证 1-12
1.5.6 配置L2TP多实例功能 1-13
1.5.7 配置ACCM消息发送 1-13
1.6 配置L2TP连接参数 1-13
1.6.1 配置隧道验证 1-13
1.6.2 配置隧道Hello报文发送时间间隔 1-14
1.6.3 开启L2TP隧道流控功能 1-14
1.6.4 强制断开指定的隧道连接 1-15
1.7 L2TP显示和维护 1-15
1.8 L2TP典型配置举例 1-15
1.8.1 NAS-Initiated VPN 1-15
1.8.2 Client-Initiated VPN 1-17
1.8.3 L2TP多域组网应用 1-19
1.8.4 复杂的组网情况 1-22
1.9 常见配置错误举例 1-22
1.9.1 错误之一 1-22
1.9.2 错误之二 1-23

1 L2TP 配置

1.1 L2TP 简介

1.1.1 VPDN 概述

VPDN（Virtual Private Dial-up Network，虚拟私有拨号网）是指利用公共网络（如 ISDN 或 PSTN）的拨号功能接入公共网络，实现虚拟专用网，从而为企业、小型 ISP、移动办公人员等提供接入服务。即，VPDN 为远端用户与私有企业网之间提供了一种经济而有效的点到点连接方式。

VPDN 采用专用的网络加密通信协议，在公共网络上为企业建立安全的虚拟专网。企业驻外机构和出差人员可从远程经由公共网络，通过虚拟加密隧道实现和企业总部之间的网络连接，而公共网络上其它用户则无法穿过虚拟隧道访问企业网内部的资源。

VPDN 有以下两种实现方式：

- 接入服务器发起 VPDN 连接

NAS（Network Access Server，网络接入服务器）通过使用 VPDN 隧道协议，将客户的 PPP 连接直接连到企业的 VPDN 网关上，从而与 VPDN 网关建立隧道。这些对于用户是透明的，用户只需要登录一次就可以接入企业网络，由企业网进行用户认证和地址分配，而不占用公共地址。该方式需要 NAS 支持 VPDN 协议、认证系统支持 VPDN 属性。

- 用户发起 VPDN 连接

客户端与 VPDN 网关建立隧道。这种方式由客户端先建立与 Internet 的连接，再通过专用的客户软件（如 Windows 2000 支持的 L2TP 客户端）与 VPDN 网关建立隧道连接。用户上网的方式和地点没有限制，不需 ISP 介入。但是，用户需要安装专用的软件（一般都是 Windows 2000 平台），限制了用户使用的平台。

VPDN 网关一般使用路由器或 VPN 专用服务器。

1.1.2 L2TP 典型组网应用

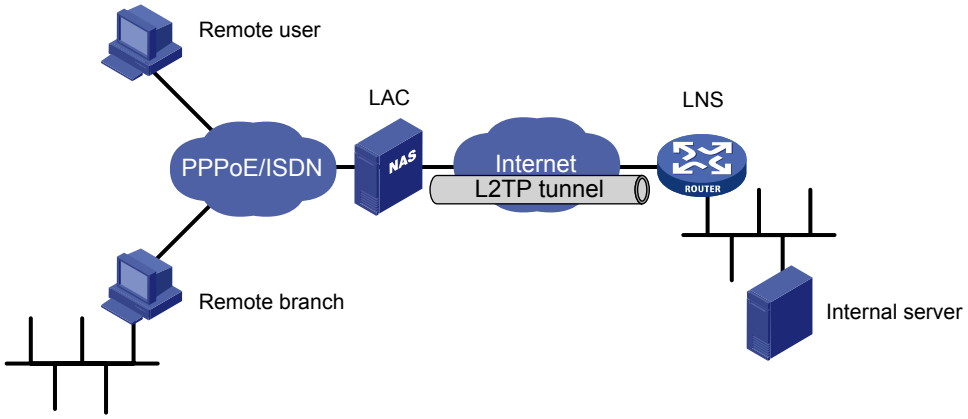
VPDN 隧道协议主要包括以下三种：

- PPTP（Point-to-Point Tunneling Protocol，点到点隧道协议）
- L2F（Layer 2 Forwarding，二层转发）
- L2TP（Layer 2 Tunneling Protocol，二层隧道协议）

目前使用最广泛的是 L2TP。

使用L2TP协议构建的VPDN应用的典型组网如 图 1-1所示。

图1-1 应用 L2TP 构建的 VPDN 服务



在 L2TP 构建的 VPDN 中，网络组件包括以下三个部分：

- 远端系统

远端系统是要接入 VPDN 网络的远地用户和远地分支机构，通常是一个拨号用户的主机或私有网络的一台路由设备。

- LAC（L2TP Access Concentrator，L2TP 访问集中器）

LAC 是附属在交换网络上的具有 PPP 端系统和 L2TP 协议处理能力的设备，通常是一个当地 ISP 的 NAS，主要用于为 PPP 类型的用户提供接入服务。

LAC 位于 LNS 和远端系统之间，用于在 LNS 和远端系统之间传递信息包。它把从远端系统收到的信息包按照 L2TP 协议进行封装并送往 LNS，同时也将从 LNS 收到的信息包进行解封装并送往远端系统。

LAC 与远端系统之间采用本地连接或 PPP 链路，VPDN 应用中通常为 PPP 链路。

- LNS（L2TP Network Server，L2TP 网络服务器）

LNS 是 PPP 端系统上用于处理 L2TP 协议服务器端部分的设备，通常是一个企业内部网的边缘设备。

LNS 作为 L2TP 隧道的另一侧端点，是 LAC 的对端设备，是 LAC 进行隧道传输的 PPP 会话的逻辑终止端点。通过在公网中建立 L2TP 隧道，将远端系统的 PPP 连接的另一端由原来的 LAC 在逻辑上延伸到了企业网内部端的 LNS。

1.1.3 L2TP 基本概念

1. L2TP 协议背景

PPP 定义了一种封装技术，可以在二层的点到点链路上传输多种协议数据包，当用户与 NAS 之间运行 PPP 协议时，二层链路端点与 PPP 会话点驻留在相同硬件设备上。

L2TP（RFC 2661）提供了对 PPP 链路层数据包的隧道传输支持，允许二层链路端点（LAC）和 PPP 会话点（LNS）驻留在不同设备上，并且采用包交换网络技术进行信息交互，从而扩展了 PPP 模型。

L2TP 结合了 L2F 和 PPTP 的各自优点，成为 IETF 有关二层隧道协议的工业标准。

2. L2TP 协议结构

图 1-2描述了PPP帧和控制通道以及数据通道之间的关系。PPP帧在不可靠的L2TP数据通道上进行传输，控制消息在可靠的L2TP控制通道内传输。

图1-2 L2TP 协议结构

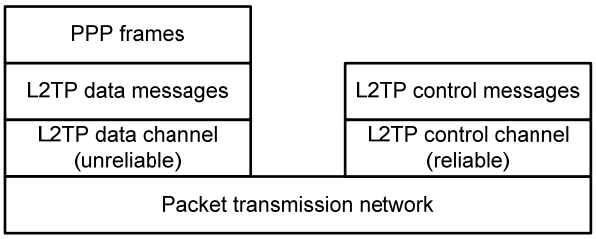


图 1-3描述了LAC与LNS之间的L2TP数据报文的封装结构。通常L2TP数据以UDP报文的形式发送。L2TP注册了UDP 1701 端口，但是这个端口仅用于初始的隧道建立过程中。L2TP隧道发起方任选一个空闲的端口（未必是 1701）向接收方的 1701 端口发送报文；接收方收到报文后，也任选一个空闲的端口（未必是 1701），给发送方的指定端口回送报文。至此，双方的端口选定，并在隧道保持连通的时间段内不再改变。

图1-3 L2TP 报文封装结构图



3. 隧道和会话

在一个 LNS 和 LAC 对之间存在着两种类型的连接。

- 隧道（Tunnel）连接：它对应了一个 LNS 和 LAC 对。
- 会话（Session）连接：它复用在隧道连接之上，用于表示承载在隧道连接中的每个 PPP 会话过程。

在同一对 LAC 和 LNS 之间可以建立多个 L2TP 隧道，隧道由一个控制连接和一个或多个会话连接组成。会话连接必须在隧道建立（包括身份保护、L2TP 版本、帧类型、硬件传输类型等信息的交换）成功之后进行，每个会话连接对应于 LAC 和 LNS 之间的一个 PPP 数据流。

控制消息和 PPP 数据报文都在隧道上传输。L2TP 使用 Hello 报文来检测隧道的连通性。LAC 和 LNS 定时向对端发送 Hello 报文，若在一段时间内未收到 Hello 报文的应答，隧道断开。

4. 控制消息和数据消息

L2TP 中存在两种消息：控制消息和数据消息。

- 控制消息用于隧道和会话连接的建立、维护以及传输控制。它的传输是可靠传输，并且支持对控制消息的流量控制和拥塞控制。
- 数据消息用于封装 PPP 帧，并在隧道上传输。它的传输是不可靠传输，若数据报文丢失，不予重传，不支持对数据消息的流量控制和拥塞控制。

控制消息和数据消息共享相同的报文头。L2TP 报文头中包含隧道标识符（Tunnel ID）和会话标识符（Session ID）信息，用来标识不同的隧道和会话。隧道标识相同、会话标识不同的报文将被复用在一个隧道上。报文头中的隧道标识符与会话标识符由对端分配。

1.1.4 L2TP 隧道模式及隧道建立过程

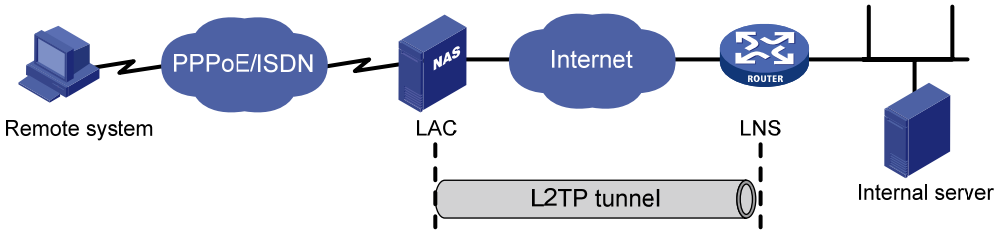
1. 两种典型的 L2TP 隧道模式

L2TP 隧道的建立包括以下两种典型模式。

- NAS-Intiated

如 图 1-4所示，由LAC端（指NAS）发起L2TP隧道连接。远程系统的拨号用户通过PPPoE//ISDN拨入LAC，由LAC通过Internet向LNS发起建立隧道连接请求。拨号用户地址由LNS分配；对远程拨号用户的验证与计费既可由LAC侧的代理完成，也可在LNS侧完成。

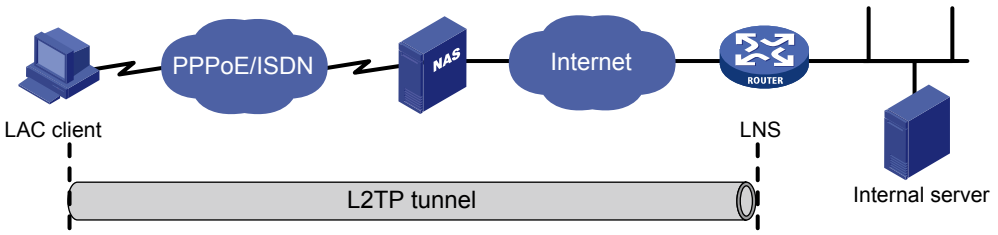
图1-4 NAS-Initiated L2TP 隧道模式



- Client-Initiated

如 图 1-5所示，直接由LAC客户（指可在本地支持L2TP协议的用户）发起L2TP隧道连接。此时LAC客户可直接向LNS发起隧道连接请求，无需再经过一个单独的LAC设备。LAC客户地址的分配由LNS来完成。

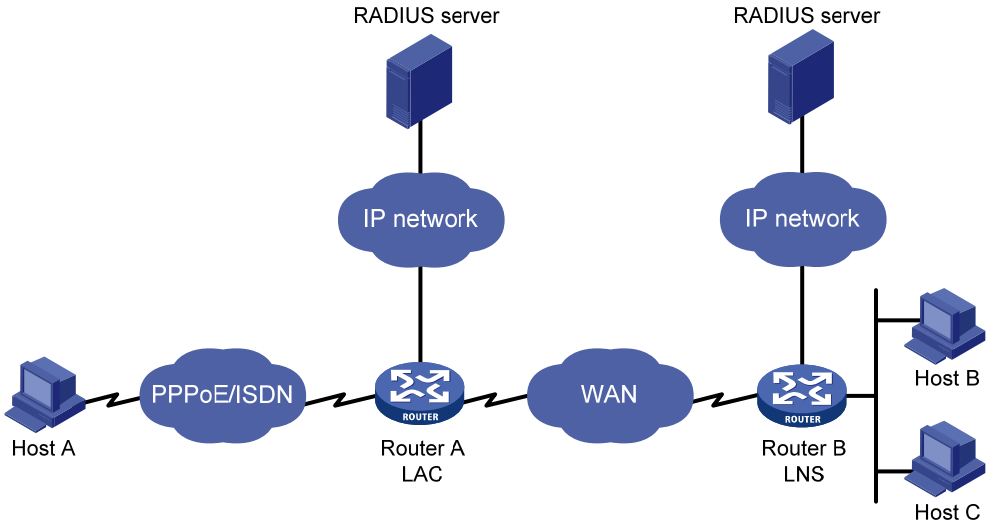
图1-5 Client-Initiated L2TP 隧道模式



2. L2TP 隧道的建立过程

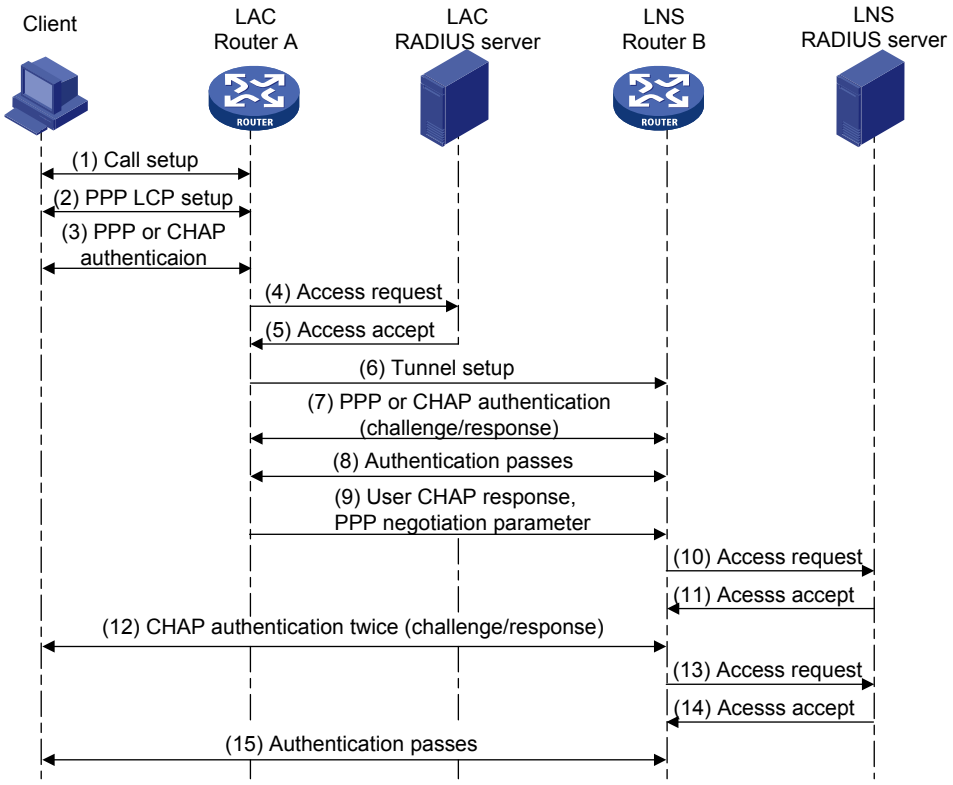
L2TP应用的典型组网如 图 1-6所示。

图1-6 L2TP 应用的典型组网



下面以 NAS-Initiated 模式的 L2TP 隧道为例，介绍 L2TP 的呼叫建立流程。

图1-7 L2TP 隧道的呼叫建立流程



如 图 1-7所示，L2TP隧道的呼叫建立流程过程为：

(1) 客户端 Host 发起呼叫连接请求；

(2) Host 和 LAC 端（RouterA）进行 PPP LCP 协商；

(3) LAC 对 Host 提供的用户信息进行 PAP 或 CHAP 认证；

(4) LAC 将认证信息（用户名、密码）发送给 RADIUS 服务器进行认证；

(5) RADIUS 服务器认证该用户，如果认证通过，LAC 准备发起 Tunnel 连接请求；

(6) LAC 端向指定 LNS 发起 Tunnel 连接请求；

(7) 在需要对隧道进行认证的情况下，LAC 端向指定 LNS 发送 CHAP challenge 信息，LNS 回送该 challenge 响应消息 CHAP response，并发送 LNS 侧的 CHAP challenge，LAC 返回该 challenge 的响应消息 CHAP response；

(8) 隧道验证通过；

(9) LAC 端将用户 CHAP response、response identifier 和 PPP 协商参数传送给 LNS；

(10) LNS 将接入请求信息发送给 RADIUS 服务器进行认证；

(11) RADIUS 服务器认证该请求信息，如果认证通过则返回响应信息；

(12) 若用户在 LNS 侧配置强制本端 CHAP 认证，则 LNS 对用户进行认证，发送 CHAP challenge，用户侧回应 CHAP response；

(13) LNS 再次将接入请求信息发送给 RADIUS 服务器进行认证；

(14) RADIUS 服务器认证该请求信息，如果认证通过则返回响应信息；

(15) 验证通过，LNS 端会给远端用户分配一个企业网内部 IP 地址，用户即可以访问企业内部资源。

1.1.5 L2TP 协议的特点

1. 灵活的身份验证机制以及高度的安全性

L2TP 协议本身并不提供连接的安全性，但它可依赖于 PPP 提供的认证（比如 CHAP、PAP 等），因此具有 PPP 所具有的所有安全特性。L2TP 可与 IPsec 结合起来实现数据安全，这使得通过 L2TP 所传输的数据更难被攻击。L2TP 还可根据特定的网络安全要求在 L2TP 之上采用隧道加密技术、端对端数据加密或应用层数据加密等方案来提高数据的安全性。

2. 多协议传输

L2TP 传输 PPP 数据包，这样就可以在 PPP 数据包内封装多种协议。

3. 支持 RADIUS 服务器的验证

LAC 端将用户名和密码发往 RADIUS 服务器进行验证申请，RADIUS 服务器负责接收用户的验证请求，完成验证。

4. 支持内部地址分配

LNS 可放置于企业网的防火墙之后，它可以对远端用户的地址进行动态的分配和管理，可支持私有地址应用（RFC 1918）。为远端用户所分配的地址不是 Internet 地址而是企业内部的私有地址，这样方便了地址的管理并可以增加安全性。

5. 网络计费的灵活性

可在 LAC 和 LNS 两处同时计费，即 ISP 处（用于产生帐单）及企业网关（用于付费及审计）。L2TP 能够提供数据传输的出入包数、字节数以及连接的起始、结束时间等计费数据，可根据这些数据方便地进行网络计费。

6. 可靠性

L2TP 协议支持备份 LNS，当主 LNS 不可达之后，LAC 可以与备份 LNS 建立连接，增加了 VPN 服务的可靠性和容错性。

下面从 LAC、LNS 两侧分别介绍 L2TP 的配置。

1.1.6 协议规范

与 L2TP 相关的协议规范有：

RFC1661：The Point-to-Point Protocol (PPP)

RFC1918：Address Allocation for Private Internets

RFC2661：Layer Two Tunneling Protocol "L2TP"

1.2 L2TP 配置任务简介

在将设备配置为 LAC 或 LNS 之前，必须首先使能 L2TP 功能、创建 L2TP 组，使设备具有基本的 L2TP 处理能力；然后根据不同的使用环境，分别进行 LAC 或 LNS 的特性配置，使设备具有独立的 L2TP 隧道端点功能。L2TP 的连接参数为可选配置，在设备作为 LAC、LNS 时均可使用，请根据实际情况配置。

表1-1 L2TP 配置任务简介

配置任务		说明	详细配置
配置 L2TP 基本功能	启用 L2TP 功能	必选	1.3
	创建 L2TP 组		
	配置隧道本端的名称		
配置 LAC 端	配置 LAC 发起 L2TP 连接请求	必选	1.4.1
	配置 AVP 数据的隐藏传输	可选	1.4.2
	配置 LAC 侧 VPN 用户认证	可选	1.4.3
配置 LNS 端	配置虚接口模板	必选	1.5.1
	配置本端地址及分配的地址池	可选	1.5.2
	配置 LNS 接收 L2TP 连接请求	必选	1.5.3
	配置 LNS 侧的用户验证	可选	1.5.4
	配置 LNS 侧 VPN 用户认证	可选	1.5.5
	配置 L2TP 多实例功能	可选	1.5.6
	配置 ACCM 消息发送	可选	1.5.7
配置 L2TP 连接参数	配置隧道验证	可选	1.6.1
	配置隧道 Hello 报文发送时间间隔		1.6.2
	开启 L2TP 隧道流控功能		1.6.3
	强制断开指定的隧道连接		1.6.4

1.3 配置 L2TP 基本功能

只有启用 L2TP 后，设备上的 L2TP 功能才能正常发挥作用；如果禁止 L2TP，则即使配置了 L2TP 的参数，设备也不会提供相关功能。

为了进行 L2TP 的相关参数配置，还需要增加 L2TP 组，这不仅可以在设备上灵活的配置 L2TP 各项功能，而且方便地实现了 LAC 和 LNS 之间一对一、一对多的组网应用。L2TP 组在 LAC 和 LNS

上独立编号，只需要保证 LAC 和 LNS 之间关联的 L2TP 组的相关配置（如隧道对端名称、LNS 地址等）保持对应关系即可。

隧道名在 LAC 和 LNS 进行隧道协商时使用。

表1-2 配置 L2TP 基本功能

配置步骤	命令	说明
进入系统视图	**system-view**	-
启用 L2TP 功能	**l2tp enable**	必选 缺省情况下，L2TP 功能处于关闭状态
创建 L2TP 组，并进入 L2TP 组视图	**l2tp-group** *group-number*	必选 缺省情况下，没有创建任何 L2TP 组
配置隧道本端的名称	**tunnel name** *name*	可选 缺省情况下，隧道本端的名称为系统的名称



说明

LAC 侧隧道名称要与 LNS 侧配置的接收 L2TP 连接请求的隧道对端名称保持一致。

1.4 配置 LAC 端

在 L2TP 的配置中，LAC 端和 LNS 端的配置有所不同，本节介绍 LAC 端基本特性的配置。LAC 的基本特性是当有 L2TP 用户上线时，和相应的 LNS 建立隧道连接，并将用户报文通过隧道发送到 LNS。

在各项配置任务中，必须先启动 L2TP、创建 L2TP 组，然后再进行其它功能特性的配置。

1.4.1 配置 LAC 发起 L2TP 连接请求

只有在满足一定的条件下，LAC 端才会向 LNS 服务器发出建立 L2TP 连接的请求。通过配置对接入用户信息的判别条件，并指定相应的 LNS 端的 IP 地址，设备可以鉴定用户是否为 VPN 用户，并决定是否向 LNS 发起连接。

LAC 发起 L2TP 连接请求时，支持两种用户名的触发：完整的用户名（**fullusername**）、带特定域名的用户名（**domain**）。

表1-3 配置 LAC 发起 L2TP 连接请求

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
配置本端作为 L2TP LAC 端时发起呼叫的触发条件	**start l2tp ip** *ip-address*&<1-5> { **domain** *domain-name* | **fullusername** *user-name* }	必选

说明

最多可以设置五个 LNS，即允许存在备用 LNS。正常运行时，LAC 按照 LNS 配置的先后顺序，依次向对端（LNS）进行 L2TP 连接请求，直到某个 LNS 接受连接请求，该 LNS 就成为 L2TP 隧道的对端。

1.4.2 配置 AVP 数据的隐藏传输

L2TP 协议的一些参数是通过 AVP（Attribute Value Pair，属性值对）数据来传输的，如果用户对这些数据的安全性要求高，可以使用本配置将 AVP 数据的传输方式配置成为隐藏传输。

表1-4 配置 AVP 数据的隐藏传输

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
配置隧道采用隐藏方式传输 AVP 数据	**tunnel avp-hidden**	可选 缺省情况下，隧道采用明文方式传输 AVP 数据

1.4.3 配置 LAC 侧 VPN 用户认证

通过在 LAC 侧配置对 VPN 用户的 AAA 认证，可以对远程拨入用户的身份信息（用户名、密码）进行验证。验证通过后，才能发起建立隧道连接的请求，否则将该用户转为其它类型的服务。

设备支持的 AAA 认证包括本地和远程两种认证方式：

- 如果选择本地认证方式，则需要在 LAC 侧配置本地用户名和密码。LAC 通过检查用户名与密码是否与本地注册用户名/密码相符合来进行用户身份验证。
- 如果选择远程认证方式，则需要在 RADIUS/HWTACACS 服务器上配置用户名和密码。LAC 将用户名和密码发往服务器进行验证申请，服务器负责进行用户身份验证。

表1-5 配置本地认证

配置步骤	命令	说明
进入系统视图	**system-view**	-
添加本地用户并进入本地用户视图	**local-user** *username*	必选 缺省情况下，LAC 侧未配置本地用户名和密码
设置本地用户的密码	**password** { **simple** | **cipher** } *password*	
授权用户可以使用 PPP 服务	**service-type ppp**	必选
退回系统视图	**quit**	-
创建一个 ISP 域，并进入 ISP 域视图	**domain** *isp-name*	必选
配置 PPP 域用户的 AAA 本地认证方案	**authentication ppp local**	可选 缺省情况下，采用 **local** 认证/授权/计费方案
	authorization ppp local	
	accounting ppp local	

说明

- 如果要使 LAC 端成功对用户进行认证，还需要在相应的 LAC 端的接口（例如接入用户的异步串口）下进行 PPP 的相关配置，具体步骤请参见“接入分册”中的“PPP 配置”。
- 配置本地验证的接口应该是接入用户的接口。
- 本小节只介绍本地认证方式的基本配置步骤，关于以上各配置命令的详细介绍及远程认证方式下的具体配置，请参见“安全分册”中的“AAA”。

1.5 配置 LNS 端

在 L2TP 的配置中，LAC 端和 LNS 端的配置有所不同，本节介绍 LNS 端基本特性的配置。LNS 的基本特性是响应 LAC 的隧道建立请求，对用户进行认证并为之分配 IP 地址。

在 LNS 的各项配置任务中，必须先启动 L2TP、创建 L2TP 组，然后再进行其它功能特性的配置。在 L2TP 支持多实例配置中，必须启用 L2TP 多实例功能。

1.5.1 配置虚接口模板

虚拟接口模板主要用于配置设备在运行过程中动态创建的虚拟访问接口的工作参数，如 MP 捆绑逻辑接口和 L2TP 逻辑接口等。

L2TP 会话连接建立之后，需要创建一个虚拟访问接口用于和对端交换数据。此时，系统将按照用户的配置，选择一个虚拟接口模板，根据该虚拟接口模板的配置参数，动态地创建一个虚拟访问接口

表1-6 配置虚接口模板

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入虚接口模板视图	**interface virtual-template** *virtual-template-number*	必选 缺省情况下，系统没有创建虚接口模板

1.5.2 配置本端地址及分配的地址池

当 LAC 与 LNS 之间的 L2TP 隧道连接建立之后，LNS 需要为 VPN 用户分配 IP 地址。地址分配包括两种方式：指定地址池或直接配置 IP 地址。在指定地址池之前，需要在系统视图或 ISP 域视图下用 **ip pool** 命令先定义一个地址池。对于需要进行认证的 VPN 用户，使用 ISP 域视图下的地址池分配 IP 地址；对于不需要进行认证的 VPN 用户，则使用系统视图下定义的全局地址池分配 IP 地址。

关于 **ip pool** 命令的详细描述，请参见“安全分册”中的“AAA RADIUS HWTACACS 命令”。

表1-7 配置本端地址及分配的地址池

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入虚接口模板视图	**interface virtual-template** *virtual-template-number*	-
配置本端 IP 地址	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	必选

配置步骤	命令	说明
配置本端对 PPP 用户进行验证	**ppp authentication-mode { chap | pap } [[call-in] domain** *isp-name*]	可选 缺省情况下，不进行认证
指定给对端分配地址所用的地址池或直接给对端分配 IP 地址	**remote address { pool** [*pool-number*] | *ip-address* }	可选 缺省情况下，从地址池 0（缺省地址池）中给对方分配地址

1.5.3 配置 LNS 接收 L2TP 连接请求

LNS 可以使用不同的虚拟接口模板接收不同 LAC 创建隧道的请求。接收到 LAC 发来的创建隧道请求后，LNS 需要检查 LAC 的名称是否与合法隧道对端名称相符合，从而决定是否允许隧道对方创建隧道。

表1-8 配置 LNS 接收 L2TP 连接请求

配置步骤		命令	说明
进入系统视图		**system-view**	-
进入 L2TP 组视图		**l2tp-group** *group-number*	-
指定接收呼叫的虚拟接口模板、隧道对端名称和域名	L2TP 组号不为 1	**allow l2tp virtual-template** *virtual-template-number* **remote** *remote-name* [**domain** *domain-name*]	二者必选其一 缺省情况下，禁止接受呼入
	L2TP 组号为 1（缺省的 L2TP 组号）	**allow l2tp virtual-template** *virtual-template-number* [**remote** *remote-name*] [**domain** *domain-name*]	

说明

- 命令 **start l2tp** 和 **allow l2tp** 是互斥的，配了一条命令之后另一条就自动失效。
- LNS 侧接收 L2TP 连接请求的隧道对端名称要与 LAC 侧配置的隧道名称保持一致。

1.5.4 配置 LNS 侧的用户验证

当 LAC 对用户进行验证后，LNS 可再次对用户进行验证，可以增加安全性。此时将对用户进行两次验证，第一次发生在 LAC 侧，第二次发生在 LNS 侧，只有两次验证全部成功后，L2TP 隧道才能建立。

在 L2TP 组网中，LNS 侧对用户的验证方式有三种：代理验证、强制 CHAP 验证和 LCP 重协商。

如果既不配置 LCP 重协商，也不配置强制 CHAP 验证，则 LNS 对用户进行的是代理验证。在这种情况下，LAC 将它从用户得到的所有验证信息及 LAC 端本身配置的验证方式发送给 LNS。

这三种验证方式中，LCP 重协商的优先级最高，如果在 LNS 上同时配置 LCP 重协商和强制 CHAP 验证，L2TP 将使用 LCP 重协商，并采用相应的虚拟模板接口上配置的验证方式。

如果只配置强制 CHAP 验证，则 LNS 对用户进行 CHAP 验证。

1. 配置强制本端 CHAP 验证

配置强制本端 CHAP 验证，则对于由 NAS 初始化（NAS-Initiated）隧道连接的 VPN 的用户端来说，会经过两次验证：一次是用户端在接入服务器端的验证，另一次是用户端在 LNS 端的 CHAP 验证。

表1-9 配置强制本端 CHAP 验证

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
配置强制 LNS 与用户端之间重新进行 CHAP 验证	**mandatory-chap**	必选 缺省情况下，系统不进行 CHAP 的重新验证



说明

- 在 LNS 使用代理验证时，如果虚拟模板接口配置的验证方式为 PAP，并且验证通过，则允许建立会话；如果虚拟模板接口配置的验证方式为 CHAP，而 LAC 端配置的验证方式为 PAP，则由于 LNS 要求的 CHAP 验证级别高于 LAC 能够提供的 PAP 验证，验证将无法通过，会话也就不能正确建立。
- 一些 PPP 用户端可能不支持进行第二次验证，这时，本端的 CHAP 验证会失败。

2. 配置强制 LCP 重新协商

对于由 NAS 初始化（NAS-Initiated）隧道连接的 VPN 的用户端，在 PPP 会话开始时，用户先和 NAS 进行 PPP 协商。若协商通过，则由 NAS 初始化 L2TP 隧道连接，并将用户信息传递给 LNS，由 LNS 根据收到的代理验证信息，判断用户是否合法。

但在某些特定的情况下（如需在 LNS 侧也要进行验证与计费），需要强制 LNS 与用户间重新进行 LCP（Link Control Protocol，链路控制协议）协商，此时将忽略 NAS 侧的代理验证信息。

启用 LCP 重协商后，如果相应的虚拟接口模板上不配置验证，则 LNS 将不对接入用户进行二次验证（这时用户只在 LAC 侧接受一次验证），直接将全局地址池的地址分配给 PPP 用户端。

表1-10 配置强制本端 LCP 重新协商

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
配置强制 LNS 与用户端之间重新进行 LCP 的协商	**mandatory-lcp**	必选 缺省情况下，系统不重新进行 LCP 协商

1.5.5 配置 LNS 侧 VPN 用户认证

在 LNS 侧，如果配置了强制 CHAP 认证，则需要在 LNS 侧配置 AAA 认证。通过在 LAC 侧配置对 VPN 用户的 AAA 认证，可以对远程拨入用户的身份信息（用户名、密码）进行验证，验证通过后就可以进行 VPN 用户和 LNS 的通信，否则将通知 L2TP 清除此 L2TP 连接。

LNS侧支持的AAA配置与LAC侧的相同，具体介绍及配置方法请参考 1.4.3 配置LAC侧VPN用户认证。

1.5.6 配置 L2TP 多实例功能

只有启用 L2TP 多实例功能，设备才能为多个企业做 LNS。L2TP 多实例功能的实现丰富了 VPN 组网方式，主要应用在 MPLS-VPN 组网中。在实际组网应用中，企业的私网设备需要通过配置 VPN 实例来实现。

关于实例的配置请参见“MPLS 分册”中的“MPLS L3VPN 配置”。

表1-11 配置 LNS

配置步骤	命令	说明
进入系统视图	**system-view**	-
启用 L2TP 多实例功能	**l2tpmoreexam enable**	必选 缺省情况下，L2TP 多实例功能处于关闭状态

说明

在 L2TP 多实例的应用中，如果 LNS 侧不同 L2TP 组中配置了相同的隧道对端名称，则 L2TP 组的隧道验证配置也必须相同，否则会因为隧道验证密码不一致造成隧道和会话建立失败。

1.5.7 配置 ACCM 消息发送

RFC 2661 中提到，LNS 可以通过 ACCM 消息，将 LNS 与 PPP 对端协商的 ACCM 情况通知 LAC。

实际应用中，各设备制造商的 LAC 对 ACCM 要求可能不同，LNS 需要根据 LAC 的要求配置是否发送 ACCM 消息。

缺省情况下，LNS 发送 ACCM 消息。如果 LAC 要求不能收到 ACCM，需要设置 LNS 端不发送 ACCM 消息。

表1-12 配置 ACCM 消息发送

配置步骤	命令	说明
进入系统视图	**system-view**	-
启用 L2TP 的 ACCM 消息发送功能	**l2tp sendaccm enable**	必选 缺省情况下，L2TP 的 ACCM 消息发送功能处于开启状态

1.6 配置 L2TP 连接参数

本节介绍 L2TP 连接的可选配置，这些配置在设备作为 LAC、LNS 时均可使用。

L2TP 连接的可选配置包括隧道验证、Hello 报文发送时间间隔、开启 L2TP 隧道流控功能等。

1.6.1 配置隧道验证

用户可根据实际需要，决定是否在创建隧道连接之前启用隧道验证。

隧道验证请求可由 LAC 或 LNS 任何一侧发起。只要有一方启用了隧道验证，则只有在对端也启用了隧道验证，两端密码完全一致并且不为空的情况下，隧道才能建立；否则本端将自动将隧道连接断开。

若隧道两端都配置了禁止隧道验证，隧道验证的密码一致与否将不起作用。

表1-13 配置隧道验证

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
启用 L2TP 的隧道验证功能	**tunnel authentication**	可选 缺省情况下，L2TP 隧道进行验证功能处于开启状态
配置隧道验证时的密码	**tunnel password** { **simple** | **cipher** } *password*	必选 缺省情况下，系统的隧道验证密码为空

说明

为了保证隧道安全，建议用户最好不要禁用隧道验证的功能。如果用户需要修改隧道验证的密码，请在隧道完全拆除后进行，否则修改的密码不生效。

1.6.2 配置隧道 Hello 报文发送时间间隔

为了检测 LAC 和 LNS 之间隧道的连通性，LAC 和 LNS 会定期向对端发送 Hello 报文，接收方接收到 Hello 报文后会进行响应。当 LAC 或 LNS 在指定时间间隔内未收到对端的 Hello 响应报文时，重复发送，如果重复发送超过 3 次都没有收到对端的响应信息则认为 L2TP 隧道已经断开，需要在 LAC 和 LNS 之间重新建立隧道连接。

表1-14 配置隧道 Hello 报文发送时间间隔

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
配置隧道 Hello 报文发送时间间隔	**tunnel timer hello** *hello-interval*	可选 缺省情况下，隧道中 Hello 报文的发送时间间隔为 60 秒

1.6.3 开启 L2TP 隧道流控功能

L2TP 隧道的流控功能应用在数据报文的接收与发送过程中。对于数据报文，可能存在接收与发送中的乱序。使用流控功能后，会对乱序报文进行缓存和调整。

表1-15 开启 L2TP 隧道流控功能

配置步骤	命令	说明
进入系统视图	**system-view**	-
进入 L2TP 组视图	**l2tp-group** *group-number*	-
开启 L2TP 隧道流控功能	**tunnel flow-control**	可选 缺省情况下，关闭 L2TP 隧道流控功能

1.6.4 强制断开指定的隧道连接

当用户数为零、网络发生故障或当管理员主动要求时，都会产生断开连接的过程。LAC 和 LNS 任何一端都可主动发起断开连接的请求，连接断开后，该隧道上的所有控制连接与会话连接也将被清除；当有新用户拨入时，还可重新建立连接。

表1-16 强制断开指定的隧道连接

配置步骤	命令	说明
强制断开指定的隧道连接	**reset l2tp tunnel** { **id** *tunnel-id* | **name** *remote-name* }	可选 该配置在用户视图下执行

1.7 L2TP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 L2TP 的运行情况，通过查看显示信息验证配置的效果。

表1-17 L2TP 显示和维护

操作	命令
显示当前 L2TP 隧道的信息	**display l2tp tunnel**
显示当前 L2TP 会话的信息	**display l2tp session**

1.8 L2TP 典型配置举例

L2TP 的呼叫可以由 NAS 主动发起，也可以由客户端发起。下面将分别针对这两种情况举例说明。

1.8.1 NAS-Initiated VPN

1. 组网需求

VPN 用户访问公司总部过程如下：

(1) 用户以普通的上网方式进行拨号上网。

(2) 在 NAS 处对此用户进行验证，发现是 VPN 用户，则由接入服务器向 LNS 发起隧道连接的请求。

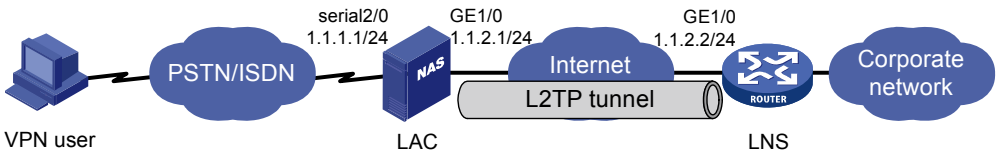
(3) 在接入服务器与 LNS 建立隧道后，接入服务器把与 VPN 用户已经协商的内容作为报文内容传给 LNS。

(4) LNS 再根据预协商的内容决定是否接受此连接。

(5) 用户与公司总部间的通信都通过接入服务器与 LNS 之间的隧道进行传输。

2. 组网图

图1-8 NAS-Initiated VPN 组网图



3. 配置步骤

(1) 用户侧的配置

在拨号网络窗口中输入用户名 vpdnuser 和密码 Hello 进行拨号。

(2) LAC 侧的配置

在 NAS 上配置：

配置各接口的 IP 地址（略）。

设置用户名及口令。

```
[LAC] local-user vpdnuser
[LAC-luser-vpdnuser] password simple Hello
[LAC-luser-vpdnuser] service-type ppp
[LAC-luser-vpdnuser] quit
```

在 Serial1/0 接口上配置 IP 地址。

```
[LAC] interface serial 2/0
[LAC-serial2/0] ip address 1.1.1.1 255.255.255.0
[LAC-serial2/0] ppp authentication-mode chap
[LAC-serial2/0] quit
```

启用 L2TP 服务。

```
[LAC] l2tp enable
```

设置一个 L2TP 组，并配置相关属性。

```
[LAC] l2tp-group 1
[LAC-l2tp1] tunnel name LAC
[LAC-l2tp1] start l2tp ip 1.1.2.2 fullusername vpdnuser
```

启用通道验证并设置通道验证密码。

```
[LAC-l2tp1] tunnel authentication
[LAC-l2tp1] tunnel password simple aabbcc
```

(3) LNS 侧的配置

配置各接口的 IP 地址（略）。

设置用户名及密码（应与用户侧的设置一致）。

```
<LNS> system-view
[LNS] local-user vpdnuser
[LNS-luser-vpdnuser] password simple Hello
[LNS-luser-vpdnuser] service-type ppp
[LNS-luser-vpdnuser] quit
```

对 VPN 用户采用本地验证。

```
[LNS] domain system
[LNS-isp-system] authentication ppp local
[LNS-isp-system] ip pool 1 192.168.0.2 192.168.0.100
[LNS-isp-system] quit
```

启用 L2TP 服务。

```
[LNS] l2tp enable
```

配置虚模板 Virtual-Template 的相关信息。

```
[LNS] interface virtual-template 1
[LNS-virtual-template1] ip address 192.168.0.1 255.255.255.0
[LNS-virtual-template1] ppp authentication-mode chap domain system
[LNS-virtual-template1] remote address pool 1
[LNS-virtual-template1] quit
```

设置一个 L2TP 组，指定接收呼叫的虚拟接口模板，配置 LNS 侧的隧道对端名称。

```
[LNS] l2tp-group 1
[LNS-l2tp1] tunnel name LNS
[LNS-l2tp1] allow l2tp virtual-template 1 remote LAC
```

启用隧道验证并设置隧道验证密码。

```
[LNS-l2tp1] tunnel authentication
[LNS-l2tp1] tunnel password simple aabbcc
```

4. 验证配置结果

通过命令 **display l2tp tunnel** 可查看建立的 L2TP 隧道。

```
[LNS] dis l2tp tunnel
Total tunnel = 1

 LocalTID RemoteTID RemoteAddress    Port   Sessions RemoteName
 1        1         1.1.2.1          1701   1        LAC
```

通过命令 **display l2tp session** 可查看建立的 L2TP 会话。

```
[LNS] display l2tp session
 Total session = 1

 LocalSID  RemoteSID  LocalTID
  23142     729        1AC
```

1.8.2 Client-Initiated VPN

1. 组网需求

VPN 用户访问公司总部过程如下：

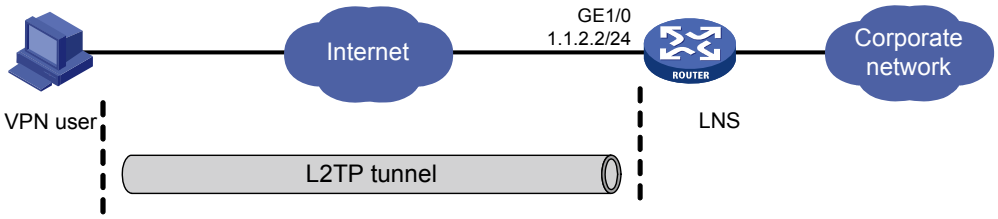
(1) 用户首先连接 Internet，之后直接由用户向 LNS 发起 Tunnel 连接的请求。

(2) 在 LNS 接受此连接请求之后，VPN 用户与 LNS 之间就建立了一条虚拟的 L2TP tunnel。

(3) 用户与公司总部间的通信都通过 VPN 用户与 LNS 之间的隧道进行传输。

2. 组网图

图1-9 Client-Initiated VPN 组网图



3. 配置步骤

(1) 用户侧的配置

在用户侧主机上必须装有 L2TP 的客户端软件，如 WinVPN Client，并且用户通过拨号方式连接到 Internet。然后再进行如下配置（设置的过程与相应的客户端软件有关，以下为设置的内容）：

在用户侧设置 VPN 用户名为 vpdnuser，密码为 Hello。

将 LNS 的 IP 地址设为安全网关的 Internet 接口地址（本例中 LNS 侧与隧道相连接的以太网接口的 IP 地址为 1.1.2.2）。

修改连接属性，将采用的协议设置为 L2TP，将加密属性设为自定义，并选择 CHAP 验证，进行隧道验证，隧道的密码为：aabbcc。

(2) LNS 侧的配置

设置用户名及密码（应与用户侧的设置一致）。

```
<LNS> system-view
[LNS] local-user vpdnuser
[LNS-luser-vpdnuser] password simple Hello
[LNS-luser-vpdnuser] service-type ppp
[LNS-luser-vpdnuser] quit
```

对 VPN 用户采用本地验证。

```
[LNS] domain system
[LNS-isp-system] authentication ppp local
[LNS-isp-system] ip pool 1 192.168.0.2 192.168.0.100
[LNS-isp-system] quit
```

启用 L2TP 服务。

```
[LNS] l2tp enable
```

配置虚模板 Virtual-Template 的相关信息。

```
[LNS] interface virtual-template 1
[LNS-virtual-template1] ip address 192.168.0.1 255.255.255.0
[LNS-virtual-template1] ppp authentication-mode chap domain system
[LNS-virtual-template1] remote address pool 1
[LNS-virtual-template1] quit
```

设置一个 L2TP 组，指定接收呼叫的虚拟接口模板。

```
[LNS] l2tp-group 1
[LNS-l2tp1] tunnel name LNS
```

```
[LNS-l2tp1] allow l2tp virtual-template 1
```

启用隧道验证并设置隧道验证密码。

```
[LNS-l2tp1] tunnel authentication
[LNS-l2tp1] tunnel password simple aabbcc
```

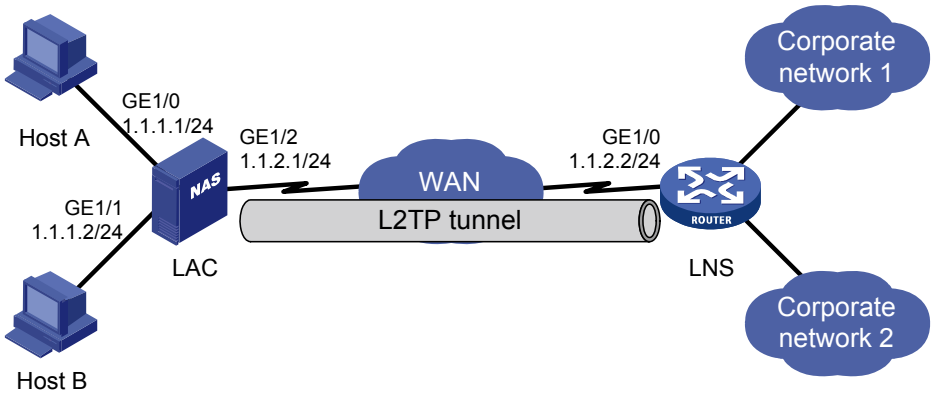
1.8.3 L2TP 多域组网应用

1. 组网需求

- 多个企业共用一个 LNS，不同的企业用户需要与自己的总部进行通讯，网络的地址采用的是私有地址。
- 一般情况下，用户无法通过 Internet 直接访问企业内部的服务器。通过建立 VPN 并支持多域，用户就可以访问自己企业内部网络的数据。

2. 组网图

图1-10 L2TP 支持多域组网图



3. 配置步骤

企业 1 的域名为 aaa.net，Host A 为企业 1 用户；

企业 2 的域名为 bbb.net，Host B 为企业 2 用户。

(1) 用户侧的配置

在用户侧配置拨号连接，拨号号码为 LAC 的接入号码，拨号用户接受由 LNS 分配的地址。

对于 Host A 而言，在弹出的拨号终端窗口中输入用户名 vpdn1@aaa.net，密码为 11111（此用户名与密码已在 LNS 中注册）。

对于 Host B 而言，在弹出的拨号终端窗口中输入用户名 vpdn2@bbb.net，密码为 22222（此用户名与密码已在 LNS 中注册）。

(2) LAC 侧的配置

（本例中 LAC 侧的 GigabitEthernet1/0 和 GigabitEthernet1/1 是用户接入接口，LAC 侧与隧道相连接的 GigabitEthernet1/2 的 IP 地址为 1.1.2.1，LNS 侧与隧道相连接的 GigabitEthernet1/0 的 IP 地址为 1.1.2.2。）

设置用户名及密码。

```
<LAC> system-view
[LAC] local-user vpdn1
```

```
[LAC-luser-vpdn1] password simple 11111
[LAC-luser-vpdn1] service-type ppp
[LAC-luser-vpdn1] quit
[LAC] local-user vpdn2
[LAC-luser-vpdn2] password simple 22222
[LAC-luser-vpdn2] service-type ppp
[LAC-luser-vpdn2] quit
```

配置域用户采用本地认证。

```
[LAC] domain aaa.net
[LAC-isp-aaa.net] authentication ppp local
[LAC-isp-aaa.net] quit
[LAC] domain bbb.net
[LAC-isp-bbb.net] authentication ppp local
[LAC-isp-bbb.net] quit
```

在 GigabitEthernet1/0 和 GigabitEthernet1/1 接口上配置 PPPoE server。

```
[LAC] interface gigabitethernet 1/0
[LAC-GigabitEthernet1/0] pppoe-server bind virtual-template 100
[LAC-GigabitEthernet1/0] quit
[LAC] interface gigabitethernet 1/1
[LAC-GigabitEthernet1/1] pppoe-server bind virtual-template 101
[LAC-GigabitEthernet1/1] quit
```

在 GigabitEthernet1/2 接口上配置 IP 地址。

```
[LAC] interface gigabitethernet 1/2
[LAC-GigabitEthernet1/2] ip address 1.1.2.1 255.255.255.0
[LAC-GigabitEthernet1/2] quit
```

在虚模板上启动 CHAP 认证。

```
[LAC] interface virtual-template 100
[LAC-Virtual-Template100] ppp authentication-mode chap domain aaa.net
[LAC-Virtual-Template100] quit
[LAC] interface virtual-template 101
[LAC-Virtual-Template101] ppp authentication-mode chap domain bbb.net
[LAC-Virtual-Template101] quit
```

设置两个 L2TP 组并配置相关属性。

```
[LAC] l2tp enable
[LAC] l2tp-group 1
[LAC-l2tp1] tunnel name LAC-1
[LAC-l2tp1] start l2tp ip 1.1.2.2 domain aaa.net
[LAC-l2tp1] quit
[LAC] l2tp-group 2
[LAC-l2tp2] tunnel name LAC-2
[LAC-l2tp2] start l2tp ip 1.1.2.2 domain bbb.net
```

启用隧道验证并设置隧道验证密码。

```
[LAC-l2tp2] tunnel authentication
[LAC-l2tp2] tunnel password simple 12345
[LAC-l2tp2] quit
```

```
[LAC] l2tp-group 1
[LAC-l2tp1] tunnel authentication
[LAC-l2tp1] tunnel password simple 12345
```

(3) LNS 侧的配置

```
<LNS> system-view
[LNS] l2tp enable
[LNS] l2tpmoreexam enable
```

创建两个用户名及密码

```
[LNS] local-user vpdn1
[LNS-luser-vpdn1] password simple 11111
[LNS-luser-vpdn1] service-type ppp
[LNS-luser-vpdn1] quit
[LNS] local-user vpdn2
[LNS-luser-vpdn2] password simple 22222
[LNS-luser-vpdn2] service-type ppp
[LNS-luser-vpdn2] quit
```

配置 LNS 侧与隧道相连接的 GigabitEthernet1/0 的 IP 地址为 1.1.2.2。

```
[LNS] interface gigabitethernet 1/0
[LNS-GigabitEthernet1/0] ip address 1.1.2.2 255.255.255.0
[LNS-GigabitEthernet1/0] quit
```

创建两个地址池。

```
[LNS] domain bbb.net
[LNS-isp-bbb.net] authentication ppp local
[LNS-isp-bbb.net] ip pool 1 10.0.1.10 10.0.1.100
[LNS-isp-bbb.net] quit
[LNS] domain aaa.net
[LNS-isp-aaa.net] authentication ppp local
[LNS-isp-aaa.net] ip pool 1 10.0.2.10 10.0.2.100
[LNS-isp-aaa.net] quit
```

创建两个相应的 virtual template。

```
[LNS]interface virtual-template 1
[LNS-Virtual-Template1] ip address 10.0.2.1 255.255.255.0
[LNS-Virtual-Template1] remote address pool 1
[LNS-Virtual-Template1] ppp authentication-mode chap domain aaa.net
[LNS-Virtual-Template1] quit
[LNS] interface virtual-template 2
[LNS-Virtual-Template2] ip address 10.0.1.1 255.255.255.0
[LNS-Virtual-Template2] remote address pool 1
[LNS-Virtual-Template2] ppp authentication-mode chap domain bbb.net
[LNS-Virtual-Template2] quit
```

创建两个相应的 L2TP-group 组。

```
[LNS] l2tp-group 3
[LNS-l2tp3] tunnel name LNS
[LNS-l2tp3] tunnel authentication
[LNS-l2tp3] allow l2tp virtual-template 1 remote LAC-1 domain aaa.net
```

```
[LNS-l2tp3] tunnel password simple 12345
[LNS-l2tp3] quit
[LNS] l2tp-group 4
[LNS-l2tp4] tunnel name LNS
[LNS-l2tp4] tunnel authentication
[LNS-l2tp4] allow l2tp virtual-template 2 remote LAC-2 domain bbb.net
[LNS-l2tp4] tunnel password simple 12345
```

上述配置中，如果 LNS 端需要采用 RADIUS 验证，请修改 AAA 配置即可。关于 AAA 的具体配置请参考“安全分册”中的“AAA 配置”。

1.8.4 复杂的组网情况

安全网关支持同时作为 LAC 及 LNS，并支持同时有多路用户呼入；只要内存及线路不受限制，L2TP 可以同时接收和发起多个呼叫。这些复杂组网的需求及配置可以综合参考以上的几种组网情况，综合应用。

特别需要注意的是静态路由的配置，许多应用是依靠路由来发起的。

1.9 常见配置错误举例

VPN 创建连接的过程比较复杂，这里就几种常见的情况进行分析。在进行 VPN 排错之前，请先确认 LAC 与 LNS 已在公共网上建立了正确的连接。

1.9.1 错误之一

1. 错误现象

用户登录失败

2. 错误排除

用户登录失败主要有以下几种原因：

(1) Tunnel 建立失败，Tunnel 不能建立的原因有：

- 在 LAC 端，LNS 的地址设置不正确。
- LNS（通常为路由器）端没有设置可以接收该隧道对端的 L2TP 组，具体可以查看 **allow** 命令的说明。
- Tunnel 验证不通过，如果配置了验证，应该保证双方都启用了隧道验证并且配置的隧道密码一致。
- 如果是本端强制挂断了连接，但由于网络传输时延等原因，对端还没有收到相应的 Disconnect 报文，此时对端发起了一个隧道连接，会连不上，因为对端必须相隔一定的时间才能侦测到链路被挂断。

(2) PPP 协商不通过，可能原因有：

- LAC 端设置的用户名与密码有误，或者是 LNS 端没有设置相应的用户。
- LNS 端不能分配地址，比如地址池设置的较小，或没有进行设置。
- 密码验证类型不一致。如 Windows 2000 所创建的 VPN 连接缺省的验证类型为 MSCHAP，如果对端不支持 MSCHAP，建议改为 CHAP。

1.9.2 错误之二

1. 错误现象

数据传输失败，在建立连接后数据不能传输，如 Ping 不通对端。

2. 错误排除

可能有如下原因：

- 用户设置的地址有误：一般情况下，由 LNS 分配地址，而用户也可以指定自己的地址。如果指定的地址和 LNS 所要分配的地址不属于同一个网段，就会发生这种情况，建议由 LNS 统一分配地址。
- 网络拥挤：Internet 主干网产生拥挤，丢包现象严重。L2TP 是基于 UDP（用户数据报文）进行传输的，UDP 不对报文进行差错控制；如果是在线路质量不稳定的情况下进行 L2TP 应用，有可能会产生 Ping 不通对端的情况。

目 录

1 QoS简介 1-1
1.1 概述 1-1
1.2 无QoS保障的网络 1-1
1.3 新业务对QoS的需求 1-1
1.4 拥塞的产生、影响和对策 1-1
1.4.1 拥塞的产生 1-1
1.4.2 拥塞的影响 1-2
1.4.3 对策 1-2
1.5 QoS技术的实现思路 1-3
1.5.1 端到端的QoS 1-3
1.5.2 流分类技术 1-3
1.5.3 报文优先级 1-4
2 两种QoS配置方式介绍 2-1
3 QoS策略配置 3-1
3.1 QoS策略简介 3-1
3.2 配置QoS策略 3-1
3.2.1 定义类 3-1
3.2.2 定义流行为 3-2
3.2.3 定义策略 3-3
3.2.4 QoS策略配置举例 3-3
3.3 应用QoS策略 3-3
3.3.1 基于接口或PVC应用QoS策略 3-4
3.3.2 基于上线用户应用QoS策略 3-4
3.4 QoS策略显示和维护 3-5
4 流量监管和流量整形配置 4-1
4.1 流量监管和流量整形简介 4-1
4.1.1 流量评估与令牌桶 4-1
4.1.2 流量监管 4-2
4.1.3 流量整形 4-3
4.1.4 物理接口限速 4-3
4.2 流量监管/流量整形/物理接口限速配置 4-4
4.2.1 流量监管配置 4-4
4.2.2 流量整形配置 4-6
4.2.3 物理接口限速配置 4-7
4.3 流量监管/流量整形/物理接口限速显示和维护 4-8
4.4 流量监管与流量整形典型配置举例 4-8
4.4.1 流量监管与流量整形典型配置举例 4-8
4.4.2 IP限速典型配置举例 4-10

5 拥塞管理配置 5-1
5.1 拥塞管理简介 5-1
5.1.1 拥塞管理策略 5-1
5.1.2 拥塞管理技术的对比 5-5
5.2 先进先出队列的配置 5-6
5.2.1 FIFO队列配置过程 5-6
5.2.2 FIFO队列配置过程举例 5-7
5.3 优先队列的配置 5-7
5.3.1 优先队列配置过程 5-7
5.3.2 优先队列配置过程举例 5-8
5.4 定制队列的配置 5-9
5.4.1 定制队列配置过程 5-9
5.4.2 定制队列配置过程举例 5-9
5.5 加权公平队列的配置 5-10
5.5.1 加权公平队列配置过程 5-10
5.5.2 加权公平队列配置过程举例 5-10
5.6 基于类的队列的配置 5-11
5.6.1 配置接口最大可用带宽 5-12
5.6.2 定义类 5-13
5.6.3 定义流行为 5-13
5.6.4 定义策略 5-17
5.6.5 应用策略 5-17
5.6.6 基于类的队列典型配置举例 5-18
5.6.7 基于类的队列的显示和维护 5-20
5.7 RTP优先队列的配置 5-20
5.7.1 RTP优先队列配置过程 5-20
5.7.2 RTP优先队列配置过程举例 5-21
5.8 QoS令牌配置 5-21
5.8.1 QoS令牌配置过程 5-21
5.8.2 QoS令牌配置举例 5-22
6 拥塞避免 6-1
6.1 拥塞避免简介 6-1
6.2 WRED配置的说明 6-2
6.2.1 WRED的配置方式 6-2
6.2.2 WRED的参数说明 6-2
6.3 以接口配置方式配置WRED 6-3
6.3.1 配置过程 6-3
6.3.2 配置举例 6-3
6.4 WRED显示和维护 6-4
6.5 WRED典型配置举例 6-4
7 MPLS QoS配置 7-1
7.1 MPLS QoS概述 7-1
7.2 MPLS QoS配置 7-1
7.2.1 配置MPLS PQ 7-1

7.2.2 配置MPLS CQ 7-2
7.2.3 配置MPLS QoS策略 7-3
7.2.4 配置MPLS CAR 7-3
7.3 MPLS QoS配置举例 7-4
7.3.1 对同一VPN内的流进行QoS配置 7-4
8 DAR配置 8-1
8.1 DAR简介 8-1
8.2 配置DAR 8-1
8.2.1 配置准备 8-1
8.2.2 配置P2P协议流量识别功能 8-1
8.2.3 配置协议的匹配规则 8-3
8.2.4 配置DAR的报文统计功能 8-3
8.3 DAR显示和维护 8-3
8.4 DAR典型配置举例 8-4
8.4.1 禁止P2P软件下载配置举例 8-4

1 QoS 简介

1.1 概述

QoS（Quality of Service，服务质量）用于评估服务方满足客户服务需求的能力。在 Internet 中，QoS 所评估的就是网络转发分组的服务能力。

由于网络提供的服务是多样的，因此 QoS 的评估可以基于不同方面。通常所说的 QoS，是对分组转发过程中带宽、延迟、抖动、丢包率等指标进行评估。

1.2 无 QoS 保障的网络

在传统的无 QoS 保障的 IP 网络中，设备无区别地对待所有的报文，设备处理报文采用的策略是 FIFO（First In First Out，先入先出），它依照报文到达时间的先后顺序分配转发所需要的资源。所有报文共享网络和设备的资源，至于得到资源的多少完全取决于报文到达的时间。这种服务策略称作 Best-Effort，它尽最大的努力将报文送到目的地，但对分组转发的延迟、抖动、丢包率等需求不提供任何承诺和保证。

传统的 Best-Effort 服务策略只适用于对带宽、延迟不敏感的 WWW（World Wide Web，万维网）、E-Mail 等业务。

1.3 新业务对 QoS 的需求

随着计算机网络的高速发展，越来越多的网络接入 Internet。Internet 无论从规模、覆盖范围还是和用户数量上来看，Internet 都扩展得非常快。

除了传统的 WWW、E-Mail 应用外，用户还尝试在 Internet 上拓展新业务，比如远程教学、远程医疗、可视电话、电视会议、视频点播等。企业用户也希望通过 VPN 技术，将分布在各地的分支机构连接起来，开展一些事务性应用：比如访问公司的数据库或通过 Telnet 管理远程设备。

这些新业务有一个共同特点，即对带宽、延迟、抖动等传输性能有着特殊的需求。比如电视会议、视频点播需要高带宽、低延迟和低抖动的保证。事务处理、Telnet 等关键任务虽然不一定要求高带宽，但非常注重低延迟，在拥塞发生时要求优先获得处理。

新业务的不断涌现对 IP 网络的服务能力提出了更高的要求，用户已不再满足于能够简单地将报文送达目的地，还希望在转发过程中得到更好的服务，诸如为用户提供专用带宽、减少报文的丢失率、管理和避免网络拥塞、调控网络的流量。所有这些都要求网络具备更为完善的服务能力。

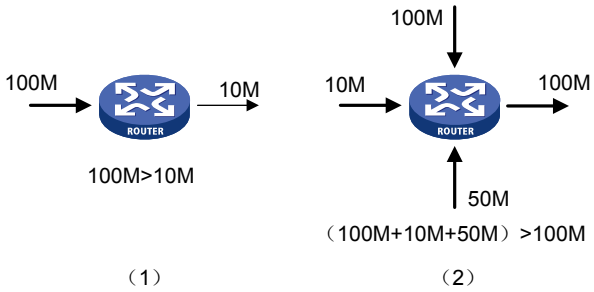
1.4 拥塞的产生、影响和对策

传统网络所面临的服务质量问题，主要是由网络拥塞引起的。所谓拥塞，是指由于供给资源的相对不足而造成转发速率下降、引入额外的延迟，从而导致服务质量下降的一种现象。

1.4.1 拥塞的产生

在复杂 Internet 分组交换环境下，拥塞极为常见。以下图中的两种情况为例：

图1-1 流量拥塞示意图



(1) 分组流从高速链路进入设备，由低速链路转发出去。

(2) 分组流从多个接口同时进入网络设备，由一个接口转发出去（多个接口的速率和大于出接口的速率）。

如果流量以线速到达，那么就会遭遇资源的瓶颈而导致拥塞。

不仅仅是链路带宽的瓶颈会导致拥塞，任何用于正常转发处理的资源的不足（如可分配的处理器时间、缓冲区、内存资源的不足）都会造成拥塞。此外，在某个时间内对所到达的流量控制不力，使之超出了可分配的网络资源，也是引发网络拥塞的一个因素。

1.4.2 拥塞的影响

拥塞有可能会引发一系列的负面影响：

- 拥塞增加了报文传输的延迟和抖动，过高的延迟会引起报文重传。
- 拥塞使网络的有效吞吐率降低，造成网络资源的利用率降低。
- 拥塞加剧会耗费大量的网络资源（特别是存储资源），不合理的资源分配甚至可能导致系统陷入资源死锁而崩溃。

可见，拥塞使流量不能及时获得资源是造成服务性能下降的源头。在分组交换以及多用户业务并存的复杂环境下，拥塞又是不可避免的，因此必须采用适当的方法来解决拥塞。

1.4.3 对策

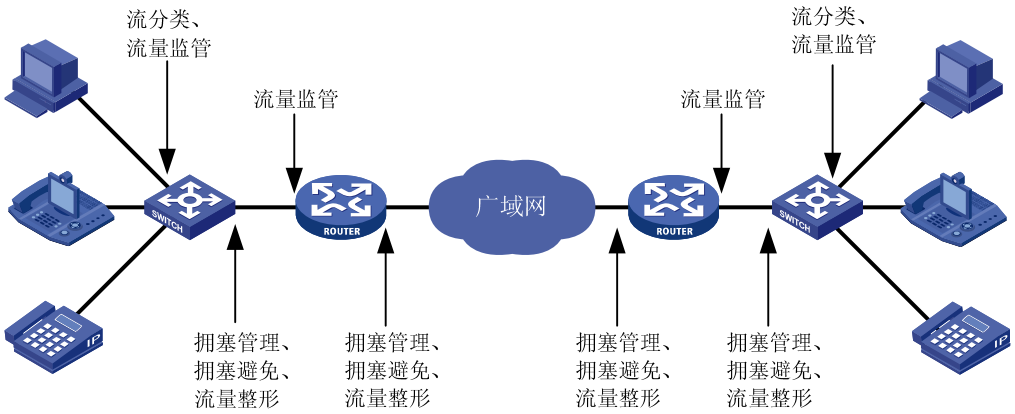
增加网络带宽是解决资源不足的一个直接途径，然而网络带宽不可能无限制的增加，因此它并不能解决所有导致网络拥塞的问题。

解决网络拥塞问题的一个更有效的办法是在网络中增加流量控制和资源分配的功能，为有不同服务需求的业务提供有差别的服务（即 Diff-Serv，Differentiated Services），更合理地分配和使用资源。在进行资源分配和流量控制的过程中，尽可能地控制好那些可能引发网络拥塞的直接或间接因素，从而减少拥塞发生的概率；在拥塞发生时，依据业务的性质及其需求权衡资源的分配，将拥塞的影响减到最小。

1.5 QoS 技术的实现思路

1.5.1 端到端的 QoS

图1-2 端到端 QoS 模型图



如 图 1-2所示，流分类、流量监管、流量整形、拥塞管理和拥塞避免是构造有区别地实施服务的基石，它们主要完成如下功能：

- 流分类：依据一定的匹配规则识别出报文。流分类是有区别地实施服务的前提，通常作用在接口入方向。
- 流量监管：对进入设备的特定流量的规格进行监管，通常作用在接口入方向。当流量超出规格时，可以采取限制或惩罚措施，以保护网络资源不受损害。
- 流量整形：一种主动调整流的输出速率的流控措施，用来使流量适配下游设备可供给的网络资源，避免不必要的报文丢弃和拥塞，通常作用在接口出方向。
- 拥塞管理：就是当拥塞发生时如何制定一个资源的调度策略，以决定报文转发的处理次序，通常作用在接口出方向。
- 拥塞避免：监督网络资源的使用情况，当发现拥塞有加剧的趋势时采取主动丢弃报文的策略，通过调整流量来解除网络的过载，通常作用在接口出方向。

在这些 QoS 技术中，流分类是基础，是有区别地实施服务的前提；而流量监管、流量整形、拥塞管理和拥塞避免从不同方面对网络流量及其分配的资源实施控制，是有区别地提供服务思想的具体体现。

本节重点介绍流分类技术，后续章节会对其他技术进行详细介绍。

1.5.2 流分类技术

流分类可以使用 IP 报文头的 ToS（Type of Service，服务类型）字段的优先级位，识别出有不同优先级特征的流量；也可以由网络管理者设置流分类的策略，例如综合源地址、目的地址、MAC 地址、IP 协议或应用程序的端口号等信息对流进行分类。

流分类的结果是没有范围限制的，它可以是一个由五元组（源地址、源端口号、协议号、目的地址、目的端口号）确定的狭小范围，也可以是到某网段的所有报文。

一般在网络边界对报文分类时，同时设置报文 IP 头的 ToS 字段中的优先级位。这样，在网络的内部就可以直接使用 IP 优先级作为分类标准。而队列技术也可以使用这个优先级来对报文进行不同的处理。下游网络可以选择接收上游网络的分类结果，也可以按照自己的标准重新进行分类。

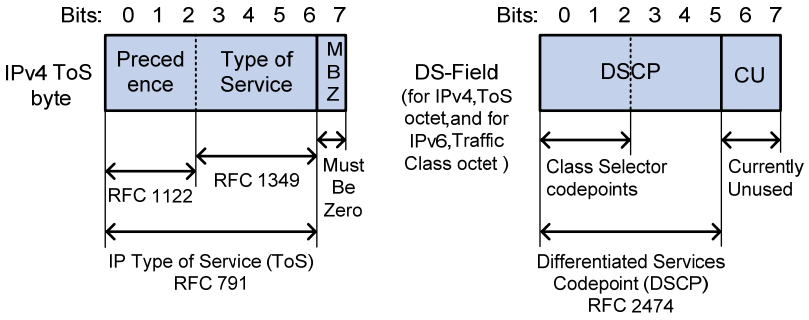
进行流分类是为了有区别地提供服务，它必须与某种流控或资源分配动作关联起来才有意义。具体采取何种流控动作，与所处的阶段以及网络当前的负载状况有关。例如，当报文进入网络时依据承诺速率对它进行监管；流出节点之前进行整形；拥塞时对队列进行调度管理；拥塞加剧时采取拥塞避免措施等。

1.5.3 报文优先级

下面介绍 IP 优先级、ToS 优先级、DSCP（Differentiated Services Codepoint，差分服务编码点）优先级、802.1p 优先级和 EXP 优先级。

(1) IP 优先级、ToS 优先级和 DSCP 优先级

图1-3 DS 域和 ToS 字节



如 图 1-3所示，IP header的ToS字段有 8 个bit，其中前 3 个bit表示的就是IP优先级，取值范围为 0～7；第 3～6 这 4 个bit表示的是ToS优先级，取值范围为 0～15；在RFC2474 中，重新定义了IP报文头部的ToS域，称之为DS（Differentiated Services，差分服务）域，其中DSCP优先级用该域的前 6 位（0～5 位）表示，取值范围为 0～63，后 2 位（6、7 位）是保留位。

表1-1 IP 优先级说明

IP 优先级（十进制）	IP 优先级（二进制）	关键字
0	000	routine
1	001	priority
2	010	immediate
3	011	flash
4	100	flash-override
5	101	critical
6	110	internet
7	111	network

Diff-Serv（Differentiated Services，差分服务）网络定义了四类流量，设备会根据报文中的 DSCP 优先级对报文执行相应的动作：

- 加速转发（Expedited Forwarding，EF）类，这种方式不用考虑其他流量是否分享其链路，适用于低时延、低丢包率、低抖动、高带宽的优先业务；
- 确保转发（Assured Forwarding，AF）类，又分为四个小类（AF1/2/3/4），每个 AF 小类又分为三个丢弃优先级，可以细分 AF 业务的等级，AF 类的 QoS 等级低于 EF 类；
- 兼容 IP 优先级（Class Selector，CS）类，是从 IP ToS 字段演变而来的，共 8 类；

- 尽力转发（Best Effort，BE）类，是 CS 中特殊一类，没有任何保证，AF 类超限后可以降级为 BE 类，现有 IP 网络流量也都默认为此类。

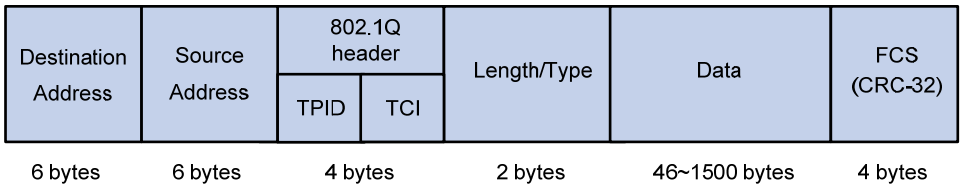
表1-2 DSCP 优先级说明

IP 优先级（十进制）	IP 优先级（二进制）	关键字
46	101110	ef
10	001010	af11
12	001100	af12
14	001110	af13
18	010010	af21
20	010100	af22
22	010110	af23
26	011010	af31
28	011100	af32
30	011110	af33
34	100010	af41
36	100100	af42
38	100110	af43
8	001000	cs1
16	010000	cs2
24	011000	cs3
32	100000	cs4
40	101000	cs5
48	110000	cs6
56	111000	cs7
0	000000	be（default）

(2) 802.1p 优先级

802.1p 优先级位于二层报文头部，适用于不需要分析三层报头，而需要在二层环境下保证 QoS 的场合。

图1-4 带有 802.1Q 标签头的以太网帧



如 图 1-4所示，4 个字节的 802.1Q标签头包含了 2 个字节的TPID（Tag Protocol Identifier，标签协议标识，取值为 0x8100）和 2 个字节的TCI（Tag Control Information，标签控制信息），图 1-5 显示了 802.1Q标签头的详细内容。

图1-5 802.1Q 标签头

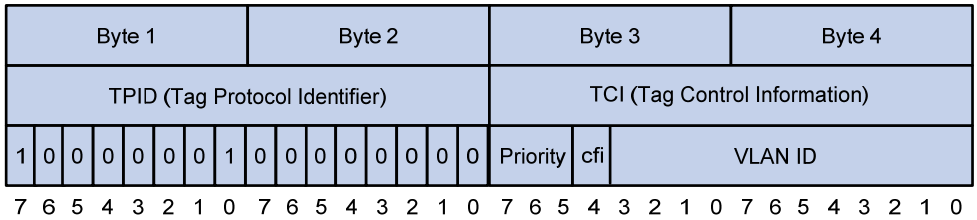


表1-3 802.1p 优先级说明

802.1p 优先级（十进制）	802.1p 优先级（二进制）	关键字
0	000	best-effort
1	001	background
2	010	spare
3	011	excellent-effort
4	100	controlled-load
5	101	video
6	110	voice
7	111	network-management

之所以称此优先级为 802.1p 优先级，是因为有关这些优先级的应用是在 802.1p 规范中被详细定义。

(3) EXP 优先级

EXP 位于 MPLS 标签内，用于 QoS。

图1-6 MPLS 标签的封装结构



在 图 1-6中，Exp字段就是EXP优先级。它由 3 个bit组成，取值范围为 0～7。

2 两种 QoS 配置方式介绍

QoS的配置方式分为QoS策略配置方式和非QoS策略配置方式两种。具体介绍如 表 2-1所示。

有些 QoS 功能可以通过 QoS 策略配置方式来配置，有些可以通过非 QoS 策略配置方式来配置，还有一些两种配置方式都可以进行配置（两种方式等价）。

在实际应用中，两种配置方式可以结合起来使用。

表2-1 两种 QoS 配置方式介绍

配置方式	配置方法		对应章节	特点
QoS 策略配置方式	配置 QoS 策略	• 定义类，并在类视图中定义一组流分类规则 • 定义流行为，并在流行为视图中定义一组 QoS 动作 • 定义策略，在策略视图下为使用的类指定对应的流行为	(1) QoS策略结构性配置过程请参见“3 QoS策略配置” (2) 可被流分类引用的功能： • 8 DAR配置 • 7.2.3 配置MPLS QoS策略 (3) 可被流行为引用的功能： • 5.6 基于类的队列的配置 • 7.2.3 配置MPLS QoS策略	定义好的类和行为可以任意组合形成不同的策略，一次配置可重复使用，配置比较灵活
	应用 QoS 策略	将定义好的策略应用到各种场合		
非 QoS 策略配置方式	-		• 4 流量监管和流量整形配置 • 5 拥塞管理配置 • 6 拥塞避免 • 7 MPLS QoS配置	配置简单，但不够灵活

3 QoS 策略配置

3.1 QoS 策略简介

QoS 策略包含了三个要素：类、流行为、策略。用户可以通过 QoS 策略将指定的类和流行为绑定起来，方便地进行 QoS 配置。

1. 类

类是用来识别流的。

类的要素包括：类的名称和类的规则。

用户可以通过命令定义一系列的规则来对报文进行分类。同时用户可以通过命令指定规则之间的关系：and 或者 or。

- and：报文只有匹配了所有的规则，设备才认为报文属于这个类。
- or：报文只要匹配了类中的一个规则，设备就认为报文属于这个类。

2. 流行为

流行为用来定义针对报文所做的 QoS 动作。

流行为的要素包括：流行为的名称和流行为中定义的动作。

用户可以通过命令在一个流行为中定义多个动作。

3. 策略

策略用来将指定的类和流行为绑定起来。

策略的要素包括：策略名称、绑定在一起的类和流行为的名称。

用户可以通过命令在一个策略中定义多个类与流行为的绑定关系。

3.2 配置 QoS 策略

QoS 策略的配置步骤如下：

- 定义类，并在类视图中定义一组流分类规则
- 定义流行为，并在流行为视图中定义一组 QoS 动作
- 定义策略，在策略视图下为使用的类指定对应的流行为

3.2.1 定义类

定义类首先要创建一个类名称，然后在此类视图下配置其匹配规则。

表3-1 定义类

操作	命令	说明
进入系统视图	**system-view**	-
定义类并进入类映射视图	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	必选 缺省为 **and**，即类视图下各匹配规则之间的关系为逻辑与
定义匹配数据包的规则	**if-match** [**not**] *match-criteria*	必选
显示类信息	**display traffic classifier** { **system-defined** | **user-defined** } [*tcl-name*]	可选 **display** 命令可以在任意视图下执行

3.2.2 定义流行为

定义流行为首先需要创建一个流行为名称，然后在此流行为视图下根据需要配置相应的流行为。

表3-2 定义流行为

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选
配置流量监管动作	**car cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]]] [**green** *action*] [**red** *action*]	可选 关于流量监管的详细介绍请参见“4 流量监管和流量整形配置”
配置流量过滤动作	**filter** { **deny** | **permit** }	可选 **deny** 表示丢弃数据包；**permit** 表示允许数据包通过
配置流量整形动作	**gts cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size* [**queue-length** *queue-length*]]]	可选 关于流量整形的详细介绍请参见“4 流量监管和流量整形配置”
配置标记报文的 DSCP 值	**remark dscp** *dscp-value*	可选
配置标记报文的 IP 优先级值	**remark ip-precedence** *ip-precedence-value*	可选
配置标记 MPLS 报文的 EXP 域的值	**remark mpls-exp** *exp-value*	可选
标记报文的 qos-local-id 值	**remark qos-local-id** *local-id-value*	可选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行



说明

如果接口或 PVC 上既应用了流行为视图下 **car** 命令配置的策略，又配置了 **qos car** 命令，那么只有前者会生效。

3.2.3 定义策略

策略定义该类和流行为的对应关系，而每个流行为由一组 QoS 动作组成。

每个策略下可以配置多组类和流行为的对应关系，过滤的时候按照配置的顺序进行。

表3-3 在策略中为类指定流行为

操作	命令	说明
进入系统视图	**system-view**	-
定义策略并进入策略视图	**qos policy** *policy-name*	必选
在策略中为类指定采用的流行为	**classifier** *tcl-name* **behavior** *behavior-name*	必选
显示指定策略中指定类及与类关联的流行为的配置信息	**display qos policy user-defined** [*policy-name* [**classifier** *tcl-name*]]	可选 **display** 命令可以在任意视图下执行



说明

如果 QoS 策略在定义流分类规则时引用了 ACL，当 **if-match** 中引用的 ACL 规则的动作为 **deny** 时，则跳出该 **if-match**，继续进行后续规则的查找。

3.2.4 QoS 策略配置举例

1. 组网需求

配置 QoS 策略 test_policy，将 IP 优先级为 6 的报文的速率限制为 100kbps。

2. 配置步骤

定义类 test_class，匹配 IP 优先级为 6 的报文。

```
<Sysname> system-view
[Sysname] traffic classifier test_class
[Sysname-classifier-test_class] if-match ip-precedence 6
[Sysname-classifier-test_class] quit
```

定义流行为 test_behavior，限制报文的承诺速率为 100kbps。

```
[Sysname] traffic behavior test_behavior
[Sysname-behavior-test_behavior] car cir 100
[Sysname-behavior-test_behavior] quit
```

定义策略 test_policy，将类和流行为结合起来。

```
[Sysname] qos policy test_policy
[Sysname-qospolicy-test_policy] classifier test_class behavior test_behavior
```

3.3 应用 QoS 策略

QoS 策略支持以下应用方式：

- 基于接口或 PVC 应用 QoS 策略：QoS 策略对接口或 PVC 接收或者发送的流量生效。
- 基于上线用户应用 QoS 策略：QoS 策略对上线用户接收或者发送的流量生效。

3.3.1 基于接口或 PVC 应用 QoS 策略

说明

QoS 策略应用后，用户仍然可以修改 QoS 策略中的流分类规则、流行为以及流分类规则和流行为的对应关系。

一个策略可以应用于多个接口或 PVC。接口或 PVC 的每个方向（出/入两个方向）只能应用一个策略。

1. 配置过程

表3-4 在接口或 PVC 上应用策略

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图或 PVC 视图	进入接口视图	**interface** *interface-type interface-number*	二者必选其一 进入接口视图后，下面进行的配置只在当前接口生效；进入 PVC 视图后，下面进行的配置只在当前 PVC 生效
	进图 PVC 视图	**interface atm** *interface-number*	
		pvc *vpi/vci*	
在接口或 PVC 上应用关联的策略		**qos apply policy** *policy-name* { **inbound** | **outbound** }	必选
显示指定接口或所有接口上策略的配置信息和运行情况		**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]	可选 **display** 命令可以在任意视图下执行
显示指定策略或所有策略中指定类或所有类及与类关联的行为的配置信息		**display qos policy** { **system-defined** | **user-defined** } [*policy-name* [**classifier** *tcl-name*]]	

说明

如果 QoS 策略应用在接口或 PVC 的出方向，则 QoS 策略对本地协议报文不起作用。（本地协议报文的含义及其作用如下：某些内部发起的报文是维持设备正常运行的重要协议报文，为了确保这些报文能够被不受影响的发送出去，遂将其定义为本地协议报文，使得 QoS 不对其进行处理，降低了因配置 QoS 而误将这些报文丢弃或进行其他处理的风险。一些常见的本地协议报文如下：链路维护报文、ISIS、OSPF、RIP、BGP、LDP、SSH 等。）

2. 配置举例

把策略 test_policy 应用到接口 GigabitEthernet1/1 入方向上。

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/1
[Sysname-GigabitEthernet1/1] qos apply policy test_policy inbound
```

3.3.2 基于上线用户应用 QoS 策略

一个策略可以应用于多个上线用户。上线用户的每个方向（发送/接收报文两个方向）只能应用一个策略，如果用户想修改某方向上应用的策略，必须先取消原先的配置，然后再配置新的策略。

1. 配置过程

表3-5 基于上线用户应用 QoS 策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 user-profile 视图	**user-profile** *profile-name* **ppp**	必选 进入 user-profile 视图后，下面进行的配置只在 User Profile 处于激活状态，且用户成功上线后才生效
应用关联的策略	**qos apply policy** *policy-name* { **inbound** | **outbound** }	必选
退回系统视图	**quit**	-
激活 User Profile	**user-profile** *profile-name* **enable**	必选 缺省情况下，User Profile 处于未激活状态



说明

- 如果 User Profile 处于激活状态，则不能删除已经应用到此 User Profile 的策略，除策略引用的 ACL 规则外也不能修改策略的其他具体内容。引用的 ACL 规则的具体内容在用户上线后也不能修改。
- user-profile 视图下应用的策略中的流行为只支持 **remark**、**car**、**filter** 三种动作。
- user-profile 视图下应用的策略不能为空策略，因为空策略的 User Profile 不能被激活。
- 关于 User Profile 的相关介绍以及配置，请参见“系统分册”中的“User Profile 配置”。

2. 配置举例

把策略 test_policy 应用到 ppp 上线用户 user 的接收报文方向上。

```
<Sysname> system-view
[Sysname] user-profile user ppp
[Sysname-user-profile-ppp-user] qos apply policy test_policy inbound
[Sysname-user-profile-DOT1X-user] quit
[Sysname] user-profile user enable
```

3.4 QoS 策略显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 QoS 策略的运行情况，通过查看显示信息验证配置的效果。

表3-6 QoS 策略显示和维护

操作	命令
显示指定策略或所有策略中指定类或所有类及与类关联的行为的配置信息	**display qos policy** { **system-defined** | **user-defined** } [*policy-name* [**classifier** *tcl-name*]]
显示指定接口或 PVC 或所有接口及 PVC 上策略的配置信息和运行情况	**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]

操作	命令
显示配置的流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]
显示配置的类信息	**display traffic classifier** { **system-defined** | **user-defined** } [*tcl-name*]

4 流量监管和流量整形配置

4.1 流量监管和流量整形简介

如果不限制用户发送的流量，那么大量用户不断突发的数据只会使网络更拥挤。为了使有限的网络资源能够更好地发挥效用，更好地为更多的用户服务，必须对用户的流量加以限制。比如限制每个时间间隔某个流只能得到承诺分配给它的那部分资源，防止由于过分突发所引发的网络拥塞。

流量监管和流量整形就是一种通过对流量规格的监督，来限制流量及其资源使用的流控策略。进行流量监管或整形有一个前提条件，就是要知道流量是否超出了规格，然后才能根据评估结果实施调控策略。一般采用令牌桶（Token Bucket）对流量的规格进行评估。

4.1.1 流量评估与令牌桶

1. 令牌桶的特点

令牌桶可以看作是一个存放一定数量令牌的容器。系统按设定的速度向桶中放置令牌，当桶中令牌满时，多出的令牌溢出，桶中令牌不再增加。

2. 用令牌桶评估流量

在用令牌桶评估流量规格时，是以令牌桶中的令牌数量是否足够满足报文的转发为依据的。如果桶中存在足够的令牌可以用来转发报文（通常用一个令牌关联一个比特的转发权限），称流量遵守或符合这个规格，否则称为不符合或超标。

评估流量时令牌桶的参数设置包括：

- 平均速率：向桶中放置令牌的速率，即允许的流的平均速度。通常设置为 CIR（Committed Information Rate，承诺信息速率）。
- 突发尺寸：令牌桶的容量，即每次突发所允许的最大的流量尺寸。通常设置为 CBS（Committed Burst Size，承诺突发尺寸），设置的突发尺寸必须大于最大报文长度。

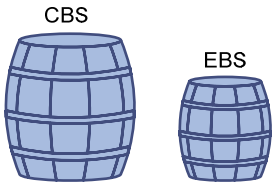
每到达一个报文就进行一次评估。每次评估，如果桶中有足够的令牌可供使用，则说明流量控制在允许的范围内，此时要从桶中取走与报文转发权限相当的令牌数量；否则说明已经耗费太多令牌，流量超标了。

3. 复杂评估

为了评估更复杂的情况，实施更灵活的调控策略，可以设置两个令牌桶（简称 C 桶和 E 桶）。例如 TP（Traffic Policing，流量监管）中有三个参数：

- CIR：表示向 C 桶中投放令牌的速率，即 C 桶允许传输或转发报文的平均速率；
- CBS：表示 C 桶的容量，即 C 桶瞬间能够通过的承诺突发流量；
- EBS（Excess Burst Size，超出突发尺寸）：表示 E 桶的容量，即 E 桶瞬间能够通过的超出突发流量。

图4-1 双令牌桶结构



双令牌桶结构如 图 4-1所示，CBS和EBS是由两个不同的令牌桶承载的。

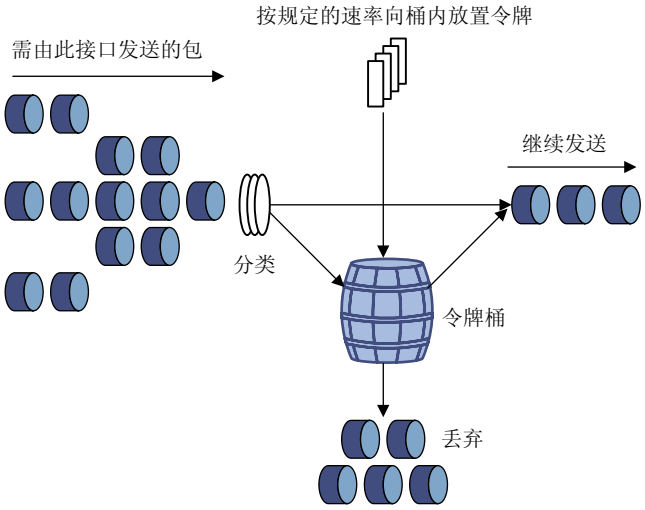
每次评估时，依据下面的的情况，可以分别实施不同的流控策略：

- 如果 C 桶有足够的令牌，报文被标记为 green，即绿色报文；
- 如果 C 桶令牌不足，但 E 桶有足够的令牌，报文被标记为 yellow，即黄色报文；
- 如果 C 桶和 E 桶都没有足够的令牌，报文被标记为 red，即红色报文。

4.1.2 流量监管

流量监管 TP（Traffic Policing）就是对流量进行控制，通过监督进入网络的流量速率，对超出部分的流量进行“惩罚”，使进入的流量被限制在一个合理的范围之内，以保护网络资源和运营商的利益。例如可以限制 HTTP 报文不能占用超过 50%的网络带宽。如果发现某个连接的流量超标，流量监管可以选择丢弃报文，或重新设置报文的优先级。

图4-2 TP 示意图



流量监管广泛的用于监管进入 Internet 服务提供商 ISP 的网络流量。流量监管还包括对所监管流量的流分类服务，并依据不同的评估结果，实施预先设定好的监管动作。这些动作可以是：

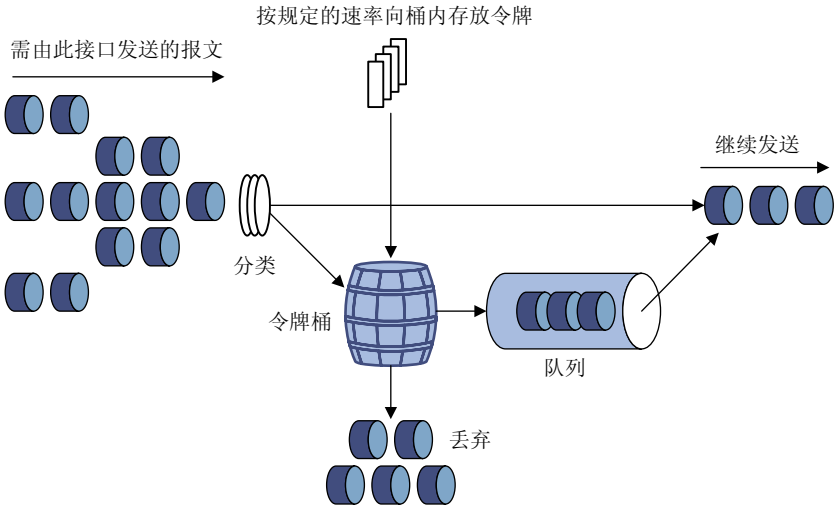
- 转发：比如对评估结果为“符合”的报文继续转发。
- 丢弃：比如对评估结果为“不符合”的报文进行丢弃。
- 改变优先级并转发：比如对评估结果为“符合”的报文，将之标记为其它的优先级后再进行转发。
- 改变优先级并进入下一级监管：比如对评估结果为“符合”的报文，将之标记为其它的优先级后再进入下一级的监管。
- 进入下一级的监管：流量监管可以逐级堆叠，每级关注和监管更具体的目标。

4.1.3 流量整形

TS（Traffic Shaping，流量整形）是一种主动调整流量输出速率的措施。一个典型应用是基于下游网络节点的 TP 指标来控制本地流量的输出。

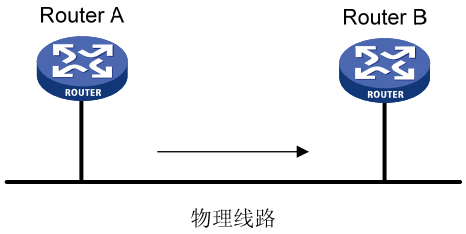
流量整形与流量监管的主要区别在于，流量整形对流量监管中需要丢弃的报文进行缓存——通常是将它们放入缓冲区或队列内，如 图 4-3所示。当令牌桶有足够的令牌时，再均匀的向外发送这些被缓存的报文。流量整形与流量监管的另一区别是，整形可能会增加延迟，而监管几乎不引入额外的延迟。

图4-3 TS 示意图



例如，在 图 4-4所示的应用中，设备Router A向Router B发送报文。Router B要对Router A发送来的报文进行TP监管，对超出规格的流量直接丢弃。

图4-4 流量整形的应用



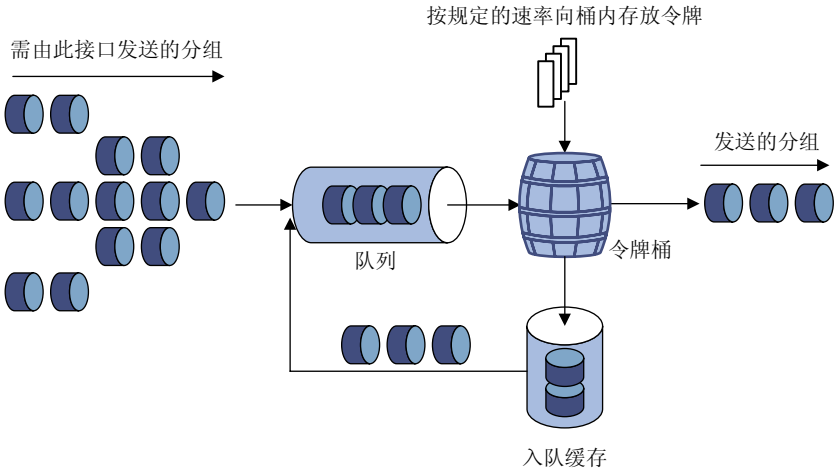
为了减少报文的无谓丢失，可以在 Router A 的出口对报文进行流量整形处理。将超出流量整形特性的报文缓存在 Router A 中。当可以继续发送下一批报文时，流量整形再从缓冲队列中取出报文进行发送。这样，发向 Router B 的报文将都符合 Router B 的流量规定。

4.1.4 物理接口限速

利用 LR（Line Rate，物理接口限速）可以在一个物理接口上限制发送报文（包括紧急报文）的总速率。

LR 也是采用令牌桶进行流量控制。如果在设备的某个接口上配置了 LR，所有经由该接口发送的报文首先要经过 LR 的令牌桶进行处理。如果令牌桶中有足够的令牌，则报文可以发送；否则，报文将进入 QoS 队列进行拥塞管理。这样，就可以对通过该物理接口的报文流量进行控制。

图4-5 LR 处理过程示意图



由于采用了令牌桶控制流量，当令牌桶中存有令牌时，可以允许报文的突发性传输；当令牌桶中没有令牌时，报文必须等到桶中生成了新的令牌后才可以继续发送。这就限制了报文的流量不能大于令牌生成的速度，达到了限制流量，同时允许突发流量通过的目的。

与流量监管相比，物理接口限速能够限制在物理接口上通过的所有报文。流量监管在 IP 层实现，可以对端口上不同的流分类进行限速，但是对于不经过 IP 层处理的报文不起作用。当用户只要求对所有报文限速时，使用物理接口限速比较简单。

4.2 流量监管/流量整形/物理接口限速配置

表4-1 配置说明

配置	配置说明	相关链接
应用 CAR 列表的流量监管	配置 CAR 列表	基于CAR列表的流量监管配置过程
	在接口应用 CAR 策略	
应用 ACL 的流量监管	配置 ACL	基于ACL的流量监管配置过程
	在接口应用 CAR 策略	
适配所有流的监管配置过程	在接口应用 CAR 策略	适配所有流的流量监管配置过程
基于 ACL 的流量整形配置	配置 ACL	基于ACL的流量整形配置
	在接口配置流量整形	
适配所有流的流量整形配置	在接口配置流量整形	适配所有流的流量整形配置
物理接口限速配置过程	设置接口限速	物理接口限速配置

4.2.1 流量监管配置

流量监管的配置包括两方面的任务，一是定义那些需要实施监管的报文的特征，二是定义对这些报文的监管策略。

1. 基于 CAR 列表的流量监管配置过程

表4-2 基于 CAR 列表的流量监管配置过程

操作	命令	说明
进入系统视图	**system-view**	-
配置 CAR 列表	**qos carl** *carl-index* { **precedence** *precedence-value* | **mac** *mac-address* | **mpls-exp** *mpls-exp-value* | **dscp** *dscp-list* | { **destination-ip-address** | **source-ip-address** } { **subnet** *ip-address mask-length* | **range** *start-ip-address* **to** *end-ip-address* } [**per-address** [**shared-bandwidth**]] }	必选
显示 CAR 列表的规则	**display qos carl** [*carl-index*]	可选 **display** 命令可以在任意视图下执行
进入接口视图	**interface** *interface-type interface-number*	-
在接口应用基于 CAR 列表的 CAR 策略	**qos car** { **inbound** | **outbound** } **carl** *carl-index* **cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**green** *action*] [**red** *action*]	必选
显示 CAR 在接口上的信息	**display qos car interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

2. 基于 ACL 的流量监管配置过程

表4-3 基于 ACL 的流量监管配置过程

操作	命令	说明
进入系统视图	**system-view**	-
配置 ACL 规则	参见 ACL 模块的描述	必选
进入接口视图	**interface** *interface-type interface-number*	-
在接口应用基于 ACL 规则的 CAR 策略	**qos car** { **inbound** | **outbound** } **acl** [**ipv6**] *acl-number* **cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**green** *action*] [**red** *action*]	必选
显示 CAR 在接口上的信息	**display qos car interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

3. 适配所有流的流量监管配置过程

表4-4 适配所有流的流量监管配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
在接口应用 CAR 策略	**qos car** { **inbound** | **outbound** } **any cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**green** *action*] [**red** *action*]	必选
显示 CAR 在接口上的信息	**display qos car interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

4. 流量监管配置举例

在 GigabitEthernet1/0 接口上进行 TP 配置，对由 GigabbitEthernet1/0 接口发送的报文进行流量控制，报文流量不能超过 1Mbps，如果超过流量限制则将违规报文丢弃。

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface gigabitethernet 1/0
```

配置监管参数。

```
[Sysname-GigabitEthernet1/0] qos car outbound any cir 1000 cbs 1000000 ebs 0 green pass red discard
```

4.2.2 流量整形配置

流量整形配置分为以下几种：

- 基于 ACL 的流量整形配置：为匹配某一 ACL 的流设置整形参数，使用不同的 ACL 可以为不同的流设置整形参数。
- 适配所有流的流量整形配置：为所有的流设置整形参数。


说明

软转发流量整形不支持 IPv6。

1. 基于 ACL 的流量整形配置

表4-5 基于 ACL 的流量整形配置

操作	命令	说明
进入系统视图	**system-view**	-
定义 ACL	参见 ACL 模块的描述	必选
进入接口视图	**interface** *interface-type interface-number*	-
在接口配置流量整形	**qos gts acl** *acl-number* **cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size* [**queue-length** *queue-length*]]]	必选
显示流量整形配置运行信息	**display qos gts interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

2. 适配所有流的流量整形配置

表4-6 适配所有流的流量整形配置

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口配置流量整形	**qos gts any cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size* [**queue-length** *queue-length*]]]	必选
显示流量整形配置运行信息	**display qos gts interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

3. 流量整形配置举例

对 GigabitEthernet1/0 接口发送的报文进行流量整形，对超过 500kbps 的报文流进行整形。

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface gigabitethernet 1/0
```

配置整形参数。

```
[Sysname-GigabitEthernet1/0] qos gts any cir 500
```

4.2.3 物理接口限速配置

1. 物理接口限速配置过程

配置物理接口限速就是限制物理接口向外发送数据的速率。

表4-7 接口限速配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
设置接口限速	**qos lr outbound cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]]	必选
显示接口的 LR 配置和统计信息	**display qos lr interface** [*interface-type interface-number*]	**display** 命令可以在任意视图下执行

2. 物理接口限速配置举例

把接口 GigabitEthernet1/0 的出速度限制为 500kbps。

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface gigabitethernet 1/0
```

配置限速参数，接口出速率限制为 500kbps。

```
[Sysname-GigabitEthernet1/0] qos lr outbound cir 500
```

4.3 流量监管/流量整形/物理接口限速显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后流量监管/流量整形/物理接口限速的运行情况，通过查看显示信息验证配置的效果。

表4-8 流量监管/流量整形/物理接口限速显示和维护

操作	命令
显示 CAR 列表的规则	**display qos carl** [*carl-index*]
显示 CAR 在接口上的信息	**display qos car interface** [*interface-type interface-number*]
显示流量整形配置运行信息	**display qos gts interface** [*interface-type interface-number*]
显示接口的 LR 配置和统计信息	**display qos lr interface** [*interface-type interface-number*]

4.4 流量监管与流量整形典型配置举例

4.4.1 流量监管与流量整形典型配置举例

1. 配置需求

- 设备 Router A 通过接口 GigabitEthernet1/0 和设备 Router B 的接口 Ethernet1/1 互连
- Server、Host A、Host B 可经由 Router A 和 Router B 访问 Internet
- Server、Host A 与 Router A 的 GigabitEthernet1/1 接口在同一网段
- Host B 与 Router A 的 GigabitEthernet1/2 接口在同一网段

要求在设备 Router A 上对接口 GigabitEthernet1/1 接收到的源自 Server 和 Host A 的报文流分别实施流量控制如下：

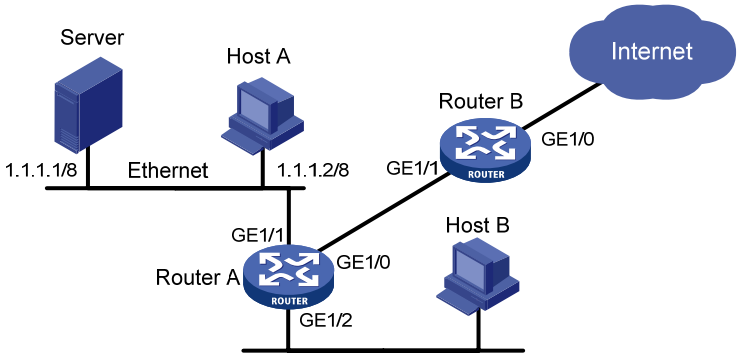
- 来自 Server 的报文流量约束为 54kbps，流量小于 54kbps 时可以正常发送，流量超过 54kbps 时则将违规报文的优先级设置为 0 后进行发送；
- 来自 Host A 的报文流量约束为 8kbps，流量小于 8kbps 时可以正常发送，流量超过 8kbps 时则丢弃违规报文；

对设备 Router B 的 GigabitEthernet1/1 和 GigabitEthernet1/0 接口收发报文有如下要求：

- Router B 的 GigabitEthernet1/1 接口接收报文的总流量限制为 500kbps，如果超过流量限制则将违规报文丢弃；
- 经由 Router B 的 GigabitEthernet1/0 接口进入 Internet 的报文流量限制为 1000kbps，如果超过流量限制则将违规报文丢弃。

2. 组网图

图4-6 流量监管、流量整形配置组网图



3. 配置步骤

(1) 配置设备 Router A：

在 Router A 的 GigabitEthernet1/0 接口上配置 TS，对该接口发送的报文进行流量整形（对超过 500kbps 的报文流进行整形），以降低在 Router B 接口 GigabitEthernet1/1 处的丢包率。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] qos gts any cir 500
[RouterA-GigabitEthernet1/0] quit
```

配置 ACL 规则列表，分别匹配来源于 Server 和 Host A 的报文流。

```
[RouterA] acl number 2001
[RouterA-acl-basic-2001] rule permit source 1.1.1.1 0
[RouterA-acl-basic-2001] quit
[RouterA] acl number 2002
[RouterA-acl-basic-2002] rule permit source 1.1.1.2 0
[RouterA-acl-basic-2002] quit
```

在 GigabitEthernet1/1 接口上进行 TP 配置，对 GigabitEthernet1/1 接收到的不同报文流进行相应流量控制。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] qos car inbound acl 2001 cir 54 cbs 4000 ebs 0 green pass red
remark-prec-pass 0
[RouterA-GigabitEthernet1/1] qos car inbound acl 2002 cir 8 cbs 1875 ebs 0 green pass red
discard
[RouterA-GigabitEthernet1/1] qos car inbound any cir 500 cbs 32000 ebs 0 green pass red discard
[RouterA-Ethernet1/1] quit
```

(2) 配置设备 Router B：

在 GigabitEthernet1/0 接口上进行 TP 配置，对由 GigabitEthernet1/0 接口发送的报文进行流量控制，报文流量不能超过 1Mbps，如果超过流量限制则将违规报文丢弃。

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] qos car outbound any cir 1000 cbs 65000 ebs 0 green pass red
discard
```

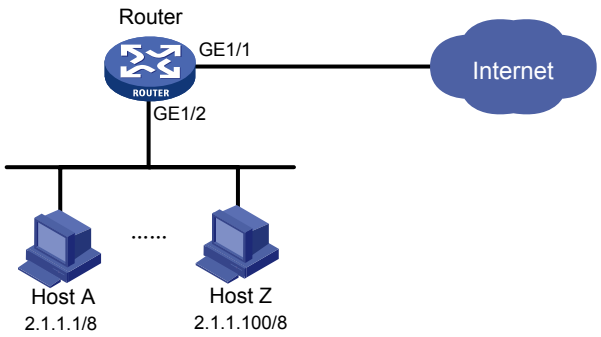
4.4.2 IP 限速典型配置举例

1. 配置需求

要求在设备 Router 上对接口 GigabitEthernet1/2 接收到的报文流进行限速：对 HostA～HostZ（源地址属于 IP 地址段 2.1.1.1～2.1.1.100）进行 IP 限速，逐 IP 地址流量限速 5kbps，网段内各 IP 地址的流量共享剩余带宽。

2. 组网图

图4-7 IP 限速配置组网图



3. 配置步骤

在接口 GigabitEthernet1/2 上进行 IP 限速，对源地址属于 IP 地址段 2.1.1.1～2.1.1.100 内所有 PC 限速 5kbps，网段内各 IP 地址的流量共享剩余带宽。

```
<Router> system-view
[Router] qos carl 1 source-ip-address range 2.1.1.1 to 2.1.1.100 per-address shared-bandwidth
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] qos car inbound carl 1 cir 5 cbs 1875 ebs 0 green pass red discard
[Router-GigabitEthernet1/2] quit
```

5 拥塞管理配置

5.1 拥塞管理简介

当分组到达的速度大于接口或 PVC 发送分组的速度时，在该接口或 PVC 处就会产生拥塞。如果没有足够的存储空间来保存这些分组，它们其中的一部分就会丢失。分组的丢失又可能会导致发送该分组的设备因超时而重传此分组，这将导致恶性循环。

拥塞管理的中心内容就是当拥塞发生时如何制定一个资源的调度策略，以决定报文转发的处理次序。拥塞管理的处理包括队列的创建、报文的分类、将报文送入不同的队列、队列调度等。

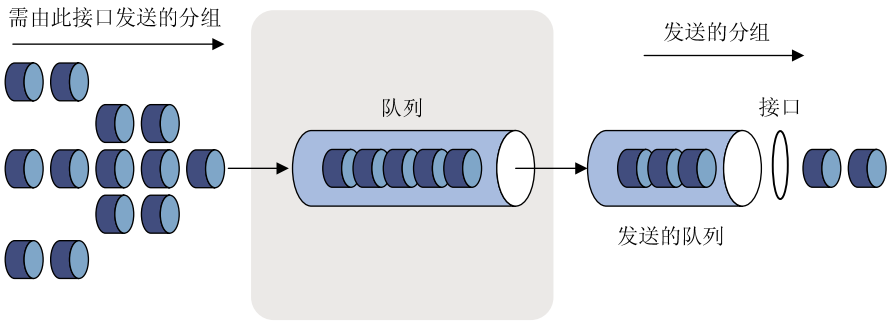
5.1.1 拥塞管理策略

对于拥塞管理，一般采用队列技术，使用一个队列算法对流量进行分类，之后用某种优先级别算法将这些流量发送出去。每种队列算法都是用以解决特定的网络流量问题，并对带宽资源的分配、延迟、抖动等有着十分重要的影响。

队列调度对不同优先级的报文进行分级处理，优先级高的会得到优先发送。这里介绍几种常用的队列调度机制。

1. FIFO（先入先出队列，First In First Out Queuing）

图5-1 先入先出队列示意图



如上图所示，FIFO 按照时间到达的先后决定分组的转发次序，先进的先出，后进的后出，不需要进行流分类和队列调度，FIFO 关心的只是队列的长度，队列的长度对延迟和丢包率的影响。用户的业务流在某个设备能够获得的资源取决于分组的到达时机及当时的负载情况。Best-Effort 报文转发方式采用的就是 FIFO 的排队策略。

如果设备的每个端口只有一个基于 FIFO 的输入或输出队列，那么恶性的应用可能会占用所有的网络资源，严重影响关键业务数据的传送。所以还需要配置一些其他的队列调度机制与 FIFO 配合对流量进行调度和拥塞控制。

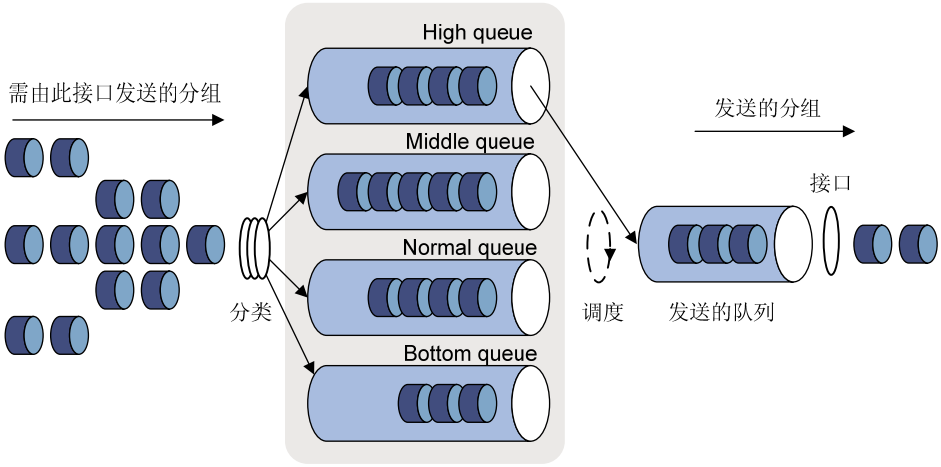
每个队列内部报文的发送次序缺省是 FIFO。

说明

配置其它的队列调度机制与 FIFO 配合对流量进行调度和拥塞管理并不是指在同一接口配置两种队列，是指其他队列技术中的一种，这些队列根据业务的需求，设置流分类规则，使不同的流进入相应的队列，从而获得不同的带宽资源和调度次序。在每一个队列内部按照 FIFO 的规则入队和出队。

2. PQ（优先队列，Priority Queuing）

图5-2 优先队列示意图



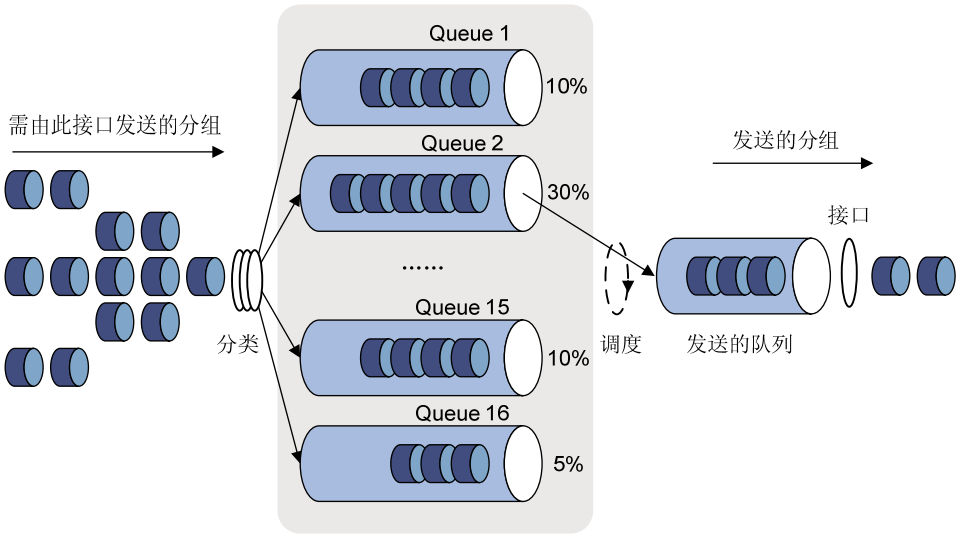
PQ 队列是针对关键业务应用设计的。关键业务有一个重要的特点，即在拥塞发生时要求优先获得服务以减小响应的延迟。PQ 可以根据网络协议（比如 IP，IPX）、数据流入接口、报文长短、源地址/目的地址等灵活地指定优先次序。优先队列将报文分成 4 类，分别为高优先队列（top）、中优先队列（middle）、正常优先队列（normal）和低优先队列（bottom），它们的优先级依次降低。缺省情况下，数据流进入 normal 队列。每个队列内部又遵循 FIFO 原则。

在队列调度时，PQ 严格按照优先级从高到低的次序，优先发送较高优先级队列中的分组，当较高优先级队列为空时，再发送较低优先级队列中的分组。这样，将关键业务的分组放入较高优先级的队列，将非关键业务的分组放入较低优先级的队列，可以保证关键业务的分组被优先传送，非关键业务的分组在处理关键业务数据的空闲间隙被传送。

PQ 的缺点是：拥塞发生时，如果较高优先级队列中长时间有分组存在，那么低优先级队列中的报文将一直得不到服务。

3. CQ（定制队列，Custom Queuing）

图5-3 定制队列示意图



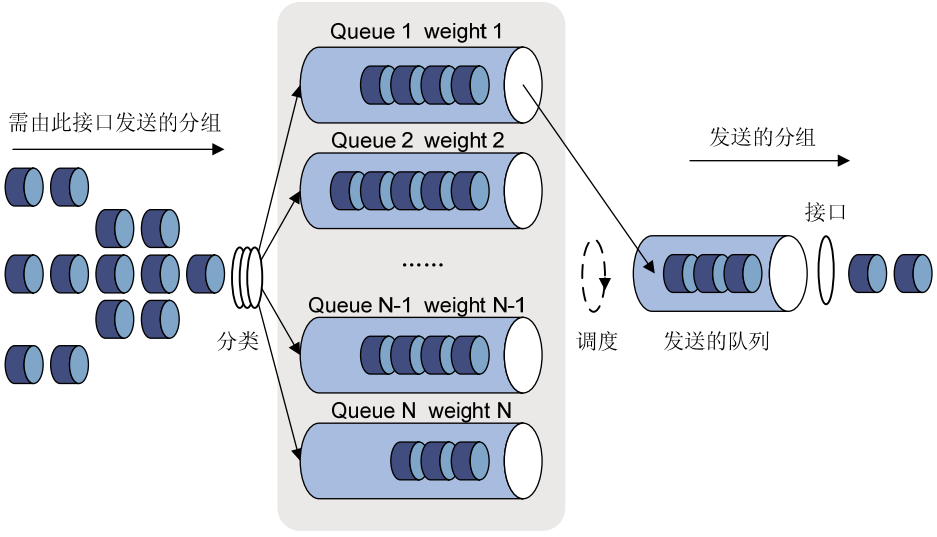
CQ 按照一定的规则将分组分成 16 类（对应于 16 个队列），分组根据自己的类别按照先进先出的策略进入相应的 CQ 队列。

CQ的 1 到 16 号队列是用户队列，如 图 5-3所示。用户可以配置流分类的规则，指定 16 个用户队列占用接口或PVC带宽的比例关系。在队列调度时，系统队列中的分组被优先发送。直到系统队列为空，再采用轮询的方式按照预先配置的带宽比例依次从 1 到 16 号用户队列中取出一定数量的分组发送出去。这样，就可以使不同业务的分组获得不同的带宽，既可以保证关键业务能获得较多的带宽，又不至于使非关键业务得不到带宽。缺省情况下，数据流进入 1 号队列。

定制队列的另一个优点是：可根据业务的繁忙程度分配带宽，适用于对带宽有特殊需求的应用。虽然 16 个用户队列的调度是轮询进行的，但对每个队列不是固定地分配服务时间片——如果某个队列为空，那么马上换到下一个队列调度。因此，当没有某些类别的报文时，CQ 调度机制能自动增加现存类别的报文可占的带宽。

4. WFQ（加权公平队列，Weighted Fair Queuing）

图5-4 加权公平队列示意图



在介绍加权公平队列前，先要理解公平队列 FQ（Fair Queuing）。FQ 是为了公平地分享网络资源，尽可能使所有流的延迟和抖动达到最优而推出的。它照顾了各方面的利益，主要表现在：

- 不同的队列获得公平的调度机会，从总体上均衡各个流的延迟。
- 短报文和长报文获得公平的调度：如果不同队列间同时存在多个长报文和短报文等待发送，应当顾及短报文的利益，让短报文优先获得调度，从而在总体上减少各个流的报文间的抖动。

与 FQ 相比，WFQ 在计算报文调度次序时增加了优先权方面的考虑。从统计上，WFQ 使高优先权的报文获得优先调度的机会多于低优先权的报文。WFQ 能够按流的“会话”信息（协议类型、源和目的 TCP 或 UDP 端口号、源和目的 IP 地址、ToS 域中的优先级位等）自动进行流分类，并且尽可能多地提供队列，以将每个流均匀地放入不同队列中，从而在总体上均衡各个流的延迟。在出队的时候，WFQ 按流的优先级（precedence）来分配每个流应占有出口的带宽。优先级的数值越小，所得的带宽越少。优先级的数值越大，所得的带宽越多。

例如：接口中当前共有 5 个流，它们的优先级分别为 0、1、2、3、4，则带宽总配额为所有（流的优先级+1）的和。即

1 + 2 + 3 + 4 + 5 = 15

每个流所占带宽比例为：（自己的优先级数 + 1）/（所有 (流的优先级 + 1) 的和）。即每个流可得的带宽分别为：1/15，2/15，3/15，4/15，5/15。

由于 WFQ 在拥塞发生时能均衡各个流的延迟和抖动，所以 WFQ 在一些特殊场合得到了有效的应用。比如在使用资源预留协议 RSVP（Resource Reservation Protocol）的保证型业务中，通常就是采用 WFQ 作为调度策略；在流量整形 TS 中，也采用 WFQ 调度缓存的报文。

5. CBQ（基于类的队列，Class Based Queuing）

CBQ 是对 WFQ 功能的扩展，为用户提供了定义类的支持。CBQ 为每个用户定义的类分配一个单独的 FIFO 预留队列，用来缓冲同一类的数据。在网络拥塞时，CBQ 对报文根据用户定义的类规则进行匹配，并使其进入相应的队列，在入队列之前必须进行拥塞避免机制（尾部丢弃或 WRED，Weighted Random Early Detection，加权随机早期检测）和带宽限制的检查。在报文出队列时，加权公平调度每个类对应的队列中的报文。

CBQ 提供一个紧急队列，紧急报文入该队列，该队列采用 FIFO 调度，没有带宽限制。这样，如果 CBQ 加权公平对待所有类的队列，语音报文这类对延迟敏感的数据流就可能得不到及时发送。为此将 PQ 特性引入 CBQ，称其为 LLQ（Low Latency Queuing，低延迟队列），为语音报文这样的对延迟敏感的数据流提供严格优先发送服务。

LLQ 将严格优先队列机制与 CBQ 结合起来使用，用户在定义类时可以指定其享受严格优先服务，这样的类称作优先类。所有优先类的报文将进入同一个优先队列，在入队列之前需对各类报文进行带宽限制的检查。报文出队列时，将首先发送优先队列中的报文，直到发送完后才发送其他类对应的队列的报文。在发送其他队列报文时将仍然按照加权公平的方式调度。

为了不让其他队列中的报文延迟时间过长，在使用 LLQ 时将会为每个优先类指定可用最大带宽，该带宽值用于拥塞发生时监管流量。如果拥塞未发生，优先类允许使用超过分配的带宽。如果拥塞发生，优先类超过分配带宽的数据包将被丢弃。LLQ 还可以指定 Burst-size。

系统在为报文匹配规则时，规则如下：

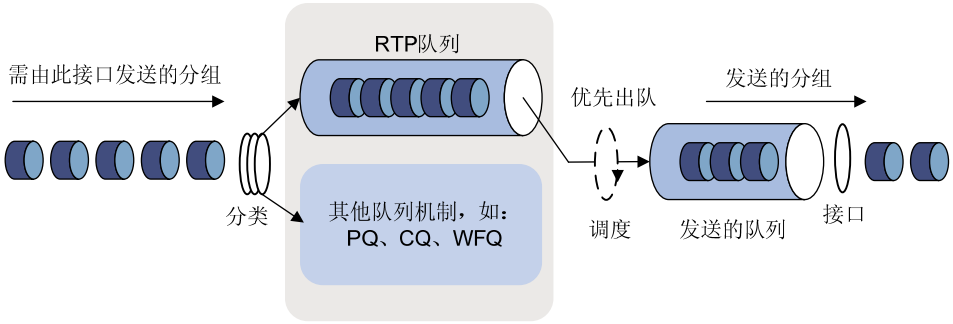
- 先匹配优先类，然后再匹配其他类；
- 对多个优先类，按照配置顺序逐一匹配；
- 对其他类，也是按照配置顺序逐一匹配；

- 对类中多个规则，按照配置顺序逐一匹配。

6. RTP（Real-time Transport Protocol）优先队列

RTP 优先队列是一种保证实时业务（包括语音与视频业务）服务质量的简单的队列技术。其原理就是将承载语音或视频的 RTP 报文送入高优先级队列，使其得到优先发送，保证时延和抖动降低为最低限度，从而保证了语音或视频这种对时延敏感业务的服务质量。

图5-5 RTP 队列示意图



如上图所示，RTP 优先队列将 RTP 报文送入一个具有较高优先级的队列。RTP 报文是端口号在一定范围内为偶数的 UDP 报文，端口号的范围可以配置。RTP 优先队列可以同任何一种队列（包括 FIFO、PQ、CQ、WFQ 与 CBQ）结合使用，而它的优先级是最高的。但由于 CBQ 中的 LLQ 也可以保证实时业务数据的转发，所以一般不推荐将 RTP 优先队列与 CBQ 结合应用。

5.1.2 拥塞管理技术的对比

设备上提供了以上拥塞管理技术，突破了传统 IP 设备的单一 FIFO 拥塞管理策略，提供了强大的 QoS 能力，使得 IP 设备可以满足不同业务所要求的不同服务质量的要求。为了用户更好地利用拥塞管理技术，现对各种队列技术做一比较。

表5-1 拥塞管理技术对比

类型	队列数	优点	缺点
FIFO	1	• 不需要配置，易于使用 • 处理简单，延迟小	• 所有的报文均进入一个“先进先出”的队列，发送报文所占用的带宽、延迟时间、丢失的概率均由报文到达队列的先后顺序决定 • 对不配合的数据源（即没有流控机制的流，如 UDP 报文发送）无约束力，不配合的数据源会造成配合的数据源（如 TCP 报文发送）带宽受损失 • 对时间敏感的实时应用（如 VoIP）的延迟得不到保证
PQ	4	• 可对不同业务的数据提供绝对的优先，对时间敏感的实时应用（如 VoIP）的延迟可以得到保证 • 对优先业务的报文的带宽占用可以绝对优先	• 需配置，处理速度慢 • 如果不对高优先级的报文的带宽加限制，可能会造成低优先级的报文得不到带宽
CQ	16	• 可对不同业务的报文按带宽比例分配带宽 • 当没有某些类别的报文时，能自动增加现存类别的报文可占的带宽	需配置，处理速度慢

类型	队列数	优点	缺点
WFQ	可配置	• 配置容易 • 可以保护配合（交互）的数据源（如TCP报文发送）的带宽 • 可以减小抖动 • 可以减小数据量小的交互式应用的延迟 • 可以为不同优先级的流分配不同的带宽 • 当流的数目减少时，能自动增加现存流可占的带宽	处理速度比FIFO要慢，但比PQ、CQ要快
CBQ	可配置（0～64）	• 可以对数据根据灵活、多样的分类规则进行划分，分别为EF（加速转发）、AF（确保转发）、BE（尽力转发）业务提供不同的队列调度机制 • 可以为AF业务提供严格、精确的带宽保证，并且保证各类AF业务之间根据权值按一定的比例关系进行队列调度 • 可以为EF业务提供绝对优先的队列调度，确保实时数据的时延满足要求；同时通过对高优先级数据流量的限制，克服了PQ的低优先级队列可能得不到服务的弊病 • 对于尽力转发的缺省类数据，提供WFQ队列调度	系统开销比较大

说明

如果流量突发较大，可以通过增加队列长度的方法来改善队列调度的准确率。

5.2 先进先出队列的配置

FIFO是接口或PVC缺省使用的队列调度机制，可以通过配置命令改变其队列长度。

5.2.1 FIFO队列配置过程

表5-2 FIFO队列配置过程

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图或PVC视图	进入接口视图	**interface** *interface-type interface-number*	-
	进入PVC视图	**interface atm** *interface-number*	-
		pvc *vpi/vci*	
设置接口或PVC队列长度		**qos fifo queue-length** *queue-length*	必选 SR6602的GE接口FIFO队列的缺省长度为1024 FIP-200、FIP-210灵活接口平台上支持的GE接口FIFO队列的缺省长度为1024 其他接口FIFO队列的缺省长度为75

说明

若是 Tunnel 接口、子接口，或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA 链路层协议，则接口需要使能 LR 功能以保证队列功能生效。

5.2.2 FIFO 队列配置过程举例

把接口的先进先出队列的长度配置为 100。

```
<Sysname> system-view
[Sysname] interface gigabitethernet1/0
[Sysname-GigabitEthernet1/0] qos fifo queue-length 100
```

5.3 优先队列的配置

可以给一个优先列表的组定义多条规则，然后把该组规则应用在某接口上。当一个分组到达该接口或 PVC 时（需要由此接口或 PVC 发送出去），系统沿规则链匹配该分组，如果匹配上某规则，则进入相应的队列，匹配结束；如果分组不与任何规则匹配，则进入缺省队列。缺省队列的缺省值为**normal**。

系统以规则被配置的顺序来匹配分组，如果发现分组与某个规则匹配，便结束整个查找过程。

5.3.1 优先队列配置过程

将一组优先列表应用到接口上。对于同一个接口，优先队列的应用命令的重复使用将为接口设定新的优先列表组。

表5-3 优先队列配置过程

操作	命令	说明
进入系统视图	**system-view**	-
配置优先列表	**qos pql** *pql-index* **protocol ip** [*queue-key key-value*] **queue** { **bottom** | **middle** | **normal** | **top** }	必选 可以根据需要选择优先列表的配置命令
设置缺省队列	**qos pql** *pql-index* **default-queue** { **bottom** | **middle** | **normal** | **top** }	可选 本配置用来指明不匹配规则的数据包的入队队列
配置队列长度	**qos pql** *pql-index* **queue** { **bottom** | **middle** | **normal** | **top** } **queue-length** *queue-length*	可选
进入接口视图	**interface** *interface-type interface-number*	-
应用优先列表	**qos pq pql** *pql-index*	必选 缺省情况下，接口不使用 PQ，而使用 FIFO
显示接口上优先列表的配置情况	**display qos pq interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行
显示优先列表的内容	**display qos pql** [*pql-number*]	可选 **display** 命令可以在任意视图下执行

说明

若是 Tunnel 接口、子接口或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA、链路层协议，则接口需要使能 LR 功能以保证队列功能生效。

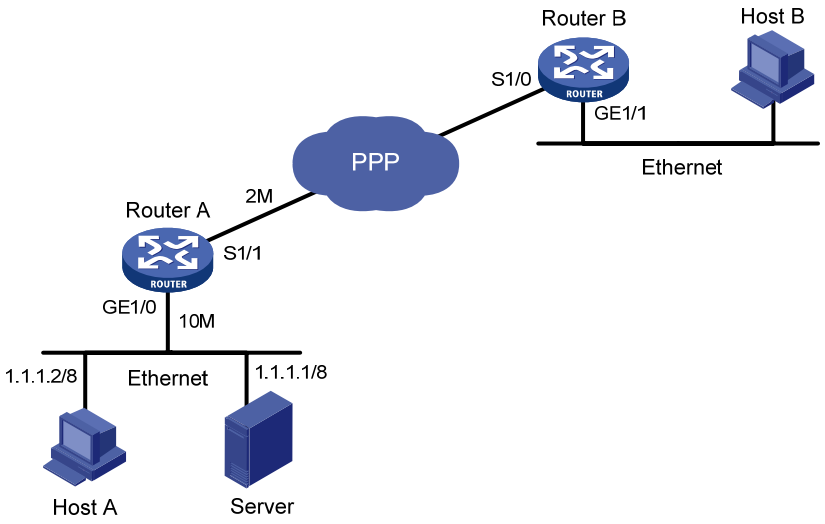
5.3.2 优先队列配置过程举例

1. 配置需求

如图所示，Server 和 Host A 通过 Router A 向 Host B 发送数据（其中 Server 发送关键业务数据，Host A 发送非关键业务数据）时，由于 Router A 入接口 GigabitEthernet1/0 的速率大于出接口 Serial1/1 的速率，在 Serial1/1 接口处可能发生拥塞，导致丢包。要求在网络拥塞时保证 Server 发送的关键业务数据得到优先处理。

2. 组网图

图5-6 优先队列配置组网图



3. 配置步骤

配置 Router A：

配置 ACL 规则列表，分别匹配来源于 Server 和 Host A 的报文。

```
[RouterA] acl number 2001
[RouterA-acl-basic-2001] rule permit source 1.1.1.1 0.0.0.0
[RouterA] acl number 2002
[RouterA-acl-basic-2002] rule permit source 1.1.1.2 0.0.0.0
```

配置优先队列规则组，使得网络拥塞发生时，源自 Server 的报文能够进入 PQ 的 top 队列缓存，源自 Host A 的报文能够进入 bottom 队列缓存，并且设定 top 队列的最大队列长度为 50、bottom 队列的最大队列长度为 100。

```
[RouterA] qos pql 1 protocol ip acl 2001 queue top
[RouterA] qos pql 1 protocol ip acl 2002 queue bottom
[RouterA] qos pql 1 queue top queue-length 50
[RouterA] qos pql 1 queue bottom queue-length 100
```

在接口 Serial1/1 上启用优先队列规则组 1。

```
[RouterA] interface serial 1/1
[RouterA-Serial1/1] qos pq pql 1
```

5.4 定制队列的配置

定制列表共可分为 16 个组（1～16），每个组指明了什么样的分组进入什么样的队列、各队列的长度和每次轮询各队列所能连续发送的字节数等信息。在一个接口或 PVC 上只能应用一个组。

5.4.1 定制队列配置过程

表5-4 定制队列配置过程

操作	命令	说明
进入系统视图	**system-view**	-
配置定制列表	**qos cql** *cql-index* **protocol ip** [*queue-key key-value*] **queue** *queue-number*	可选 可以根据需要选择优先列表的配置命令
设置缺省队列	**qos cql** *cql-index* **default-queue** *queue-number*	可选 本配置指明不匹配规则的数据包的入队队列。
配置队列长度	**qos cql** *cql-index* **queue** *queue-number* **queue-length** *queue-length*	可选
配置连续发送字节数	**qos cql** *cql-index* **queue** *queue-number* **serving** *byte-count*	可选
进入接口视图	**interface** *interface-type interface-number*	-
应用定制列表	**qos cq cql** *cql-index*	必选 缺省情况下，接口不使用 CQ，而使用 FIFO
显示定制列表在接口上应用情况	**display qos cq interface** [*interface-type interface-number*] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]	可选 **display** 命令可以在任意视图下执行
显示定制列表的内容	**display qos cql**	可选 **display** 命令可以在任意视图下执行

说明

若是 Tunnel 接口、子接口或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA、链路层协议，则接口需要使能 LR 功能以保证队列功能生效。

5.4.2 定制队列配置过程举例

1. 组网需求

配置定制队列，使得匹配访问控制列表 2001 的数据进入队列 1，队列 1 每次轮询所发送数据包的字节数为 2000。

2. 配置步骤

进入系统视图。

```
<Sysname> system-view
```

配置 ACL 规则列表

```
[RouterA] acl number 2001
[RouterA-acl-basic-2001] rule permit source 1.1.1.1 0.0.0.0
```

配置定制规则。

```
[Sysname] qos cql 1 protocol ip acl 2001 queue 1
[Sysname] qos cql 1 queue 1 serving 2000
```

在接口 Serial2/0 上启用定制队列规则组 1。

```
[Sysname] interface serial 2/0
[Sysname-Serial 2/0] qos cq cql 1
```

5.5 加权公平队列的配置

5.5.1 加权公平队列配置过程

当接口没有使用 WFQ 策略时，使用 **qos wfq** 命令可以使接口使用 WFQ 策略，同时指定 WFQ 的参数。如果接口已经使用了 WFQ 策略，使用 **qos wfq** 命令可以修改 WFQ 的参数。

表5-5 加权公平队列配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置加权队列	**qos wfq** [**precedence** | **dscp**] [**queue-length** *max-queue-length*] [**queue-number** *total-queue-number*]]	必选 缺省情况下，接口不使用 WFQ，而使用 FIFO
显示接口加权公平队列的配置统计信息	**display qos wfq interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行



说明

若是 Tunnel 接口、子接口或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA、链路层协议，则接口需要使能 LR 功能以保证队列功能生效。

5.5.2 加权公平队列配置过程举例

1. 组网需求

在接口 Serial2/0 上应用 WFQ，队列长度设置为 100，总队列个数设置为 512 个。

2. 配置步骤

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface serial 2/0
```

在接口 Serial2/0 上应用 WFQ，并设置队列长度为 100，总队列个数设置为 512 个。

```
[Sysname-Serial2/0] qos wfq queue-length 100 queue-number 512
```

5.6 基于类的队列的配置

基于类的队列 CBQ 的配置步骤如下：

(1) 定义类，并在类视图中定义一组流分类规则

(2) 定义流行为，并在流行为视图中定义一组 QoS 特性

(3) 定义策略，在策略视图下为使用的类中指定对应的流行为

(4) 在接口视图或 PVC 视图下应用 QoS 策略

为方便用户使用，系统预定义了一些类、流行为以及策略，具体如下。

1. 系统预定的类

系统预定义了一些类，并为这些类定义了通用的规则，用户定义策略时可直接使用这些类，这些类包括：

(1) 缺省类

default-class：匹配的是缺省数据流。

(2) 基于 DSCP 的预定义类

ef、af1、af2、af3、af4：分别匹配 IP DSCP 值 ef、af1、af2、af3、af4

(3) 基于 IP 优先级的预定义类

ip-prec0，ip-prec1，…ip-prec7：分别匹配 IP 优先级 0，1，…7

(4) 基于 MPLS EXP 的预定义类

mpls-exp0，mpls-exp1,…mpls-exp7：分别匹配 MPLS EXP 值 0，1，…7

2. 系统预定义的流行为

系统预定义了一些流行为，并为这些流行为定义了 QoS 特性：

- ef：定义了一个特性为入 EF 队列，占用带宽为接口或 PVC 可用带宽的 20%
- af：定义了一个特性为入 AF 队列，占用带宽为接口或 PVC 可用带宽的 20%
- be：不定义任何特性
- be-flow-based：定义了一个特性为入 WFQ 队列，丢弃策略为 WRED，其中 WFQ 默认有 256 条队列，WRED 为基于 IP 优先级丢弃的 WRED

3. 系统预定义的策略

系统预定义了一个策略，为该策略指定了使用的预定义类，并为这些类指定预定义的动作。该策略名为 default，具有缺省的 CBQ 动作。

default 策略的具体规则如下：

- 预定义类 ef，采用预定义流行为 ef
- 预定义类 af1～af4，采用预定义流行为 af
- default-class 类，采用预定义流行为 be

5.6.1 配置接口最大可用带宽

1. 配置接口最大可用带宽

最大可用带宽指 CBQ 中报文入队列带宽检查时使用的最大接口带宽，并非指物理接口的实际带宽。

表5-6 配置接口最大可用带宽

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口最大可用带宽	**qos max-bandwidth** *bandwidth*	必选

在未配置各种接口的最大可用带宽的条件下，计算 CBQ 时实际使用的基准带宽如下：

- 对于物理接口，其取值为物理接口实际的波特率或速率；
- 对于 E1、MFR 等通过绑定生成的逻辑串口，其取值为绑定通道的总带宽；
- 对于 VT 等模板类型的接口，取值为 1000000Kbps；
- 对于其他虚接口如 Tunnel 接口，取值为 0Kbps。

说明

- 建议最大可用带宽的取值小于物理接口或逻辑链路的实际可用带宽。
- 对于 VT 等主通道型接口，如果配置了 **qos max-bandwidth** 命令，**af**、**ef** 按照 **qos max-bandwidth** 的配置值进行队列带宽检测及计算，同步到 VA、B 通道等子通道类型接口上的 **af**、**ef** 也按照该值进行检测及计算，忽略子通道接口带宽，此种情况主通道接口及子通道接口 QoS 配置相同，仅输出主通道接口的提示信息；如果未配置 **qos max-bandwidth** 命令，**af**、**ef** 按照 1Gbps 带宽进行计算，同步到子通道的 **af**、**ef** 按照 VA、B 通道实际带宽进行队列计算，此种情况下，若子通道接口因带宽变化导致队列失效，将输出子通道接口提示信息。
- 对于 MP-group 和 MFR 接口，如果配置了 **qos max-bandwidth** 命令，**af**、**ef** 按照 **qos max-bandwidth** 的配置值进行队列带宽检测及计算。如果未配置 **qos max-bandwidth** 命令，当绑定子通道的总带宽足够时（子通道带宽之和大于等于 **af**、**ef** 的带宽之和），**af**、**ef** 按照接口的实际带宽进行计算；当绑定子通道的总带宽不够时（子通道带宽之和小于 **af**、**ef** 的带宽之和），**af**、**ef** 按照 1Gbps 带宽进行计算，同时输出实际带宽不够的提示信息，这种情况下不能保证队列功能有效。最大预留带宽占可用带宽的百分比可以通过命令 **qos reserved-bandwidth** 来配置。
- 若是 Tunnel 接口、子接口，或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA 链路层协议，上述接口需要配置该命令以提供 CBQ 计算的基准带宽。

2. 配置接口最大可用带宽举例

(1) 组网需求

配置接口的最大可用带宽是 60Kbps。

(2) 配置步骤

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface gigabitethernet 1/0
```

在接口 GigabitEthernet1/0 上配置最大可用带宽。

```
[Sysname-GigabitEthernet1/0] qos max-bandwidth 60
```

5.6.2 定义类

定义类首先要创建一个类名称，然后在此类视图下配置其匹配规则。

表5-7 定义类

操作	命令	说明
进入系统视图	**system-view**	-
定义类并进入类映射视图	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	必选 缺省为 **and**，即类视图下各匹配规则之间的关系为逻辑与
定义匹配数据包的规则	**if-match** [**not**] *match-criteria*	必选
显示类信息	**display traffic classifier** { **system-defined** | **user-defined** } [*tcl-name*]	可选 **display** 命令可以在任意视图下执行

5.6.3 定义流行为

定义流行为首先需要创建一个流行为名称，然后在此流行为视图下配置其特性。

1. 配置确保转发（AF），并配置最小可保证带宽

表5-8 配置确保转发（AF），并配置最小可保证带宽

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置确保转发（AF），并配置最小可保证带宽	**queue af bandwidth** { *bandwidth* | **pct** *percentage* }	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行



说明

- 该行为只能应用在接口的出方向。
- 在同一流行为下需要用同一单位配置 **queue ef** 和 **queue af**，或者用 *bandwidth*，或者用百分比进行配置。

2. 配置加速转发（EF），并配置最大带宽

表5-9 配置加速转发（EF），并配置最大带宽

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置加速转发（EF），并配置最大带宽	**queue ef bandwidth** { *bandwidth* [**cbs** *burst*] | **pct** *percentage* [**cbs-ratio** *ratio*] }	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行

说明

- 本配置在流行为视图下不能与 **queue af**、**queue-length** 和 **wred** 同时使用。
- 缺省类不能与包含加速转发的行为关联。
- 在同一策略下各个类需用同一单位配置 **queue ef** 和 **queue af**，或者用 *bandwidth*，或者用百分比进行配置。

3. 配置采用公平队列

表5-10 配置采用公平队列

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置采用公平队列	**queue wfq** [**queue-number** *total-queue-number*]	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行

说明

配置了公平队列的流行为仅可以与缺省类关联使用。

4. 配置最大队列长度

配置最大队列长度，丢弃方式为尾丢弃。

当发现低优先级业务抢占 AF 业务带宽，导致 AF 业务带宽无法得到保障，可以尝试增大 AF 业务的队列长度。

表5-11 配置最大队列长度

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置最大队列长度	**queue-length** *queue-length*	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行



说明

最大队列长度命令必须在配置了 **queue af** 或 **queue wfq** 后使用；执行 **undo queue af** 或 **undo queue wfq** 命令，则 **queue-length** 也同时被取消。

5. 配置丢弃方式为随机丢弃方式

表5-12 配置丢弃方式为随机丢弃方式

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置丢弃方式为随机丢弃方式	**wred** [**dscp** | **ip-precedence**]	必选 **dscp**：表明在为一个包计算丢弃概率时使用的是 DSCP 值 **ip-precedence**：表明在为一个包计算丢弃概率时使用的是 IP 优先级值，作为缺省配置
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行



说明

- **wred** [**dscp** | **ip-precedence**]命令必须在配置了 **queue af** 或 **queue wfq** 后使用。
- **wred** 和 **queue-length** 是互斥的。
- 删除 WRED 时将删除在该随机丢弃下的其他配置。
- 当接口上应用了包含 WRED 特性的 QoS 策略后，原有的接口级的 WRED 配置失效。

6. 设置 WRED 计算平均队列长度的指数

表5-13 设置 WRED 计算平均队列长度的指数

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置WRED计算平均队列长度的指数	**wred weighting-constant** *exponent*	必选 缺省情况下，WRED 计算平均队列长度的指数为 9
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行

说明

wred weighting-constant 必须在配置了 **queue af** 或 **queue wfq**，并已用 **wred** 使能了 WRED 丢弃方式后才可以进行配置。

7. 配置 WRED 各 DSCP 的下限、上限和丢弃概率分母

在进行本配置前需已用 **wred dscp** 使能了基于 DSCP 的 WRED 丢弃方式。

表5-14 设置 WRED 各 DSCP 的下限、上限和丢弃概率分母

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
配置 WRED 各 DSCP 的下限、上限和丢弃概率分母	**wred dscp** *dscp-value* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行

dscp-value：DSCP 值，取值范围为 0～63，或可以是如下关键字 ef，af11，af12，af13，af21，af22，af23，af31，af32，af33，af41，af42，af43，cs1，cs2，cs3，cs4，cs5，cs6，cs7 或 default。

说明

- 取消 **wred** 配置，**wred dscp** 配置同时被取消。
- 取消 **queue af** 或 **queue wfq** 配置，丢弃参数的配置同时被取消。

8. 配置 WRED 各 IP 优先级的下限、上限和丢弃概率分母

进行本配置前需已用 **wred ip-precedence** 使能了基于 IP 优先级的 WRED 丢弃方式。

表5-15 配置 WRED 各 IP 优先级的下限、上限和丢弃概率分母

操作	命令	说明
进入系统视图	**system-view**	-
定义一个流行为并进入流行为视图	**traffic behavior** *behavior-name*	必选 *behavior-name*：流行为名，不允许为系统预定义流行为
设置 WRED 各优先级的下限、上限和丢弃概率分母	**wred ip-precedence** *precedence* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]	必选
显示流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]	可选 **display** 命令可以在任意视图下执行

说明

- 取消 WRED 配置，**wred ip-precedence** 配置同时被取消。
- 取消 **queue af** 或 **queue wfq** 配置，丢弃参数的配置同时被取消。

5.6.4 定义策略

策略定义该策略中类和流行为的对应关系，而每个流行为由一组 QoS 动作组成。例如队列调度，包括 EF，AF，WFQ，流量监管，流量整形，WRED、标记等等。

表5-16 在策略中为类指定流行为

操作	命令	说明
进入系统视图	**system-view**	-
定义策略并进入策略视图	**qos policy** *policy-name*	-
在策略中为类指定采用的流行为	**classifier** *tcl-name* **behavior** *behavior-name*	必选 *tcl-name*：类名，必须是已经定义的类，可以是系统定义或用户定义类 *behavior-name*：必须是已定义的流行为名，可以是系统定义或用户定义流行为
显示指定策略中指定类及与类关联的流行为的配置信息	**display qos policy** { **system-defined** | **user-defined** } [*policy-name* [**classifier** *tcl-name*]]	可选 **display** 命令可以在任意视图下执行

5.6.5 应用策略

qos apply policy 命令将一个策略映射到具体的物理接口或 ATM PVC。一个策略映射可以在多个物理接口或 ATM PVC 上得到应用。

表5-17 将接口或 ATM PVC 与所设置的策略相关联

操作		命令	说明
进入系统视图		**system-view**	-
进入接口视图或 PVC 视图	进入接口视图	**interface** *interface-type interface-number*	二者必选其一 进入接口视图后，下面进行的配置只在当前接口生效；进入 PVC 视图后，下面进行的配置只在当前 PVC 生效
	进入 PVC 视图	**interface atm** *interface-number*	
		pvc *vpi/vci*	
在接口或 PVC 上应用关联的策略		**qos apply policy** *policy-name* { **inbound** | **outbound** }	必选
显示指定接口、指定 PVC 或所有接口与 PVC 上的策略的配置信息和运行情况		**display qos policy interface** [*interface-type interface-number* [**inbound** | **outbound**] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]	可选 **display** 命令可以在任意视图下执行

策略在接口或 ATM PVC 视图下应用的规格如下：

- 普通物理接口和 MP 引用的 VT，可以应用配置了各种特性（包括 remark、car、gts、queue af、queue ef、queue wfq、wred 等）的策略。
- 配置了流量整形（gts）和队列（queue ef、queue af、queue wfq）特性的策略，不能作为入方向策略应用在入接口上。
- 若是 Tunnel 接口、子接口，或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA 链路层协议，则接口需要使能 LR 功能以保证 CBQ 队列功能生效，同时需要配置 **qos max-bandwidth** 命令以提供 CBQ 计算的基准带宽。

5.6.6 基于类的队列典型配置举例

1. 组网需求

在下面的组网图中，从 Router C 发出的数据流经过 Router A 和 Router B 到达 Router D，需求如下：

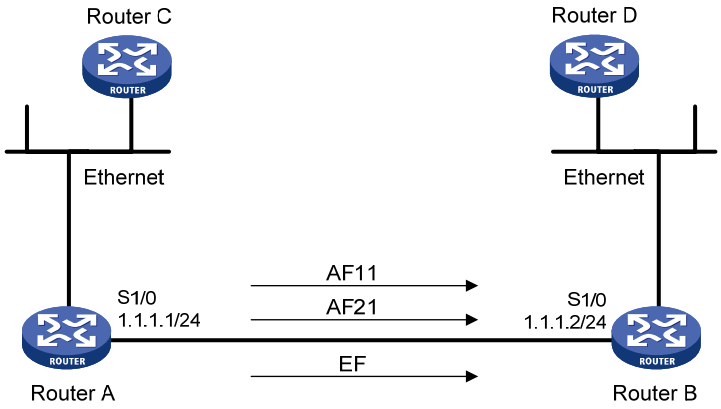
- Router C 发出的数据流根据 IP 报文的 DSCP 域分为 3 类，要求配置 QoS 策略，对于 DSCP 域为 AF11 和 AF21 的流进行确保转发（AF），最小带宽为 5%；
- 对于 DSCP 域为 EF 的流进行加速转发（EF），最大带宽为 30%。

在进行配置之前，应保证：

- Router C 发出的流能够通过 Router A 和 Router B 可达 Router D。
- 报文的 DSCP 域在进入 Router A 之前已经设置完毕。

2. 组网图

图5-7 基于类的队列配置组网图



3. 配置步骤

Router A 上的配置如下。

定义三个类，分别匹配 DSCP 域为 AF11、AF21 和 EF 的 IP 报文。

```
[RouterA] traffic classifier af11_class
[RouterA-classifier-af11_class] if-match dscp af11
[RouterA-classifier-af11_class] quit
[RouterA]traffic classifier af21
[RouterA-classifier-af21] if-match dscp af21
[RouterA-classifier-af21] quit
[RouterA] traffic classifier ef_class
[RouterA-classifier-ef_class] if-match dscp ef
[RouterA-classifier-ef_class] quit
```

定义流行为，配置 AF，并分配最小可用带宽。

```
[RouterA] traffic behavior af11_behav
[RouterA-behavior-af11_behav] queue af bandwidth pct 5
[RouterA-behavior-af11_behav] quit
[RouterA] traffic behavior af21_behav
[RouterA-behavior-af21_behav] queue af bandwidth pct 5
[RouterA-behavior-af21_behav] quit
```

定义流行为，配置 EF，并分配最大可用带宽（对于 EF 流，将同时保证带宽和时延）。

```
[RouterA] traffic behavior ef_behav
[RouterA-behavior-ef_behav] queue ef bandwidth pct 30
[RouterA-behavior-ef_behav] quit
```

定义 QoS 策略，将已配置的流行为指定给不同的类。

```
[RouterA] qos policy dscp
[RouterA-qospolicy-dscp] classifier af11_class behavior af11_behav
[RouterA-qospolicy-dscp] classifier af21_class behavior af21_behav
[RouterA-qospolicy-dscp] classifier ef_class behavior ef_behav
[RouterA-qospolicy-dscp] quit
```

将已定义的 QoS 策略应用在 Router A 的出方向。

```
[RouterA] interface serial 1/0
[RouterA-Serial1/0] ip address 1.1.1.1 255.255.255.0
[RouterA-Serial1/0] qos apply policy dscp outbound
```

配置完成后，当发生拥塞时，可以观察到 EF 流被以较高的优先级转发。

5.6.7 基于类的队列的显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示基于类的队列的运行情况，通过查看显示信息验证配置的效果。

表5-18 基于类的队列的显示和维护

操作	命令
显示设备配置的类信息	**display traffic classifier** { **system-defined** | **user-defined** } [*tcl-name*]
显示设备配置的流行为信息	**display traffic behavior** { **system-defined** | **user-defined** } [*behavior-name*]
显示指定策略中指定类及与类关联的流行为的配置信息	**display qos policy** { **system-defined** | **user-defined** } [*policy-name* [**classifier** *tcl-name*]]
显示指定接口、指定 PVC 或所有接口与 PVC 上策略的配置信息和运行情况	**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]
显示指定接口、指定 PVC 或所有接口与 PVC 的基于类的队列配置信息和运行情况。	**display qos cbq interface** [*interface-type interface-number*] [**pvc** { *pvc-name* [*vpi/vci*] | *vpi/vci* }]

5.7 RTP 优先队列的配置

5.7.1 RTP 优先队列配置过程

表5-19 RTP 队列配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
应用 RTP 队列	**qos rtpq start-port** *first-rtp-port-number* **end-port** *last-rtp-port-number* **bandwidth** *bandwidth* [**cbs** *burst*]	必选
配置最大预留带宽占接口实际可用带宽的百分比	**qos reserved-bandwidth pct** *percent*	可选 缺省值为 80
显示接口上 RTP 队列的配置统计信息	**display qos rtpq interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行



说明

若是 Tunnel 接口、子接口或是 VT 接口用于 PPPoE、PPPoA、PPPoEoA、链路层协议，则接口需要使能 LR 功能以保证队列功能生效。

5.7.2 RTP 优先队列配置过程举例

1. 组网需求

在接口上配置 RTP 队列，设置为 RTP 队列的应用分配的最大预留带宽为可用带宽的 70%。

配置串口 Serial1/0 下启动 RTP 队列特性，起始端口号为 16384，结束端口为 32767，RTP 报文占用 64kbps 的带宽，如果输出接口拥塞，进入 RTP 队列。

2. 配置步骤

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface serial 2/0
```

设置为 RTP 队列的应用分配的最大预留带宽为可用带宽的 70%。

```
[Sysname-Serial2/0] qos reserved-bandwidth pct 70
```

配置串口 Serial2/0 下启动 RTP 队列特性，起始端口号为 16384，结束端口为 32767，RTP 报文占用 64kbps 的带宽，如果输出接口拥塞，进入 RTP 队列。

```
[Sysname-Serial2/0] qos rtpq start-port 16384 end-port 32767 bandwidth 64
```

5.8 QoS 令牌配置

5.8.1 QoS 令牌配置过程

当进行 FTP 等数据传输工作时，如果配置了 QoS 的 CQ 和 WFQ，由于上层协议（如 TCP）提供了流控功能，可能会导致 QoS 的队列失效。设备提供了 QoS 令牌功能来解决这种问题。QoS 的发送令牌功能提供了一种底层队列的流量控制机制，它可以根据令牌的数量控制向底层接口队列发送的报文数量。

通常，在进行 FTP 传输时，建议将接口的发送令牌数量设置为 1。

如果上层协议（如 UDP）没有流控功能，建议不要使用 QoS 令牌功能，以提高数据传输的效率。

请在接口视图下进行下列配置。

表5-20 QoS 令牌配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 QoS 令牌功能	**qos qmtoken** *token-number*	必选 缺省情况下，未启用 QoS 令牌功能



说明

- 在进行了本配置后，需要用 **shutdown / undo shutdown** 功能将接口重新启动，才能使能 QoS 的发送令牌功能。
- 目前只有串口支持本配置。

5.8.2 QoS 令牌配置举例

1. 组网需求

在接口上配置令牌参数

2. 配置步骤

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface serial 2/0
```

设置 QoS 的发送令牌数量为 1。

```
[Sysname-Serial2/0] qos qmtoken 1
```

6 拥塞避免

6.1 拥塞避免简介

过度的拥塞会对网络资源造成极大危害，必须采取某种措施加以解除。拥塞避免（Congestion Avoidance）是一种流量控制机制，它通过监视网络资源（如队列或内存缓冲区）的使用情况，在拥塞产生或有加剧的趋势时主动丢弃报文，通过调整网络的流量来解除网络过载。

与端到端的流量控制相比，这里的流量控制具有更广泛的意义，它影响到设备中更多的业务流的负载。设备在丢弃报文时，需要与源端的流量控制动作（比如 TCP 流量控制）的配合，调整网络的流量到一个合理的负载状态。丢包策略和源端流控机制有效的组合，可以使网络的吞吐量和利用效率最大化，并且使报文丢弃和延迟最小化。

1. 传统的丢包策略

传统的丢包策略采用尾部丢弃（Tail-Drop）的方法。当队列的长度达到最大值后，所有新到来的报文都将被丢弃。

这种丢弃策略会引发 TCP 全局同步现象：当队列同时丢弃多个 TCP 连接的报文时，将造成多个 TCP 连接同时进入拥塞避免和慢启动状态以降低并调整流量，而后又会在某个时间同时出现流量高峰。如此反复，使网络流量忽大忽小，网络不停震荡。

2. RED 与 WRED

为避免 TCP 全局同步现象，可使用 RED（Random Early Detection，随机早期检测）或 WRED（Weighted Random Early Detection，加权随机早期检测）。

RED 和 WRED 通过随机丢弃报文避免了 TCP 的全局同步现象，使得当某个 TCP 连接的报文被丢弃、开始减速发送的时候，其他的 TCP 连接仍然有较高的发送速度。这样，无论什么时候，总有 TCP 连接在进行较快的发送，提高了线路带宽的利用率。

在 RED 类算法中，为每个队列都设定上限和下限，对队列中的报文进行如下处理：

- 当队列的长度小于下限时，不丢弃报文；
- 当队列的长度超过上限时，丢弃所有到来的报文；
- 当队列的长度在上限和下限之间时，开始随机丢弃到来的报文。队列越长，丢弃概率越高，但有一个最大丢弃概率。

与 RED 不同，WRED 生成的随机数是基于优先权的，它引入 IP 优先权区别丢弃策略，考虑了高优先权报文的利益，使其被丢弃的概率相对较小。

直接采用队列的长度和上限、下限比较并进行丢弃，将会对突发性的数据流造成不公正的待遇，不利于数据流的传输。WRED 采用平均队列和设置的队列上限、下限比较来确定丢弃的概率。

队列平均长度既反映了队列的变化趋势，又对队列长度的突发变化不敏感，避免了对突发性数据流的不公正待遇。计算队列平均长度的公式为：平均队列长度=（以前的平均队列长度×（1-1/（2 的 n 次方））） +（当前队列长度×（1/（2 的 n 次方））） 。其中 n 可以通过命令 **qos wred weighting-constant** 进行配置。

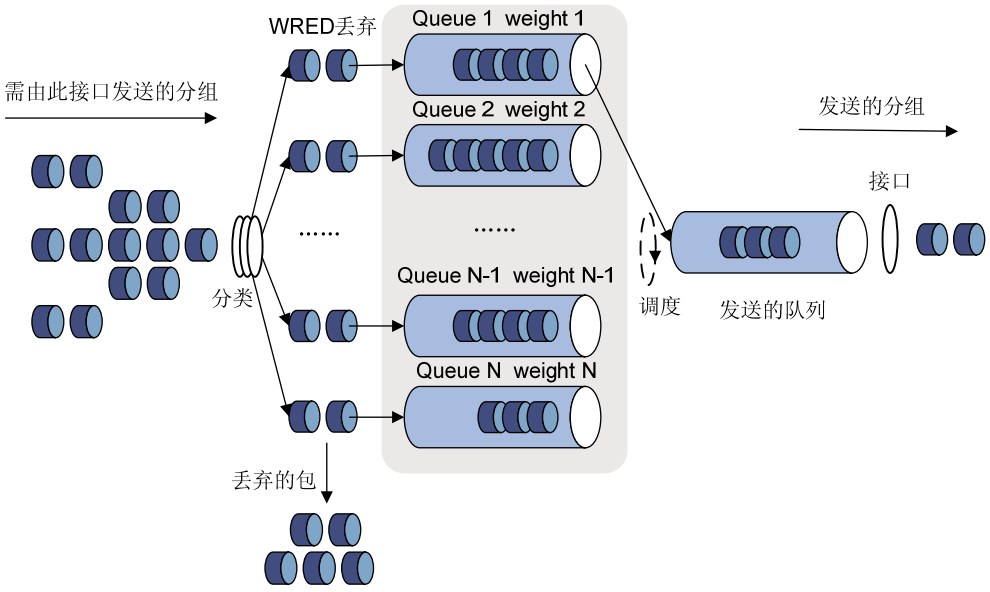
当队列机制采用 WFQ 时，可以为不同优先级的报文设定计算队列平均长度时的指数、上限、下限、丢弃概率，从而对不同优先级的报文提供不同的丢弃特性。

当队列机制采用 FIFO、PQ、CQ 时，可以为每个队列设定计算队列平均长度时的指数、上限、下限、丢弃概率，为不同类别的报文提供不同的丢弃特性。

3. WRED 和队列机制的关系

WRED 和队列机制的关系如下图所示。

图6-1 WRED 和队列机制关系示意图



当 WRED 和 WFQ 配合使用时，可以实现基于流的 WRED。在进行分类的时候，不同的流有自己的队列，对于流量小的流，由于其队列长度总是比较小，所以丢弃的概率将比较小。而流量大的流将会有较大的队列长度，从而丢弃较多的报文，保护了流量较小的流的利益。

6.2 WRED 配置的说明

6.2.1 WRED 的配置方式

SR6600 路由器支持的 WRED 配置方式为接口配置方式，即在接口上配置 WRED 的各种参数，并使能 WRED。

6.2.2 WRED 的参数说明

在进行 WRED 配置时，需要事先确定如下参数：

- 队列上限和下限：当队列平均长度小于下限时，不丢弃报文。当队列平均长度在上限和下限之间时，设备随机丢弃报文，队列越长，丢弃概率越高。当队列平均长度超过上限时，丢弃所有到来的报文。
- 计算平均队列长度的指数：指数越大，计算平均队列长度时对队列的实时变化越不敏感。
- 计算丢弃概率的分母：在计算丢弃概率的公式中作为分母。取值越大，计算出的丢弃概率越小。

6.3 以接口配置方式配置 WRED

6.3.1 配置过程

表6-1 以接口配置方式配置 WRED 配置过程

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 WRED	**qos wred** [**dscp** | **ip-precedence**] **enable**	必选
设置计算平均队列长度的指数	**qos wred weighting-constant** *exponent*	可选 缺省情况下，该指数取值为 9
设置各优先级的对应的参数	**qos wred ip-precedence** *ip-precedence* | **dscp** *dscp-value* } **low-limit** *low-limit* **high-limit** *high-limit* **discard-probability** *discard-prob*	可选 缺省情况下，*low-limit* 为 10，*high-limit* 为 30，*discard-prob* 为 10
显示接口的 WRED 配置情况和统计信息	**display qos wred interface** [*interface-type interface-number*]	可选 **display** 命令可以在任意视图下执行

注意

qos wred enable 命令需要先在接口上应用 WFQ 队列。

6.3.2 配置举例

1. 组网需求

- 在接口上配置基于 IP 优先级的 WRED
- 设置 IP 优先级为 3 的报文的队列下限为 20、上限为 40、丢弃概率分母为 15
- 设置计算平均队列长度的指数为 6

2. 配置步骤

进入系统视图。

```
<Sysname> system-view
```

进入接口视图。

```
[Sysname] interface gigabitethernet 1/0
```

使能基于 IP 优先级的 WRED。

```
[Sysname-GigabitEthernet1/0] qos wred ip-precedence enable
```

设置优先级为 3 的报文的队列下限为 20，上限为 40，丢弃概率分母为 15。

```
[Sysname-GigabitEthernet1/0] qos wred ip-precedence 3 low-limit 20 high-limit 40 discard-probability 15
```

配置计算平均队列长度的指数为 6。

```
[Sysname-GigabitEthernet1/0] qos wred weighting-constant 6
```

6.4 WRED 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 WRED 的运行情况，通过查看显示信息验证配置的效果。

表6-2 WRED 显示和维护

操作	命令
显示接口或 PVC 的 WRED 配置情况和统计信息	**display qos wred interface** [*interface-type interface-number*]

6.5 WRED 典型配置举例

1. 组网需求

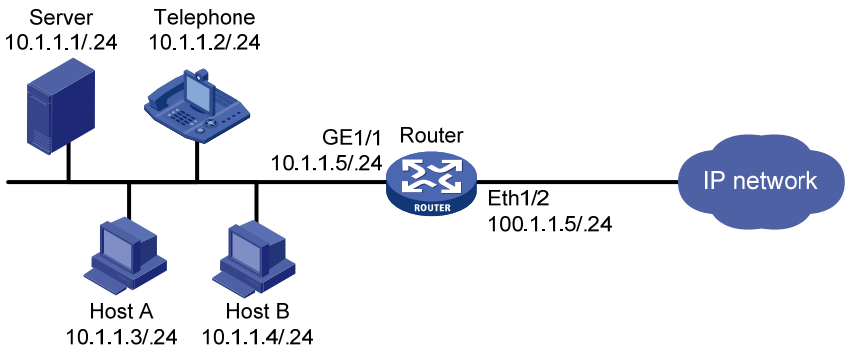
如 图 6-2所示，，Server、Telephone、Host A和Host B通过Router向IP网络发送数据，其中Server发送关键业务数据，Telephone发送语音数据，Host A和Host B发送非关键业务数据。由于Router入接口GE1/1 的速率大于出接口Eth1/2 的速率，在Eth1/2 接口处可能发生拥塞，并且可能出现拥塞加剧。

要求在网络拥塞时保证 Server 和 Telephone 发送的业务数据得到优先发送。同时，Host A 和 Host B 的数据在发送的过程中也需要一定的带宽保证，可以有少量延迟，但不希望延迟过大。在拥塞加剧的时候需要考虑优先权来丢弃报文。

需要配置 WFQ 和 WRED 配合调度和丢弃。

2. 组网图

图6-2 WRED 典型配置举例组网图



3. 配置步骤

配置 ACL，分别匹配来源于 Server、Telephone、Host A 和 Host B 的报文。

```
<Router> system-view
[Router] acl number 2001
[Router-acl-basic-2001] rule 1 permit source 10.1.1.1 0
[Router-acl-basic-2001] quit
[Router] acl number 2002
```

```
[Router-acl-basic-2002] rule 2 permit source 10.1.1.2 0
[Router-acl-basic-2002] quit
[Router] acl number 2003
[Router-acl-basic-2003] rule 3 permit source 10.1.1.3 0
[Router-acl-basic-2003] quit
[Router] acl number 2004
[Router-acl-basic-2004] rule 1 permit source 10.1.1.4 0
[Router-acl-basic-2004] quit
```

标记各个流的优先级。

```
[Router] traffic classifier class1
[Router-classifier-class1] if-match acl 2001
[Router-classifier-class1] quit
[Router] traffic classifier class2
[Router-classifier-class2] if-match acl 2002
[Router-classifier-class2] quit
[Router] traffic classifier class3
[Router-classifier-class3] if-match acl 2003
[Router-classifier-class3] quit
[Router] traffic classifier class4
[Router-classifier-class4] if-match acl 2004
[Router-classifier-class4] quit
[Router] traffic behavior behavior1
[Router-behavior-behavior1] remark ip-precedence 5
[Router-behavior-behavior1] quit
[Router]traffic behavior behavior2
[Router-behavior-behavior2] remark ip-precedence 4
[Router-behavior-behavior2] quit
[Router] traffic behavior behavior3
[Router-behavior-behavior3] remark ip-precedence 3
[Router-behavior-behavior3] quit
[Router] traffic behavior behavior4
[Router-behavior-behavior4] remark ip-precedence 2
[Router-behavior-behavior4] quit
[Router] qos policy aa
[Router-qospolicy-aa] classifier class1 behavior behavior1
[Router-qospolicy-aa] classifier class2 behavior behavior2
[Router-qospolicy-aa] classifier class3 behavior behavior3
[Router-qospolicy-aa] classifier class4 behavior behavior4
[Router-qospolicy-aa] quit
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] qos apply policy aa inbound
[Router-GigabitEthernet1/1] quit
```

配置 WFQ，保证各流量公平调度的同时又考虑到优先权；配置 WRED，在拥塞加剧时考虑优先权进行丢弃。

```
[Router] interface ethernet 1/2
[Router-Ethernet1/2] qos wfq queue-length 100 queue-number 16
[Router-Ethernet1/2] qos wred enable
```

```
[Router-Ethernet1/2] qos wred ip-precedence 5 low-limit 10 high-limit 250
discard-probability 11
[Router-Ethernet1/2] qos wred ip-precedence 4 low-limit 10 high-limit 200
discard-probability 12
[Router-Ethernet1/2] qos wred ip-precedence 3 low-limit 10 high-limit 180
discard-probability 15
[Router-Ethernet1/2] qos wred ip-precedence 2 low-limit 10 high-limit 180
discard-probability 15
[Router-Ethernet1/2] quit
```

7 MPLS QoS 配置

7.1 MPLS QoS 概述

说明

理解 MPLS QoS 需要用户有 MPLS 相关知识背景。MPLS 基本概念以及相关配置请参见“MPLS 分册”中的“MPLS 基本配置”。

MPLS QoS 主要完成以下功能：

- 根据需要在 CE 上或 PE 上对业务流进行分类。例如，MPLS QoS 可以把业务流分为三类：语音、视频和数据。
- PE 在给报文加 Label 时，把 IP 报文携带的 IP 优先级标记映射到标签的 EXP 域，这样原来由 IP 携带的类型信息，现在由标签携带。
- 在 PE 之间，根据标签的 EXP 域，进行有差别的调度（如 PQ、WFQ、CBQ 等），即在一条 LSP 上为携带标签的业务流提供有差别的 QoS。

说明

MPLS 标签中 EXP 字段的处理采用如下原则：

- 任何一台转发设备，如果配置了 QoS 处理都可以将最外层标签的 EXP 字段重新赋值（Remark）；
- 给 IP 报文封装 MPLS 标签时，直接将 IP 报文的 ToS 字段转换成 MPLS 标签的 EXP 字段；
- 标签交换（swap）操作时，EXP 字段保持不变；
- 标签压栈（push）操作时，新压入的外层标签 EXP 字段继承内层标签的 EXP 字段；
- 标签弹栈（pop）操作时，如果弹栈之后是 MPLS 报文，不会将弹出标签的 EXP 字段复制到内层标签上；如果弹栈之后为 IP 报文，不会将 EXP 字段复制到 IP 报文的 ToS 字段。

7.2 MPLS QoS 配置

配置 MPLS QoS 之前，首先要进行 MPLS 的相关配置并指定为显示路由转发方式。MPLS 的详细配置请参见本书的 MPLS 配置部分。本章仅介绍 MPLS QoS 的配置。

MPLS QoS 可以分为四种：

- MPLS PQ，配置过程请参见“7.2.1 配置MPLS PQ”。
- MPLS CQ，配置过程请参见“7.2.2 配置MPLS CQ”。
- MPLS QoS策略，配置过程请参见“7.2.3 配置MPLS QoS策略”。
- MPLS CAR，配置过程请参见“7.2.4 配置MPLS CAR”。

7.2.1 配置 MPLS PQ

1. 配置准备

- 进行了 MPLS 的相关配置

- 确定了 MPLS PQ 中优先列表和 MPLS EXP 值之间的关系
- 确定了应用优先列表的接口

2. 配置过程

表7-1 配置 MPLS PQ

操作	命令	说明
进入系统视图	**system-view**	-
配置优先列表	**qos pql** *pql-index* **protocol mpls-exp** *exp-value-list* **queue** { **bottom** | **middle** | **normal** | **top** }	必选
进入接口视图	**interface** *interface-type interface-number*	-
在接口上应用优先列表组	**qos pq pql** *pql-index*	必选

3. 配置举例

- 配置基于 MPLS 协议的优先列表 10 的分类规则，规则为队列 top 对应 EXP 值 5
- 在 GigabitEthernet1/0 上应用优先列表 10

配置过程如下：

```
<Sysname> system-view
[Sysname] qos pql 10 protocol mpls exp 5 queue top
[Sysname] interface gigabitEthernet1/0
[Sysname-GigabitEthernet1/0] qos pq pql 10
```

7.2.2 配置 MPLS CQ

1. 配置准备

- 进行了 MPLS 的相关配置
- 确定了 MPLS CQ 中定制列表和 MPLS EXP 值之间的关系
- 确定了应用定制列表的接口

2. 配置过程

表7-2 配置 MPLS CQ

操作	命令	说明
进入系统视图	**system-view**	-
根据 MPLS EXP 配置定制列表	**qos cql** *cql-index* **protocol mpls-exp** *exp-value-list* **queue** *queue-number*	必选
进入接口视图	**interface** *interface-type interface-number*	-
在接口上应用定制列表	**qos cq cql** *cql-index*	必选

3. 配置举例

- 配置基于 MPLS 的定制列表 10 的分类规则，设置队列 2 对应 EXP 值 1
- 在 GigabitEthernet1/0 上应用定制列表 10

配置过程如下：

```
<Sysname> system-view
[Sysname] qos cql 10 protocol mpls exp 1 queue 2
[Sysname] interface gigabitEthernet1/0
[Sysname-GigabitEthernet1/0] qos cq cql 10
```

7.2.3 配置 MPLS QoS 策略

1. 配置准备

- 进行了 MPLS 的相关配置
- 确定了 MPLS QoS 策略中类的定义
- 确定了 MPLS QoS 策略中流行为的定义
- 确定了 MPLS QoS 策略中策略的定义
- 确定了应用 MPLS QoS 策略的 QoS 策略的接口

2. 配置过程

表7-3 配置 MPLS QoS 策略

操作	命令	说明
进入系统视图	**system-view**	-
定义类	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	必选 用户定义的类名 *tcl-name* 不允许为系统预定义类 缺省为 **and**，即类视图下各匹配规则之间的关系为逻辑与
	if-match [**not**] **mpls-exp** *exp-value-list*	必选 该规则仅对 MPLS 报文有效，对 IP 报文无效
退出当前视图	**quit**	-
定义流行为	**traffic behavior** *behavior-name*	必选
	remark mpls-exp *exp-value*	必选
退出当前视图	**quit**	-
定义 QoS 策略	**qos policy** *policy-name*	必选
	classifier *tcl-name* **behavior** *behavior-name*	必选 将 MPLS QoS 策略的类和流行为进行绑定
退出当前视图	**quit**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口上应用 QoS 策略	**qos apply policy** *policy-name* { **inbound** | **outbound** }	必选

7.2.4 配置 MPLS CAR

1. 配置准备

- 进行了 MPLS 的相关配置
- 确定了 MPLS CAR 的策略

- 确定了应用 MPLS CAR 的接口

2. 配置过程

表7-4 配置 MPLS CAR

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口上应用 MPLS CAR	**qos car** { **inbound** | **outbound** } { **any** | **acl** *acl-number* | **carl** *carl-index* } **cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**green** *action*] [**red** *action*]	必选

action 中关于 MPLS 的动作如下：

- **remark-mpls-exp-continue** *new-exp*：设置新的 MPLS 报文的 EXP 标志位的值，并继续由下一个 CAR 策略处理，取值范围 0～7。
- **remark-mpls-exp-pass** *new-exp*：设置新的 MPLS 报文的 EXP 标志位的值，并允许数据包通过，取值范围 0～7。

只能在 PE 的入接口上才可以设置 MPLS 报文的 EXP 域。如果入方向为 IP 报文，出方向封装为 MPLS 报文，则配置的 MPLS CAR 策略生效。

7.3 MPLS QoS 配置举例

7.3.1 对同一 VPN 内的流进行 QoS 配置

1. 组网需求

如 图 7-1，网络描述如下：

- CE 1 和 CE 2 都属于 VPN 1
- PE 1 与 P 之间的链路带宽为 2M
- P 与 PE 2 之间的链路带宽为 2M

组网需求为对 VPN 1 中优先级不同的流给予不同的 QoS 保证。

配置可分为两个主要部分：

首先在 CE 1、PE 1、P、PE 2 和 CE 2 进行 MPLS VPN 的配置：

- PE 1、P、PE 2 之间运行 OSPF
- PE 与 CE 之间建立 MP-EBGP 邻居
- PE 与 PE 之间建立 MP-IBGP 邻居

然后是在 PE 1 和 P 上进行 MPLS QoS 的配置：

- 在 PE 1 的入接口 GigabitEthernet1/1 上配置 QoS 策略，根据 MPLS 报文的 DSCP 属性，设置报文 EXP 域的值。

- 在 P 设备上，根据 MPLS 报文的 EXP 域值区分流，配置基于流的队列 CBQ：对于 EXP 为 1 的流保证 10%的带宽；对于 EXP 为 2 的流保证 20%的带宽；对于 EXP 为 3 的流保证 30%的带宽；而对于 EXP 为 4 的流，保证 40%的带宽并保证时延。

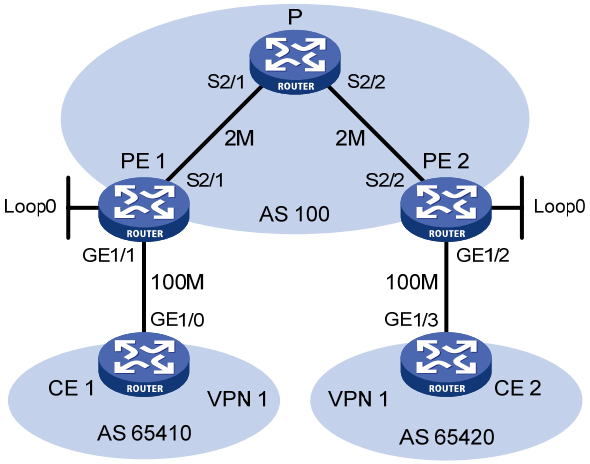
说明

MPLS VPN 的配置请参考本手册“MPLS 分册”中的“MPLS L3VPN 配置”，下面只描述 MPLS QoS 部分的配置。

---

2. 组网图

图7-1 MPLS QoS 组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	GE1/0	10.1.1.2/24	CE 2	GE1/3	10.2.1.2/24
PE 1	GE1/1	10.1.1.1/24	PE 2	GE1/2	10.2.1.1/24
	S2/1	12.1.1.1/24		S2/2	12.2.1.1/24
	Loop0	1.1.1.1/32		Loop0	1.1.1.2/32
P	S2/1	12.1.1.2/24			
	S2/2	12.2.1.2/24			

3. 配置步骤

(1) PE 1 的配置

定义四个类，分别用来匹配同一 VPN 内 DSCP 分别为 AF11、AF21、AF31 和 EF 的 MPLS 报文。

```
<PE1> system-view
[PE1] traffic classifier af11
[PE1-classifier-af11] if-match dscp af11
[PE1-classifier-af11] traffic classifier af21
[PE1-classifier-af21] if-match dscp af21
[PE1-classifier-af21] traffic classifier af31
[PE1-classifier-af31] if-match dscp af31
[PE1-classifier-af31] traffic classifier efclass
[PE1-classifier-efclass] if-match dscp ef
[PE1-classifier-efclass] quit
```

定义四个流行为，设置 MPLS 报文 EXP 域的值。

```
[PE1] traffic behavior exp1
```

```
[PE1-behavior-exp1] remark mpls-exp 1
[PE1-behavior-exp1] traffic behavior exp2
[PE1-behavior-exp2] remark mpls-exp 2
[PE1-behavior-exp2] traffic behavior exp3
[PE1-behavior-exp3] remark mpls-exp 3
[PE1-behavior-exp3] traffic behavior exp4
[PE1-behavior-exp4] remark mpls-exp 4
[PE1-behavior-exp4] quit
```

定义 QoS 策略，为不同类的报文指定流行为，即对不同类的报文标记不同的 EXP 值。

```
[PE1] qos policy REMARK
[PE1-qospolicy-REMARK] classifier af11 behavior exp1
[PE1-qospolicy-REMARK] classifier af21 behavior exp2
[PE1-qospolicy-REMARK] classifier af31 behavior exp3
[PE1-qospolicy-REMARK] classifier efclass behavior exp4
[PE1-qospolicy-REMARK] quit
```

在 MPLS 网络的 PE 入接口上应用已定义的 QoS 策略。

```
[PE1] interface gigabitethernet 1/1
[PE1-GigabitEthernet1/1] qos apply policy REMARK inbound
[PE1-GigabitEthernet1/1] quit
```

(2) P 的配置

定义四个类，分别用来匹配 EXP 域为 1、2、3 和 4 的 MPLS 报文。

```
<P> system-view
[P] traffic classifier EXP1
[P-classifier-EXP1] if-match mpls-exp 1
[P-classifier-EXP1] traffic classifier EXP2
[P-classifier-EXP2] if-match mpls-exp 2
[P-classifier-EXP2] traffic classifier EXP3
[P-classifier-EXP3] if-match mpls-exp 3
[P-classifier-EXP3] traffic classifier EXP4
[P-classifier-EXP4] if-match mpls-exp 4
[P-classifier-EXP4] quit
```

定义流行为，为不同的流设置不同的带宽和时延保证。

```
[P] traffic behavior AF11
[P-behavior-AF11] queue af bandwidth pct 10
[P-behavior-AF11] traffic behavior AF21
[P-behavior-AF21] queue af bandwidth pct 20
[P-behavior-AF21] traffic behavior AF31
[P-behavior-AF31] queue af bandwidth pct 30
[P-behavior-AF31] traffic behavior EF
[P-behavior-EF] queue ef bandwidth pct 40
[P-behavior-EF] quit
```

定义 QoS 策略，使符合流分类策略的 4 条数据流满足：EXP1 流保证 10%带宽；EXP2 流保证 20%带宽；EXP3 流保证 30%带宽；EXP4 流保证 40%带宽并保证时延。

```
[P] qos policy QUEUE
[P-qospolicy-QUEUE] classifier EXP1 behavior AF11
```

```
[P-qospolicy-QUEUE] classifier EXP2 behavior AF21
[P-qospolicy-QUEUE] classifier EXP3 behavior AF31
[P-qospolicy-QUEUE] classifier EXP4 behavior EF
[P-qospolicy-QUEUE] quit
```

将 QoS 策略应用在 P 设备 Serial2/2 接口的出方向上。

```
[P] interface serial 2/2
[P-Serial2/2] qos apply policy QUEUE outbond
```

配置完成后，VPN 1 中的流量在发生拥塞时，DSCP 域为 af11、af21、af31、ef 的流量接收比例为 1:2:3:4，其中，ef 流比其他 3 条流的延时都小。

8 DAR 配置

说明

- DAR 功能只对通过 IP 转发的报文生效。
- DAR 应用目前仅 SR6602 路由器的以太网接口支持。

8.1 DAR 简介

今天，Internet 已经成为企业实施各种业务的主要媒介，各种基于业务的应用也不断出现。因此，目前只对报文中的 IP 报头进行检查的简单机制已经无法适用于当今复杂的网络，故提出了基于业务的深度应用识别的概念。

DAR（Deeper Application Recognition，深度应用识别）是一个智能的识别分类工具，它可以对报文中第 4 层到第 7 层的内容和一些动态协议（如 BT）进行检查和识别，以区分出各种基于应用的协议，并能进行相应的控制（如限流、阻断），从而弥补了原先只能对报文进行简单分类的不足。

通过对报文进行深度的识别和分类，大大加强了用户对网络应用的控制力度，可以更有效的为关键业务数据实施高优先级的策略，进一步保护用户的投资。

随着 P2P（Peer to Peer，对等网络）技术的不断发展和广泛应用，P2P 数据流量在网络流量中所占的比重越来越大，这不仅对网络带宽资源造成了浪费，还对其他业务的正常运行造成了严重影响。因此，需要对 P2P 流量进行单独的识别。

DAR 对 P2P 协议的识别，是将 P2P 报文的数据特征以特征文件（*.mtd 文件）的方式提供给用户，用户在设备上加载这些特征文件后，设备将接收到的报文内容与这些数据特征进行比较，如果相同，则认为该报文为 P2P 报文。

8.2 配置 DAR

8.2.1 配置准备

一台支持 DAR 的设备。

8.2.2 配置 P2P 协议流量识别功能

对 P2P 流量的控制可以分为 P2P 识别和流量控制两步：

- P2P识别：对P2P流量进行识别，起配置过程如 表 8-1所示。
- 流量控制：利用QoS提供的各种流行为来对P2P流量进行控制。具体配置过程请参考“2 两种QoS配置方式介绍”。

P2P识别的配置过程如 表 8-1所示。

表8-1 P2P 协议流量识别功能配置任务简介

配置任务	说明	详细配置
加载 P2P 特征文件	必选 只有特征文件中包含的协议才能被识别出来	1.
配置 P2P 协议组	必选 只有特征文件中包含的协议才能加入 P2P 协议组	2.
使能 P2P 协议流量识别功能	必选 使能该功能后，接口才对 P2P 协议流量进行识别	3.

1. 加载 P2P 特征文件

用户需要对 P2P 协议进行识别时，首先需要将特征文件加载到系统中。

表8-2 加载 P2P 特征文件

操作	命令	说明
进入系统视图	**system-view**	-
加载 P2P 特征文件	**dar p2p signature-file** *filename*	必选 缺省情况下，未加载 P2P 特征文件



说明

系统只能加载根目录下的特征文件，请将特征文件放在系统根目录下。

2. 配置 P2P 协议组

配置 P2P 协议组后，用户可以通过引用一个 P2P 协议组，来同时对多个 P2P 协议的流量进行识别。

表8-3 配置 P2P 协议组

操作	命令	说明
进入系统视图	**system-view**	-
创建 P2P 协议组并进入协议组视图	**dar protocol-group** *group-id*	必选 缺省情况下，系统中不存在协议组
向协议组中添加协议	**protocol** *protocol-name*	必选 缺省情况下，协议组中未配置协议

3. 使能 P2P 协议流量识别功能

对 P2P 协议报文的识别，将会消耗大量的系统资源，为了不影响其他模块的正常工作，系统默认不对 P2P 协议报文进行识别。如果用户想对某个接口收发的 P2P 协议报文进行识别，可以使能该接口的 P2P 协议识别功能。

表8-4 使能 P2P 协议流量识别功能

操作	命令	说明
进入系统视图	**system-view**	-
进入三层以太网接口视图	**interface** *interface-type interface-number*	-
使能 P2P 协议流量识别功能	**dar enable**	必选 缺省情况下，接口的 P2P 协议流量识别功能处于关闭状态

8.2.3 配置协议的匹配规则

当对相应数据流应用各种策略时（如设置报文的优先级，为数据流分配带宽等），首要需要使用 DAR 的深度识别功能对数据流进行分类。

表8-5 配置协议的匹配规则

操作	命令	说明
进入系统视图	**system-view**	-
进入类视图	**traffic classifier** *tcl-name* **[operator { and | or }]**	必选
配置协议组匹配规则	**lf-match [not] protocol-group** *protocol-group-id*	可选 缺省情况下，未配置协议组匹配规则

8.2.4 配置 DAR 的报文统计功能

通过使能 DAR 的报文统计功能，用户可以及时对各个接口上的应用协议的报文个数、数据流量进行监控，便于对数据流实施相应的策略。

表8-6 配置 DAR 的报文统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入以太网接口视图	**interface** *interface-type interface-number*	-
使能 DAR 的报文统计功能	**dar protocol-statistic**	必选 缺省情况下，DAR 的报文统计功能关闭

8.3 DAR 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 DAR 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 DAR 的统计信息。

表8-7 DAR 显示和维护

操作	命令
显示 DAR 的报文统计信息	**display dar protocol-statistic p2p** [**interface** *interface-type interface-number*] [**direction** { **in** | **out** }]
清除 DAR 的协议统计信息	**reset dar protocol-statistic p2p** [**interface** *interface-type interface-number*] **reset dar protocol-statistic interface** *interface-type interface-number* **p2p**

8.4 DAR 典型配置举例

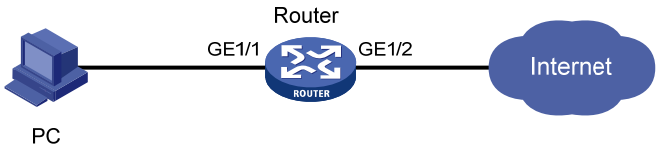
8.4.1 禁止 P2P 软件下载配置举例

1. 配置需求

- PC 通过设备 Router 分别与外网相连。
- 通过在 Router 上进行配置，禁止 PC 通过 BT 客户端和 eMule/eDonkey 客户端下载文件。

2. 组网图

图8-1 配置禁止 P2P 软件下载组网图



3. 配置步骤

加载特征文件 meta.mtd。

```
<Router> system-view
[Router] dar p2p signature-file meta.mtd
```

配置协议组。

```
[Router] dar protocol-group 1
[Router-protocol-group-1] protocol bittorrent
[Router-protocol-group-1] protocol eMule/eDonkey
[Router-protocol-group-1] quit
```

配置对协议组的匹配规则。

```
[Router] traffic classifier p2p
[Router-classifier-p2p] if-match protocol-group 1
[Router-classifier-p2p] quit
```

配置过滤丢弃的流行为。

```
[Router] traffic behavior deny
[Router-behavior-deny] filter deny
[Router-behavior-deny] quit
```

配置对协议组 1 的协议进行匹配过滤的 QoS 策略。

```
[Router] qos policy p2p
```

```
[Router-qospolicy-p2p] classifier bt behavior deny
[Router-qospolicy-p2p] quit
```

在 GigabitEthernet1/1 接口使能 P2P 协议流量识别功能，并在入方向应用 QoS 策略。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] dar enalbe
[Router-GigabitEthernet1/1] qos apply policy p2p inbound
```

在 PC 上运行 BT 客户端和 eMule/eDonkey 客户端软件，开始下载文件。

查看客户端软件，可以看到无法进行下载。

目 录

1 帧中继QoS配置 1-1
1.1 帧中继QoS简介 1-1
1.1.1 帧中继QoS 1-1
1.1.2 重要参数 1-1
1.1.3 帧中继QoS的实现 1-1
1.2 帧中继QoS的配置 1-6
1.2.1 帧中继QoS配置任务简介 1-6
1.2.2 创建和配置帧中继类 1-7
1.2.3 配置帧中继流量整形 1-8
1.2.4 配置帧中继流量监管 1-8
1.2.5 配置帧中继拥塞管理 1-9
1.2.6 配置帧中继DE规则列表 1-10
1.2.7 配置帧中继队列管理 1-10
1.2.8 配置帧中继分片 1-11
1.3 帧中继QoS显示和维护 1-11
1.4 帧中继QoS典型配置举例 1-12
1.4.1 帧中继流量整形配置举例 1-12
1.4.2 帧中继分片配置举例 1-13

1 帧中继 QoS 配置

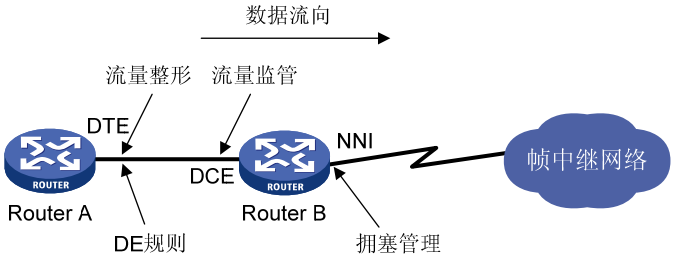
1.1 帧中继 QoS 简介

1.1.1 帧中继 QoS

在帧中继接口上，可以使用通用的 QoS 服务为用户提供接口上的流量监管、流量整形、拥塞管理、拥塞避免等服务。除此之外，帧中继网络还拥有自己的 QoS 服务机制，包括帧中继流量整形、帧中继流量监管、帧中继拥塞管理、帧中继 DE（Discard Eligibility）规则列表、帧中继队列管理等。

与通用的 QoS 相比，帧中继 QoS 能够在接口的每条虚电路上提供 QoS 服务，而通用 QoS 只能在整个接口上提供 QoS 服务，因此帧中继 QoS 能够为用户提供更灵活的服务。

图1-1 帧中继 QoS 应用示意图



说明

帧中继相关知识的介绍请参见“接入分册”中的“帧中继配置”。

1.1.2 重要参数

下面列出了对帧中继（FR，Frame Relay）进行控制的一些重要参数：

- 允许的承诺信息速率（CIR ALLOW，Committed Information Rate ALLOW）：FR 正常情况下可以提供的传送速率。网络不拥塞时，可以保证用户以此速率传送数据。
- 承诺信息速率（CIR，Committed Information Rate）：虚电路可以提供的最小传送速率。即使网络拥塞，也可以保证用户以此速率传送数据。
- 承诺突发尺寸（CBS，Committed Burst Size）：帧中继网络承诺的在 Tc 时间内传送的包流量。在网络拥塞时，帧中继网络保证这部分可以成功地传送。
- 超出突发尺寸（EBS，Excess Burst Size）：在 Tc 时间内，帧中继网络可以超过承诺突发尺寸的最大值。在网络拥塞时，这部分流量将被首先丢弃。因此，帧中继网络不保证这部分流量能被成功地传送。

1.1.3 帧中继 QoS 的实现

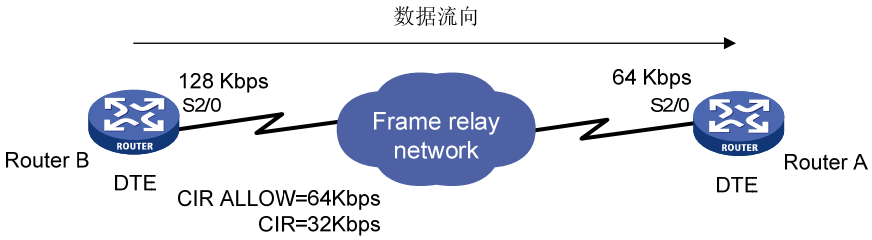
1. 帧中继流量整形

- 帧中继流量整形的作用

帧中继流量整形（FRTS，Frame Relay Traffic Shaping）能够限制从某一虚电路发出的报文流量和突发报文流量，使这类报文能够以比较均匀的速度向外发送。

在帧中继网络中，如果不同网段的带宽不匹配，往往会在网段的交界处产生瓶颈。如 图 1-2所示，如果Router B以 128Kbps的速率向Router A发送报文，而Router A的接口速率最大只能达到64Kbps，这时就会在帧中继网络与Router A的连接处产生瓶颈，从而发生拥塞，导致数据无法正常传送。但是，如果在Router B的出接口Serial2/0 应用帧中继流量整形，可以使它以相对比较均匀的速率 64Kbps发送报文，从而避免网络拥塞的发生。即使网络发生拥塞，Router B仍能够以 32Kbps的速率发送报文。

图1-2 帧中继流量整形示意图

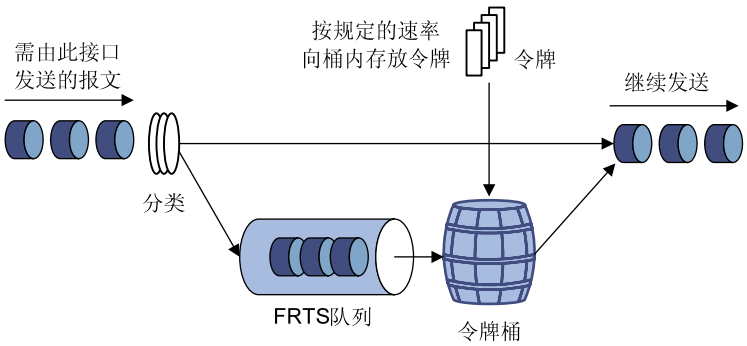


帧中继流量整形应用在设备的输出接口，它可以向用户提供 CIR ALLOW、CIR、CBS、EBS 等参数。帧中继虚电路能够以 CIR ALLOW 的速率发送报文。帧中继流量整形还允许虚电路在突发情况下以超过 CIR ALLOW 的速率发送报文。

- 帧中继流量整形的工作原理

帧中继流量整形是用令牌桶算法实现的，根据实际的算法原理对协议中相关参数的含义进行了相应的修改。令牌桶原理如 图 1-3所示。

图1-3 令牌桶原理



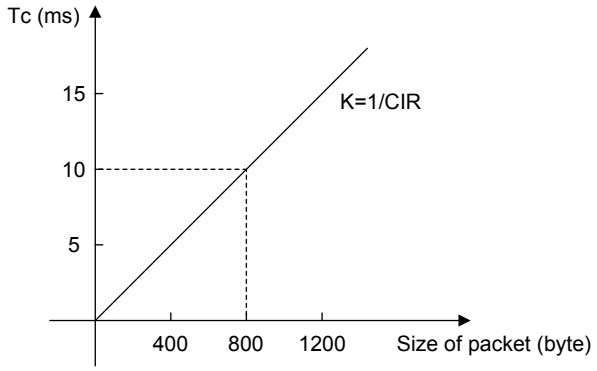
令牌桶原理是：需要进行流量控制的报文在发送前进入令牌桶中进行处理。如果令牌桶中有足够的令牌可以用来发送报文，则允许报文通过，即报文被正常发送。如果令牌桶中的令牌数不满足报文发送条件，则报文会放入帧中继类队列（在帧中继流量整形中，帧中继类队列即为 FRTS 队列），等令牌桶中有足够的令牌再从帧中继类队列中取报文发送。这样，就可以对某类报文的流量进行控制。令牌的单位为 bit。

协议规定的相关参数和实际帧中继流量整形原理参数的对应关系如下：

- CBS 与 EBS 的和为令牌桶的大小；
- CIR ALLOW 为每秒向令牌桶中添加的令牌数。

帧中继流量整形方案中为了提高效率，提出了动态 Tc 的概念。Tc 参数的大小在 10 毫秒到 100 毫秒之间，可根据每次发送的报文大小进行动态调节，Tc=size of packet/CIR。即无论待发送报文大小为多少（小于 1500 字节），设备都会在当前计算出的 Tc 时间内，将所需要的全部令牌分发给当前报文。

图1-4 Tc 与 CIR 关系示意图



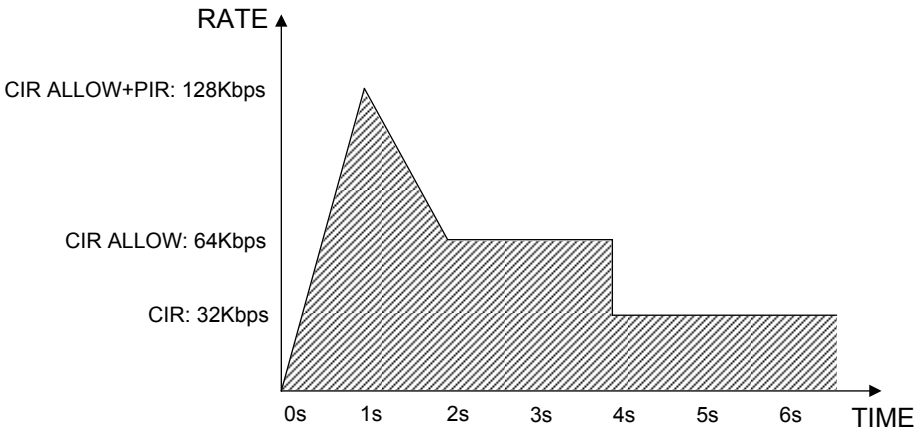
例如，发送一个 800 字节的报文，需要 800×8=6400bit 的令牌。设 CIR 为 64000bps，则向令牌桶中添加发送该报文令牌需要的时间为 6400/64000=0.1s，即对于该报文来说 Tc 为 100ms。100ms 之内令牌桶中添加了 6400bit 的令牌，该报文即被送出。特殊情况下，此时如果 CIR 为 8000bps，则该报文的令牌时间为 6400/8000 = 800ms>100ms，但是由于 Tc 参数大小被定义在 10ms 到 100ms 之间滑动，此时 Tc 取其上限值为 100ms，而不是计算得到的 800ms。同样的，如果此时 CIR 为 1024000bps，则该报文的令牌时间为 6400/1024000=6.25ms<10ms，但是由于 Tc 参数大小被定义在 10ms 到 100ms 之间滑动，所以该报文的 Tc 取其下限值 10ms。

由于 CBS 与 EBS 的和是令牌桶的大小，同时设备是一次性分配令牌给待发送报文，为了使令牌桶中的令牌能够满足任何大小报文的发送，特别是大报文的发送，比如 1500byte 的报文，该报文发送需要 1500×8=12Kbit 的令牌。故 CBS 不得小于 15Kbit，同时建议用户配置 CBS 的大小等同 CIR 的大小。

按照标准协议中规定，若 Tc=20ms，CIR 为 64000bps，则每个 Tc 时间内只能添加 0.02×64000=1280bit 的令牌。如发送长度为 800 字节的报文需要 5 个 Tc，即添加 5 次令牌才能将该报文发送出去。与标准协议的实现方式相比，此方案在一个动态 Tc 中只需添加一次令牌就可将该报文发送出，大大提高了工作效率。

当网络发生拥塞时，如果帧中继交换设备的出接口配置了拥塞管理功能（关于拥塞管理的介绍请参见“4. 帧中继拥塞管理”），它会通知设备网络发生了拥塞。设备接收到了网络拥塞的通知后，会使发送速率逐渐下降到CIR，以缓解网络的拥塞状况，此时用户仍然能够以CIR的速率发送数据。此后如果设备在一定时间内没有再收到网络拥塞的通知，发送速率又会逐渐由CIR上升回CIR ALLOW。

图1-5 帧中继流量整形原理示意图

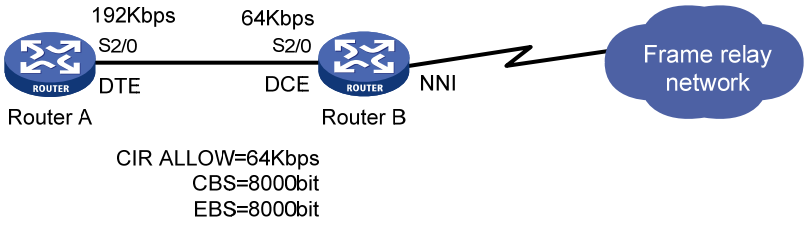


如图1-5所示，设置帧中继流量整形参数分别为：CIR ALLOW=64Kbps，CIR=32Kbps，CBS=64000bit，EBS=64000bit，令牌桶的大小为128000，在4s之前令牌桶添加令牌的速率为64Kbps，虚电路以64Kbps的速率发送报文。当在4s时，设备收到了带有BECN标志位的帧中继报文，表明网络发生拥塞，令牌桶添加令牌的速率下降为CIR 32Kbps，虚电路以32Kbps的速率发送报文。

2. 帧中继流量监管

帧中继流量监管可以对每个虚电路上进入网络的流量进行监督，限制流量在一个允许的范围之内。若某条虚电路的报文流量超过了用户设置的范围，设备会采取丢弃报文等措施，以保护网络资源。

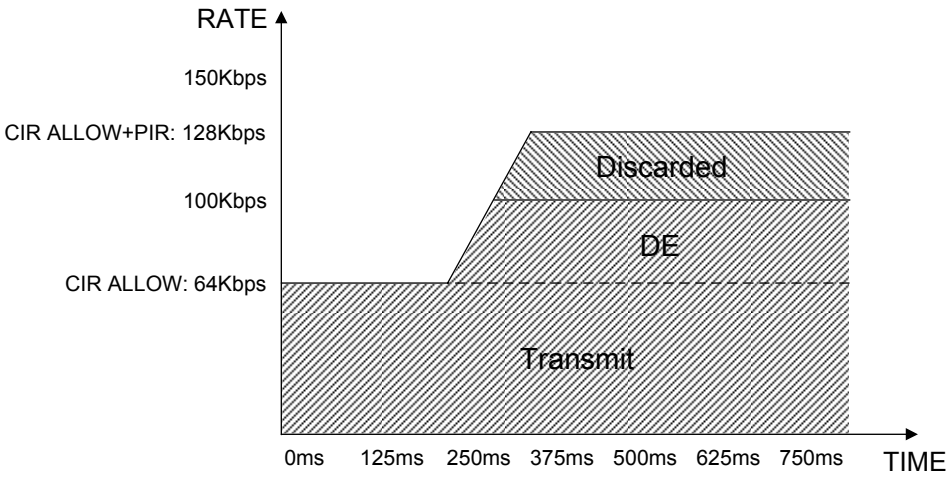
图1-6 帧中继流量监管示意图



如图1-6所示，用户端设备Router A以192Kbps的速率向交换端设备Router B发送报文，但设备Router B希望只为设备Router A提供64Kbps的带宽。这时就需要在设备Router B的DCE端配置帧中继流量监管。

帧中继流量监管只能应用于设备的 DCE 接口上，它可以对从 DTE 端发送来的报文流量进行监视。当报文流量小于 CBS 时，报文可以正常发送，设备对报文不做处理；当报文流量大于 CBS 并小于 EBS 与 CBS 之和时，报文可以正常发送，但设备将把流量大于 CBS 的报文的帧中继报文头中的 DE 标志位置为 1；当报文流量大于 CBS＋EBS 时，设备将以 CBS＋EBS 的流量进行发送，对超过 CBS＋EBS 的流量进行丢弃，对于超过 CBS 的流量，设备将把帧中继报文头中的 DE 标志位置为 1。

图1-7 帧中继流量监管原理示意图



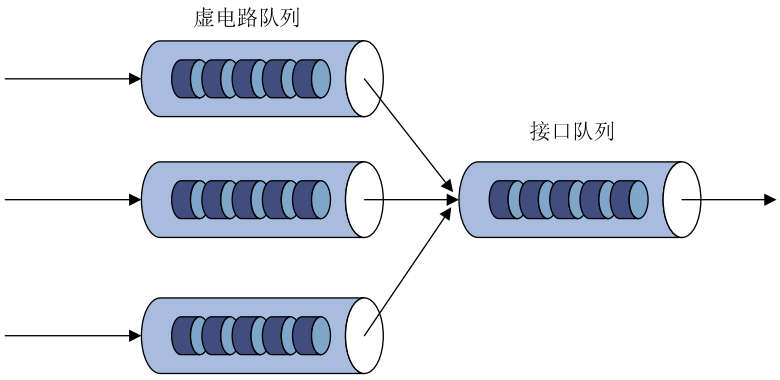
如 图 1-7所示，设置帧中继流量监管参数分别为：CIR ALLOW=64kbps，CBS=8000bit，EBS=8000bit，时间间隔Tc=CBS/CIR ALLOW=125ms。在 0～250ms时，DTE以速率 64Kbps向DCE发送报文，DCE对这部分报文正常转发，转发的速率为 64Kbps。在 250～500ms时，DTE以速率 100Kbps向DCE发送报文，DCE的转发速率为 100Kbps，但是超过CBS流量报文的DE标志位被置为 1。在 500ms以后，DTE以速率 150Kbps向DCE发送报文，DCE的转发速率为 128Kbps，超过CBS但小于CBS＋EBS流量报文的DE标志位被置为 1，超过CBS＋EBS流量的报文被直接丢弃。

3. 帧中继队列管理

除了帧中继虚电路队列之外，帧中继接口也拥有接口队列。帧中继流量整形没有使能的情况下，只有帧中继接口队列起作用，即预先配置的帧中继 PVC 队列只有在帧中继流量整形使能的情况下才能生效。

虚电路队列和接口队列的关系如 图 1-8所示。

图1-8 帧中继队列示意图



在帧中继接口上能够使用的队列有：

- FIFO：First In First Out Queuing，先进先出队列
- PQ：Priority Queuing，优先队列
- CQ：Custom Queuing，定制队列
- WFQ：Weighted Fair Queuing，加权公平队列

- CBQ：Class Based Queuing，基于类的队列
- RTPQ：Real-time Transport Protocol Priority Queuing，实时传输协议优先队列
- PVC PQ：PVC Priority Queuing，虚电路优先级队列

其中，FIFO、PQ、CQ、WFQ、CBQ 和 RTPQ 队列属于通用的队列，关于它们的详细介绍请参见“QoS 分册”中的“QoS 配置”。

PVC PQ 队列被应用于帧中继接口之上，它拥有四种队列类型：top、middle、normal、bottom，队列优先级依次降低。同一虚电路发送的报文只能入 PVC PQ 中的一种队列，不同虚电路上的报文根据虚电路所处的优先级入接口下的 PVC PQ 队列。PVC PQ 的队列发送策略是：按照队列优先级，在发送完高优先级队列中的报文之后，再发送低优先级队列中的报文。

帧中继虚电路的队列包括：FIFO、PQ、CQ、WFQ、CBQ 和 RTPQ。RTPQ 可以与其它队列共存，而其它队列相互之间不能共存。目前 SR6600 路由器仅支持 FIFO 队列。

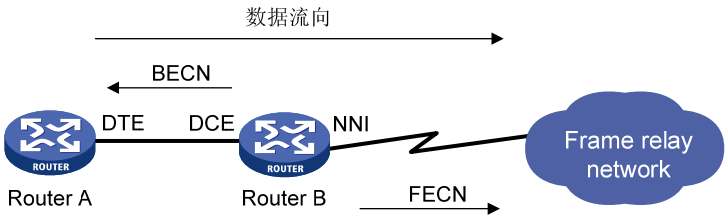
当接口实行帧中继流量整形时，接口队列类型只能是：FIFO、RTPQ 或 PVC PQ。

4. 帧中继拥塞管理

帧中继拥塞管理可以在网络发生拥塞时对帧中继报文进行处理：丢弃 DE 标志位为 1 的报文；同时发送拥塞通知至网络上的其他设备，告知其发生了拥塞。

帧中继拥塞管理应用于帧中继交换设备的出接口上。当拥塞没有发生时，对于转发的帧中继报文，设备将正常转发，不作特殊处理。当拥塞发生时，DE 标志位为 1 的报文将会被丢弃；对于转发的正向报文，设备把帧中继报文头的 FECN 标志位置为 1；对于同一虚电路上反方向发送的报文，设备把帧中继报文头的 BECN 标志位置为 1。若在一段时间内没有反方向的报文发送，设备将自动发送 BECN 标志位为 1 的 Q.922A Test Reponse 报文给主叫 DTE。

图1-9 帧中继拥塞管理示意图



5. 帧中继 DE 规则列表

在帧中继网络中，DE 标志位为 1 的报文会在拥塞发生时被优先丢弃。DE 规则列表应用于设备的帧中继虚电路上，每个规则列表内包含多条 DE 规则。虚电路中发送的报文如果符合 DE 规则列表中的规则，则它的 DE 标志位将会被置 1，这类报文在网络发生拥塞时将会被优先丢弃。

1.2 帧中继 QoS 的配置

1.2.1 帧中继 QoS 配置任务简介

表1-1 帧中继 QoS 配置任务简介

配置任务	说明	详细配置
创建和配置帧中继类	必选	1.2.2
配置帧中继流量整形	可选	1.2.3

配置任务	说明	详细配置
配置帧中继流量监管	可选	1.2.4
配置帧中继拥塞管理	可选	1.2.5
配置帧中继 DE 规则列表	可选	1.2.6
配置帧中继队列管理	可选	1.2.7
配置帧中继分片	可选	1.2.8

1.2.2 创建和配置帧中继类

设备把帧中继虚电路上的 QoS 服务整合在帧中继类上，提供了一个灵活、完整的帧中继流量控制和质量服务解决方案。在配置帧中继流量整形等 QoS 服务时，需要先创建一个帧中继类，然后在这个帧中继类上配置各种 QoS 参数，这样一个帧中继类就相当于一套 QoS 网络服务方案。然后将帧中继类关联到一个帧中继虚电路上，就相当于将一套 QoS 服务方案应用到帧中继虚电路上。不同的虚电路可以使用不同的帧中继类，也可以使用同一个帧中继类。

当帧中继虚电路提供 QoS 服务时，它将按照下面的顺序寻找对应的帧中继类：

- 使用和此帧中继虚电路相关联的帧中继类。
- 使用帧中继虚电路所在帧中继接口的帧中继类。

表1-2 创建和配置帧中继类

操作			命令	说明
进入系统视图			**system-view**	-
创建帧中继类并进入帧中继类视图			**fr class** *class-name*	必选 缺省情况下，系统没有创建帧中继类
退出帧中继类视图			**quit**	-
将帧中继类同帧中继接口或虚电路相关联	将帧中继类同帧中继接口相关联	进入帧中继接口视图	**interface** *interface-type interface-number*	二者必选其一或全选 缺省情况下，没有帧中继类同帧中继接口或虚电路相关联
		将帧中继类同帧中继接口相关联	**fr-class** *class-name*	
	将帧中继类同帧中继虚电路相关联	进入帧中继接口视图	**interface** *interface-type interface-number*	
		进入帧中继虚电路视图	**fr dlci** *dlci*	
		将帧中继类同帧中继虚电路相关联	**fr-class** *class-name*	

说明

使用 **fr class** 命令创建帧中继类之后，将进入帧中继类视图，用户可以在此视图下为帧中继流量整形、帧中继流量监管、帧中继拥塞管理、帧中继队列管理等 QoS 服务配置不同的参数。详细的参数设置请见下面各节里的内容。

1.2.3 配置帧中继流量整形

表1-3 配置帧中继流量整形

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继接口视图	**interface** *interface-type interface-number*	-
使能帧中继流量整形功能	**fr traffic-shaping**	必选 缺省情况下，禁止帧中继流量整形功能
退出帧中继接口视图	**quit**	-
进入帧中继类视图	**fr class** *class-name*	-
设置帧中继虚电路的承诺突发尺寸	**cbs** [**outbound**] *committed-burst-size*	可选 缺省情况下，承诺突发尺寸为56000bps
设置帧中继虚电路的超出突发尺寸	**ebs** [**outbound**] *excess-burst-size*	可选 缺省情况下，超出突发尺寸为 0bit
设置帧中继虚电路的允许承诺信息速率	**cir allow** [**outbound**] *committed-information-rate*	可选 缺省情况下，允许承诺信息速率为56000bps
设置帧中继虚电路的承诺信息速率	**cir** *committed-information-rate*	可选 缺省情况下，承诺信息速率为56000bps
使能帧中继流量整形的自适应流量调节功能	**traffic-shaping adaptation** { **becn** *percentage* | **interface-congestion** *number* }	可选 缺省情况下，使能对带 BECN 标志的报文进行自适应流量调节功能，每次调节的比例为 25

说明

- 帧中继流量整形功能应用于设备帧中继报文的发送接口上。通常情况下，它被应用于帧中继网络的 DTE 端。
- 命令 **cbs**、**ebs**、**cir allow** 可以设置虚电路上输入和输出（**inbound** 和 **outbound**）两个方向的参数，但对帧中继流量整形只有 **outbound** 方向的参数有效。
- 在配置 CIR ALLOW 和 CBS 时，CBS 之值不应小于 CIR ALLOW 之值，否则可能会出现大报文无法发送的问题。

1.2.4 配置帧中继流量监管

表1-4 配置帧中继流量监管

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能帧中继流量监管功能	**fr traffic-policing**	必选 缺省情况下，禁止接口的帧中继流量监管功能
退出帧中继接口视图	**quit**	-
进入帧中继类视图	**fr class** *class-name*	-
设置帧中继虚电路的承诺突发尺寸	**cbs** [**inbound**] *committed-burst-size*	可选 缺省情况下，承诺突发尺寸为56000bps
设置帧中继虚电路的超出突发尺寸	**ebs** [**inbound**] *excess-burst-size*	可选 缺省情况下，超出突发尺寸为 0bit
设置帧中继虚电路的允许承诺信息速率	**cir allow** [**inbound**] *committed-information-rate*	可选 缺省情况下，允许承诺信息速率为56000bps

说明

- 帧中继流量监管功能应用于设备帧中继报文的接收接口上。它只能被应用于帧中继网络的 DCE 端。
- 命令 **cbs**、**ebs**、**cir allow** 可以设置虚电路上输入和输出（**inbound** 和 **outbound**）两个方向的参数，但对帧中继流量监管只有 **inbound** 方向的参数有效。

1.2.5 配置帧中继拥塞管理

帧中继拥塞管理包括帧中继接口上的拥塞管理和帧中继虚电路上的拥塞管理两部分。对拥塞门限的设置，也可以采用两种方式：既可以在指定的帧中继类里配置某条虚电路的拥塞门限，也可以在接口视图下配置整个接口的拥塞门限。设备根据帧中继接口或虚电路当前的队列长度占接口队列总长度的比例来判断接口是否发生拥塞。如果接口或虚电路当前队列长度与队列总长度之比大于用户设置的门限值，则设备认为发生了拥塞，从而对报文进行丢弃等相应处理。

1. 配置帧中继接口的拥塞管理策略

表1-5 配置帧中继接口的拥塞管理策略

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继接口视图	**interface** *interface-type interface-number*	-
使能帧中继接口的拥塞管理功能	**fr congestion-threshold** { **de** | **ecn** } *queue-percentage*	必选 缺省情况下，禁止帧中继接口的拥塞管理功能

2. 配置帧中继虚电路的拥塞管理策略

表1-6 配置帧中继虚电路的拥塞管理策略

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继类视图	**fr class** *class-name*	-
使能帧中继虚电路的拥塞管理功能	**congestion-threshold** { **de** | **ecn** } *queue-percentage*	必选 缺省情况下，禁止帧中继虚电路的拥塞管理功能



说明

- 当帧中继接口使能拥塞管理时，接口的队列类型只能为 FIFO 或 PVC PQ。
- 当帧中继虚电路使能拥塞管理时，虚电路的队列类型只能为 FIFO。
- 只有当虚电路所在接口使能了帧中继流量整形功能之后，虚电路的拥塞管理功能才能够起作用。

1.2.6 配置帧中继 DE 规则列表

表1-7 配置帧中继 DE 规则列表

操作		命令	说明
进入系统视图		**system-view**	-
配置 DE 规则列表	配置基于接口的 DE 规则列表	**fr del** *list-number* **inbound-interface** *interface-type interface-number*	二者必选其一 缺省情况下，没有创建 DE 规则列表
	配置基于 IP 协议的 DE 规则列表	**fr del** *list-number* **protocol ip** [**acl** *acl-number* | **fragments** | **greater-than** *bytes* | **less-than** *bytes* | **tcp** *ports* | **udp** *ports*]	
进入帧中继接口视图		**interface** *interface-type interface-number*	-
将 DE 规则列表应用到帧中继虚电路上		**fr de del** *list-number* **dlci** *dlci-number*	必选 缺省情况下，帧中继虚电路上没有应用 DE 规则列表



说明

一台设备最多可以支持 10 个 DE 规则列表，在每个 DE 规则列表内最多可以拥有 100 条 DE 规则。

1.2.7 配置帧中继队列管理

1. 配置帧中继虚电路队列

当帧中继接口使能了帧中继流量整形之后，在此接口下的每条虚电路都将拥有一个独立的虚电路队列。

表1-8 配置帧中继虚电路队列

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继类视图	**fr class** *class-name*	-
设置帧中继虚电路的 FIFO 队列长度	**fifo queue-length** *queue-length*	可选 缺省情况下，FIFO 队列长度为 40



说明

- 缺省情况下，帧中继虚电路的队列类型为 FIFO。
- 当帧中继虚电路使能拥塞管理时，接口的队列类型只能为 FIFO。

1.2.8 配置帧中继分片

设备支持帧中继论坛 FRF.12 标准的 end-to-end 分片功能。

在低速帧中继线路上，大数据报文将会造成传输时延的增大。帧中继分片特性可以将大的帧中继报文分割成几个小报文，从而可以保证在低速线路上数据也可以被低延迟的发送。

当语音与数据同传时，大数据报文的发送将长时间占用带宽，会造成语音报文被延时甚至丢弃，影响语音质量。配置帧中继分片的目的是尽量减少语音的延时，保证语音的实时性。配置分片后大的数据报文将被拆分为较小的数据分片，语音报文与拆分后的分片交替发送，保证语音报文及时均匀地得到处理，降低时延。

表1-9 配置帧中继分片

操作	命令	说明
进入系统视图	**system-view**	-
进入帧中继类视图	**fr class** *class-name*	-
使能帧中继分片功能	**fragment** [*fragment-size*]	必选 缺省情况下，禁止帧中继分片功能



说明

- 配置完成后，将需要应用分片功能的帧中继虚电路同此帧中继类关联，并使能帧中继流量整形，这时帧中继分片才会生效。
- MFR接口上不支持FRF.12的分片。如果两边都是MFR接口，且配置了 FRF.12 分片，此时 FRF.12 分片不起作用，本端报文会完整发出，对端可以接收，在本端执行 ping 命令可以得到响应；如果本端是 MFR 接口，对端是普通帧中继接口（即封装了 FR 的 Serial 接口），则本端 FRF.12 分片不起作用，本端报文会完整发出，而对端的 FRF.12 分片起作用。

1.3 帧中继 QoS 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令显示配置后帧中继 QoS 的运行情况，通过查看显示信息验证配置的效果。

表1-10 帧中继 QoS 显示和维护

操作	命令
显示帧中继类与接口（包括属于该接口的 dlci、该接口下的子接口以及子接口下的 dlci）的映射关系	**display fr class-map** { **fr-class** *class-name* | **interface** *interface-type interface-number* }
查看 FR QoS 的配置和统计信息	**display fr pvc-info** [**interface** *interface-type interface-number*] [*dlci-number*]
查看配置的帧中继交换 PVC 的信息	**display fr switch-table** { **all** | **name** *switch-name* | **interface** *interface-type interface-number* }
查看接口应用 CBQ 的信息	**display qos policy interface** [*interface-type interface-number* [**dlci** *dlci-number* [**outbound**] | **inbound** | **outbound**]]
查看帧中继分片信息	**display fr fragment-info** [**interface** *interface-type interface-number*] [*dlci-number*]
查看帧中继收发数据统计信息	**display fr statistics** [**interface** *interface-type interface-number*]

1.4 帧中继 QoS 典型配置举例

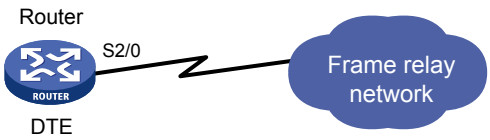
1.4.1 帧中继流量整形配置举例

1. 组网需求

- 设备 Router 通过接口 Serial2/0 连接至帧中继网络。
- 要求设备的平均发送速率为 96Kbps，最大发送速率为 128Kbps，最小发送速率为 32Kbps，设备具备流量自适应调整功能。

2. 组网图

图1-10 帧中继流量整形组网图



3. 配置步骤

创建帧中继类并配置帧中继流量整形参数。

```
[Router] fr class 96k
[Router-fr-class-96k] cir allow 96000
[Router-fr-class-96k] cir 32000
[Router-fr-class-96k] cbs 96000
[Router-fr-class-96k] ebs 32000
[Router-fr-class-96k] traffic-shaping adaptation becn 20
[Router-fr-class-96k] quit
```

配置接口 Serial2/0，使能帧中继流量整形。

```
[Router] interface serial 2/0
[Router-Serial2/0] link-protocol fr
[Router-Serial2/0] ip address 1.1.1.1 255.255.255.0
```

```
[Router-Serial2/0] fr traffic-shaping
```

创建帧中继虚电路，将帧中继类同虚电路相关联。

```
[Router-Serial2/0] fr dlci 16
[Router-fr-dlci-Serial2/0-16] fr-class 96k
```

1.4.2 帧中继分片配置举例

1. 组网需求

Router A 和 Router B 通过帧中继网络相连，使能帧中继分片功能（FRF.12）。

2. 组网图

图1-11 FRF.12 典型组网图



3. 配置步骤

(1) 配置 Router A

创建并配置帧中继类 test1。

```
<RouterA> system-view
[RouterA] fr class test1
[RouterA-fr-class-test1] fragment 80
[RouterA-fr-class-test1] quit
```

配置接口 Serial2/0，使能帧中继流量整形。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] link-protocol fr
[RouterA-Serial2/0] ip address 10.1.1.2 255.0.0.0
[RouterA-Serial2/0] fr traffic-shaping
```

创建 DLCI 16。

```
[RouterA-Serial2/0] fr dlci 16
```

将名为 test1 的帧中继类应用到 DLCI 16 上。

```
[RouterA-fr-dlci-Serial2/0-16] fr-class test1
```

(2) 配置 Router B

配置帧中继类 test1。

```
<RouterB> system-view
[RouterB] fr class test1
[RouterB-fr-class-test1] fragment 80
[RouterB-fr-class-test1] quit
```

配置接口 Serial2/0，使能帧中继流量整形。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] link-protocol fr
[RouterB-Serial2/0] ip address 10.1.1.1 255.0.0.0
```

```
[RouterB-Serial2/0] fr traffic-shaping
```

创建 DLCI 16。

```
[RouterB-Serial2/0] fr dlci 16
```

将名为 test1 的帧中继类应用到 DLCI 16 上。

```
[RouterB-fr-dlci-Serial2/0-16] fr-class test1
```

目 录

1 AAA配置 1-1
1.1 AAA简介 1-1
1.2 RADIUS协议简介 1-2
1.2.1 客户端/服务器模式 1-2
1.2.2 安全和认证机制 1-2
1.2.3 RADIUS的基本消息交互流程 1-3
1.2.4 RADIUS报文结构 1-3
1.2.5 RADIUS扩展属性 1-6
1.3 HWTACACS协议简介 1-7
1.3.1 HWTACACS协议与RADIUS协议的区别 1-7
1.3.2 HWTACACS的基本消息交互流程 1-7
1.4 协议规范 1-9
1.5 AAA RADIUS HWTACACS配置任务简介 1-9
1.6 配置AAA 1-11
1.6.1 配置准备 1-11
1.6.2 创建ISP域 1-11
1.6.3 配置ISP域的属性 1-12
1.6.4 配置ISP域的AAA认证方案 1-12
1.6.5 配置ISP域的AAA授权方案 1-14
1.6.6 配置ISP域的AAA计费方案 1-15
1.6.7 配置本地用户的属性 1-16
1.6.8 配置用户组的属性 1-18
1.6.9 配置强制切断用户连接 1-19
1.6.10 AAA显示和维护 1-19
1.7 配置RADIUS 1-19
1.7.1 创建RADIUS方案 1-20
1.7.2 配置RADIUS认证/授权服务器 1-20
1.7.3 配置RADIUS计费服务器及相关参数 1-21
1.7.4 配置RADIUS报文的共享密钥 1-21
1.7.5 配置RADIUS报文的超时重传次数最大值 1-22
1.7.6 配置支持的RADIUS服务器的类型 1-22
1.7.7 配置RADIUS服务器的状态 1-23
1.7.8 配置发送给RADIUS服务器的数据相关属性 1-24
1.7.9 配置RADIUS服务器的定时器 1-24
1.7.10 配置RADIUS的accouting-on功能 1-25
1.7.11 配置RADIUS服务器的安全策略服务器 1-26
1.7.12 使能RADIUS客户端的监听端口 1-27
1.7.13 RADIUS显示和维护 1-27
1.8 配置HWTACACS 1-28
1.8.1 创建HWTACACS方案 1-28
1.8.2 配置HWTACACS认证服务器 1-28

1.8.3 配置HWTACACS授权服务器 1-29
1.8.4 配置HWTACACS计费服务器 1-29
1.8.5 配置HWTACACS报文的共享密钥 1-30
1.8.6 配置发送给HWTACACS服务器的数据相关属性 1-30
1.8.7 配置HWTACACS服务器的定时器 1-31
1.8.8 HWTACACS显示和维护 1-31
1.9 AAA典型配置举例 1-32
1.9.1 Telnet/SSH用户通过RADIUS服务器认证、授权、计费的应用配置 1-32
1.9.2 FTP/Telnet用户本地认证、授权、计费配置 1-35
1.9.3 PPP用户通过HWTACACS服务器认证、授权、计费的应用配置 1-36
1.10 AAA RADIUS HWTACACS常见配置错误举例 1-37
1.10.1 RADIUS认证/授权失败 1-37
1.10.2 RADIUS报文传送失败 1-38
1.10.3 RADIUS计费功能异常 1-38
1.10.4 HWTACACS常见配置错误举例 1-38

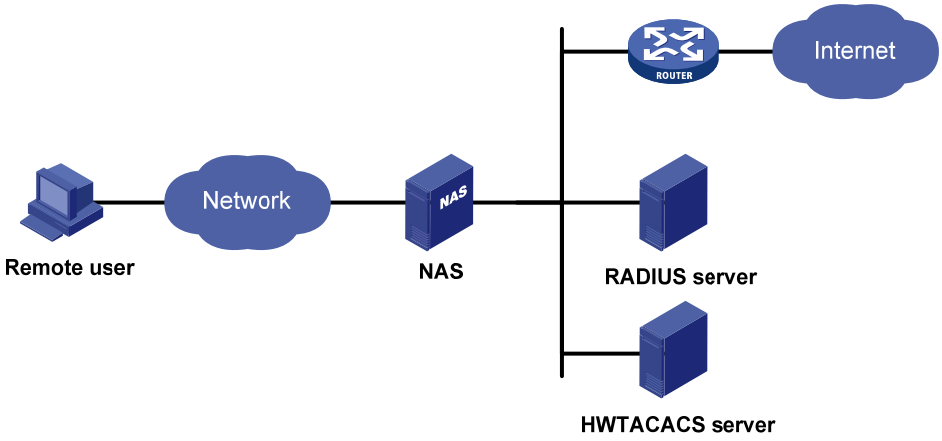
1 AAA 配置

1.1 AAA 简介

AAA 是 Authentication、Authorization、Accounting（认证、授权、计费）的简称，是网络安全的一种管理机制，提供了认证、授权、计费三种安全功能。

AAA一般采用客户机/服务器结构，客户端运行于NAS（Network Access Server，网络接入服务器）上，服务器上则集中管理用户信息。NAS对于用户来讲是服务器端，对于服务器来说是客户端。AAA的基本组网结构如 图 1-1。

图1-1 AAA 基本组网结构示意图



当用户想要通过某网络与 NAS 建立连接，从而获得访问其它网络的权利或取得某些网络资源的权利时，NAS 起到了验证用户或对应连接的作用。NAS 负责把用户的认证、授权、计费信息透传给服务器（RADIUS 服务器或 HWTACACS 服务器），RADIUS 协议或 HWTACACS 协议规定了 NAS 与服务器之间如何传递用户信息。

图 1-1的AAA基本组网结构中有两台服务器，用户可以根据实际组网需求来决定认证、授权、计费功能分别由使用哪种协议类型的服务器来承担。例如，可以选择RADIUS服务器实现为认证、授权，HWTACACS服务器来实现计费。

这三种安全服务功能的具体作用如下：

- 认证：确认远端访问用户的身份，判断访问者是否为合法的网络用户；
- 授权：对不同用户赋予不同的权限，限制用户可以使用的服务。例如用户成功登录服务器后，管理员可以授权用户对服务器中的文件进行访问和打印操作；
- 计费：记录用户使用网络服务中的所有操作，包括使用的服务类型、起始时间、数据流量等，它不仅是一种计费手段，也对网络安全起到了监视作用。

当然，用户可以只使用 AAA 提供的一种或两种安全服务。例如，公司仅仅想让员工在访问某些特定资源的时候进行身份认证，那么网络管理员只要配置认证服务器就可以了。但是若希望对员工使用网络的情况进行记录，那么还需要配置计费服务器。

如上所述，AAA 是一种管理框架，它提供了授权部分实体去访问特定资源，同时可以记录这些实体操作行为的一种安全机制，因其具有良好的可扩展性，并且容易实现用户信息的集中管理而被广泛使用。

AAA 可以通过多种协议来实现，目前设备支持基于 RADIUS 协议、HWTACACS 协议来实现 AAA，在实际应用中，最常使用 RADIUS 协议。

1.2 RADIUS 协议简介

RADIUS（Remote Authentication Dial-In User Service，远程认证拨号用户服务）是一种分布式的、客户端/服务器结构的信息交互协议，能保护网络不受未授权访问的干扰，常应用在既要求较高安全性、又允许远程用户访问的各种网络环境中。该协议定义了基于 UDP 的 RADIUS 帧格式及其消息传输机制，并规定 UDP 端口 1812、1813 分别作为认证、计费端口。

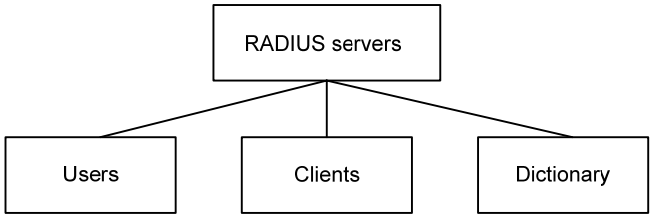
RADIUS 最初仅是针对拨号用户的 AAA 协议，后来随着用户接入方式的多样化发展，RADIUS 也适应多种用户接入方式，如以太网接入、ADSL 接入。它通过认证授权来提供接入服务，通过计费来收集、记录用户对网络资源的使用。

1.2.1 客户端/服务器模式

- 客户端：RADIUS 客户端一般位于 NAS 设备上，可以遍布整个网络，负责传输用户信息到指定的 RADIUS 服务器，然后根据从服务器返回的信息进行相应处理（如接受/拒绝用户接入）。
- 服务器：RADIUS 服务器运行在中心计算机或工作站上，维护相关的用户认证和网络服务访问信息，负责接收用户连接请求并认证用户，然后给客户端返回所有需要的信息（如接受/拒绝认证请求）。

RADIUS服务器通常要维护三个数据库，如 图 1-2所示：

图1-2 RADIUS 服务器的组成



- “Users”：用于存储用户信息（如用户名、口令以及使用的协议、IP 地址等配置信息）。
- “Clients”：用于存储 RADIUS 客户端的信息（如接入设备的共享密钥、IP 地址等）。
- “Dictionary”：用于存储 RADIUS 协议中的属性和属性值含义的信息。

1.2.2 安全和认证机制

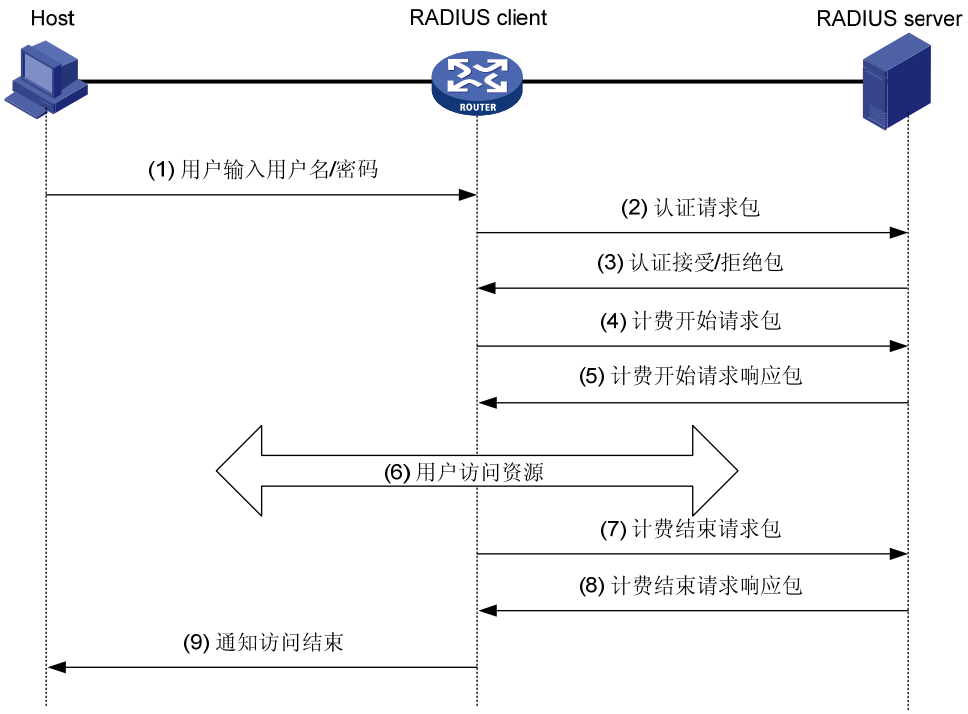
RADIUS 客户端和 RADIUS 服务器之间认证消息的交互是通过共享密钥的参与来完成的，并且共享密钥不能通过网络来传输，增强了信息交互的安全性。另外，为防止用户密码在不安全的网络上传递时被窃取，在传输过程中对密码进行了加密。

RADIUS 服务器支持多种方法来认证用户，如基于 PPP 的 PAP、CHAP 认证。另外，RADIUS 服务器还可以为其它类型的认证服务器提供代理客户端的功能，向其提出认证请求。

1.2.3 RADIUS 的基本消息交互流程

用户、RADIUS客户端和RADIUS服务器之间的交互流程如 图 1-3所示。

图1-3 RADIUS 的基本消息交互流程



消息交互流程如下：

(1) 用户发起连接请求，向 RADIUS 客户端发送用户名和密码。

(2) RADIUS 客户端根据获取的用户名和密码，向 RADIUS 服务器发送认证请求包（Access-Request），其中的密码在共享密钥的参与下由 MD5 算法进行加密处理。

(3) RADIUS 服务器对用户名和密码进行认证。如果认证成功，RADIUS 服务器向 RADIUS 客户端发送认证接受包（Access-Accept）；如果认证失败，则返回认证拒绝包（Access-Reject）。由于 RADIUS 协议合并了认证和授权的过程，因此认证接受包中也包含了用户的授权信息。

(4) RADIUS 客户端根据接收到的认证结果接入/拒绝用户。如果允许用户接入，则 RADIUS 客户端向 RADIUS 服务器发送计费开始请求包（Accounting-Request）。

(5) RADIUS 服务器返回计费开始响应包（Accounting-Response），并开始计费。

(6) 用户开始访问网络资源。

(7) 用户请求断开连接，RADIUS 客户端向 RADIUS 服务器发送计费停止请求包（Accounting-Request），状态类型 Status-Type 取值为 stop；

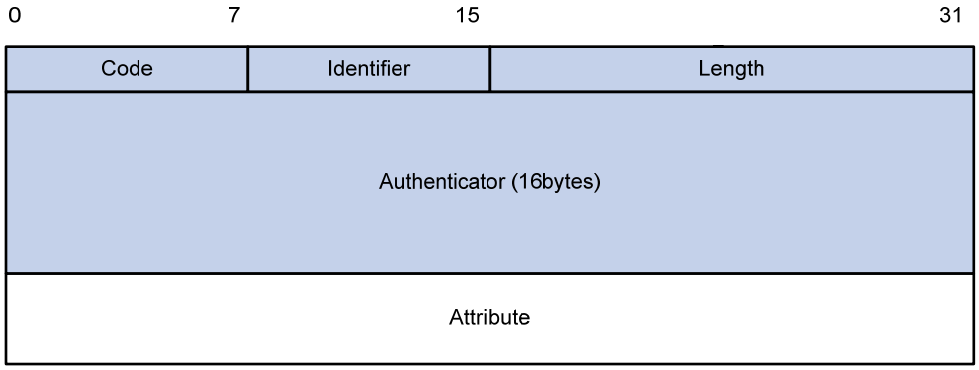
(8) RADIUS 服务器返回计费结束响应包（Accounting-Response），并停止计费。

(9) 用户结束访问网络资源。

1.2.4 RADIUS 报文结构

RADIUS采用UDP报文来传输消息，通过定时器管理机制、重传机制、备用服务器机制，确保RADIUS服务器和客户端之间交互消息的正确收发。RADIUS报文结构如 图 1-4所示。

图1-4 RADIUS 报文结构



各字段的解释如下：

(1) Code 域

长度为 1 个字节，用于说明RADIUS报文的类型，如 表 1-1所示。

表1-1 Code 域的主要取值说明

Code	报文类型	报文说明
1	Access-Request 认证请求包	方向 Client->Server，Client 将用户信息传输到 Server，由 Server 判断是否接入该用户。该报文中必须包含 User-Name 属性，可选包含 NAS-IP-Address、User-Password、NAS-Port 等属性
2	Access-Accept 认证接受包	方向 Server->Client，如果 Access-Request 报文中的所有 Attribute 值都可以接受（即认证通过），则传输该类型报文
3	Access-Reject 认证拒绝包	方向 Server->Client，如果 Access-Request 报文中存在任何无法被接受的 Attribute 值（即认证失败），则传输该类型报文
4	Accounting-Request 计费请求包	方向 Client->Server，Client 将用户信息传输到 Server，请求 Server 开始/停止计费，由该报文中的 Acct-Status-Type 属性区分计费开始请求和计费结束请求
5	Accounting-Response 计费响应包	方向 Server->Client，Server 通知 Client 已经收到 Accounting-Request 报文，并且已经正确记录计费信息

(2) Identifier 域

长度为 1 个字节，用于匹配请求包和响应包，以及检测在一段时间内重发的请求包。类型一致的请求包和响应包的 Identifier 值相同。

(3) Length 域

长度为 2 个字节，表示 RADIUS 数据包（包括 Code、Identifier、Length、Authenticator 和 Attribute）的长度，范围从 20～4096。超过 Length 域的字节将作为填充字符被忽略。如果接收到的包的实际长度小于 Length 域的值时，则包会被丢弃。

(4) Authenticator 域

长度为 16 个字节，用于验证 RADIUS 服务器的应答，另外还用于用户密码的加密。Authenticator 包括两种类型：Request Authenticator 和 Response Authenticator。

(5) Attribute 域

不定长度，用于携带专门的认证、授权和计费信息，提供请求和响应报文的配置细节。Attribute 可包括多个属性，每一个属性都采用（Type、Length、Value）三元组的结构来表示。

- 类型（Type），1 个字节，取值为 1～255，用于表示属性的类型，表 1-2列出了RADIUS认证、授权、计费常用的属性。
- 长度（Length），表示该属性（包括类型、长度和属性）的长度，单位为字节。
- 属性值（Value），表示该属性的信息，其格式和内容由类型和长度决定，最大长度为 253 字节。

表1-2 RADIUS 属性

属性编号	属性名称	属性编号	属性名称
1	User-Name	45	Acct-Authentic
2	User-Password	46	Acct-Session-Time
3	CHAP-Password	47	Acct-Input-Packets
4	NAS-IP-Address	48	Acct-Output-Packets
5	NAS-Port	49	Acct-Terminate-Cause
6	Service-Type	50	Acct-Multi-Session-Id
7	Framed-Protocol	51	Acct-Link-Count
8	Framed-IP-Address	52	Acct-Input-Gigawords
9	Framed-IP-Netmask	53	Acct-Output-Gigawords
10	Framed-Routing	54	(unassigned)
11	Filter-ID	55	Event-Timestamp
12	Framed-MTU	56-59	(unassigned)
13	Framed-Compression	60	CHAP-Challenge
14	Login-IP-Host	61	NAS-Port-Type
15	Login-Service	62	Port-Limit
16	Login-TCP-Port	63	Login-LAT-Port
17	(unassigned)	64	Tunnel-Type
18	Reply_Message	65	Tunnel-Medium-Type
19	Callback-Number	66	Tunnel-Client-Endpoint
20	Callback-ID	67	Tunnel-Server-Endpoint
21	(unassigned)	68	Acct-Tunnel-Connection
22	Framed-Route	69	Tunnel-Password
23	Framed-IPX-Network	70	ARAP-Password
24	State	71	ARAP-Features
25	Class	72	ARAP-Zone-Access
26	Vendor-Specific	73	ARAP-Security
27	Session-Timeout	74	ARAP-Security-Data
28	Idle-Timeout	75	Password-Retry
29	Termination-Action	76	Prompt
30	Called-Station-Id	77	Connect-Info
31	Calling-Station-Id	78	Configuration-Token
32	NAS-Identifier	79	EAP-Message

属性编号	属性名称	属性编号	属性名称
33	Proxy-State	80	Message-Authenticator
34	Login-LAT-Service	81	Tunnel-Private-Group-id
35	Login-LAT-Node	82	Tunnel-Assignment-id
36	Login-LAT-Group	83	Tunnel-Preference
37	Framed-AppleTalk-Link	84	ARAP-Challenge-Response
38	Framed-AppleTalk-Network	85	Acct-Interim-Interval
39	Framed-AppleTalk-Zone	86	Acct-Tunnel-Packets-Lost
40	Acct-Status-Type	87	NAS-Port-Id
41	Acct-Delay-Time	88	Framed-Pool
42	Acct-Input-Octets	89	(unassigned)
43	Acct-Output-Octets	90	Tunnel-Client-Auth-id
44	Acct-Session-Id	91	Tunnel-Server-Auth-id

说明

表 1-2中所列的属性由RFC 2865、RFC 2866、RFC2867 和RFC2568 分别定义。

1.2.5 RADIUS 扩展属性

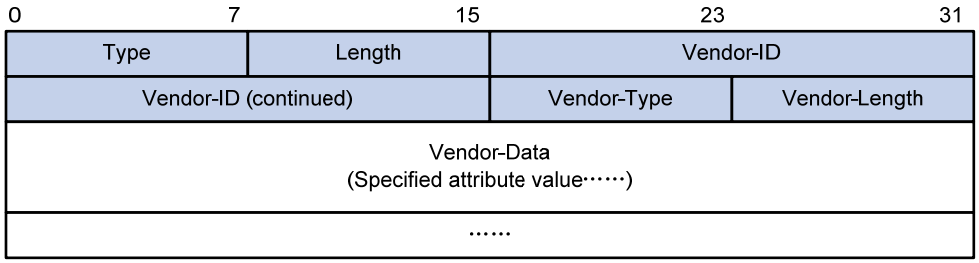
RADIUS 协议具有良好的可扩展性，协议（RFC 2865）中定义的 26 号属性（Vendor-Specific）用于设备厂商对 RADIUS 进行扩展，以实现标准 RADIUS 没有定义的功能。

设备厂商可以封装多个自定义的“（Type、Length、Value）”子属性来扩展RADIUS。如 图 1-5 所示，26 号属性报文内封装的子属性包括以下四个部分：

- Vendor-ID 域占 4 字节，表示厂商代号，最高字节为 0，其余 3 字节的编码见 RFC 1700。H3C 公司的 Vendor-ID 是 2011。
- Vendor-Type，表示扩展属性的子属性类型。
- Vendor-Length，表示该子属性长度。

Vendor-Data，表示该子属性的内容。

图1-5 包括扩展属性的 RADIUS 报文片断



1.3 HWTACACS 协议简介

HWTACACS（HW Terminal Access Controller Access Control System，HW 终端访问控制器控制系统协议）是在 TACACS（RFC 1492）基础上进行了功能增强的安全协议。该协议与 RADIUS 协议类似，采用客户端/服务器模式实现 NAS 与 HWTACACS 服务器之间的通信。

HWTACACS 协议主要用于 PPP（Point-to-Point Protocol，点对点协议）和 VPDN（Virtual Private Dial-up Network，虚拟私有拨号网络）接入用户及终端用户的认证、授权和计费。其典型应用是对需要登录到设备上进行操作的终端用户进行认证、授权、计费。设备作为 HWTACACS 的客户端，将用户名和密码发给 HWTACACS 服务器进行验证。用户验证通过并得到授权之后可以登录到设备上进行操作。

1.3.1 HWTACACS 协议与 RADIUS 协议的区别

HWTACACS协议与RADIUS协议都实现了认证、授权、计费的功能，它们有很多相似点：结构上都采用客户端/服务器模式；都使用公共密钥对传输的用户信息进行加密；都有较好的灵活性和可扩展性。两者之间存在的主要区别如 表 1-3所示。

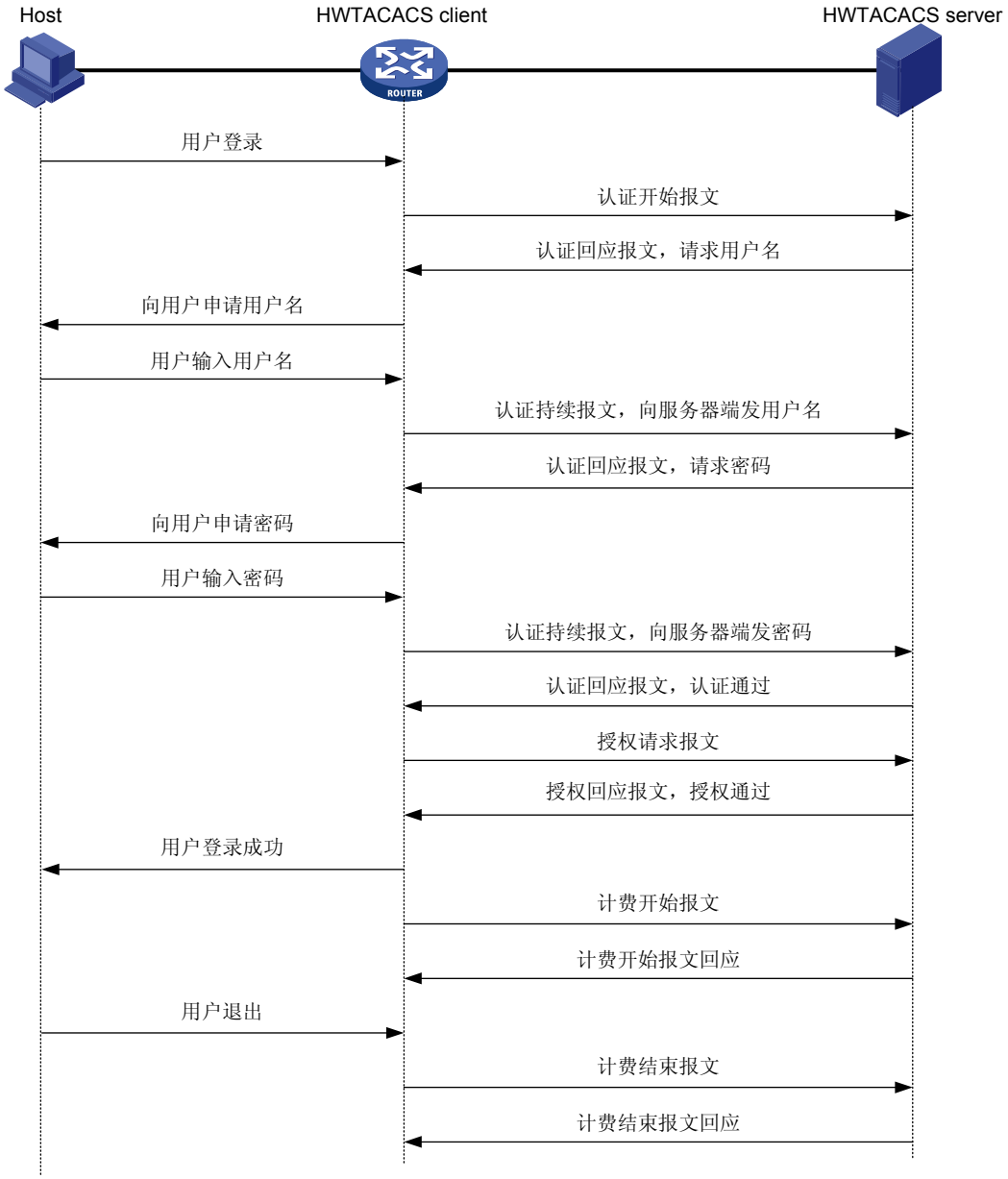
表1-3 HWTACACS 协议和 RADIUS 协议区别

HWTACACS 协议	RADIUS 协议
使用 TCP，网络传输更可靠	使用 UDP，网络传输效率更高
除了 HWTACACS 报文头，对报文主体全部进行加密	只对验证报文中的密码字段进行加密
协议报文较为复杂，认证和授权分离，使得认证、授权服务可以分离在不同的安全服务器上实现。例如，可以用一个 HWTACACS 服务器进行认证，另外一个 HWTACACS 服务器进行授权	协议报文比较简单，认证和授权结合，难以分离
支持对设备的配置命令进行授权使用。例如，可以授权一个通过安全认证登录设备的用户对设备进行配置	不支持

1.3.2 HWTACACS 的基本消息交互流程

下面以Telnet用户为例，说明使用HWTACACS对用户进行认证、授权和计费的过程。基本消息交互流程图如 图 1-6所示。

图1-6 Telnet 用户认证、授权和计费流程图



在整个过程中的基本消息交互流程如下：

(1) Telnet 用户请求登录设备。

(2) HWTACACS 客户端收到请求之后，向 HWTACACS 服务器发送认证开始报文。

(3) HWTACACS 服务器发送认证回应报文，请求用户名。

(4) HWTACACS 客户端收到回应报文后，向用户询问用户名。

(5) 用户输入用户名。

(6) HWTACACS 客户端收到用户名后，向 HWTACACS 服务器发送认证持续报文，其中包括了用户名。

(7) HWTACACS 服务器发送认证回应报文，请求登录密码。

(8) HWTACACS 客户端收到回应报文，向用户询问登录密码。

(9) 用户输入密码。

(10) HWTACACS 客户端收到登录密码后，向 HWTACACS 服务器发送认证持续报文，其中包括了登录密码。

(11) HWTACACS 服务器发送认证回应报文，指示用户通过认证。

(12) HWTACACS 客户端向 HWTACACS 服务器发送授权请求报文。

(13) HWTACACS 服务器发送授权回应报文，指示用户通过授权。

(14) HWTACACS 客户端收到授权回应成功报文，向用户输出路由器的配置界面。

(15) HWTACACS 客户端向 HWTACACS 服务器发送计费开始报文。

(16) HWTACACS 服务器发送计费回应报文，指示计费开始报文已经收到。

(17) 用户请求断开连接。

(18) HWTACACS 客户端向 HWTACACS 服务器发送计费结束报文。

(19) HWTACACS 服务器发送计费结束报文，指示计费结束报文已经收到。

1.4 协议规范

与 AAA RADIUS HWTACACS 相关的协议规范有：

- RFC 2865：Remote Authentication Dial In User Service (RADIUS)
- RFC 2866：RADIUS Accounting
- RFC 2867：RADIUS Accounting Modifications for Tunnel Protocol Support
- RFC 2868：RADIUS Attributes for Tunnel Protocol Support
- RFC 2869：RADIUS Extensions
- RFC 1492：An Access Control Protocol, Sometimes Called TACACS

1.5 AAA RADIUS HWTACACS 配置任务简介

AAA 的基本配置思路如下：

(1) 配置 AAA 方案。根据需要配置本地或远程认证方案。

- 本地认证：需要配置本地用户，即 local user 的相关属性，包括手动添加认证的用户名和密码等；
- 远程认证：需要配置 RADIUS、HWTACACS 方案，并在服务器上配置相应的用户属性。

(2) 配置实现 AAA 的方法。在用户所属的 ISP 域中引用已配置的 AAA 方案。

- 认证方法：可选择不认证（none）、本地认证（local）或远程认证（scheme）；
- 授权方法：可选择不授权（none）、本地授权（local）或远程授权（scheme）；
- 计费方法：可选择不计费（none）、本地计费（local）或远程计费（scheme）。


说明

对于 Login 用户，在以上配置之前还需要配置登录用户界面的认证方式为 **scheme**。有关登录用户界面认证方式的详细介绍请参见“系统分册/用户界面配置”。

表1-4 AAA 配置任务简介

配置任务	说明	详细配置
创建 ISP 域	必选	1.6.2
配置 ISP 域的属性	可选	1.6.3
配置 ISP 域的 AAA 认证方案	必选	1.6.4
配置 ISP 域的 AAA 授权方案	可选	1.6.5
配置 ISP 域的 AAA 计费方案	可选	1.6.6
配置本地用户的属性	可选	1.6.7
配置用户组的属性	可选	1.6.8
配置强制切断用户连接	可选	1.6.9
AAA 显示和维护	可选	1.6.10

表1-5 RADIUS 配置任务简介

配置任务	说明	详细配置
创建 RADIUS 方案	必选	1.7.1
配置 RADIUS 认证/授权服务器	必选	1.7.2
配置 RADIUS 计费服务器及相关参数	可选	1.7.3
配置 RADIUS 报文的共享密钥	必选	1.7.4
配置 RADIUS 报文的超时重传次数的最大值	可选	1.7.5
配置支持的 RADIUS 服务器的类型	可选	1.7.6
配置 RADIUS 服务器的状态	可选	1.7.7
配置发送给 RADIUS 服务器的数据相关属性	可选	1.7.8
配置 RADIUS 服务器的定时器	可选	1.7.9
配置 RADIUS 的 accounting-on 功能	可选	1.7.10
配置 RADIUS 服务器的安全策略服务器	可选	1.7.11
使能 RADIUS 客户端的监听端口	可选	1.7.12
RADIUS 显示和维护	可选	1.7.13

表1-6 HWTACACS 配置任务简介

配置任务	说明	详细配置
创建 HWTACACS 方案	必选	1.8.1
配置 HWTACACS 认证服务器	必选	1.8.2
配置 HWTACACS 授权服务器	必选	1.8.3
配置 HWTACACS 计费服务器	可选	1.8.4
配置 HWTACACS 报文的共享密钥	必选	1.8.5
配置发送给 HWTACACS 服务器的数据相关属性	可选	1.8.6

配置任务	说明	详细配置
配置 HWTACACS 服务器的定时器	可选	1.8.7
HWTACACS 显示和维护	可选	1.8.8

1.6 配置 AAA

AAA 功能可以在为合法用户提供网络接入服务的同时，对网络设备进行保护，防止非授权访问和抵赖行为。同时，通过配置 ISP 域可以对接入用户进行 AAA 等管理。

AAA 将用户类型分为： login 用户（如 SSH、Telnet、FTP、终端接入用户）、Portal 用户、PPP 用户、命令行用户（命令行授权用户）。

对于除命令行用户之外的各类型用户请求，按照实际需求，可以单独配置认证/授权/计费的策略。命令行用户可以单独配置授权策略，用来支持对设备的配置命令进行授权使用。

1.6.1 配置准备

进行远端认证、授权或计费时，需要已经创建 RADIUS 或 HWTACACS 方案。

- RADIUS方案（**radius-scheme**）：通过引用已配置的RADIUS方案来实现认证、授权、计费。有关RADIUS方案的配置请参见“1.7 配置RADIUS”。
- HWTACACS方案（**hwtacacs-scheme**）：通过引用已配置的HWTACACS方案来实现认证、授权、计费。有关HWTACACS方案的配置请参见“1.8 配置HWTACACS”。

1.6.2 创建 ISP 域

一个 ISP（Internet Service Provider，Internet 服务提供者）域是由属于同一个 ISP 的用户构成的群体。

在“*userid@isp-name*”形式的用户名中，“*userid*”为用于身份认证的用户名，“*isp-name*”为域名。

在多 ISP 的应用环境中，不同 ISP 域的用户有可能接入同一台设备。而且各 ISP 用户的用户属性（例如用户名及密码构成、服务类型/权限等）有可能不相同，因此有必要通过设置 ISP 域把它们区分开，并为每个 ISP 域单独配置包括 AAA 策略（一组不同的认证/授权/计费方案）在内的属性集。

对于设备来说，每个接入用户都属于一个 ISP 域。系统中最多可以配置 16 个 ISP 域。如果某个用户在登录时没有提供 ISP 域名，系统将把它归于缺省的 ISP 域。

表1-7 创建 ISP 域

操作	命令	说明
进入系统视图	**system-view**	-
创建 ISP 域并进入其视图	**domain** *isp-name*	必选
返回系统视图	**quit**	-
手工配置缺省的 ISP 域	**domain default** *isp-name*	可选 缺省情况下，系统缺省的 ISP 域为 system

说明

- 配置为缺省的 ISP 域不能够被删除，除非首先使用命令 **undo domain default** 将其修改为非缺省 ISP 域。
- 若用户登录时输入的用户名未携带 ISP 域名，则用户使用缺省 ISP 域下的认证方案。

1.6.3 配置 ISP 域的属性

表1-8 配置 ISP 域的属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 ISP 域视图	**domain** *isp-name*	-
设置 ISP 域的状态	**state** { **active** | **block** }	可选 缺省情况下，当一个 ISP 域被创建以后，其状态为 **active**，即允许任何属于该域的用户请求网络服务
设置当前 ISP 域可容纳接入用户数的限制	**access-limit enable** *max-user-number*	可选 缺省情况下，不对当前 ISP 域可容纳的接入用户数作限制
设置用户闲置切断功能	**idle-cut enable** *minute flow*	可选 缺省情况下，用户闲置切断功能处于关闭状态 此命令目前只对 Portal、PPP 服务类型的用户有效
设置自助服务器定位功能	**self-service-url enable** *url-string*	可选 缺省情况下，自助服务器定位功能处于关闭状态
定义为 PPP 用户分配 IP 地址的地址池	**ip pool** *pool-number low-ip-address* [*high-ip-address*]	可选 缺省情况下，没有定义为 PPP 用户分配 IP 地址的 IP 地址池

说明

自助服务器定位功能需要与支持自助服务的 RADIUS 服务器配合使用，如 CAMS/iMC。安装自助服务软件的服务器即自助服务器。通过使用自助服务，用户可以对自己的帐号或卡号进行管理和控制。

1.6.4 配置 ISP 域的 AAA 认证方案

在 AAA 中，认证、授权和计费是三个独立的业务流程。认证的职责是完成各接入或服务请求的用户名 / 密码 / 用户信息的交互认证过程，它不会下发授权信息给请求用户，也不会触动计费的流程。

AAA 支持以下认证方法：

- 不认证：对用户非常信任，不对其进行合法性检查，一般情况下不采用这种方式。
- 本地认证：认证过程在接入设备上完成，用户信息（包括用户名、密码和各种属性）配置在接入设备上。优点是速度快，可以降低运营成本；缺点是存储信息量受设备硬件条件限制。

远端认证：认证过程在接入设备和远端的服务器之间完成，接入设备和远端服务器之间通过RADIUS协议或 HWTACACS 协议通信。对于 RADIUS 协议，可以采用标准或扩展的 RADIUS 协议，与CAMS/iMC 等系统配合完成认证。优点是用户信息集中在服务器上统一管理，可实现大容量、高可靠性、支持多设备的集中式统一认证。当远端服务器无效时，可配置本地认证作为备选认证方式完成认证。

在 AAA 中，可以只使用认证，而不使用授权或计费。如果用户没有进行任何配置，则 ISP 域的缺省认证方案为 **local**。

配置认证有三个步骤：

(1) 如果用户使用 RADIUS 或 HWTACACS 认证，则需要先配置要引用的 RADIUS 方案或 HWTACACS 方案；如果使用 **local** 或 **none** 认证，则不需要配置方案。

(2) 确定要配置的接入方式或者服务类型。AAA 可以对不同的接入方式和服务类型配置不同的认证方案，并且从配置上限制了用户接入时使用的认证协议。

(3) 确定是否为所有的接入方式或服务类型配置认证方案。

表1-9 配置 ISP 域的 AAA 认证方案

操作	命令	说明
进入系统视图	**system-view**	-
进入 ISP 域视图	**domain** *isp-name*	-
为所有类型的用户配置缺省的认证方案	**authentication default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，所有类型的用户采用 **local** 认证方案
为 login 用户配置认证方案	**authentication login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，login 用户采用缺省的认证方案
为 Portal 用户配置认证方案	**authentication portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，Portal 用户采用缺省的认证方案
为 PPP 用户配置认证方案	**authentication ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，PPP 用户采用缺省的认证方案

说明

- **authentication default** 命令配置的认证方案不区分用户类型，即对所有类型的用户都起作用。此配置的优先级低于具体接入方式的配置。
- 当 RADIUS 被配置选择为认证方案时，AAA 只接受 RADIUS 服务器的认证结果，RADIUS 授权的信息虽然在认证成功回应的报文中携带，但在认证回应的处理流程中不会被处理。
- 如果配置了 **radius-scheme** *radius-scheme-name* **local** 或 **hwtacacs-scheme** *hwtacacs-scheme-name* **local**，则 **local** 为远端服务器没有正常响应后的备选认证方案。
- 如果 **local** 或者 **none** 作为第一认证方案，那么只能采用本地认证或者不进行认证，不能再同时采用 RADIUS 或 HWTACACS 方案。

1.6.5 配置 ISP 域的 AAA 授权方案

在 AAA 中，授权是一个和认证、计费同级别的独立流程，其职责是发送授权请求给所配置的授权服务器，授权通过后向用户下发授权信息。在 ISP 域的 AAA 配置中，授权方案为可选配置。

AAA 支持以下授权方法：

- 直接授权：对用户非常信任，直接授权用户的权限为接入设备允许用户所使用默认权限。
- 本地授权：授权过程在接入设备上进行，根据接入设备上为本地用户配置的相关属性进行授权。
- 远端授权：授权过程在接入设备和远端服务器之间完成。RADIUS 协议的认证和授权是绑定在一起的，不能单独使用 RADIUS 进行授权。RADIUS 认证成功后，才能进行授权，RADIUS 授权信息携带在认证回应报文中下发给用户。HWTACACS 授权与认证相分离，在认证成功后，HWTACACS 授权信息通过授权报文进行交互。当远端服务器无效时，可配置本地授权或直接授权作为备选授权方式完成授权。

如果用户没有进行任何配置，则 ISP 域的缺省授权方案为 **local**。如果用户配置的授权方案为 **none**，则意味着用户是不需要授权的，此时，认证通过的用户只有系统所给予的默认权限。EXEC 用户（控制台用户，可以通过 console/aux/异步串口或者 Telnet 连接设备，每个连接称为一个 EXEC 用户，如 Telnet 用户、SSH 用户）的默认级别为最低权限的访问级。FTP 用户被默认授权使用设备的根目录。

配置授权有三个步骤：

(1) 如果用户要使用 HWTACACS 授权，则需要先配置要引用的 HWTACACS 方案；如果是 RADIUS 授权，只有将认证和授权的 RADIUS 方案配置相同，授权才起作用。

(2) 确定要配置的接入方式或者服务类型，AAA 可以按照不同的接入方式和服务类型进行 AAA 授权的配置，并且从配置上正确限制了接入所能使用的授权协议。

(3) 确定是否为所有的接入方式或服务类型配置授权方案。

表1-10 配置 ISP 域的 AAA 授权方案

操作	命令	说明
进入系统视图	**system-view**	-
进入 ISP 域视图	**domain** *isp-name*	-
为所有类型的用户配置缺省的授权方案	**authorization default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，所有类型的用户采用 **local** 授权方案
为命令行用户配置授权方案	**authorization command** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local** | **none**] | **local** | **none** }	可选 缺省情况下，命令行用户采用缺省的授权方案
为 login 用户配置授权方案	**authorization login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，login 用户采用缺省的授权方案
为 Portal 用户配置授权方案	**authorization portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，Portal 用户采用缺省的授权方案

操作	命令	说明
为 PPP 用户配置授权方案	**authorization ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，PPP 用户采用缺省的授权方案

说明

- **authorization default** 命令配置的授权方案不区分用户类型，即对所有类型的用户都起作用。此配置的优先级低于具体接入方式的配置。
- RADIUS 授权是特殊的流程，只是在认证和授权的 RADIUS 方案相同的条件下，RADIUS 授权才起作用。对于所有 RADIUS 授权失败的情况，授权失败返回给 NAS 的原因为服务器没有响应。
- 如果配置了 **radius-scheme** *radius-scheme-name* **local** 或 **hwtacacs-scheme** *hwtacacs-scheme-name* **local**，则 **local** 为 RADIUS 服务器或 HWTACACS 服务器没有正常响应后的备选授权方案。
- 如果 **local** 或者 **none** 作为第一授权方案，那么只能采用本地授权或者不进行授权，不能再同时采用 RADIUS 或 HWTACACS 方案。
- RADIUS 服务器的授权信息是随认证应答报文发给 RADIUS 客户端的，故不能单独指定授权服务器。因此，如果指定的认证和授权方案是 RADIUS 方案，必须保证认证和授权的方案相同，否则系统会给出错误提示。

1.6.6 配置 ISP 域的 AAA 计费方案

在 AAA 中，计费是一个和认证、授权同级别的独立流程，其职责为发送计费开始 / 更新 / 结束请求给所配置的计费服务器。计费不是必须使用的。在 ISP 域的 AAA 配置中，允许不配置计费方案。如果不配置计费方案，则 ISP 域的缺省计费方案为 **local**。

AAA 支持以下计费方法：

- 不计费：不对用户计费。
- 本地计费：计费过程在接入设备上完成，实现了本地用户连接数的统计和限制，并没有实际的费用统计功能。
- 远端计费：计费过程在接入设备和远端的服务器之间完成。当远端服务器无效时，可配置本地计费作为备选计费方式完成计费。

配置计费有三个步骤：

(1) 如果用户要使用 RADIUS 或 HWTACACS 计费，则需要先配置要引用的 RADIUS 方案或 HWTACACS 方案；如果要使用 **local** 或 **none** 计费，则不需要配置方案。

(2) 确定要配置的接入方式或者服务类型，AAA 可以支持为按照不同的接入方式和服务类型进行 AAA 计费的配置，并且从配置上正确限制了接入所能使用的计费协议。

(3) 确定是否为所有的接入方式或服务类型配置计费方案。

表1-11 配置 ISP 域的 AAA 计费方案

操作	命令	说明
进入系统视图	**system-view**	-
进入 ISP 域视图	**domain** *isp-name*	-

操作	命令	说明
打开计费可选开关	**accounting optional**	可选 缺省情况下，ISP 域的计费开关处于关闭状态
为所有类型的用户配置缺省的计费方案	**accounting default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，所有类型的用户采用 **local** 计费方案
为 login 用户配置计费方案	**accounting login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，login 用户采用缺省的计费方案
为 Portal 用户配置计费方案	**accounting portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，Portal 用户采用缺省的计费方案
为 PPP 用户配置计费方案	**accounting ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	可选 缺省情况下，PPP 用户采用缺省的计费方案

说明

- 在对用户实施计费时，如果发现没有可用的计费服务器或与计费服务器通信失败，若配置了 **accounting optional** 命令，则用户可以继续使用网络资源，否则用户连接将被切断。
- **accounting default** 命令配置的计费方案不区分用户类型，即对所有类型的用户都起作用。此配置的优先级低于具体接入方式的配置。
- 如果配置了 **radius-scheme** *radius-scheme-name* **local** 或 **hwtacacs-scheme** *hwtacacs-scheme-name*，则 **local** 为服务器没有正常响应后的备选计费方案。
- 如果 **local** 或者 **none** 作为第一计费方案，那么只能采用本地认证或者不进行计费，不能再同时采用 RADIUS 或 HWTACACS 方案。
- login 类型的接入中 FTP 方式不支持计费流程。

1.6.7 配置本地用户的属性

当 AAA 方案选择了本地认证方案（**local**）时，应在设备上创建本地用户并配置相关属性。

所谓本地用户，是指在本地设备上设置的一组用户属性的集合。该集合以用户名为用户的唯一标识。为使某个请求网络服务的用户可以通过本地认证，需要在设备上的本地用户数据库中添加相应的表项。

每一个本地用户都属于一个本地用户组，并继承组中的所有属性（密码安全策略和用户授权属性）。关于本地用户组的介绍和配置请参见 1.6.8 配置用户组的属性。

在配置本地用户和用户组时，需要首先明确用户组属性和部分用户属性的生效范围及优先级关系，具体介绍如下：

- 密码管理属性

本地用户的密码管理属性在系统视图（具有全局性）、用户组视图和本地用户视图下都可以配置，其生效的优先级顺序由高到底依次为本地用户、用户组、全局。全局配置对所有本地用户生效，用户组的配置对组内所有本地用户生效。系统优先采用本地用户视图下的配置，其次是本地用户所属用户组的配置，最后才会采用全局的配置。

- 用户授权属性

本地用户的授权属性在用户组和本地用户视图下都可以配置，且本地用户视图下的配置优先级高于用户组视图下的配置。用户组的配置对组内所有本地用户生效

表1-12 配置本地用户的属性

操作	命令	说明
进入系统视图	**system-view**	-
设置本地用户密码的显示方式	**local-user password-display-mode** { **auto** | **cipher-force** }	可选 缺省情况下，所有接入用户的密码显示方式为**auto**，表示按照用户设置密码时指定的密码显示方式显示
添加本地用户，并进入本地用户视图	**local-user** *user-name*	必选 缺省情况下，系统中没有任何本地用户
设置指定用户的密码	**password** { **cipher** | **simple** } *password*	可选
设置指定用户的状态	**state** { **active** | **block** }	可选 缺省情况下，当一个本地用户被创建以后，其状态为 **active**，允许该用户请求网络服务
设置本地用户名可容纳的最大接入用户数	**access-limit** *max-user-number*	可选 缺省情况下，不限制当前本地用户名可容纳的接入用户数
设置用户可以使用的服务类型	**service-type** { { **ftp**| { **ssh** | **telnet** | **terminal** } * | **portal** | **ppp**] }	可选 缺省情况下，系统不对本地用户授权任何服务
设置本地用户的绑定属性	**bind-attribute** { **call-number** *call-number* [: *subcall-number*] | **ip** *ip-address* | **location port** *slot-number subslot-number port-number* | **mac** *mac-address* | **vlan** *vlan-id* } *	可选 缺省情况下，未设置本地用户的任何绑定属性
设置本地用户的授权属性	**authorization-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	可选 缺省情况下，未设置本地用户的任何授权属性
设置本地用户的有效期	**expiration-date** *time*	可选 缺省情况下，未设置本地用户的有效期 有用户临时需要接入网络时，可以建立来宾帐户来管理用户的临时访问，并通过该命令对用户帐户的有效期进行控制
设置本地用户所属的用户组	**group** *group-name*	可选 缺省情况下，本地用户属于系统默认用户组 system

说明

- 支持 Password Control 特性的设备上不显示本地用户密码，而且 **local-user password-display-mode** 命令也不生效。
- 当采用 **local-user password-display-mode cipher-force** 命令后，即使用户通过 **password** 命令指定密码明文显示（即 **simple** 方式），密码也会显示为密文。同样，若强制用户密码密文显示并使用 **save** 命令保存当前配置，重启设备后，即使恢复为 **auto** 方式，原来配置为明文显示的密码仍然显示为密文。
- 本地用户的 **access-limit** 命令只在配置了本地计费方案的情况下生效。
- 本地用户的 **service-type** 是本地认证的检测项。如果没有用户可以使用的服务类型，则该用户无法正常认证通过。
- 如果配置登录设备的认证方式需要用户名和密码（包括本地认证、RADIUS 认证及 HWTACACS 认证），则用户登录系统后所能访问的命令级别由用户的级别确定，其它认证方式下则由用户界面所能访问的命令级别确定。对于 SSH 用户，使用 RSA 公钥认证时，其所能使用的命令以用户界面上设置的级别为准。关于登录设备的认证方式与用户界面所能访问的命令级别的详细介绍请参见“系统分册”中的“用户界面配置”。
- 配置的绑定属性是本地用户进行认证时需要检测的属性，如果检测不通过，则认证失败，因此在配置绑定属性时要考虑该用户是否需要绑定某些属性。
- 可配置的授权属性都有其明确使用环境和用途，因此配置下发的授权属性时要考虑该用户是否需要某些属性。例如，PPP 接入用户不需要授权的目录，因此就不要设置 PPP 用户的 **work-directory** 属性。

1.6.8 配置用户组的属性

为了简化本地用户的配置，增强本地用户的可管理性，引入了用户组的概念。用户组是一个本地用户属性的集合，某些需要集中管理属性可在用户组中统一配置和管理，用户组内的所有本地用户都可以继承这些属性。目前，用户组中可以管理的内容为用户的授权属性。

每个新增的本地用户都默认属于一个系统自动创建的用户组 system，且继承该组的所有属性。本地用户所属的用户组可以通过使用本地用户视图下的 **group** 命令来修改。

表1-13 配置用户组的属性

操作	命令	说明
进入系统视图	**system-view**	-
创建用户组，并进入用户组视图	**user-group** *group-name*	必选
设置用户组的授权属性	**authorization-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	可选 缺省情况下，未设置用户组的授权属性

说明

设备暂不支持本地用户密码的控制策略属性。

1.6.9 配置强制切断用户连接

表1-14 配置强制切断用户连接

操作		命令	说明
进入系统视图		**system-view**	-
强制切断指定AAA用户的连接	集中式设备	**cut connection** { **access-type portal** | **all** | **domain** *isp-name* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name* }	必选 此命令目前只对PPP、Portal服务类型的用户有效
	分部试设备	**cut connection** { **access-type portal** | **all** | **domain** *isp-name* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name* } [**slot** *slot-number*]	

1.6.10 AAA 显示和维护

完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 AAA 的运行情况，通过查看显示信息验证配置的效果。

表1-15 AAA 显示和维护

操作	命令
显示所有或指定 ISP 域的配置信息	**display domain** [*isp-name*]
显示 AAA 用户连接的相关信息（集中式设备）	**display connection** [**access-type portal** | **domain** *isp-name* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name*]
显示 AAA 用户连接的相关信息（分布式设备）	**display connection** [**access-type portal** | **domain** *isp-name* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name*] [**slot** *slot-number*]
显示本地用户相关信息(集中式设备)	**display local-user** [**service-type** { **ftp** | **portal** | **ppp** | **ssh** | **telnet** | **terminal** } | **state** { **active** | **block** } | **user-name** *user-name*]
显示本地用户相关信息（分布式设备）	**display local-user** [**service-type** { **ftp** | **portal** | **ppp** | **ssh** | **telnet** | **terminal** } | **state** { **active** | **block** } | **user-name** *user-name*] [**slot** *slot-number*]
显示本地用户组的相关信息	**display user-group** [*group-name*]

1.7 配置 RADIUS

RADIUS 配置是以 RADIUS 方案为单位进行的。当创建一个新的 RADIUS 方案之后，需要对属于此方案的 RADIUS 服务器的 IP 地址和 UDP 端口号进行设置，这些服务器包括认证/授权和计费服务器，而每种服务器又有主服务器和从服务器的区别。每个 RADIUS 方案的属性包括：主服务器的 IP 地址、从服务器的 IP 地址、共享密钥以及 RADIUS 服务器类型等。

RADIUS配置仅仅定义了设备和RADIUS服务器之间进行信息交互所必须的一些参数。为了使这些参数能够生效，还必须在某个ISP域视图下指定该域引用的RADIUS方案。具体配置细节，请参见“1.6 配置AAA”。

说明

对于 RADIUS，在有用户在线的情况下，除报文重传次数和定时器的之外的配置均不能修改。

1.7.1 创建 RADIUS 方案

在进行 RADIUS 的其它配置之前，必须先创建 RADIUS 方案并进入其视图。

表1-16 创建 RADIUS 方案

操作	命令	说明
进入系统视图	**system-view**	-
创建 RADIUS 方案并进入其视图	**radius scheme** *radius-scheme-name*	必选 缺省情况下，未定义 RADIUS 方案

说明

一个 RADIUS 方案可以同时被多个 ISP 域引用。

1.7.2 配置 RADIUS 认证/授权服务器

表1-17 配置 RADIUS 认证/授权服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
配置主 RADIUS 认证/授权服务器	**primary authentication** { *ip-address* | **ipv6** *ipv6-address* } [*port-number*]	二者必选其一 缺省情况下，未配置主认证/授权服务器和从认证/授权服务器 两者同时配置的情况下，当主服务器不可达时，设备将使用从服务器进行认证
配置从 RADIUS 认证/授权	**secondary authentication** { *ip-address* | **ipv6** *ipv6-address* } [*port-number*]	

说明

- 建议在不需要备份的情况下，只配置主 RADIUS 认证/授权服务器即可
- 在实际组网环境中，可以指定两台 RADIUS 服务器分别作为主、从认证/授权服务器；也可以指定一台服务器既作为某个方案的主认证/授权服务器，又作为另一个方案的从认证/授权服务器。
- 在同一个方案中指定的主认证/授权服务器和从认证/授权服务器的 IP 地址不能相同，否则将提示配置不成功。
- 主服务器和从服务器的 IP 地址版本必须保持一致；认证/授权服务器和计费服务器的 IP 地址版本也必须保持一致。

1.7.3 配置 RADIUS 计费服务器及相关参数

表1-18 配置 RADIUS 计费服务器及相关参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
配置主 RADIUS 计费服务器	**primary accounting** { *ip-address* | **ipv6** *ipv6-address* } [*port-number*]	二者必选其一 缺省情况下，未配置主/从计费服务器 两者同时配置的情况下，当主服务器不可达时，设备将使用从服务器进行计费
配置从 RADIUS 计费服务器	**secondary accounting** { *ip-address* | **ipv6** *ipv6-address* } [*port-number*]	
使能停止计费报文缓存功能	**stop-accounting-buffer enable**	可选 缺省情况下，使能停止计费报文缓存功能
设置当出现没有得到响应的停止计费请求时，将该报文存入设备缓存后，允许停止计费请求无响应的最大次数	**retry stop-accounting** *retry-times*	可选 缺省情况下，允许停止计费请求无响应的最大次数为 500
设置允许实时计费请求无响应的最大次数	**retry realtime-accounting** *retry-times*	可选 缺省情况下，最多允许 5 次实时计费请求无响应

说明

- 建议在不需要备份的情况下，只配置主 RADIUS 计费服务器即可。
- 在实际组网环境中，可以指定两台 RADIUS 服务器分别作为主、从计费服务器；也可以指定一台服务器既作为某个方案的主计费服务器，又作为另一个方案的从计费服务器。此外，由于 RADIUS 协议采用不同的 UDP 端口来收发认证/授权和计费报文，因此必须将认证/授权端口号和计费端口号配置得不同。
- 如果 RADIUS 计费服务器对设备发出的停止计费请求报文没有响应，且设备使能了停止计费报文缓存功能，设备应将其缓存在本机上，然后重新发送直到 RADIUS 计费服务器产生响应，或者在重新发送的次数达到指定的次数限制后将其丢弃。
- 设备提供对连续实时计费请求无响应次数限制的设置——在设备向 RADIUS 服务器发出的实时计费请求没有得到响应的次数超过所设定的限度时，设备将切断用户连接。
- 主计费服务器和从计费服务器的 IP 地址不能相同，否则将提示配置不成功。
- 目前 RADIUS 不支持对 FTP 用户进行计费。
- 主服务器和从服务器的 IP 地址版本必须保持一致；计费服务器和认证/授权服务器的 IP 地址版本也必须保持一致。

1.7.4 配置 RADIUS 报文的共享密钥

RADIUS 客户端与 RADIUS 服务器使用 MD5 算法来加密 RADIUS 报文，双方通过设置共享密钥来验证报文的合法性。只有在密钥一致的情况下，彼此才能接收对方发来的报文并作出响应。

表1-19 配置 RADIUS 报文的共享密钥

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
配置 RADIUS 认证/授权或计费报文的共享密钥	**key** { **accounting** | **authentication** } *string*	必选 缺省情况下，无密钥



必须保证设备上设置的共享密钥与 RADIUS 服务器上的完全一致。

1.7.5 配置 RADIUS 报文的超时重传次数最大值

由于 RADIUS 协议采用 UDP 报文来承载数据，因此其通信过程是不可靠的。如果 RADIUS 服务器在应答超时定时器规定的时长内没有响应设备，则设备有必要向 RADIUS 服务器重传 RADIUS 请求报文。如果累计的传送次数超过最大传送重传次数而 RADIUS 服务器仍旧没有响应，则设备将认为本次认证失败。

表1-20 配置 RADIUS 报文的超时重传次数最大值

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案并进入其视图	**radius scheme** *radius-scheme-name*	-
设置 RADIUS 报文超时重传次数的最大值	**retry** *retry-times*	可选 缺省情况下，RADIUS 报文超时重传次数的最大值为 3 次



- RADIUS 报文超时重传次数的最大值与 RADIUS 服务器应答超时时间的乘积不能超过 75 秒。
- RADIUS 服务器应答超时时间的配置请参考命令 **timer response-timeout**。

1.7.6 配置支持的 RADIUS 服务器的类型

表1-21 配置支持的 RADIUS 服务器的类型

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
设置支持何种类型的 RADIUS 服务器	**server-type** { **extended** | **standard** }	可选 缺省情况下，设备支持的 RADIUS 服务器类型为 **standard**

说明

- 当设备系统支持的 RADIUS 服务器类型的配置被更改时，会将原有的发送到 RADIUS 服务器的数据流的单位恢复为缺省情况。
- 使用第三方 RADIUS 服务器时，RADIUS 服务器类型可以选择 **standard** 类型或 **extended** 类型；使用 CAMS/iMC 服务器时，RADIUS 服务器类型应选择 **extended** 类型。

1.7.7 配置 RADIUS 服务器的状态

对于某个 RADIUS 方案中的主、从服务器（无论是认证/授权服务器还是计费服务器），当主服务器与设备的通信中断时，设备会主动地转而与从服务器交互报文。在主备服务器均可达的条件下，设备遵从主服务器优先的原则发送请求报文。

当主服务器不可达时，状态变为 **block**，设备与已配置了 IP 地址的从服务器交互。

- 若从服务器可达，设备将开启超时定时器，在 **timer quiet** 设定的时间达到后主服务器状态立即恢复为 **active**，从服务器状态不变；
- 若从服务器不可达，设备立即将主服务器状态恢复为 **active**。

之后，如果主服务器恢复正常，设备会立即恢复与其通信，而中断与从服务器通信。而计费开始后，客户端与从计费服务器之间的通信不会因为主计费服务器的恢复而切换。

表1-22 配置 RADIUS 服务器的状态

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
设置主 RADIUS 认证/授权服务器的状态	**state primary authentication { active | block }**	可选 缺省情况下，RADIUS 方案中配置了 IP 地址的各 RADIUS 服务器的状态均为 **active**
设置主 RADIUS 计费服务器的状态	**state primary accounting { active | block }**	
设置从 RADIUS 认证/授权服务器的状态	**state secondary authentication { active | block }**	
设置从 RADIUS 计费服务器的状态	**state secondary accounting { active | block }**	

说明

- 当主服务器与从服务器的状态都为 **block** 时，若希望使用从服务器进行认证，必须将从服务器的状态手工配置为 **active**，否则无法完成主从服务器的切换。
- 在一个服务器状态为 **active**、另外一个服务器状态为 **block** 的情况下，即使状态为 **active** 的服务器不可达，设备也不会进行主从服务器的切换。
- **state** 命令设置的服务器状态属于动态信息，不能被保存在配置文件中，设备重启后，服务器状态恢复为缺省状态 **active**。

1.7.8 配置发送给 RADIUS 服务器的数据相关属性

表1-23 配置发送给 RADIUS 服务器的数据相关属性

操作		命令	说明
进入系统视图		**system-view**	-
使能 RADIUS trap 功能		**radius trap { accounting-server-down | authentication-server-down }**	可选 缺省情况下，RADIUS trap 功能关闭
进入 RADIUS 方案视图		**radius scheme** *radius-scheme-name*	-
设置发送给 RADIUS 服务器的用户名格式		**user-name-format { keep-original | with-domain | without-domain }**	可选 缺省情况下，设备发送给 RADIUS 服务器的用户名携带有 ISP 域名
设置发送给 RADIUS 服务器的数据流的单位		**data-flow-format { data { byte | giga-byte | kilo-byte | mega-byte } | packet { giga-packet | kilo-packet | mega-packet | one-packet } } ***	可选 缺省情况下，数据的单位为 **byte**，数据包的单位为 **one-packet**
设置设备发送 RADIUS 报文使用的源地址	RADIUS 方案视图	**nas-ip** { *ip-address* | **ipv6** *ipv6-address* }	两者可选其一 缺省情况下，RADIUS 方案使用系统视图下指定的源地址；系统中未指定发送 RADIUS 报文使用的源地址，即以发送报文的接口地址作为源地址
	系统视图	**quit**	
		radius nas-ip { *ip-address* | **ipv6** *ipv6-address* }	

说明

- 由于有些较早期的 RADIUS 服务器不能接受携带有 ISP 域名的用户名，因此必需将用户名的域名去除后再传送给 RADIUS 服务器。可以通过命令 **user-name-format without-domain** 不携带有 ISP 域名。
- 如果指定某个 RADIUS 方案不允许用户名中携带有 ISP 域名，那么请不要在两个乃至两个以上的 ISP 域中同时设置使用该 RADIUS 方案，否则，会出现虽然实际用户不同（在不同的 ISP 域中）、但 RADIUS 服务器认为用户相同（因为传送到它的用户名相同）的错误。
- 设备向 RADIUS 服务器发送数据流的单位应与 RADIUS 服务器上的流量统计单位保持一致，否则无法正确计费。
- RADIUS 方案视图下的命令 **nas-ip** 只对本 RADIUS 方案有效，系统视图下的命令 **radius nas-ip** 对所有 RADIUS 方案有效。RADIUS 方案视图下的设置具有更高的优先级。
- 设备发送 RADIUS 报文使用的源地址的 IP 地址版本必须与当前使用的 RADIUS 服务器的 IP 地址的版本保持一致，否则该源地址的配置不能生效。

1.7.9 配置 RADIUS 服务器的定时器

在与 RADIUS 服务器交互的过程中，设备上可启动的定时器主要有以下几种：

- 服务器响应超时定时器（**response-timeout**）：如果在 RADIUS 请求报文（认证/授权请求或计费请求）传送出去一段时间后，设备还没有得到 RADIUS 服务器的响应，则有必要重传 RADIUS 请求报文，以保证用户确实能够得到 RADIUS 服务，这段时间被称为 RADIUS 服务器响应超时时长。
- 主服务器恢复激活状态定时器（**timer quiet**）：当主服务器不可达时，状态变为 **block**，设备会与已配置了 IP 地址的从服务器交互。若从服务器可达，设备与从服务器通信，并开启超时

定时器，在设定的一定时间间隔之后，将主服务器的状态恢复为 **active**，并尝试与主服务器交互，同时保持从服务器的状态不变。若主服务器功能恢复正常，设备会立即恢复与其通信，而中断与从服务器通信。这段时间被称为 RADIUS 主服务器恢复激活状态时长。

- 实时计费间隔定时器（**realtime-accounting**）：为了对用户实施实时计费，有必要设置实时计费的时间间隔。在设置了该属性以后，每隔设定的时间，交换机会向 RADIUS 服务器发送一次在线用户的计费信息。

表1-24 设置 RADIUS 服务器的定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
设置 RADIUS 服务器应答超时时间	**timer response-timeout** *seconds*	可选 缺省情况下，RADIUS 服务器应答超时定时器为 3 秒
设置主服务器恢复激活状态的时间	**timer quiet** *minutes*	可选 缺省情况下，主服务器恢复激活状态前需要等待 5 分钟
设置实时计费间隔	**timer realtime-accounting** *minutes*	可选 缺省情况下，实时计费间隔为 12 分钟 为了对用户实施实时计费，有必要设置实时计费的时间间隔。在设置了该属性以后，每隔设定的时间，设备会向 RADIUS 服务器发送一次在线用户的计费信息

说明

- RADIUS 各类报文重传次数的最大值与 RADIUS 服务器应答超时时间的乘积不能超过 75 秒。此乘积具体为不同接入模块超时时间的上限。
- 不同接入模块配置的 RADIUS 服务器应答超时时间与请求报文的最大传送次数乘积不能大于等于接入模块的超时时间。比如对于语音接入，因为其超时时间为 10 秒，所以应答超时时间与请求报文最大传送次数的乘积不能大于等于 10 秒；对于 Telnet 接入，其超时时间为 30 秒，所以此乘积不能大于等于 30 秒。否则会造成计费停止报文不能被缓存，以及主备服务器不能切换的问题。具体接入模块超时时间请参考相关接入手册。
- RADIUS 各类报文重传次数的最大值配置请参考命令 **retry**。

1.7.10 配置 RADIUS 的 accouting-on 功能

说明

本特性只有集中式设备支持，分布式设备不支持。

accounting-on 功能用于设备重启后，通知 RADIUS 服务器该设备已经重启，要求 RADIUS 服务器强制该设备的用户下线，用来解决设备重启后，重启前原在线用户无法再次登录的问题

由于设备重启动并且配置恢复完成后会立即执行 accounting-on 功能，如果多个方案（目前设备上支持最多 16 个）中配置的 RADIUS 服务器多数不响应，或设置的 accounting-on 报文重发次数太大、重发时间间隔太长，则设备需要等待较长的时间来处理 accounting-on 回应报文和重发

accounting-on 报文，直到所配置的方案全被置为 accounting-on 已响应后才会处理用户 AAA 业务。这样会对用户的接入时间有所影响。

表1-25 配置 RADIUS 的 accouting-on 功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
使能 accouting-on 功能	**accouting-on enable**	必选 缺省情况下，accouting-on 功能处于关闭状态
设置 accouting-on 报文重发次数	**accouting-on enable send** *send-times*	可选 缺省情况下，accouting-on 报文重发次数为 5 次
设置 accouting-on 报文重发时间间隔	**accouting-on enable interval** *seconds*	可选 缺省情况下，accouting-on 报文重发时间间隔为 3 秒



说明

accounting-on 功能需要和 H3C CAMS/iMC 网管系统配合使用。

1.7.11 配置 RADIUS 服务器的安全策略服务器

EAD 方案的核心是整合与联动，其中的安全策略服务器是 EAD 方案中的管理与控制中心，它作为一个软件的集合可运行在 Windows、Linux 平台下，兼具用户管理、安全策略管理、安全状态评估、安全联动控制以及安全事件审计等功能。该配置用来设定该安全策略服务器的 IP 地址。

表1-26 设置 RADIUS 服务器的安全策略服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入 RADIUS 方案视图	**radius scheme** *radius-scheme-name*	-
设置 RADIUS 服务器的安全策略服务器	**security-policy-server** *ip-address*	可选 缺省情况下，未指定 RADIUS 服务器的安全策略服务器



警告

- 当接入设备上有多个端口口配置了用户接入认证功能时，各接口所对应的安全策略服务器可能不同，因此一个 RADIUS 方案中可以配置多个安全策略服务器 IP 地址，最多不能超过 8 个。
- 如果安全策略服务器和 RADIUS 服务器为同一台服务器时，安全策略服务器 IP 地址也可以不用设置。
- 指定的安全策略服务器必须为通过配置的安全策略服务器或 RADIUS 服务器，否则，设备认为其非法。

1.7.12 使能 RADIUS 客户端的监听端口

通过以下配置可以打开 RADIUS 客户端的 RADIUS 报文监听端口，使能后的端口可以接收和发送 RADIUS 报文。

表1-27 使能 RADIUS 客户端的监听端口

操作	命令	说明
进入系统视图	**system-view**	-
使能 RADIUS 客户端的监听端口	**radius client enable**	可选 缺省情况下，监听端口处于使能状态

1.7.13 RADIUS 显示和维护

完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 RADIUS 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除相关统计信息。

表1-28 RADIUS 显示和维护

操作	命令
显示所有或指定 RADIUS 方案的配置信息（集中式设备）	**display radius scheme** [*radius-scheme-name*]
显示所有或指定 RADIUS 方案的配置信息（分布式设备）	**display radius scheme** [*radius-scheme-name*] [**slot** *slot-number*]
显示 RADIUS 报文的统计信息（集中式设备）	**display radius statistics**
显示 RADIUS 报文的统计信息（分布式设备）	**display radius statistics** [**slot** *slot-number*]
显示缓存的没有得到响应的停止计费请求报文（集中式设备）	**display stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* }
显示缓存的没有得到响应的停止计费请求报文（分布式设备）	**display stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]
清除 RADIUS 协议的统计信息（集中式设备）	**reset radius statistics**
清除 RADIUS 协议的统计信息（分布式设备）	**reset radius statistics** [**slot** *slot-number*]
清除缓存中的没有得到响应的停止计费请求报文（集中式设备）	**reset stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* }
清除缓存中的没有得到响应的停止计费请求报文（分布式设备）	**reset stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]

1.8 配置 HWTACACS

说明

与 RADIUS 不同，对于 HWTACACS，除删除 HWTACACS 方案和修改 HWTACACS 服务器 IP 地址之外，其它配置改变时均不检查当前是否有用户在线。

1.8.1 创建 HWTACACS 方案

HWTACACS 的配置是以 HWTACACS 方案为单位进行的。在进行 HWTACACS 的其他相关配置之前，必须先创建 HWTACACS 方案并进入其视图。

表1-29 创建 HWTACACS 方案

操作	命令	说明
进入系统视图	**system-view**	-
创建 HWTACACS 方案并进入其视图	**hwtacacs scheme** *hwtacacs-scheme-name*	必选 缺省情况下，未定义 HWTACACS 方案

说明

系统最多支持配置 16 个 HWTACACS 方案。只有未被引用的方案才可以被删除。

1.8.2 配置 HWTACACS 认证服务器

表1-30 配置 HWTACACS 认证服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入 HWTACACS 方案视图	**hwtacacs scheme** *hwtacacs-scheme-name*	-
配置主 HWTACACS 认证服务器	**primary authentication** *ip-address* [*port-number*]	二者必选其一 缺省情况下，未配置主/从认证服务器 两者同时配置的情况下，当主服务器不可达时，设备将使用从服务器进行认证
配置从 HWTACACS 认证服务器	**secondary authentication** *ip-address* [*port-number*]	

说明

- 建议在不需要备份的情况下，只配置主认证服务器即可。
- 主认证服务器和从认证服务器的 IP 地址不能相同，否则将提示配置不成功。
- 只有当没有活跃的用于发送认证报文的 TCP 连接使用该认证服务器时，才允许删除该服务器。

1.8.3 配置 HWTACACS 授权服务器

表1-31 配置 HWTACACS 授权服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入 HWTACACS 方案视图	**hwtacacs scheme** *hwtacacs-scheme-name*	-
设置主 HWTACACS 授权服务器	**primary authorization** *ip-address* [*port-number*]	二者必选其一 缺省情况下，未配置主/从授权服务器 两者同时配置的情况下，当主服务器不可达时，设备将使用从服务器进行授权
设置从 HWTACACS 授权服务器	**secondary authorization** *ip-address* [*port-number*]	

说明

- 建议在不需要备份的情况下，只配置主授权服务器即可。
- 主授权服务器和从授权服务器的 IP 地址不能相同，否则将提示配置不成功。
- 只有当没有活跃的用于发送授权报文的 TCP 连接使用该授权服务器时，才允许删除该服务器。

1.8.4 配置 HWTACACS 计费服务器

表1-32 配置 HWTACACS 计费服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入 HWTACACS 方案视图	**hwtacacs scheme** *hwtacacs-scheme-name*	-
设置 HWTACACS 主计费服务器	**primary accounting** *ip-address* [*port-number*]	二者必选其一 缺省情况下，未配置主计费服务器 两者同时配置的情况下，当主服务器不可达时，设备将使用从服务器进行计费
设置 HWTACACS 从计费服务器	**secondary accounting** *ip-address* [*port-number*]	
使能停止计费报文缓存功能	**stop-accounting-buffer enable**	可选 缺省情况下，使能停止计费报文缓存功能
设置当出现没有得到响应的停止计费请求时，将该报文存入设备缓存后，停止计费请求报文的最大发送次数	**retry stop-accounting** *retry-times*	可选 缺省情况下，*retry-times* 的值为 100

说明

- 建议在不需要备份的情况下，只配置主计费服务器即可。
- 主计费服务器和从计费服务器的 IP 地址不能相同，否则将提示配置不成功。
- 只有当没有活跃的用于发送计费报文的 TCP 连接使用该计费服务器时，才允许删除该服务器。
- 目前 HWTACACS 不支持对 FTP 用户进行计费。

1.8.5 配置 HWTACACS 报文的共享密钥

使用 HWTACACS 服务器作为 AAA 服务器时，可设置密钥以提高设备与 HWTACACS 服务器通信的安全性。

HWTACACS 客户端（即设备系统）与 HWTACACS 服务器使用 MD5 算法来加密交互的 HWTACACS 报文，双方通过设置共享密钥来验证报文的合法性。只有在密钥一致的情况下，双方才能彼此接收对方发来的报文并作出响应。因此，必须保证设备上设置的共享密钥与 HWTACACS 服务器上的完全一样。

表1-33 配置 HWTACACS 报文的共享密钥

操作	命令	说明
进入系统视图	**system-view**	-
进入 HWTACACS 方案视图	**hwtacacs scheme** *hwtacacs-scheme-name*	-
配置 HWTACACS 认证、授权计费报文的共享密钥	**key { accounting | authentication | authorization }** *string*	必选 缺省情况下，无密钥

1.8.6 配置发送给 HWTACACS 服务器的数据相关属性

表1-34 配置发送给 HWTACACS 服务器的数据相关属性

操作		命令	说明
进入系统视图		**system-view**	-
进入 HWTACACS 方案，并进入 HWTACACS 视图		**hwtacacs scheme** *hwtacacs-scheme-name*	必选 缺省情况下，未定义 HWTACACS 方案
设置发送给 HWTACACS 服务器的用户名格式		**user-name-format { keep-original | with-domain | without-domain }**	可选 缺省情况下，发往 HWTACACS 服务器的用户名带域名
设置发送给 HWTACACS 服务器的数据流的单位		**data-flow-format { data { byte | giga-byte | kilo-byte | mega-byte } | packet { giga-packet | kilo-packet | mega-packet | one-packet } }** *	可选 缺省情况下，数据的单位为 **byte**，数据包的单位为 **one-packet**
配置发送 HWTACACS 报文使用的源地址	HWTACACS 视图	**nas-ip** *ip-address*	二者可选其一 缺省情况下，不指定源地址，即以发送报文的接口地址作为源地址
	系统视图	**quit**	
		hwtacacs nas-ip *ip-address*	



说明

- 如果 HWTACACS 服务器不接受带域名的用户名，可以配置将用户名的域名去除后再传送给 HWTACACS 服务器。
- HWTACACS 方案视图下的命令 **nas-ip** 只对本 HWTACACS 方案有效，系统视图下的命令 **hwtacacs nas-ip** 对所有 HWTACACS 方案有效。HWTACACS 方案视图下的设置具有更高的优先级。

1.8.7 配置 HWTACACS 服务器的定时器

表1-35 配置 HWTACACS 服务器的定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 HWTACACS 方案视图	**hwtacacs scheme** *hwtacacs-scheme-name*	-
设置 HWTACACS 服务器应答超时时间	**timer response-timeout** *seconds*	可选 缺省情况下，应答超时时间为 5 秒
设置主服务器恢复激活状态的时间	**timer quiet** *minutes*	可选 缺省情况下，主服务器恢复激活状态前需要等待 5 分钟
设置实时计费的时间间隔	**timer realtime-accounting** *minutes*	可选 缺省情况下，实时计费间隔为 12 分钟

说明

- 为了对用户实施实时计费，有必要设置实时计费的时间间隔。在设置了该属性以后，每隔设定的时间，设备会向 HWTACACS 服务器发送一次在线用户的计费信息。按照协议，如果服务器对实时计费报文没有正常响应，设备也不会强制切断在线用户。
- 实时计费间隔时间取值必须为 3 的整数倍。
- 实时计费间隔的取值对设备和 HWTACACS 服务器的性能有一定的相关性要求，取值越小，对设备和 HWTACACS 服务器的性能要求越高。

1.8.8 HWTACACS 显示和维护

完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 HWTACACS 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除相关统计信息。

表1-36 HWTACACS 协议显示和维护

操作	命令
查看所有或指定 HWTACACS 方案的配置信息或统计信息（集中式设备）	**display hwtacacs** [*hwtacacs-server-name* [**statistics**]]
查看所有或指定 HWTACACS 方案的配置信息或统计信息（分布式设备）	**display hwtacacs** [*hwtacacs-server-name* [**statistics**]] [**slot** *slot-number*]
显示缓存的没有得到响应的停止计费请求报文（集中式设备）	**display stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name*
显示缓存的没有得到响应的停止计费请求报文（分布式设备）	**display stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name* [**slot** *slot-number*]
清除 HWTACACS 协议的统计信息（集中式设备）	**reset hwtacacs statistics** { **accounting** | **all** | **authentication** | **authorization** }
清除 HWTACACS 协议的统计信息（分布式设备）	**reset hwtacacs statistics** { **accounting** | **all** | **authentication** | **authorization** } [**slot** *slot-number*]

操作	命令
清除缓存中的没有得到响应的停止计费请求报文（集中式设备）	**reset stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name*
清除缓存中的没有得到响应的停止计费请求报文（分布式设备）	**reset stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name* [**slot** *slot-number*]

1.9 AAA 典型配置举例

1.9.1 Telnet/SSH 用户通过 RADIUS 服务器认证、授权、计费的应用配置

说明

SSH 用户和 Telnet 用户通过 RADIUS 服务器进行认证、授权、计费的配置方法类似，下面的描述以 Telnet 用户的远端认证为例。

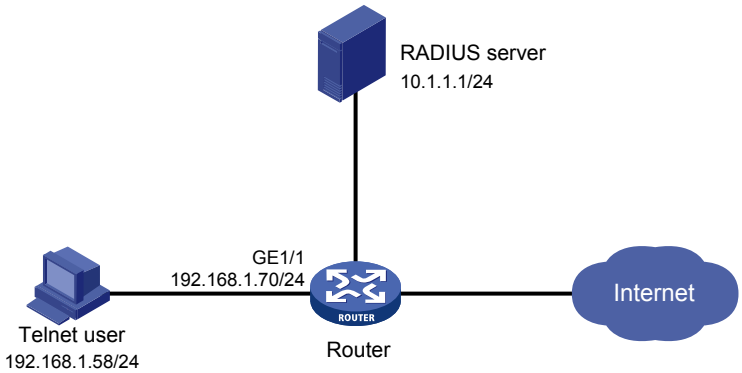
1. 组网需求

如 图 1-7所示，配置Router实现RADIUS服务器对登录Router的Telnet用户进行认证、授权和计费。

- 一台 RADIUS 服务器（其担当认证 RADIUS 服务器和计费 RADIUS 服务器的职责）与 Router 相连，服务器 IP 地址为 10.1.1.1。
- Router 与认证、计费 RADIUS 服务器交互报文时的共享密钥均为 expert，发送给 RADIUS 服务器的用户名带域名。
- RADIUS 服务器使用 CAMS 服务器。

2. 组网图

图1-7 Telnet 用户远端 RADIUS 认证、授权和计费配置组网图



3. 配置步骤

(1) 配置 RADIUS server。

说明

本文以 CAMS 服务器 V2.10 为例，说明该例中 RADIUS server 的基本配置。

增加接入设备。

登录进入 CAMS 管理平台，点击左侧菜单树中[系统管理]->[系统配置]的“接入设备配置”->“修改”->“增加”后，进入接入设备配置页面。

- 添加 Router 的 IP 地址 192.168.1.70；
- 设置与 Router 交互报文时的认证、计费共享密钥为 expert；
- 选择协议类型为 LAN 接入业务；
- 设置验证及计费的端口号分别为 1812 和 1813；
- 选择 RADIUS 协议类型为扩展协议；
- 选择 RADIUS 报文类型为标准报文。

图1-8 接入设备配置页面



增加设备管理用户。

点击左侧菜单树中[用户管理]--> [设备管理用户]的 “增加”后，进入设备管理用户配置页面。

- 添加用户名 hello@bbb 和密码；
- 选择服务类型为 Telnet；
- 添加所管理主机 IP 地址范围“192.168.1.0～192.168.1.255”。

图1-9 设备管理用户配置页面



(2) 配置 Router

配置接口 GigabitEthernet1/1 的 IP 地址，Telnet 用户将通过该地址连接 Router。

```
<Router> system-view
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.1.70 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

开启 Router 的 Telnet 服务器功能。

```
 [Router] telnet server enable
```

配置 Telnet 用户登录采用 AAA 认证方式。

```
[Router] user-interface vty 0 4
[Router-ui-vty0-4] authentication-mode scheme
[Router-ui-vty0-4] quit
```

配置 RADIUS 方案。

```
[Router] radius scheme rad
[Router-radius-rad] primary authentication 10.1.1.1 1812
[Router-radius-rad] primary accounting 10.1.1.1 1813
[Router-radius-rad] key authentication expert
[Router-radius-rad] key accounting expert
[Router-radius-rad] server-type extended
[Router-radius-rad] user-name-format with-domain
[Router-radius-rad] quit
```

配置 ISP 域的 AAA 方案。在该应用配置中，因为接入用户对认证、授权和计费三个过程都需要，所以需要在域中对认证、授权和计费方案全部配置。

```
[Router] domain bbb
[Router-isp-bbb] authentication login radius-scheme rad
[Router-isp-bbb] authorization login radius-scheme rad
[Router-isp-bbb] accounting login radius-scheme rad
[Router-isp-bbb] quit
```

或者不区分用户类型，配置缺省的 AAA 方案。

```
[Router] domain bbb
```

```
[Router-isp-bbb] authentication default radius-scheme rad
[Router-isp-bbb] authorization default radius-scheme rad
[Router-isp-bbb] accounting default radius-scheme rad
```

使用 Telnet 登陆时输入用户名为 *userid*@bbb，以使用 bbb 域进行认证。

(3) Telnet 用户建立与 Router 的连接

在 Telnet 客户端按照提示输入用户名 hello@bbb 及密码，即可进入 Router 的用户界面。Telnet 用户登录系统后所能访问的命令级别由 CAMS 服务器授权，可通过设备管理用户界面的 EXEC 权限级别来设置。

1.9.2 FTP/Telnet 用户本地认证、授权、计费配置

说明

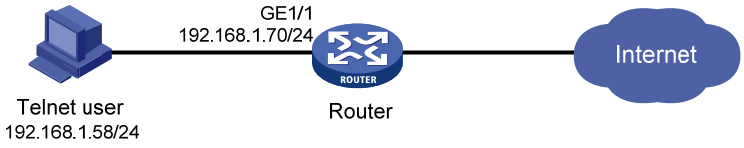
FTP 用户除了没有计费过程外，与 Telnet 用户通过本地认证、授权的配置方法类似，下面描述仅以 Telnet 用户为例。

1. 组网需求

如 图 1-10所示，配置Router实现对登录Router的Telnet用户进行本地认证、授权和计费。

2. 组网图

图1-10 Telnet 用户的本地认证、授权和计费配置组网图



3. 配置步骤

配置接口 GigabitEthernet1/1 的 IP 地址，Telnet 用户将通过该地址连接 Router。

```
<Router> system-view
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.1.70 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

开启设备的 Telnet 服务器功能。

```
[Router] telnet server enable
```

配置 Telnet 用户登录采用 AAA 认证方式。

```
[Router] user-interface vty 0 4
[Router-ui-vty0-4] aut hentication-mode scheme
[Router-ui-vty0-4] quit
```

创建本地用户 telnet。

```
[Router] local-user telnet
[Router-luser-telnet] service-type telnet
[Router-luser-telnet] password simple aabbcc
[Router-luser-telnet] quit
```

配置 ISP 域的 AAA 方案为本地认证、授权和计费。

```
[Router] domain system
[Router-isp-system] authentication login local
[Router-isp-system] authorization login local
[Router-isp-system] accounting login local
[Router-isp-system] quit
```

或者不区分用户类型，配置缺省的 AAA 方案。

```
[Router-isp-system] authentication default local
[Router-isp-system] authorization default local
[Router-isp-system] accounting default local
```

使用 Telnet 登陆时输入用户名为 *telnet@system*，以使用 system 域进行认证。

1.9.3 PPP 用户通过 HWTACACS 服务器认证、授权、计费的应用配置

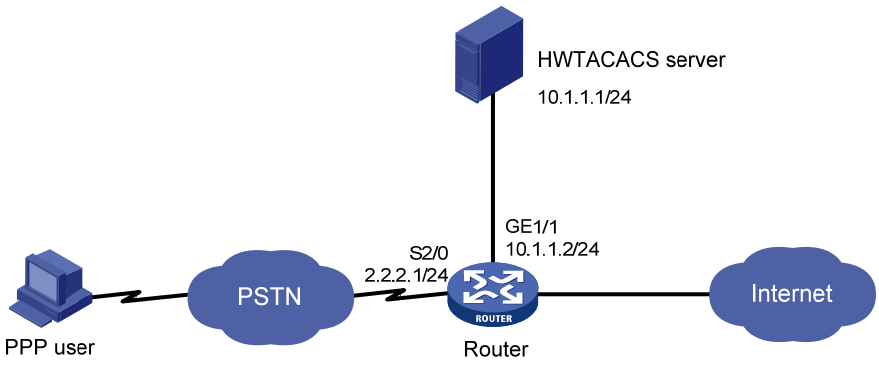
1. 组网需求

如 图 1-11所示，配置Router实现HWTACACS服务器对登录Router的PPP用户分配IP地址，并进行认证、授权、计费。

- 一台 HWTACACS 服务器（其担当认证、授权、计费服务器的职责）与 Router 相连，服务器 IP 地址为 10.1.1.1。
- Router 与认证、授权、计费 HWTACACS 服务器交互报文时的共享密钥均为 expert，发送给 HWTACACS 服务器的用户名中不带域名。
- 在 HWTACACS 服务器上设置与 Router 交互报文时的共享密钥为 expert；添加 PPP 用户名及密码。

2. 组网图

图1-11 PPP 用户远端 HWTACACS 认证、授权和计费配置组网图



3. 配置步骤

配置 HWTACACS 方案。

```
<Router> system-view
[Router] hwtacacs scheme hwtac
[Router-hwtacacs-hwtac] primary authentication 10.1.1.1 49
[Router-hwtacacs-hwtac] primary authorization 10.1.1.1 49
[Router-hwtacacs-hwtac] primary accounting 10.1.1.1 49
[Router-hwtacacs-hwtac] key authentication expert
```

```
[Router-hwtacacs-hwtac] key authorization expert
[Router-hwtacacs-hwtac] key accounting expert
[Router-hwtacacs-hwtac] user-name-format without-domain
[Router-hwtacacs-hwtac] quit
```

配置 ISP 的 AAA 方案。

```
[Router] domain bbb
[Router-isp-bbb] authentication ppp hwtacacs-scheme hwtac
[Router-isp-bbb] authorization ppp hwtacacs-scheme hwtac
[Router-isp-bbb] accounting ppp hwtacacs-scheme hwtac
[Router-isp-bbb] ip pool 1 200.1.1.1 200.1.1.99
[Router-isp-bbb] quit
```

或者不区分用户类型，配置缺省的 AAA 方案。

```
[Router] domain bbb
[Router-isp-bbb] authentication default hwtacacs-scheme hwtac
[Router-isp-bbb] authorization default hwtacacs-scheme hwtac
[Router-isp-bbb] accounting default hwtacacs-scheme hwtac
[Router-isp-bbb] ip pool 1 200.1.1.1 200.1.1.99
[Router-isp-bbb] quit
```

配置串口。

```
[Router] interface serial 2/0
[Router-Serial2/0] link-protocol ppp
[Router-Serial2/0] ppp authentication-mode pap domain bbb
[Router-Serial2/0] ip address 2.2.2.1 255.255.255.0
[Router-Serial2/0] remote address pool 1
[Router-Serial2/0] quit
```

配置以太网接口。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 10.1.1.2 255.255.255.0
```

1.10 AAA RADIUS HWTACACS 常见配置错误举例

1.10.1 RADIUS 认证/授权失败

1. 故障现象

用户认证/授权总是失败。

2. 故障分析

(1) 设备与 RADIUS 服务器之间存在通信故障。

(2) 用户名不是“*userid@isp-name*”的形式，或设备没有指定缺省的 ISP 域。

(3) RADIUS 服务器的数据库中没有配置该用户。

(4) 用户侧输入的密码不正确。

(5) RADIUS 服务器和设备的报文共享密钥不同。

3. 处理过程

(1) 使用 **ping** 命令检查设备与 RADIUS 服务器是否可达。

(2) 使用正确形式的用户名或在设备中设定缺省的 ISP 域。

(3) 检查 RADIUS 服务器的数据库以保证该用户的配置信息确实存在。

(4) 确保接入用户输入正确的密码。

(5) 检查两端的共享密钥，并确保两端一致。

1.10.2 RADIUS 报文传送失败

1. 故障现象

RADIUS 报文无法传送到 RADIUS 服务器。

2. 故障分析

(1) 设备与 RADIUS 服务器之间的通信线路不通（物理层/链路层）。

(2) 设备上没有设置相应的 RADIUS 服务器 IP 地址。

(3) 认证/授权和计费服务的 UDP 端口设置不正确。

(4) RADIUS 服务器的认证/授权和计费端口被其它应用程序占用。

3. 处理过程

(1) 确保线路通畅。

(2) 确保正确设置 RADIUS 服务器的 IP 地址。

(3) 确保与 RADIUS 服务器提供的端口号一致。

(4) 确保 RADIUS 服务器上的认证/授权和计费端口可用。

1.10.3 RADIUS 计费功能异常

1. 故障现象

用户认证通过并获得授权，但是计费功能出现异常。

2. 故障分析

(1) 计费端口号设置不正确。

(2) 计费服务器和认证/授权服务器不是同一台机器，设备却要求认证/授权和计费功能属于同一个服务器（IP 地址相同）。

3. 处理过程

(1) 正确设置 RADIUS 计费端口号。

(2) 确保设备的认证/授权和计费服务器的设置与实际情况相同。

1.10.4 HWTACACS 常见配置错误举例

HWTACACS 的常见配置错误举例与 RADIUS 基本相似，可以参考以上内容。

目 录

1 防火墙配置 1-1
1.1 防火墙简介 1-1
1.1.1 包过滤防火墙简介 1-1
1.1.2 ASPF简介 1-1
1.2 配置包过滤防火墙 1-3
1.2.1 包过滤防火墙配置任务简介 1-3
1.2.2 启用防火墙功能 1-4
1.2.3 配置缺省过滤方式 1-4
1.2.4 配置接口的报文过滤功能 1-5
1.2.5 包过滤防火墙显示和维护 1-6
1.2.6 包过滤防火墙典型配置举例 1-6
1.3 配置ASPF 1-8
1.3.1 ASPF配置任务简介 1-8
1.3.2 启用防火墙功能 1-8
1.3.3 配置ASPF策略 1-8
1.3.4 在接口上应用ASPF策略 1-8
1.3.5 配置端口映射 1-9
1.3.6 ASPF显示和维护 1-9
1.3.7 ASPF典型配置举例 1-9

1 防火墙配置

1.1 防火墙简介

防火墙一方面可以阻止来自因特网的、对受保护网络的未授权访问，另一方面允许内部网络用户对因特网进行 Web 访问或收发 E-mail 等。防火墙也可以作为一个访问因特网的权限控制关口，如允许组织内的特定主机可以访问因特网。现在，许多防火墙同时还具有一些其它特点，如进行身份鉴别、对信息进行安全（加密）处理等。

防火墙不单用于控制因特网连接，也可以用来在组织网络内部保护大型机和重要的资源（如数据）。对受保护数据的访问都必须经过防火墙的过滤，即使网络内部用户要访问受保护的数据，也要经过防火墙。

目前设备中的防火墙主要指以下三种：

- 包过滤防火墙，即基于 ACL（Access Control List，访问控制列表）的包过滤
- 状态防火墙，即 ASPF（Application Specific Packet Filter，基于应用层状态的包过滤）
- 地址转换



说明

关于地址转换的详细介绍，请参见“安全分册”中的“NAT 配置”，本章主要介绍包过滤防火墙和状态防火墙。

1.1.1 包过滤防火墙简介

包过滤实现了对 IP 数据包的过滤。对设备需要转发的数据包，先获取其包头信息（包括 IP 层所承载的上层协议的协议号、数据包的源地址、目的地址、源端口和目的端口等），然后与设定的 ACL 规则进行比较，根据比较的结果对数据包进行相应的处理。



说明

关于 ACL 的详细介绍，请参见“安全分册”中的“ACL 配置”

1.1.2 ASPF 简介

包过滤防火墙属于静态防火墙，目前存在的问题如下：

- 对于多通道的应用层协议（如 FTP 等），部分安全策略配置无法预知。
- 无法检测某些来自传输层和应用层的攻击行为（如 TCP SYN、Java Applets 等）。
- 无法识别来自网络中伪造的 ICMP 差错报文，从而无法避免 ICMP 的恶意攻击。
- 对于 TCP 连接均要求其首报文为 SYN 报文，非 SYN 报文的 TCP 首包将被丢弃。在这种处理方式下，当设备首次加入网络时，网络中原有 TCP 连接的非首包在经过新加入的防火墙设备时均被丢弃，这会中断已有的连接。

因此，提出了状态防火墙——ASPF 的概念。ASPF 能够实现的检测有：

- 应用层协议检测，包括 FTP、HTTP、SMTP、RTSP 检测；
- 传输层协议检测，包括 TCP 和 UDP 检测，即通用 TCP/UDP 检测。

1. ASPF 的功能

ASPF 的主要功能如下：

- 能够检查应用层协议信息，如报文的协议类型和端口号等信息，并且监控基于连接的应用层协议状态。对于所有连接，每一个连接状态信息都将被 ASPF 维护，并用于动态地决定数据包是否被允许通过防火墙进入内部网络，以阻止恶意的入侵。
- 能够检测传输层协议信息（即通用 TCP/UDP 检测），能够根据源、目的地址及端口号决定 TCP 或 UDP 报文是否可以通过防火墙进入内部网络。

ASPF 的其它功能有：

- 增强的会话日志功能。可以对所有的连接进行记录，包括：连接的时间、源地址、目的地址、使用的端口和传输的字节数。
- 支持 PAM（Port to Application Map，应用协议端口映射），允许用户自定义应用层协议使用非通用端口。
- 支持 ICMP 差错报文检测。正常 ICMP 差错报文中均携带有本报文对应连接的相关信息，根据这些信息可以匹配到相应的连接。如果匹配失败，则根据当前配置决定是否丢弃该 ICMP 报文。
- 支持 TCP 连接首包检测。对 TCP 连接的首报文进行检测，查看是否为 SYN 报文，如果不是 SYN 报文则根据当前配置决定是否丢弃该报文。

在网络边界，ASPF 和包过滤防火墙协同工作，能够为企业内部网络提供更全面的、更符合实际需求的安全策略。

2. ASPF 基本概念

(1) PAM

应用层协议使用通用的端口号进行通信，PAM 允许用户对不同的应用定义一组新的端口号，并提供了一些机制来维护和使用用户定义的端口配置信息。

PAM 支持两类映射机制：

- 通用端口映射：是将用户自定义端口号与应用层协议建立映射关系。例如：将 8080 端口映射为 HTTP 协议，这样所有目的端口是 8080 的 TCP 报文将被认为是 HTTP 报文。
- 主机端口映射：是对去往或来自某些特定主机的报文建立自定义端口号和应用协议的映射。例如：将目的地址为 10.110.0.0/16 网段的、使用 8080 端口的 TCP 报文映射为 HTTP 报文。主机的范围可由基本 ACL 指定。

(2) 单通道协议和多通道协议

- 单通道协议：从会话建立到删除的全过程中，只有一个通道参与数据交互，如 SMTP、HTTP。
- 多通道协议：包含一个控制通道和若干其它控制或数据通道，即控制信息的交互和数据的传送是在不同的通道上完成的，如 FTP、RTSP。

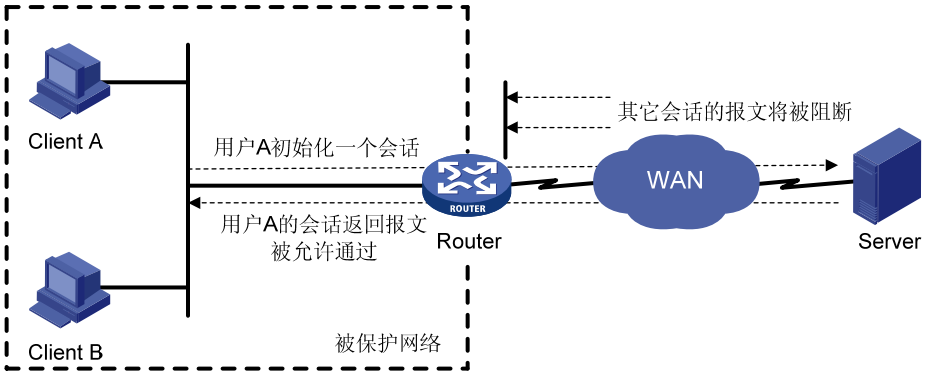
(3) 内部接口和外部接口

如果设备连接了内部网和互联网，并且设备要通过部署 ASPF 来保护内部网的服务器，则设备上与内部网连接的接口就是内部接口，与互联网相连的接口就是外部接口。

当 ASPF 应用于设备外部接口的出方向时，可以在防火墙上为内部网用户访问互联网的返回报文打开一个临时通道。

3. 应用层协议检测基本原理

图1-1 应用层协议检测基本原理示意图



如 图 1-1所示，为了保护内部网络，一般情况下需要在路由器上配置访问控制列表，以允许内部网的主机访问外部网络，同时拒绝外部网络的主机访问内部网络。但访问控制列表会将用户发起连接后返回的报文过滤掉，导致连接无法正常建立。

ASPF 的应用层协议检测功能由会话管理及 ALG 功能协作实现。ASPF 将检测到的应用层会话的首报文与配置的策略进行匹配，匹配的结果交由会话管理用来建立会话信息数据库以及维护会话状态。之后，ASPF 根据会话管理返回的会话状态信息决定后续报文的处理方式。

关于会话管理及 ALG 功能的具体介绍请参见“安全分册”中的“会话管理配置”和“ALG 配置”。

4. 传输层协议检测基本原理

这里的传输层协议检测是指通用 TCP/UDP 检测。通用 TCP/UDP 检测与应用层协议检测不同，是对报文的传输层信息进行的检测，如源、目的地址及端口号等。

通用 TCP/UDP 检测要求返回到 ASPF 外部接口的报文要与前面从 ASPF 外部接口发出去的报文完全匹配，即源、目的地址及端口号恰好对应，否则返回的报文将被阻塞。因此对于 FTP 这样的多通道应用层协议，在不配置应用层检测而直接配置 TCP 检测的情况下会导致数据连接无法建立。

1.2 配置包过滤防火墙

1.2.1 包过滤防火墙配置任务简介

表1-1 包过滤防火墙配置任务简介

配置任务	说明	详细配置
启用防火墙功能	必选	1.2.2
配置缺省过滤方式	可选	1.2.3
配置接口的报文过滤功能	必选	1.2.4

1.2.2 启用防火墙功能

1. 启用 IPv4 防火墙功能

表1-2 启用 IPv4 防火墙功能

操作	命令	说明
进入系统视图	**system-view**	-
启用 IPv4 防火墙功能（集中式设备）	**firewall enable**	必选 缺省情况下，IPv4 防火墙功能处于关闭状态
启用 IPv4 防火墙功能（分布式设备）	**firewall enable** { **all** | **slot** *slot-number* }	必选 缺省情况下，IPv4 防火墙功能处于关闭状态

2. 启用 IPv6 防火墙功能

表1-3 启用 IPv6 防火墙功能

操作	命令	说明
进入系统视图	**system-view**	-
启用 IPv6 防火墙功能	**firewall ipv6 enable**	必选 缺省情况下，IPv6 防火墙功能处于关闭状态

1.2.3 配置缺省过滤方式

缺省过滤方式用来定义对访问控制列表以外数据包的处理方式，即在没有规则去判定用户数据包是否可以通过的时候，防火墙采取的策略是允许还是禁止该数据包通过。

1. 配置 IPv4 防火墙缺省过滤方式

表1-4 配置 IPv4 防火墙缺省过滤方式

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv4 防火墙缺省过滤方式(集中式设备)	**firewall default** { **deny** | **permit** }	可选 缺省情况下，IPv4 防火墙缺省过滤方式为允许报文通过（**permit**）
配置 IPv4 防火墙缺省过滤方式(分布式设备)	**firewall default** { **deny** | **permit** } { **all** | **slot** *slot-number* }	可选 缺省情况下，IPv4 防火墙缺省过滤方式为允许报文通过（**permit**）

2. 配置 IPv6 防火墙缺省过滤方式

表1-5 配置 IPv6 防火墙缺省过滤方式

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 防火墙缺省过滤方式	**firewall ipv6 default { deny | permit }**	可选 缺省情况下，IPv6 防火墙缺省过滤方式为允许报文通过（**permit**）

1.2.4 配置接口的报文过滤功能

一般采用在接口上应用访问控制列表来实现报文过滤功能。将 ACL 规则应用到接口时，同时会遵循时间段过滤原则，另外可以对接口收发报文分别指定 ACL 规则：

- 基本访问控制列表只根据三层的源地址信息制定规则，对数据包进行相应的分析处理；
- 高级访问控制列表可以使用数据包的源地址信息、目的地址信息、IP 承载的协议类型、TCP 或 UDP 的源端口、目的端口等内容定义规则。

高级访问控制列表提供一种匹配方式：

- 标准匹配，即三层信息的匹配，该匹配将忽略三层以外的信息。



说明

接口的报文过滤功能与接口加入聚合组及接口加入业务环回组互斥。

1. 配置接口的 IPv4 报文过滤功能

表1-6 配置接口的 IPv4 报文过滤功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置接口的 IPv4 报文过滤功能	**firewall packet-filter** { *acl-number* | **name** *acl-name* } { **inbound** | **outbound** }	必选 缺省情况下，不对通过接口的 IPv4 报文进行过滤

2. 配置接口的 IPv6 报文过滤功能

IPv6 报文过滤为提供了基于 IPv6 ACL 的基本防火墙功能。可在每个接口的进方向或出方向的 IPv6 报文进行过滤。在接口的一个方向上，用户只能配置一个 IPv6 ACL。

表1-7 配置 IPv6 防火墙报文过滤

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置接口的 IPv6 报文过滤功能	**firewall packet-filter ipv6** { *acl6-number* | **name** *acl6-name* } { **inbound** | **outbound** }	必选 缺省情况下，不对通过接口的 IPv6 报文进行过滤

1.2.5 包过滤防火墙显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后包过滤防火墙的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除包过滤防火墙的统计信息。

表1-8 包过虑防火墙显示和维护

操作	命令
查看 IPv4 防火墙过滤报文的统计信息	**display firewall-statistics** { **all** | **interface** *interface-type interface-number* }
查看 IPv6 防火墙的过滤报文统计信息	**display firewall ipv6 statistics** { **all** | **interface** *interface-type interface-number* }
清除 IPv4 防火墙过滤报文的统计信息	**reset firewall-statistics** { **all** | **interface** *interface-type interface-number* }
清除 IPv6 防火墙的过滤报文统计信息	**reset firewall ipv6 statistics** { **all** | **interface** *interface-type interface-number* }

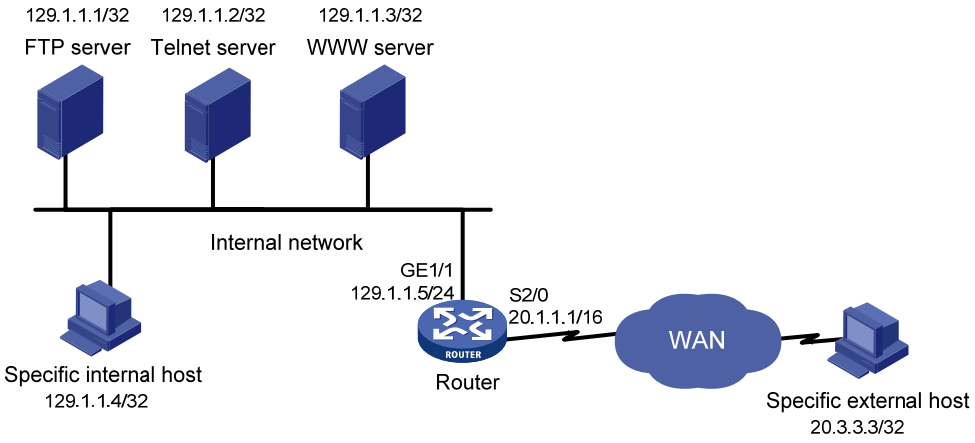
1.2.6 包过滤防火墙典型配置举例

1. 组网需求

- 某公司通过 Router 的接口 Serial2/0 访问 Internet，Router 与内部网通过接口 GigabitEthernet1/1 连接；
- 公司内部对外提供 WWW、FTP 和 Telnet 服务：公司内部子网为 129.1.1.0。其中，内部 FTP 服务器地址为 129.1.1.1，内部 Telnet 服务器地址为 129.1.1.2，内部 WWW 服务器地址为 129.1.1.3；公司对外地址为 20.1.1.1。在 Router 上配置了地址转换，这样公司内部主机可以访问 Internet，公司外部主机可以访问公司内部的服务器；
- 通过配置防火墙，希望实现以下要求：外部网络只有特定用户可以访问内部服务器；内部网络只有特定主机可以访问外部网络。
- 假定外部特定用户的 IP 地址为 20.3.3.3。

2. 组网图

图1-2 包过滤防火墙典型配置组网图



3. 配置步骤

在 Router 上启用防火墙功能。

```
<Router> system-view
[Router] firewall enable
```

创建高级访问控制列表 3001。

```
[Router] acl number 3001
```

配置规则允许特定主机访问外部网，允许内部服务器访问外部网。

```
[Router-acl-adv-3001] rule permit ip source 129.1.1.1 0
[Router-acl-adv-3001] rule permit ip source 129.1.1.2 0
[Router-acl-adv-3001] rule permit ip source 129.1.1.3 0
[Router-acl-adv-3001] rule permit ip source 129.1.1.4 0
```

配置规则禁止所有 IP 包通过。

```
[Router-acl-adv-3001] rule deny ip
```

创建高级访问控制列表 3002。

```
[Router-acl-adv-3001] quit
[Router] acl number 3002
```

配置规则允许特定用户从外部网访问内部服务器。

```
[Router-acl-adv-3002] rule permit tcp source 20.3.3.3 0 destination 129.1.1.0 0.0.0.255
```

配置规则允许外部特定数据进入内部网络（只允许端口大于 1024 的包）。

```
[Router-acl-adv-3002] rule permit tcp destination 20.1.1.1 0 destination-port gt 1024
[Router-acl-adv-3002] rule deny ip
```

将 ACL 3001 作用于从接口 GigabitEthernet1/1 进入的包。

```
[Router-acl-adv-3002] quit
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] firewall packet-filter 3001 inbound
```

将 ACL 3002 作用于从接口 Serial2/0 进入的包。

```
[Router-GigabitEthernet1/1] quit
```

```
[Router] interface serial 2/0
[Router-Serial2/0] firewall packet-filter 3002 inbound
```

1.3 配置 ASPF

1.3.1 ASPF 配置任务简介

表1-9 ASPF 配置任务简介

配置任务	说明	详细配置
启用防火墙功能	必选	1.3.2
配置 ASPF 策略	必选	1.3.3
在接口上应用 ASPF 策略	必选	1.3.4
配置端口映射	可选	1.3.5

1.3.2 启用防火墙功能

表1-10 启用防火墙功能

操作	命令	说明
进入系统视图	**system-view**	-
启用 IPv4 防火墙功能（集中式设备）	**firewall enable**	必选 缺省情况下，IPv4 防火墙功能处于关闭状态
启用 IPv4 防火墙功能（分布式设备）	**firewall enable** { **all** | **slot** *slot-number* }	必选 缺省情况下，IPv4 防火墙功能处于关闭状态

1.3.3 配置 ASPF 策略

表1-11 配置 ASPF 策略

操作	命令	说明
进入系统视图	**system-view**	-
创建 ASPF 策略，并进入 ASPF 策略视图	**aspf-policy** *aspf-policy-number*	必选
配置 ICMP 差错报文丢弃功能	**icmp-error drop**	可选 缺省情况下，不丢弃 ICMP 差错报文
配置非 SYN 的 TCP 首报文丢弃功能	**tcp syn-check**	可选 缺省情况下，不丢弃非 SYN 的 TCP 首报文

1.3.4 在接口上应用 ASPF 策略

对于 ASPF 策略，有两个概念：内部接口和外部接口。如果设备连接了内部网和互联网，并且设备要通过部署 ASPF 来保护内部网的服务器，则设备上与内部网连接的接口就是内部接口，与互联网相连的接口就是外部接口。当 ASPF 与包过滤防火墙协同工作，并应用于外部接口时，可以拒绝互联网上的用户对内部网的访问，但内部网的用户访问互联网时，返回的报文可以通过 ASPF。

只有将定义好 ASPF 策略应用到接口上，才能对通过接口的流量进行检测。由于 ASPF 对于应用层协议状态的保存和维护都是基于接口的，因此在实际应用中，必须保证报文入口的一致性，即必须保证连接发起报文和返回报文基于同一接口。

表1-12 在接口上应用 ASPF 策略

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在接口上应用 ASPF 策略	**firewall aspf** *aspf-policy-number* { **inbound** | **outbound** }	必选 缺省情况下，接口上没有应用 ASPF 策略

1.3.5 配置端口映射

有如下两类映射机制：

- 通用端口映射：是将用户自定义的端口号与应用层协议建立起映射关系。例如，在建立端口 8080 到 HTTP 协议的映射后，所有目的端口为 8080 的 TCP 报文都将被认为是 HTTP 报文。
- 基于基本访问控制列表的主机端口映射：是对来自某些特定主机的报文建立自定义的端口号与应用协议的映射。例如，可以把去往网段 1.1.0.0 的、使用 8080 端口的 TCP 报文映射为 HTTP 报文。主机的范围由基本 ACL 指定。

表1-13 配置端口映射

操作	命令	说明
进入系统视图	**system-view**	-
配置端口到应用层协议的映射	**port-mapping** *application-name* **port** *port-number* [**acl** *acl-number*]	必选 缺省情况下，没有端口到应用层协议的映射关系

1.3.6 ASPF 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 ASPF 的运行情况，通过查看显示信息验证配置的效果。

表1-14 ASPF 显示和维护

操作	命令
查看所有 ASPF 策略信息	**display aspf all**
查看接口上的 ASPF 策略信息	**display aspf interface**
查看指定 ASPF 策略的信息	**display aspf policy** *aspf-policy-number*
查看端口映射信息	**display port-mapping** [*application-name* | **port** *port-number*]

1.3.7 ASPF 典型配置举例

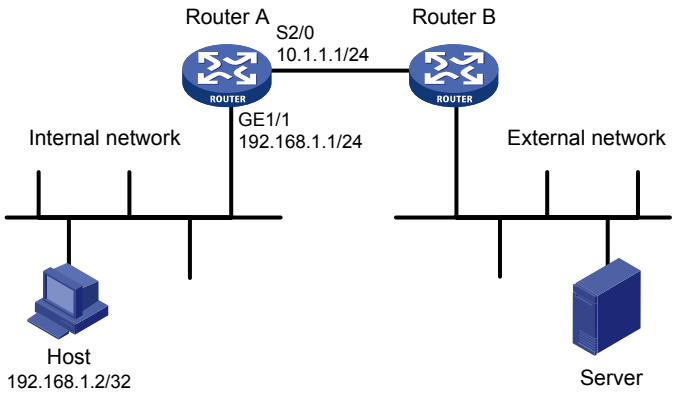
1. 组网需求

在 Router A 上配置一个 ASPF 策略，实现以下功能。

- ICMP 差错报文检测及 TCP 非 SYN 首报文丢弃。
- 本例可以应用在本地用户需要访问远程网络服务的情况下。

2. 组网图

图1-3 ASPF 典型配置组网图



3. 配置步骤

在 Router A 上启用防火墙功能。

```
<RouterA> system-view
[RouterA] firewall enable
```

配置 ACL 3111，以拒绝所有 IP 流量进入内部网络，ASPF 会为允许通过的流量创建临时的访问控制列表。

```
[RouterA] acl number 3111
[RouterA-acl-adv-3111] rule deny ip
[RouterA-acl-adv-3111] quit
```

创建 ASPF 策略 1。

```
[RouterA] aspf-policy 1
[RouterA-aspf-policy-1] icmp-error drop
[RouterA-aspf-policy-1] tcp syn-check
[RouterA-aspf-policy-1] quit
```

在接口 Serial2/0 上应用 ASPF 策略和 ACL 3111。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] firewall aspf 1 outbound
[RouterA-Serial2/0] firewall packet-filter 3111 inbound
```

目 录

1 ALG配置 1-1
1.1 ALG简介 1-1
1.2 配置使能ALG功能 1-3
1.3 ALG典型配置举例 1-3
1.3.1 ALG支持FTP典型配置举例 1-3
1.3.2 ALG支持SIP/H.323 典型配置举例 1-4
1.3.3 ALG支持NBT典型配置举例 1-5

1 ALG 配置

1.1 ALG 简介

ALG（Application Level Gateway，应用层网关）主要完成对应用层报文的处理。通常情况下，NAT 只对报文头中的 IP 地址和端口信息进行转换，不对应用层数据载荷中的字段进行分析。然而一些特殊协议，它们报文的数据载荷中可能包含 IP 地址或端口信息，这些内容不能被 NAT 进行有效的转换，就可能导致问题。

例如，FTP 应用就由数据连接和控制连接共同完成，而且数据连接的建立动态地由控制连接中的载荷字段信息决定，这就需要 ALG 来完成载荷字段信息的转换，以保证后续数据连接的正确建立。

ALG 在与 NAT（Network Address Translation，网络地址转换）、ASPF（Application Specific Packet Filter，基于应用层状态的包过滤）配合使用的情况下，可以实现地址转换、数据通道检测和应用层状态检查的功能。

- 地址转换

对报文应用层数据载荷中携带的 IP 地址、端口、协议类型（TCP 或者 UDP）、对端地址（在数据载荷中带有对端的地址）进行地址转换。

- 数据通道检测

提取数据通道信息，为后续的报文连接建立数据通道。此处的数据通道为相对于用户的控制连接而言的数据连接。

- 应用层状态检查

对报文的应用层协议状态进行检查，若正确则更新报文状态机进行下一步处理，否则丢弃报文。

本特性支持对多种应用层协议的 ALG 处理，不同的协议对以上三种功能的支持情况有所不同，实际中根据具体需要选择支持全部或部分功能。

目前实现 ALG 功能的常用应用层协议包括：

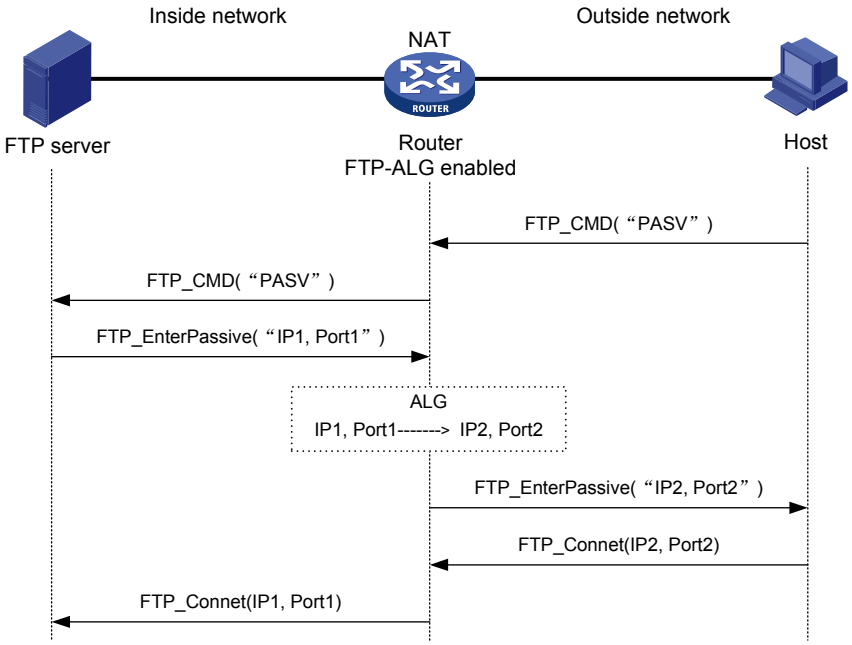
- DNS（Domain Name System，域名系统）
- FTP（File Transfer Protocol，文件传输协议）
- H.323（包括 RAS、H.225、H.245），一种多媒体会话协议
- HTTP（Hyper Text Transport Protocol，超级文本传输协议）
- ICMP（Internet Control Message Protocol，Internet 控制报文协议）
- ILS（Internet Locator Server，Internet 定位服务）
- MSN/QQ，两种常见的语音视频通讯协议
- PPTP（Point to Point Tunneling Protocol，点到点隧道协议）
- NBT（Network Basic Input/Output System，网络基本输入/输出系统）
- RTSP（Real-Time Streaming Protocol，实时流协议）
- SIP（Session Initiation Protocol，会话发起协议）
- SQLNET，一种 Oracle 数据库语言
- TFTP（Trivial File Transfer Protocol，简单文件传输协议）

说明

对 DNS、FTP、H.323、ILS、MSN、NBT、PPTP、QQ、RTSP、SIP、SQLNET、TFTP 的处理可以通过命令行进行使能和去使能的配置，对 HTTP、ICMP 差错报文的处理不支持命令行的配置，系统默认使能。

下面以FTP协议为例，简单描述使能ALG特性的设备在网络中的工作过程。如 图 1-1所示，位于外部网络的客户端Host以PASV方式访问内部网络的FTP服务器，经过中间的设备Router进行NAT转换，该设备上使能了ALG特性。

图1-1 PASV 方式的 FTP-ALG



整个通信过程包括如下四个阶段：

(1) 建立控制通道

客户端向服务器发送 TCP 连接请求。TCP 连接建立成功后，服务器和客户端进入用户认证阶段。若 TCP 连接失败，服务器会断开与客户端的连接。

(2) 用户认证

客户端向服务器发送认证请求，报文中包含 FTP 命令（USER、PASSWORD）及命令所对应的内容。

客户端发送的认证请求报文在通过配置了 ALG 的设备时，报文载荷中携带的命令字将会被解析出来，用于进行状态机转换过程是否正确的检查。若状态机转换发生错误，则丢弃报文。这样可防止客户端发送状态机错误的报文攻击服务器或者非法登录服务器，起到保护服务器的作用。

客户端的认证请求报文通过 ALG 处理之后，到达服务器端，服务器将对其进行响应。

(3) 创建数据通道

认证状态正确且用户是服务器已经授权的客户端，才能和服务器建立数据连接，进行数据的交互。

如 图 1-1所示，当客户端发送“PASV”命令发起连接时，服务器会在发送给客户端的PASV响应报文中携带自己的私网地址和端口号（IP1，Port1），响应报文经过ALG设备时被解析，其中携带的

服务器的私网地址和端口号被转换成其对应的公网地址和端口号（IP2，Port2），之后再利用该地址和端口与客户端的地址和端口之间建立起数据通道。

(4) 数据交互

客户端和服务器之间的数据交互可以直接通过数据通道来进行。

1.2 配置使能 ALG 功能

表1-1 使能 ALG 功能

操作	命令	说明
进入系统视图	**system-view**	-
使能指定协议的 ALG 功能	**alg { all | dns | ftp| h323 | ils | msn | nbt | pptp | qq | rtsp | sip | sqlnet | tftp }**	可选 缺省情况下，所有协议的 ALG 功能均处于使能状态

1.3 ALG 典型配置举例

说明

以下典型配置举例中，需保证服务器和客户端的配置已经完成，本节只介绍设备端的 ALG 配置，其它配置略。

1.3.1 ALG 支持 FTP 典型配置举例

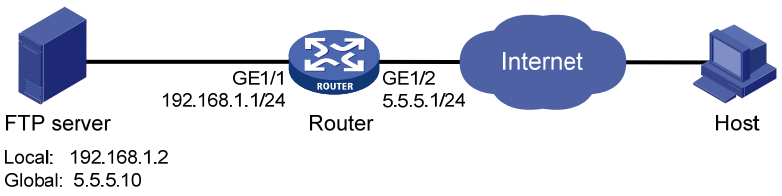
1. 组网需求

如 图 1-2所示，某公司通过启用了NAT和ALG功能的设备连接到Internet。公司内部对外提供FTP服务。公司内部网址为 192.168.1.0/24。其中，内部FTP服务器的IP地址为 192.168.1.2。通过配置NAT和ALG，满足如下要求：

- 外部网络的 Host 可以访问内部的 FTP 服务器。
- 公司具有 5.5.5.1、5.5.5.9～5.5.5.11 四个合法的公网 IP 地址。FTP 服务器使用 5.5.5.10 作为对外的 IP 地址。

2. 组网图

图1-2 ALG 应用组网图



3. 配置步骤

配置地址池和访问控制列表。

```
<Router> system-view
```

```
[Router] nat address-group 1 5.5.5.9 5.5.5.11
[Router] acl number 2001
[Router-acl-basic-2001] rule permit
[Router-acl-basic-2001] quit
```

使能 FTP 协议的 ALG 功能。

```
[Router] alg ftp
```

配置 NAT 转换。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] nat outbound 2001 address-group 1
```

配置内部 FTP 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 5.5.5.10 ftp inside 192.168.1.2
ftp
```

1.3.2 ALG 支持 SIP/H.323 典型配置举例

说明

ALG 支持 H.323 的典型配置方法与支持 SIP 的类似，下面仅以 SIP 为例。

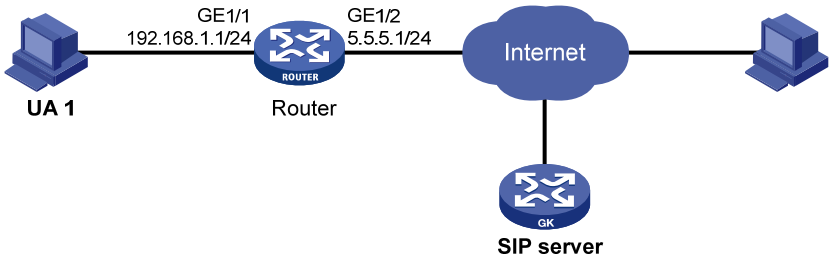
1. 组网需求

如 图 1-3所示，某公司通过启用了NAT和ALG功能的设备连接到Internet。公司内部网址为192.168.1.0/24。通过配置NAT和ALG，满足如下要求：

- 公司内部的 SIP UA 1 和外部网络的 SIP UA 2 均可以通过别名与对方成功建立通信。
- 公司具有 5.5.5.1、5.5.5.9～5.5.5.11 四个合法的公网 IP 地址。公司内部 SIP UA 1 在向外部的 SIP server 注册时选择 5.5.5.9～5.5.5.11 中的一个地址作为其公网地址。

2. 组网图

图1-3 ALG 支持 SIP 应用组网图



3. 配置步骤

配置地址池和访问控制列表。

```
<Router> system-view
[Router] nat address-group 1 5.5.5.9 5.5.5.11
[Router] acl number 2001
[Router-acl-basic-2001] rule permit source 192.168.1.0 0.0.0.255
[Router-acl-basic-2001] rule deny
```

```
[Router-acl-basic-2001] quit
```

使能 SIP 协议的 ALG 功能。

```
[Router] alg sip
```

配置 NAT 转换。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] nat outbound 2001 address-group 1
```

1.3.3 ALG 支持 NBT 典型配置举例

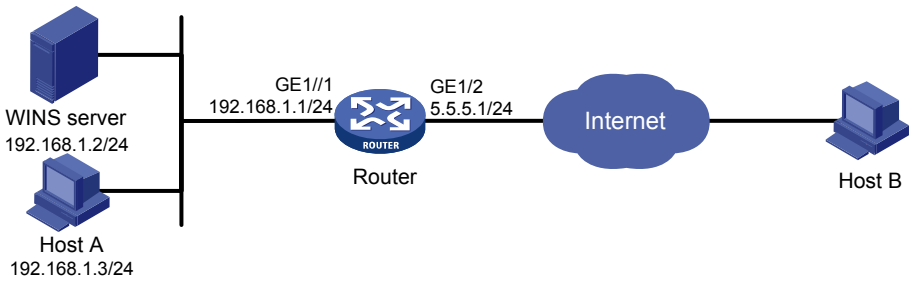
1. 组网需求

如 图 1-4所示，某公司通过启用了NAT和ALG功能的设备连接到Internet。公司内部对外提供NBT服务。公司内部网址为 192.168.1.0/24。其中，内部服务器通过配置NAT和ALG，满足如下要求：

- 外部网络的主机可以通过主机名访问内部的 WINS 服务器和主机。
- Host A 使用 5.5.5.9 作为对外的 IP 地址。WINS 服务器使用 5.5.5.10 作为对外的 IP 地址。

2. 组网图

图1-4 ALG 支持 NBT 应用组网图



3. 配置步骤

配置静态地址转换。

```
<Router> system-view
[Router] nat static 192.168.1.3 5.5.5.9
```

使能 NBT 协议的 ALG 功能。

```
[Router] alg nbt
```

配置 NAT 转换。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] nat outbound static
```

配置内部 WINS 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol udp global 5.5.5.10 137 inside 192.168.1.2
137
[Router-GigabitEthernet1/2] nat server protocol udp global 5.5.5.10 138 inside 192.168.1.2
138
[Router-GigabitEthernet1/2] nat server protocol tcp global 5.5.5.10 139 inside 192.168.1.2
139
```

目 录

1 PKI配置 1-1
1.1 PKI简介 1-1
1.1.1 概述 1-1
1.1.2 相关术语 1-1
1.1.3 体系结构 1-2
1.1.4 主要应用 1-2
1.1.5 PKI的工作过程 1-3
1.2 PKI配置任务简介 1-3
1.3 配置实体DN 1-4
1.4 配置PKI域 1-5
1.5 PKI证书申请 1-7
1.5.1 自动申请证书 1-7
1.5.2 手工申请证书 1-7
1.6 手工获取证书 1-8
1.7 配置PKI证书验证 1-9
1.8 销毁本地RSA密钥对 1-10
1.9 删除证书 1-10
1.10 配置证书属性的访问控制策略 1-10
1.11 PKI显示和维护 1-11
1.12 PKI典型配置举例 1-11
1.12.1 PKI实体向CA申请证书（方式一） 1-12
1.12.2 PKI实体向CA申请证书（方式二） 1-15
1.12.3 使用PKI证书体系的RSA证书签名方法进行IKE协商认证 1-18
1.13 常见配置错误举例 1-21
1.13.1 获取CA证书失败 1-21
1.13.2 本地证书申请失败 1-21
1.13.3 CRL获取失败 1-22

1 PKI 配置

1.1 PKI 简介

1.1.1 概述

PKI（Public Key Infrastructure，公钥基础设施）是一个利用公共密钥理论和技术来实现并提供信息安全服务的具有通用性的安全基础设施。

公共密钥体制也称为非对称密钥体制，是目前应用最广泛的一种加密体制，在这一体制中，它使用一个非对称的密钥对，分别是一个公开的加密密钥（公钥）和一个保密的解密密钥（私钥），用公钥加密的信息只能用私钥机密，反之亦然。由于公钥是公开的，需要在网上传送，故公钥的管理问题就是公共密钥体制所需要解决的关键问题。

目前，PKI 系统中引出的数字证书机制就是一个很好的解决方案。基于公共密钥技术的数字证书是一个用户的身份和他所持有的公钥的结合，它是使用 PKI 系统的用户建立安全通信的信任基础。

基于数字证书的 PKI 系统，能够为网络通信和网络交易，特别是电子政务和电子商务业务，透明地提供一整套安全服务，主要包括身份认证、保密、数据完整性和不可否认性。

1.1.2 相关术语

1. 数字证书

数字证书是一个经 CA（Certificate Authority，证书机构）签名的、包含公开密钥及相关的用户身份信息的文件，它建立了用户身份信息与用户公钥的关联。CA 对数字证书的签名保证了证书的合法性和权威性。数字证书的格式遵循 ITU-T X.509 国际标准，目前最常用的为 X.509 V3 标准。一个数字证书中包含多个字段，包括证书签发者的名称、主体的公钥信息、CA 对证书的数字签名、证书的有效期等。

本手册中涉及两类证书：本地（local）证书和 CA（Certificate Authority）证书。本地证书是 CA 签发给用户的数字证书；CA 证书是 CA 自身的证书。若 PKI 系统中存在多个 CA，则会形成一个 CA 层次结构，最上层的 CA 是根 CA，它拥有一个 CA “自签名”的证书。

目前，我司的 PKI 可为安全协议 IPsec、SSL 提供证书管理机制。

2. 证书废除列表（CRL，Certificate Revocation List）

由于用户姓名的改变、私钥泄漏或业务中止等原因，需要存在一种方法将现行的证书撤消，即撤消公开密钥及相关的用户身份信息的绑定关系。在 PKI 中，所使用的这种方法为证书废除列表。任何一个证书被废除以后，CA 就要发布 CRL 来声明该证书是无效的，并列出所有被废除的证书的序列号。CRL 提供了一种检验证书有效性的方式。

当一个 CRL 的撤消信息过多时会导致 CRL 的发布规模变得非常庞大，且随着 CRL 大小的增加，网络资源的使用性能也会随之下降。为了避免这种情况，允许一个 CA 的撤消信息通过多个 CRL 发布出来，并且使用 CRL 发布点来指出这些小 CRL 的位置。。

3. CA 策略

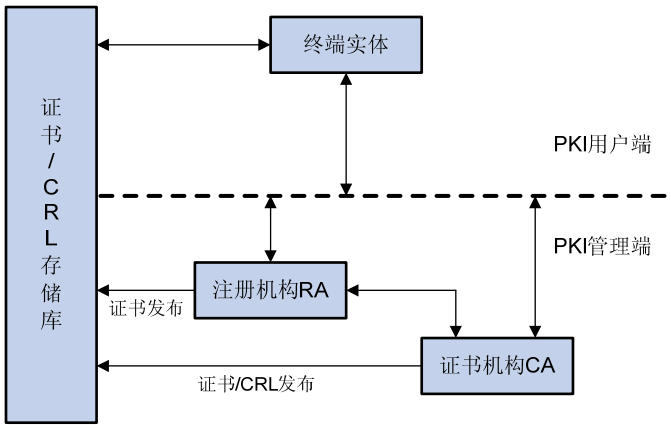
CA 在受理证书请求、颁发证书、吊销证书和发布 CRL 时所采用的一套标准被称为 CA 策略。通常，CA 以一种叫做证书惯例声明（CPS，Certification Practice Statement）的文档发布其策略，CA 策

略可以通过带外（如电话、磁盘、电子邮件等）或其他方式获取。由于不同的 CA 使用不同的方法验证公开密钥与实体之间的绑定，所以在选择信任的 CA 进行证书申请之前，必须理解 CA 策略，从而指导对实体进行相应的配置。

1.1.3 体系结构

一个PKI体系由终端实体、证书机构、注册机构和PKI存储库四类实体共同组成，如下 图 1-1。

图1-1 PKI 体系结构图



1. 终端实体

终端实体是 PKI 产品或服务的最终使用者，可以是个人、组织、设备（如路由器、交换机）或计算机中运行的进程。

2. 证书机构（CA，Certificate Authority）

CA 是 PKI 的信任基础，是一个用于签发并管理数字证书的可信实体。其作用包括：发放证书、规定证书的有效期和通过发布 CRL 确保必要时可以废除证书。

3. 注册机构（RA，Registration Authority）

RA 是 CA 的延伸，可作为 CA 的一部分，也可以独立。RA 功能包括个人身份审核、CRL 管理、密钥对产生和密钥对备份等。PKI 国际标准推荐由一个独立的 RA 来完成注册管理的任务，这样可以增强应用系统的安全性。

4. PKI 存储库

PKI 存储库包括 LDAP（Lightweight Directory Access Protocol，轻量级目录访问协议）服务器和普通数据库，用于对用户申请、证书、密钥、CRL 和日志等信息进行存储和管理，并提供一定的查询功能。

LDAP 提供了一种访问 PKI 存储库的方式，通过该协议来访问并管理 PKI 信息。LDAP 服务器负责将 RA 服务器传输过来的用户信息以及数字证书进行存储，并提供目录浏览服务。用户通过访问 LDAP 服务器获取自己和其他用户的数字证书。

1.1.4 主要应用

PKI 技术的广泛应用能满足人们对网络交易安全保障的需求。作为一种基础设施，PKI 的应用范围非常广泛，并且在不断发展之中，下面给出几个应用实例。

1. 虚拟专用网络（VPN，Virtual Private Network）

VPN 是一种构建在公用通信基础设施上的专用数据通信网络，利用网络层安全协议（如 IPsec）和建立在 PKI 上的加密与数字签名技术来获得机密性保护。

2. 安全电子邮件

电子邮件的安全也要求机密、完整、认证和不可否认，而这些都可以利用 PKI 技术来实现。目前发展很快的安全电子邮件协议 S/MIME(Secure/Multipurpose Internet Mail Extensions，安全/多用途 Internet 邮件扩充协议)，是一个允许发送加密和有签名邮件的协议。该协议的实现需要依赖于 PKI 技术。

3. Web 安全

为了透明地解决 Web 的安全问题，在两个实体进行通信之前，先要建立 SSL（Secure Sockets Layer，安全套接字层）连接，以此实现对应用层透明的安全通信。利用 PKI 技术，SSL 协议允许在浏览器和服务器之间进行加密通信。此外，服务器端和浏览器端通信时双方可以通过数字证书确认对方的身份。

1.1.5 PKI 的工作过程

针对一个使用 PKI 的网络，配置 PKI 的目的就是为指定的实体向 CA 申请一个本地证书，并由设备对证书的有效性进行验证。下面是 PKI 的工作过程：

(1) 实体向 RA 提出证书申请；

(2) RA 审核实体身份，将实体身份信息和公开密钥以数字签名的方式发送给 CA；

(3) CA 验证数字签名，同意实体的申请，颁发证书；

(4) RA 接收 CA 返回的证书，发送到 LDAP 服务器以提供目录浏览服务，并通知实体证书发行成功；

(5) 实体获取证书，利用该证书可以与其它实体使用加密、数字签名进行安全通信；

(6) 实体希望撤消自己的证书时，向 CA 提交申请。CA 批准实体撤消证书，并更新 CRL，发布到 LDAP 服务器。

1.2 PKI 配置任务简介

表1-1 PKI 配置任务简介

配置任务		说明	详细配置
配置实体 DN		必选	1.3
配置 PKI 域		必选	1.4
PKI 证书申请	自动申请证书	二者必选其一	1.5.1
	手工申请证书		1.5.2
手工获取证书		可选	1.6
配置 PKI 证书验证		可选	1.7
销毁本地 RSA 密钥对		可选	1.8

配置任务	说明	详细配置
删除证书	可选	1.9
配置证书属性的访问控制策略	可选	1.10

1.3 配置实体 DN

一份证书是一个公开密钥与一个身份的绑定，而身份必须与一个特定的 PKI 实体相关联。实体 DN（Distinguished Name，可识别名称）的参数是实体的身份信息，CA 根据实体提供的身份信息来唯一标识证书申请者。

实体 DN 的参数包括：

- 实体通用名
- 实体所属国家代码，用标准的 2 字符代码表示。例如，“CN”是中国的合法国家代码，“US”是美国的合法国家代码
- 实体 FQDN（Fully Qualified Domain Name，合格域名），是实体在网络中的唯一标识，由一个主机名和域名组成，可被解析为 IP 地址。例如，www 是一个主机名，whatever.com 是一个域名，则 www.whatever.com 就是一个 FQDN。
- 实体 IP 地址
- 实体所在地理区域名称
- 实体所属组织名称
- 实体所属部门名称
- 实体所属州省

说明

实体 DN 的配置必须与 CA 证书颁发策略相匹配，以确认实体 DN 的配置任务，如哪些实体参数为必选配置，哪些为可选配置。申请者的身份信息必须符合 CA 证书颁发策略，否则证书申请可能会失败。

表1-2 配置实体 DN

操作	命令	说明
进入系统视图	**system-view**	-
创建一个实体，并进入该实体视图	**pki entity** *entity-name*	必选 缺省情况下，无实体存在
配置实体通用名	**common-name** *name*	可选 缺省情况下，未配置实体的通用名
配置实体所属国家代码	**country** *country-code-str*	可选 缺省情况下，未配置实体所属国家代码
配置实体 FQDN	**fqdn** *name-str*	可选 缺省情况下，未配置实体 FQDN

操作	命令	说明
配置实体 IP 地址	**ip** *ip-address*	可选 缺省情况下，未配置实体 IP 地址
配置实体所在地理区域名称	**locality** *locality-name*	可选 缺省情况下，未配置实体所在地理区域
配置实体所属组织名称	**organization** *org-name*	可选 缺省情况下，未配置实体所属组织
配置实体所属部门名称	**organization-unit** *org-unit-name*	可选 缺省情况下，未配置实体所属部门
配置实体所属州省	**state** *state-name*	可选 缺省情况下，未配置实体所属州省

说明

- 目前设备上最多可以创建两个实体。
- Windows 2000 CA 服务器对证书申请的数据长度有一定的限制。实体 DN 配置项超过一定数据长度时，申请证书没有回应。

1.4 配置 PKI 域

实体在进行 PKI 证书申请操作之前需要配置一些注册信息来配合完成申请的过程。这些信息的集合就是一个实体的 PKI 域。

PKI 域是一个本地概念，因此创建 PKI 域的目的是便于其它应用引用 PKI 的配置，比如 IKE、SSL 等。一个设备上配置的 PKI 域对 CA 和其它设备是不可见的，每一个 PKI 域有单独的域参数配置信息。

PKI 域中包括以下参数：

- 信任的 CA 名称

在申请证书时，是通过一个可信实体认证机构，来完成实体证书的注册颁发，因此必须指定一个信任的 CA 名称。

- 实体名称

向 CA 发送证书申请请求时，必须指定所使用的实体名，以向 CA 表明自己的身份。

- 证书申请的注册机构

证书申请的受理一般由一个独立的注册机构（即 RA）来承担，它接收用户的注册申请，审查用户的申请资格，并决定是否同意 CA 给其签发数字证书。注册机构并不给用户签发证书，而只是对用户进行资格审查。有时 PKI 把注册管理的职能交给 CA 来完成，而不设立独立运行的 RA，但这并不是取消了 PKI 的注册功能，而只是将其作为 CA 的一项功能而已。PKI 推荐独立使用 RA 作为注册审理机构。

- 注册服务器的 URL

证书申请之前必须指定注册服务器的 URL，随后实体可通过简单证书注册协议（SCEP，Simple Certification Enrollment Protocol）向该服务器提出证书申请，SCEP 是专门用于与认证机构进行通信的协议。

- 证书申请状态查询的周期和次数

实体在发送证书申请后，如果 CA 采用手工验证申请，证书的发布会需要很长时间。在此期间，客户端需要定期发送状态查询，以便在证书签发后能及时获取到证书。客户端可以配置证书申请状态的查询周期和次数。

- LDAP 服务器 IP 地址

在 PKI 系统中，用户的证书和 CRL 信息的存储是一个非常核心的问题。一般采用 LDAP 服务器来存储证书和 CRL，这时就需要指定 LDAP 服务器的位置。

- 验证根证书时使用的指纹

当设备从 CA 获得根证书时，需要验证 CA 根证书的指纹，即根证书内容的散列值，该值对于每一个证书都是唯一的。如果 CA 根证书的指纹与在 PKI 域中配置的指纹不同，则设备将拒绝接收根证书。

表1-3 配置 PKI 域

配置任务	命令	说明
进入系统视图	**system-view**	-
创建一个 PKI 域，并进入 PKI 域视图	**pki domain** *domain-name*	必选
配置信任的 CA 名称	**ca identifier** *name*	必选 缺省情况下，未配置信任的 CA 名称
指定实体名称	**certificate request entity** *entity-name*	必选 缺省情况下，未指定实体名称 指定的实体名称必须已创建
配置证书申请的注册受理机构	**certificate request from** { **ca** | **ra** }	必选 缺省情况下，未指定证书申请的注册受理机构
配置注册服务器 URL	**certificate request url** *url-string*	必选 缺省情况下，未指定注册服务器 URL
配置证书申请状态查询的周期和次数	**certificate request polling** { **count** *count* | **interval** *minutes* }	可选 缺省情况下，证书申请状态查询周期为 20 分钟，每一个周期内查询 50 次
配置 LDAP 服务器	**ldap-server ip** *ip-address* [**port** *port-number*] [**version** *version-number*]	可选 缺省情况下，未指定 LDAP 服务器
配置验证根证书时使用的指纹	**root-certificate fingerprint** { **md5** | **sha1** } *string*	当证书申请方式为自动方式时，此配置必选；当证书申请方式为手工方式时，此配置可选，若不配置，需要用户自行验证根证书指纹 缺省情况下，未指定验证根证书时使用的指纹

说明

- 目前一个设备只支持同时创建两个 PKI 域。
- CA 的名称只是在获取 CA 证书时使用，申请本地证书时不会用到。
- 目前，注册服务器 URL 的配置不支持域名解析。

1.5 PKI 证书申请

证书申请就是实体向 CA 自我介绍的过程。实体向 CA 提供身份信息，以及相应的公开密钥，这些信息将成为颁发给该实体证书的主要组成部分。实体向 CA 提出证书申请，有离线和在线两种方式。离线申请方式下，CA 允许申请方通过带外方式（如电话、磁盘、电子邮件等）向 CA 提供申请信息。

在线证书申请有手工发起和自动发起两种方式。

1.5.1 自动申请证书

配置证书申请方式为自动方式后，在没有本地证书时实体自动通过 SCEP 协议进行申请，而且在证书即将过期时自动申请新的证书。

表1-4 自动申请证书

操作	命令	说明
进入系统视图	**system-view**	-
进入 PKI 域视图	**pki domain** *domain-name*	-
配置证书申请为自动方式	**certificate request mode auto [key-length** *key-length* **| password { cipher | simple }** *password* **] ***	必选 缺省情况下，证书申请为手工方式

1.5.2 手工申请证书

配置证书申请方式为手工方式后，需要手工完成获取 CA 证书、生成密钥对、申请本地证书的工作。

获取 CA 证书的目的是用来验证本地证书的真实性和合法性。

密钥对的产生是证书申请过程中重要的一步。申请过程使用了一对主机密钥：私钥和公钥。私钥由用户保留，公钥和其他信息则交由 CA 中心进行签名，从而产生证书。

有关 RSA 密钥对的具体配置请参见“安全分册”中的“公钥管理配置”。

表1-5 手工申请证书

操作	命令	说明
进入系统视图	**system-view**	-
进入 PKI 域视图	**pki domain** *domain-name*	-
配置证书申请为手工方式	**certificate request mode manual**	可选 缺省情况下，证书申请为手工方式
退回系统视图	**quit**	-
手工获取 CA 证书	请参见 1.6	必选

操作	命令	说明
生成本地 RSA 密钥对	**public-key local create rsa**	必选 缺省情况下，本地没有 RSA 密钥对
手工申请本地证书	**pki request-certificate domain** *domain-name* [*password*] [**pkcs10** [**filename** *filename*]]	必选

说明

- 若本地证书已存在，为保证密钥对与现存证书的一致性，不应执行创建密钥对命令，必须在删除本地证书后再执行 **public-key local create** 命令生成新的密钥对。有关 **public-key local create** 命令的详细介绍，请参见“安全分册”中的“公钥管理配置”。
- 创建的新密钥对将覆盖旧密钥对。若本地已有 RSA 密钥对，执行 **public-key local create** 命令时，系统会提示是否替换原有密钥对。
- 如果本地证书已存在，则不允许再执行证书申请操作，以避免因相关配置的修改使得证书与注册信息不匹配。若想重新申请，请先使用 **pki delete-certificate** 命令删除存储于本地的 CA 证书与本地证书，然后再执行 **pki request-certificate domain** 命令。
- 当无法通过 SCEP 协议向 CA 在线申请证书时，可以使用可选参数 **pkcs10** 打印出本地的证书申请信息。用户保存证书申请信息，并将其通过带外方式发送给 CA 进行证书申请。
- 证书申请之前必须保证实体时钟与 CA 的时钟同步，否则申请证书的有效期会出现异常。
- **pki request-certificate domain** 配置不能被保存在配置文件中。

1.6 手工获取证书

用户通过此配置可以将已存在的 CA 证书或本地证书获取至本地。获取证书有两种方式：离线方式和在线方式。离线方式下获取证书需要通过带外方式（如 FTP、磁盘、电子邮件等）取得证书，然后将其导入至本地。

获取证书的目的有两个：

- 将 CA 签发的与实体所在安全域有关的证书存放到本地，以提高证书的查询效率，减少向 PKI 证书存储库查询的次数。
- 为证书的验证做好准备。

在线本地证书之前必须完成 LDAP 服务器的配置。

表1-6 手工获取证书

操作		命令	说明
进入系统视图		**system-view**	-
手工获取证书	在线方式	**pki retrieval-certificate** { **ca** | **local** } **domain** *domain-name*	二者必选其一
	离线方式导入证书	**pki import-certificate** { **ca** | **local** } **domain** *domain-name* { **der** | **p12** | **pem** } [**filename** *filename*]	

注意

- 如果本地已有 CA 证书存在，则不允许执行手工获取 CA 证书的操作，避免因相关配置的修改使得证书与注册信息不匹配。若想重新获取，请先使用 **pki delete-certificate** 命令删除 CA 证书与本地证书后，再执行此命令。
- **pki retrieval-certificate** 配置不能被保存在配置文件中。

1.7 配置 PKI 证书验证

在使用每一个证书之前，必须对证书进行验证。证书验证包括对签发时间、签发者信息以及证书的有效性几方面进行验证。证书验证的核心是检查 CA 在证书上的签名，并确定证书仍在有效期内，而且未被废除，因此在进行证书验证操作之前必须首先获取 CA 证书。

配置证书验证时可以设置是否必须进行 CRL 检查。如果配置为使能 CRL 检查，则检验证书的有效性，必须通过 CRL 判断。

表1-7 配置使能 CRL 检查的证书验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 PKI 域视图	**pki domain** *domain-name*	-
配置 CRL 发布点的 URL	**crl url** *url-string*	可选 缺省情况下，未指定 CRL 发布点的 URL
配置 CRL 的更新周期	**crl update-period** *hours*	可选 缺省情况下，根据 CRL 文件中的下次更新域进行更新
使能 CRL 检查	**crl check enable**	可选 缺省情况下，CRL 检查处于开启状态
退回系统视图	**quit**	-
获取 CA 证书	请参见 1.6	必选
获取 CRL 并下载至本地	**pki retrieval-crl domain** *domain-name*	必选
检验证书的有效性	**pki validate-certificate** { **ca** | **local** } **domain** *domain-name*	必选

表1-8 配置不使能 CRL 检查的 PKI 证书验证

操作	命令	说明
进入系统视图	**system-view**	-
进入 PKI 域视图	**pki domain** *domain-name*	-
禁止 CRL 检查	**crl check disable**	必选 缺省情况下，CRL 检查处于开启状态
退回系统视图	**quit**	-
获取 CA 证书	请参见 1.6	必选

操作	命令	说明
检验证书的有效性	**pki validate-certificate** { **ca** | **local** } **domain** *domain-name*	必选

说明

- CRL 的更新周期是指本地从 CRL 存储服务器下载 CRL 的时间间隔。手工配置的 CRL 更新周期将优先于 CRL 文件中指定的更新时间。
- **pki retrieval-crl domain** 命令不能被保存在配置文件中。
- 目前，CRL 发布点 URL 的配置不支持域名解析。

1.8 销毁本地 RSA 密钥对

由 CA 颁发的证书都会设置有效期，证书生命周期的长短由签发证书的 CA 中心来确定。当用户的私钥被泄漏或证书的有效期快到时，用户应该删除旧的密钥对，产生新的密钥对，重新申请新的证书。通过此配置用户可以销毁本地 RSA 密钥对。

表1-9 销毁本地 RSA 密钥对

操作	命令	说明
进入系统视图	**system-view**	-
销毁本地 RSA 密钥对	**public-key local destroy rsa**	必选

说明

命令 **public-key local destroy rsa** 的详细介绍可参考“安全分册/公钥管理配置”。

1.9 删除证书

证书过期或希望重新申请证书，可以通过此配置删除一个已经存在的本地证书或 CA 证书。

表1-10 配置删除证书

操作	命令	说明
进入系统视图	**system-view**	-
配置删除证书	**pki delete-certificate** { **ca** | **local** } **domain** *domain-name*	必选

1.10 配置证书属性的访问控制策略

通过配置证书属性的访问控制策略，可以对用户的访问权限进行进一步的控制，保证了与之通信的服务器的安全性。

表1-11 配置证书属性的访问控制策略

操作	命令	说明
进入系统视图	**system-view**	-
创建证书属性组，并进入证书属性组视图	**pki certificate attribute-group** *group-name*	必选 缺省情况下，不存在证书属性组
配置证书颁发者名、证书主题名及备用主题名的属性规则	**attribute** *id* { **alt-subject-name** { **fqdn** | **ip** } | { **issuer-name** | **subject-name** } { **dn** | **fqdn** | **ip** } } { **ctn** | **equ** | **nctn** | **nequ**} *attribute-value*	可选 缺省情况下，对证书颁发者名、证书主题名及备用主题名没有限制
退回系统视图	**quit**	-
创建证书属性访问控制策略，并进入证书属性访问控制策略视图	**pki certificate access-control-policy** *policy-name*	必选 缺省情况下，不存在证书属性访问控制策略
配置证书属性控制规则	**rule** [*id*] { **deny** | **permit** } *group-name*	必选 缺省情况下，不存在证书属性控制规则

注意

配置证书属性控制规则时，*group-name* 必须是已存在的证书属性组的名称。

1.11 PKI 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 PKI 的运行情况，通过查看显示信息验证配置的效果。

表1-12 PKI 显示和维护

操作	命令
显示证书内容或证书申请状态	**display pki certificate** { { **ca** | **local** } **domain** *domain-name* | **request-status** }
显示 CRL 内容	**display pki crl domain** *domain-name*
显示证书属性组的信息	**display pki certificate attribute-group** { *group-name* | **all** }
显示证书属性访问控制策略的信息	**display pki certificate access-control-policy** { *policy-name* | **all** }

1.12 PKI 典型配置举例

注意

- 当采用 Windows Server 作为 CA 时，需要安装 SCEP 插件。此时，配置 PKI domain 时，需要使用 **certificate request from ra** 命令指定实体从 RA 注册申请证书。
- 当采用 RSA Keon 软件时，不需要安装 SCEP 插件。此时，配置 PKI domain 时，需要使用 **certificate request from ca** 命令指定实体从 CA 注册申请证书。

1.12.1 PKI 实体向 CA 申请证书（方式一）

说明

本配置举例中，CA 服务器上采用 RSA Keon 软件。

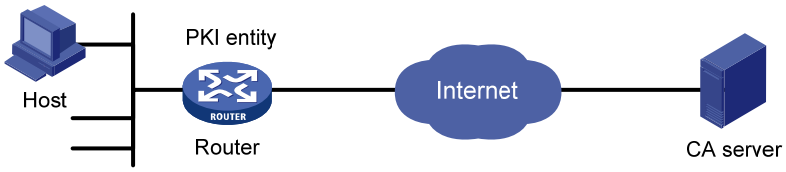
1. 组网需求

在作为 PKI 实体的设备 Router 上进行相关配置，实现以下需求：

- 设备向 CA 服务器申请本地证书
- 获取 CRL 为证书验证做准备

2. 组网图

图1-2 PKI 实体向 CA 申请证书组网图



3. 配置步骤

- CA 服务器端的配置

(1) 创建 CA 服务器 myca

在本例中，CA 服务器上首先需要进行基本属性 Nickname 和 Subject DN 的配置。其它属性选择默认值。其中，Nickname 为可信任的 CA 名称，Subject DN 为 CA 的 DN 属性，包括 CN、OU、O 和 C。

(2) 配置扩展属性

基本属性配置完毕之后，还需要在生成的 CA 服务器管理页面上对“Jurisdiction Configuration”进行配置，主要内容包括：根据需要选择合适的扩展选项；启动自动颁发证书功能；添加可以自动颁发证书的地址范围。

(3) 配置 CRL 发布

CA 服务器的基本配置完成之后，需要进行 CRL 的相关配置。

本例中选择 CRL 的发布方式为 HTTP，自动生成 CRL 发布点的 URL 为 http://4.4.4.133:447/myca.crl。

以上配置完成之后，还需要保证设备的系统时钟与 CA 的时钟同步才可以正常使用设备来申请证书和获取 CRL。

- 设备 Router 上的配置

(1) 配置实体命名空间

配置实体名称为 aaa，通用名为 router。

```
<Router> system-view
[Router] pki entity aaa
```

```
[Router-pki-entity-aaa] common-name router
[Router-pki-entity-aaa] quit
```

(2) 配置 PKI 域参数

创建并进入 PKI 域 torsa。

```
[Router] pki domain torsa
```

配置可信任的 CA 名称为 myca。

```
[Router-pki-domain-torsa] ca identifier myca
```

配置注册服务器 URL，格式为 http://host:port/Issuing Jurisdiction ID。其中的 Issuing Jurisdiction ID 为 CA 服务器上生成的 16 进制字符串。

```
[Router-pki-domain-torsa] certificate request url
http://4.4.4.133:446/c95e970f632d27be5e8cbf80e971d9c4a9a93337
```

配置证书申请的注册受理机构为 CA。

```
[Router-pki-domain-torsa] certificate request from ca
```

指定实体名称为 aaa。

```
[Router-pki-domain-torsa] certificate request entity aaa
```

配置 CRL 发布点位置。

```
[Router-pki-domain-torsa] crl url http://4.4.4.133:447/myca.crl
[Router-pki-domain-torsa] quit
```

(3) 用 RSA 算法生成本地密钥对

```
[Router] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits in the modulus [default = 1024]:
Generating Keys...
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++
```

(4) 证书申请

获取 CA 证书并下载至本地。

```
[Router] pki retrieval-certificate ca domain torsa
Retrieving CA/RA certificates. Please wait a while......
The trusted CA's finger print is:
    MD5  fingerprint:EDE9 0394 A273 B61A F1B3 0072 A0B1 F9AB
    SHA1 fingerprint: 77F9 A077 2FB8 088C 550B A33C 2410 D354 23B2 73A8

Is the finger print correct?(Y/N):y

Saving CA/RA certificates chain, please wait a moment......
CA certificates retrieval success.
```

获取 CRL 并下载至本地。

```
[Router] pki retrieval-crl domain torsa
Connecting to server for retrieving CRL. Please wait a while.....
CRL retrieval success!
```

手工申请本地证书。

```
[Router] pki request-certificate domain torsa challenge-word
Certificate is being requested, please wait......
[Router]
Enrolling the local certificate,please wait a while......
Certificate request Successfully!
Saving the local certificate to device......
Done!
```

(5) 验证配置结果

通过以下显示命令可以查看获取的本地证书信息。

```
<Router> display pki certificate local domain torsa
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            9A96A48F 9A509FD7 05FFF4DF 104AD094
        Signature Algorithm: sha1WithRSAEncryption
        Issuer:
            C=cn
            O=org
            OU=test
            CN=myca
        Validity
            Not Before: Jan  8 09:26:53 2007 GMT
            Not After : Jan  8 09:26:53 2008 GMT
        Subject:
            CN=router
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00D67D50 41046F6A 43610335 CA6C4B11
                    F8F89138 E4E905BD 43953BA2 623A54C0
                    EA3CB6E0 B04649CE C9CDDD38 34015970
                    981E96D9 FF4F7B73 A5155649 E583AC61
                    D3A5C849 CBDE350D 2A1926B7 0AE5EF5E
                    D1D8B08A DBF16205 7C2A4011 05F11094
                    73EB0549 A65D9E74 0F2953F2 D4F0042F
                    19103439 3D4F9359 88FB59F3 8D4B2F6C
                    2B
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 CRL Distribution Points:
```

```
            URI:http://4.4.4.133:447/myca.crl

    Signature Algorithm: sha1WithRSAEncryption
        836213A4 F2F74C1A 50F4100D B764D6CE
        B30C0133 C4363F2F 73454D51 E9F95962
        EDE9E590 E7458FA6 765A0D3F C4047BC2
        9C391FF0 7383C4DF 9A0CCFA9 231428AF
        987B029C C857AD96 E4C92441 9382E798
        8FCC1E4A 3E598D81 96476875 E2F86C33
        75B51661 B6556C5E 8F546E97 5197734B
        C8C29AC7 E427C8E4 B9AAF5AA 80A75B3C
```

关于获取的 CA 证书及 CRL 文件的详细信息可以通过相应的显示命令来查看，此处略。具体内容请参考命令 **display pki certificate ca domain** 和 **display pki crl domain**。

1.12.2 PKI 实体向 CA 申请证书（方式二）

说明

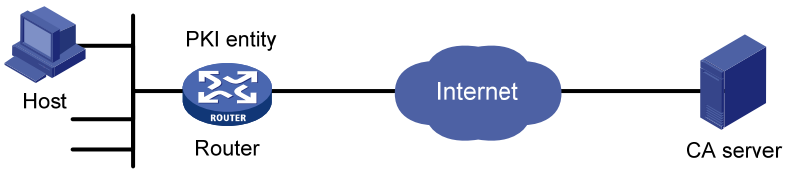
本配置举例中，采用 Windows 2003 server 做 CA 服务器。

1. 组网需求

配置 PKI 实体 Router 向 CA 服务器申请本地证书。

2. 组网图

图1-3 PKI 实体向 CA 申请证书组网图



3. 配置步骤

- CA 服务器端的配置

(1) 安装证书服务器组件

打开[控制面板]/[添加/删除程序]，选择[添加/删除 Windows 组件]中的“证书服务”进行安装。

(2) 安装 SCEP 插件

由于 Windows 2003 server 作为 CA 服务器时，缺省情况下不支持 SCEP，所以需要安装 SCEP 插件，才能使设备具备证书自动注册、获取等功能。插件安装完毕后，弹出提示框，提示框中的 URL 地址即为设备上配置的注册服务器地址。

(3) 修改证书服务的属性

完成上述配置后，打开[控制面板/管理工具]中的[证书颁发机构]，如果安装成功，在[颁发的证书]中将存在两个 CA 颁发给 RA 的证书。选择[CA server 属性]中的“策略模块”的属性为“如果可以的话，按照证书模板中的设置。否则，将自动颁发证书(F)。”

(4) 修改 IIS 服务的属性

打开[控制面板/管理工具]中的[Internet 信息服务(IIS)管理器]，将[默认网站 属性]中“主目录”的本地路径修改为证书服务保存的路径。另外，为了避免与已有的服务冲突，建议修改默认网站的 TCP 端口号为未使用的端口号。

以上配置完成之后，还需要保证设备的系统时钟与 CA 的时钟同步才可以正常使用设备来申请证书。

- 设备 Router 上的配置

(1) 配置实体命名空间

配置实体名称为 aaa，通用名为 router。

```
<Router> system-view
[Router] pki entity aaa
[Router-pki-entity-aaa] common-name router
[Router-pki-entity-aaa] quit
```

(2) 配置 PKI 域参数

创建并进入 PKI 域 torsa。

```
[Router] pki domain torsa
```

配置可信任的 CA 名称为 myca。

```
[Router-pki-domain-torsa] ca identifier myca
```

配置注册服务器 URL，格式为 http://host:port/ certsrv/mscep/mscep.dll。其中，host:port 为 CA 服务器的主机地址和端口号。

```
[Router-pki-domain-torsa] certificate request url
http://4.4.4.1:8080/certsrv/mscep/mscep.dll
```

配置证书申请的注册受理机构为 RA。

```
[Router-pki-domain-torsa] certificate request from ra
```

指定实体名称为 aaa。

```
[Router-pki-domain-torsa] certificate request entity aaa
```

(3) 用 RSA 算法生成本地密钥对

```
[Router] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It may take a few minutes.
Press CTRL+C to abort.
Input the bits in the modulus [default = 1024]:
Generating Keys...
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++
```

(4) 证书申请

获取 CA 证书并下载至本地。

```
[Router] pki retrieval-certificate ca domain torsa
```

```
Retrieving CA/RA certificates. Please wait a while......
The trusted CA's finger print is:
    MD5  fingerprint:766C D2C8 9E46 845B 4DCE 439C 1C1F 83AB
    SHA1 fingerprint:97E5 DDED AB39 3141 75FB DB5C E7F8 D7D7 7C9B 97B4

Is the finger print correct?(Y/N):y

Saving CA/RA certificates chain, please wait a moment......
CA certificates retrieval success.
```

手工申请本地证书。

```
[Router] pki request-certificate domain torsa challenge-word
Certificate is being requested, please wait......
[Router]
Enrolling the local certificate,please wait a while......
Certificate request Successfully!
Saving the local certificate to device......
Done!
```

4. 验证配置结果

通过以下显示命令可以查看获取的本地证书信息。

```
<Router> display pki certificate local domain torsa
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            48FA0FD9 00000000 000C
        Signature Algorithm: sha1WithRSAEncryption
        Issuer:
            CN=CA server
        Validity
            Not Before: Nov 21 12:32:16 2007 GMT
            Not After : Nov 21 12:42:16 2008 GMT
        Subject:
            CN=router
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00A6637A 8CDEA1AC B2E04A59 F7F6A9FE
                    5AEE52AE 14A392E4 E0E5D458 0D341113
                    0BF91E57 FA8C67AC 6CE8FEBB 5570178B
                    10242FDD D3947F5E 2DA70BD9 1FAF07E5
                    1D167CE1 FC20394F 476F5C08 C5067DF9
                    CB4D05E6 55DC11B6 9F4C014D EA600306
                    81D403CF 2D93BC5A 8AF3224D 1125E439
                    78ECEFE1 7FA9AE7B 877B50B8 3280509F
                    6B
```

```
            Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Subject Key Identifier:
            B68E4107 91D7C44C 7ABCE3BA 9BF385F8 A448F4E1
            X509v3 Authority Key Identifier:
            keyid:9D823258 EADFEFA2 4A663E75 F416B6F6 D41EE4FE

            X509v3 CRL Distribution Points:
            URI:http://l00192b/CertEnroll/CA%20server.crl
            URI:file://\\l00192b\CertEnroll\CA server.crl

            Authority Information Access:
            CA Issuers - URI:http://l00192b/CertEnroll/l00192b_CA%20server.crt
            CA Issuers - URI:file://\\l00192b\CertEnroll\l00192b_CA server.crt

            1.3.6.1.4.1.311.20.2:
                .0.I.P.S.E.C.I.n.t.e.r.m.e.d.i.a.t.e.O.f.f.l.i.n.e
    Signature Algorithm: sha1WithRSAEncryption
        81029589 7BFA1CBD 20023136 B068840B
```

……（略）

关于获取的 CA 证书的详细信息可以通过相应的显示命令来查看，此处略。具体内容请参考命令**display pki certificate ca domain**。

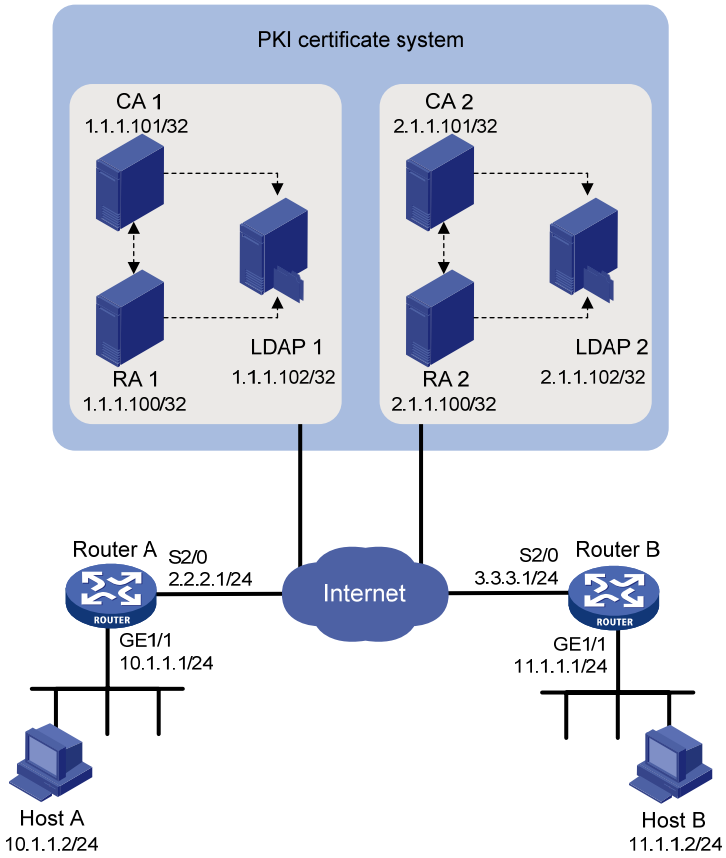
1.12.3 使用 PKI 证书体系的 RSA 证书签名方法进行 IKE 协商认证

1. 组网需求

- 在路由器 A 和路由器 B 之间建立一个 IPsec 安全隧道对子网 10.1.1.0/24 上的主机 A 与子网 11.1.1.0/24 上的主机 B 之间的数据流进行安全保护。
- 在路由器 A 和路由器 B 之间使用 IKE 自动协商建立安全通信，IKE 认证策略采用 PKI 证书体系的 RSA 证书签名方法进行身份认证。
- 图 1-4中路由器A和路由器B使用不同的CA（可以相同，根据实际情况确定）。

2. 组网图

图1-4 PKI 进行 IKE 协商认证的组网图



3. 配置步骤

(1) 配置 Route A

配置实体命名空间。

```
<RouterA> system-view
[RouterA] pki entity en
[RouterA-pki-entity-en] ip 2.2.2.1
[RouterA-pki-entity-en] common-name routera
[RouterA-pki-entity-en] quit
```

配置 PKI 域参数。（证书申请的注册机构服务器的 URL 根据所使用的 CA 服务器的不同而有所不同，这里的配置只作为示例，请根据具体情况配置。）

```
[RouterA] pki domain 1
[RouterA-pki-domain-1] ca identifier CA1
[RouterA-pki-domain-1] certificate request url http://1.1.1.100/certsrv/mscep/mscep.dll
[RouterA-pki-domain-1] certificate request entity en
[RouterA-pki-domain-1] ldap-server ip 1.1.1.102
```

配置通过 RA 注册申请证书。

```
[RouterA-pki-domain-1] certificate request from ra
```

配置 CRL 发布点位置（若禁止 CRL 检查，则无需配置）。

```
[RouterA-pki-domain-1] crl url ldap://1.1.1.102
[RouterA-pki-domain-1] quit
```

用 RSA 算法生成本地的密钥对。

```
[RouterA] public-key local create rsa
```

证书申请。

```
[RouterA] pki retrieval-certificate ca domain 1
[RouterA] pki retrieval-crl domain 1
[RouterA] pki request-certificate domain 1
```

在 Router A 上使用 IKE 策略 1，配置使用数字签名（rsa-signature）方法为身份认证策略。

```
[RouterA] ike proposal 1
[RouterA-ike-proposal-1] authentication-method rsa-signature
[RouterA-ike-proposal-1] quit
```

在 IKE 对等体中引用 PKI 域的配置。

```
[RouterA]ike peer peer
[RouterA-ike-peer-peer]certificate domain 1
```

(2) 配置 Router B

配置实体命名空间。

```
<RouterB> system-view
[RouterB] pki entity en
[RouterB-pki-entity-en] ip 3.3.3.1
[RouterB-pki-entity-en] common-name routerb
[RouterB-pki-entity-en] quit
```

配置 PKI 域参数。（证书申请的注册机构服务器的 URL 根据所使用的 CA 服务器的不同而有所不同，这里的配置只作为示例，请根据具体情况配置。）

```
[RouterB] pki domain 1
[RouterB-pki-domain-1] ca identifier CA2
[RouterB-pki-domain-1] certificate request url http://2.1.1.100/certsrv/mscep/mscep.dll
[RouterB-pki-domain-1] certificate request entity en
[RouterB-pki-domain-1] ldap-server ip 2.1.1.102
```

配置通过 RA 注册申请证书。

```
[RouterB-pki-domain-1] certificate request from ra
```

配置 CRL 发布点（若禁止 CRL 检查，则无需配置）。

```
[RouterB-pki-domain-1] crl url ldap://2.1.1.102
[RouterB-pki-domain-1] quit
```

用 RSA 算法生成本地的密钥对。

```
[RouterB] public-key local create rsa
```

证书申请。

```
[RouterB] pki retrieval-certificate ca domain 1
[RouterB] pki retrieval-crl domain 1
[RouterB] pki request-certificate domain 1
```

在 Router B 上使用 IKE 策略 1，配置使用 rsa-signature 方法为身份认证策略。

```
[RouterB] ike proposal 1
[RouterB-ike-proposal-1] authentication-method rsa-signature
```

```
[RouterB-ike-proposal-1] quit
```

在 IKE 对等体中引用 PKI 域的配置。

```
[RouterB]ike peer peer
[RouterB-ike-peer-peer]certificate domain 1
```



说明

以上是对 IKE 协商采用 RSA 数字签名认证方法的配置，若希望建立 IPsec 安全通道进行安全通信，还需要进行 IPsec 的相应配置，具体内容请参见“安全分册”中“IPsec 配置”。

1.13 常见配置错误举例

1.13.1 获取 CA 证书失败

1. 故障现象

获取 CA 证书时失败。

2. 故障分析

可能有以下原因：

- 网络连接故障，如网线折断，接口松动；
- 没有设置信任的 CA 名称；
- 证书申请的注册服务器 URL 位置不正确或未配置；
- 没有配置证书申请注册受理机构；
- 设备的系统时钟与 CA 的时钟不同步。

3. 处理过程

- 排除物理连接故障；
- 查看各必配项是否都正确配置；
- 可通过 **ping** 命令测试注册服务器是否连接正常；
- 配置证书申请注册受理机构；
- 保持系统时钟与 CA 同步。

1.13.2 本地证书申请失败

1. 故障现象

手工证书请求失败。

2. 故障分析

可能有以下原因：

- 网络连接故障，如网线折断，接口松动；
- 申请之前没有先获取 CA 证书；
- 当前的密钥对已经绑定证书；
- 没有设置信任的 CA 名称；

- 证书申请的注册服务器 URL 位置不正确或未配置；
- 没有配置证书申请注册受理机构；
- 没有配置实体 DN 中必配参数。

3. 处理过程

- 排除物理连接故障；
- 获取 CA 证书；
- 重新创建密钥对；
- 配置信任的 CA 名称；
- 可通过 **ping** 命令测试注册服务器是否连接正常；
- 配置证书申请注册受理机构；
- 可通过查看 CA/RA 注册策略选择相关的实体 DN 属性进行配置。

1.13.3 CRL 获取失败

1. 故障现象

获取 CRL 发生失败。

2. 故障分析

可能有以下原因：

- 网络连接故障，如网线折断，接口松动；
- 获取 CRL 之前未先取得 CA 证书；
- 未设置 LDAP 服务器的 IP 地址；
- 未设置 CRL 分布点位置；
- LDAP 服务器版本配置错误。

3. 处理过程

- 排除物理连接故障；
- 获取 CA 证书；
- 设置 LDAP 服务器的 IP 地址；
- 设置 CRL 分布点位置；
- 重新配置 LDAP 版本。

目 录

1 会话管理配置 1-1
1.1 会话管理简介 1-1
1.1.1 会话管理的工作原理 1-1
1.1.2 会话管理在设备上的实现 1-1
1.2 配置会话管理 1-2
1.2.1 配置协议状态会话超时时间 1-2
1.2.2 配置应用层协议会话超时时间 1-2
1.2.3 使能校验和检查 1-3
1.2.4 配置长连接会话规则 1-3
1.2.5 删除会话 1-4
1.3 配置会话日志 1-4
1.3.1 使能会话日志功能 1-4
1.3.2 配置会话日志阈值 1-5
1.3.3 配置会话日志的输出 1-5
1.4 会话管理显示和维护 1-6

1 会话管理配置

1.1 会话管理简介

会话管理是为了实现如 NAT、ASPF 等基于会话进行处理的业务而抽象出来的公共功能。此功能把传输层报文之间的交互关系抽象为会话，并根据发起方或响应方的报文信息对会话进行状态更新和超时老化。

会话管理支持多个业务特性分别对同一个业务报文进行处理，实现的主要功能包括：

- 报文到会话的快速匹配；
- 传输层协议状态的管理；
- 报文应用层协议类型的识别；
- 支持会话按照协议状态或应用层协议类型进行老化；
- 支持指定会话维持永久连接；
- 会话的传输层协议报文校验和检查；
- 为需要进行端口协商的应用层协议提供特殊的报文匹配；
- 支持对 ICMP 差错控制报文的解析以及根据解析结果进行会话的匹配。

1.1.1 会话管理的工作原理

会话管理主要基于传输层协议对报文进行检测。其实质是通过检测传输层协议信息（即通用 TCP 协议和 UDP 协议）来对连接的状态进行跟踪，并对所有连接的状态信息进行统一维护和管理。

在实际应用中，会话管理配合 ASPF 特性，可实现根据连接状态信息动态地决定数据包是否被允许通过防火墙进入内部区域，以便阻止恶意的入侵。

需要注意的是，会话管理作为基础特性，只是实现连接跟踪，模块本身并不阻止潜在的攻击报文通过。

1.1.2 会话管理在设备上的实现

目前会话管理在设备上实现的具体功能如下：

- 支持 TCP、UDP、ICMP、Raw IP 等 IPv4 报文的会话创建、会话状态更新以及根据协议状态设置超时时间。
- 支持应用层协议的端口映射，允许为应用层协议自定义对应的非通用端口号，同时可以根据应用层协议设置不同会话超时时间。
- 支持 TCP、UDP、ICMP 报文的校验和检查。在校验和检查失败的情况下，不进行会话的匹配或创建，而是由基于会话管理的其他业务来处理。
- 支持 ICMP 差错报文的映射，可以根据 ICMP 差错报文的内层报文查找原始的会话。另外，由于差错报文都是由于某主机出错后产生的，因而可以加速该原始会话的超时老化。
- 支持设置长连接会话，保证指定的会话永不超时。
- 支持应用层协议（如 FTP 等协议）的控制通道和动态数据通道的会话管理。

- 支持对基于会话的连接数目进行限制。相关介绍及配置请参见“安全分册”中的“连接限制配置”。

1.2 配置会话管理

会话管理可支持的配置包括：协议状态超时时间、应用层协议会话超时时间、使能会话校验和检查、长连接会话及删除会话。这些配置可根据实际应用需求选择进行，配置无先后，相互不关联。

1.2.1 配置协议状态会话超时时间

说明

- 此会话超时时间只对正在建立过程中的会话有效。
- 如果会话所属的应用层协议支持会话超时时间的配置，则该会话处于READY/ESTABLISH状态时，会话的超时时间就为对应的应用层协议会话超时时间，具体配置请参见 1.2.2 配置应用层协议会话超时时间。

已经建立的会话表项如果在一定时间内未被任何报文匹配，则会由于超时而被系统自动删除。以下配置用于实现根据会话所处协议状态来设置会话表项的超时时间。

表1-1 配置各协议状态的会话超时时间

操作	命令	说明
进入系统视图	**system-view**	-
配置各协议状态的会话超时时间	**session aging-time** { **accelerate** | **fin** | **icmp-closed** | **icmp-open** | **rawip-open** | **rawip-ready** | **syn** | **tcp-est** | **udp-open** | **udp-ready** } *time-value*	必选 缺省情况下，各协议状态的会话超时时间为： • **accelerate**：10s • **fin**：30s • **icmp-closed**：30s • **icmp-open**：60s • **rawip-open**：30s • **rawip-ready**：60s • **syn**：30s • **tcp-est**：3600s • **udp-open**：30s • **udp-ready**：60s

注意

当会话数目过多时（大于 80 万条），建议不要将协议状态超时时间设置得过短。否则会造成控制台响应速度过慢。

1.2.2 配置应用层协议会话超时时间

说明

此会话超时时间只对已经建立并处于 READY/ESTABLISH 状态的会话有效。

对于处于已建立状态的会话（TCP 握手建立连接成功或 UDP 进入 READY 状态），用户可以根据会话所属的应用层协议类型设置超时时间。

表1-2 配置应用层协议会话超时时间

操作	命令	说明
进入系统视图	**system-view**	-
配置应用层协议的会话超时时间	**application aging-time** { **dns** | **ftp** | **msn** | **qq** } *time-value*	必选 缺省情况下，各应用层协议的会话超时时间为： • **dns**：60 秒 • **ftp**：3600 秒 • **msn**：3600 秒 • **qq**：60 秒

注意

当会话数目过多时（大于 80 万条），建议不要将应用层协议会话的超时时间设置得过短。否则会造成控制台响应速度过慢。

1.2.3 使能校验和检查

为保证对会话的连接跟踪不被校验和发生错误的报文所干扰，可以使能对协议报文的校验和检查功能。此功能可以确保会话管理只处理校验和正确的报文，而校验和错误的报文由基于会话管理的其它业务来处理。

表1-3 配置使能校验和检查

操作	命令	说明
进入系统视图	**system-view**	-
配置使能校验和检查	**session checksum** { **all** | { **icmp** | **tcp** | **udp** } * }	必选 缺省情况下，不进行校验和检查

注意

启用此功能后可能会导致设备处理性能下降，请慎用。

1.2.4 配置长连接会话规则

用户可以根据需要将一些符合给定特征的会话设置为永不超时。即这些会话的老化时间不会随着状态的变迁而更改，同时也不会由于没有报文命中而被删除。此类会话一旦建立，将永久连接，直到会话的发起方或响应方主动发起关闭连接请求或管理员手动删除会话。

1. 配置准备

需要根据长连接会话特征配置基本或高级 ACL（Access Control List，访问控制列表），满足该 ACL 规则的会话为长连接会话。

说明

基本或高级访问控制列表的具体配置请参见“安全分册”中的“ACL 配置”。

2. 配置长连接会话规则

表1-4 配置长连接会话规则

操作	命令	说明
进入系统视图	**system-view**	-
配置长连接会话规则	**session persist acl** *acl-number*	必选 缺省情况下，无长连接会话规则

说明

一个长连接会话规则只能引用一个 ACL。

1.2.5 删除会话

通过以下配置可以手动删除会话。

表1-5 删除会话

操作		命令	说明
删除会话	集中式设备	**reset session** [**source-ip** *source-ip*] [**destination-ip** *destination-ip*] [**protocol-type** *protocol-type*] [**source-port** *source-port*] [**destination-port** *destination-port*] [**vpn-instance** *vpn-instance-name*]	必选 此命令在用户视图下执行
	分布式设备	**reset session** [**slot** *slot-num*] [**source-ip** *source-ip*] [**destination-ip** *destination-ip*] [**protocol-type** *protocol-type*] [**source-port** *source-port*] [**destination-port** *destination-port*] [**vpn-instance** *vpn-instance-name*]	

1.3 配置会话日志

会话日志是为满足网络管理员安全审计的需要，对用户的访问信息、用户 IP 地址的转换信息、用户的网络流量信息等进行的记录，可以通过一定的格式发给日志服务器主机。

1.3.1 使能会话日志功能

表1-6 使能会话日志功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能会话日志功能	**session log enable** [**acl** *acl-number*] { **inbound** | **outbound** }	必选 缺省情况下，会话日志功能处于关闭状态

1.3.2 配置会话日志阈值

存活时间或收发数目达到一定阈值的会话才会以日志的形式进行记录并输出，该阈值包括以下两种类型：

- 时间阈值：当一个会话存在的时间达到设定的时间阈值时，输出会话日志。
- 流量阈值：分为报文数阈值和字节数阈值两种。当一个会话收发的报文数或字节数达到设定的流量阈值时，输出会话日志。

表1-7 配置会话日志阈值

操作		命令	说明
进入系统视图		**system-view**	-
配置输出会话日志的时间阈值		**session log time-active** *time-value*	可选 缺省情况下，时间阈值为 0，表示不依据时间阈值发送会话日志
配置输出会话日志的流量阈值	配置报文数流量阈值	**session log packets-active** *packets-value*	可选 缺省情况下，报文数流量阈值为 0，表示不依据报文数流量阈值发送会话日志
	配置字节数流量阈值	**session log bytes-active** *bytes-value*	可选 缺省情况下，字节数流量阈值为 0，表示不依据字节数流量阈值发送会话日志

说明

- 同时配置了时间阈值和流量阈值的情况下，只要有一个阈值首先到达，就会输出相应的会话日志，并将会话的所有统计信息清零。
- 同时仅有一种流量阈值有效，以最后一次配置的阈值类型为准，例如，先配置报文数阈值再配置字节数阈值，则当前有效的阈值是字节数阈值，只会输出达到字节数阈值的会话日志。

1.3.3 配置会话日志的输出

会话日志采用 Flow 日志的形式输出，下面为具体的配置。关于 Flow 日志各项功能的具体介绍请参见“系统分册”中的“Flow 日志配置”，此处不做详细描述。

表1-8 配置会话日志的输出

操作		命令	说明
进入系统视图		**system-view**	-
配置 Flow 日志版本信息		**userlog flow export version** *version-number*	可选 缺省情况下，Flow 日志报文的版本号为 1.0
设置承载 Flow 日志的 UDP 报文的源 IP 地址		**userlog flow export source-ip** *ip-address*	可选 缺省情况下，承载 Flow 日志的 UDP 报文的源 IP 地址为发送该报文的接口的 IP 地址
设置 Flow 日志服务器的 IP 地址和 UDP 端口号	集中式设备	**userlog flow export** [**vpn-instance** *vpn-instance-name*] **host** *ip-address udp-port*	必选 缺省情况下，没有配置 Flow 日志服务器的 IP 地址和 UDP 端口号
	分布式设备	**userlog flow export slot** *slot-number* [**vpn-instance** *vpn-instance-name*] **host** *ip-address udp-port*	
配置 Flow 日志输出到信息中心		**userlog flow syslog**	可选 缺省情况下，Flow 日志输出到 Flow 日志服务器

说明

以上 Flow 日志的相关命令请参见“系统分册”中的“Flow 日志命令”。

1.4 会话管理显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后会话的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除会话统计信息。

表1-9 会话管理显示和维护

操作	命令
显示会话表的信息（集中式设备）	**display session table** [**source-ip** *source-ip*] [**destination-ip** *destination-ip*] [**verbose**]
显示会话表的信息（分布式设备）	**display session table** [**slot** *slot-num*] [**source-ip** *source-ip*] [**destination-ip** *destination-ip*] [**verbose**]
显示会话统计信息（集中式设备）	**display session statistics**
显示会话统计信息（分布式设备）	**display session statistics** [**slot** *slot-num*]
显示会话关联表信息（集中式设备）	**display session relation-table**
显示会话关联表信息（分布式设备）	**display session relation-table** [**slot** *slot-num*]
查看日志的配置和统计信（集中式设备）	**display userlog export**
查看日志的配置和统计信（分布式设备）	**display userlog export slot** *slot-number*

操作	命令
清除会话统计信息（集中式设备）	**reset session statistics**
清除会话统计信息（分布式设备）	**reset session statistics** [**slot** *slot-num*]
清除所有日志的统计信息（集中式设备）	**reset userlog flow export**
清除所有日志的统计信息（分布式设备）	**reset userlog flow export slot** *slot-number*
清除 FLOW 日志缓存中的记录（集中式设备）	**reset userlog flow logbuffer**
清除 FLOW 日志缓存中的记录（分布式设备）	**reset userlog flow logbuffer slot** *slot-number*



说明

display userlog export 命令的具体介绍请参见“安全分册”中的“NAT 命令”；**reset userlog flow logbuffer** 和 **reset userlog flow export** 命令的具体介绍请参见“系统分册”中的“Flow 日志命令”。

目 录

1 NAT配置 1-1
1.1 NAT简介 1-1
1.1.1 NAT概述 1-1
1.1.2 NAT实现的功能 1-2
1.2 NAT配置任务简介 1-5
1.3 配置地址池 1-5
1.4 配置地址转换 1-6
1.4.1 地址转换介绍 1-6
1.4.2 配置静态地址转换 1-6
1.4.3 配置动态地址转换 1-7
1.5 配置内部服务器 1-8
1.5.1 内部服务器介绍 1-8
1.5.2 配置普通内部服务器 1-8
1.6 配置DNS-map 1-9
1.7 配置基于IP地址的连接数限制 1-9
1.8 NAT显示和维护 1-9
1.9 NAT典型配置举例 1-10
1.9.1 私网访问公网典型配置举例 1-10
1.9.2 普通内部服务器典型配置举例 1-11
1.9.3 NAT DNS-mapping典型配置举例 1-12
1.10 NAT常见配置错误举例 1-13
1.10.1 故障之一：地址转换不正常。 1-13
1.10.2 故障之二：内部服务器工作不正常。 1-13

1 NAT 配置

1.1 NAT 简介

1.1.1 NAT 概述

NAT（Network Address Translation，网络地址转换）是将 IP 数据报报头中的 IP 地址转换为另一个 IP 地址的过程。在实际应用中，NAT 主要用于实现私有网络访问公共网络的功能。这种通过使用少量的公有 IP 地址代表较多的私有 IP 地址的方式，将有助于减缓可用 IP 地址空间的枯竭。

说明

私有 IP 地址是指内部网络或主机的 IP 地址，公有 IP 地址是指在因特网上全球唯一的 IP 地址。

RFC 1918 为私有网络预留出了三个 IP 地址块，如下：

A 类：10.0.0.0 ~ 10.255.255.255

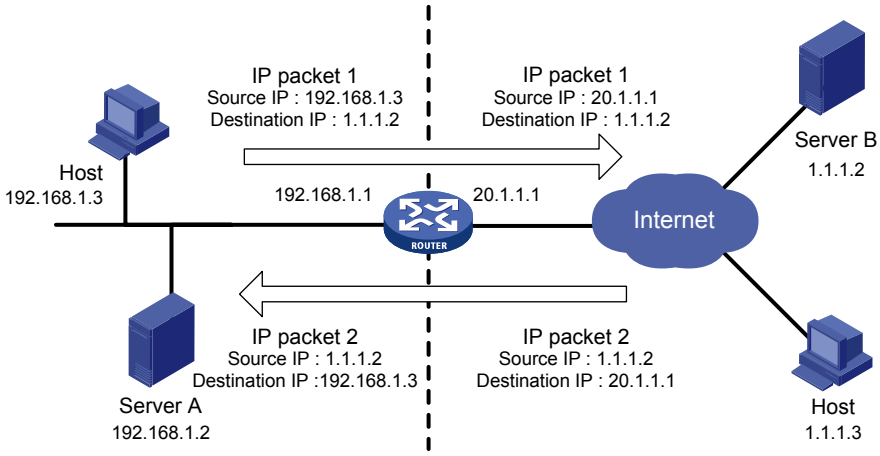
B 类：172.16.0.0 ~ 172.31.255.255

C 类：192.168.0.0 ~ 192.168.255.255

上述三个范围内的地址不会在因特网上被分配，因此可以不必向 ISP 或注册中心申请而在公司或企业内部自由使用。

图 1-1描述了一个基本的NAT应用。

图1-1 地址转换的基本过程



- NAT 网关处于私有网络和公有网络的连接处。
- 当内部 PC（192.168.1.3）向外部服务器（1.1.1.2）发送一个数据报 1 时，数据报将通过 NAT 网关。
- NAT 网关查看报头内容，发现该数据报是发往外网的，那么它将数据报 1 的源地址字段的私有地址 192.168.1.3 换成一个可在 Internet 上选路的公有地址 20.1.1.1，并将该数据报发送到外部服务器，同时在 NAT 网关的网络地址转换表中记录这一映射。

- 外部服务器给内部 PC 发送的应答报文 2（其初始目的地址为 20.1.1.1）到达 NAT 网关后，NAT 网关再次查看报头内容，然后查找当前网络地址转换表的记录，用内部 PC 的私有地址 192.168.1.3 替换初始的目的地址。

上述的 NAT 过程对终端（如图中的 Host 和 Server）来说是透明的。对外部服务器而言，它认为内部 PC 的 IP 地址就是 20.1.1.1，并不知道有 192.168.1.3 这个地址。因此，NAT“隐藏”了企业的私有网络。

地址转换的优点在于，在为内部主机提供了“隐私”保护的前提下，实现了内部网络的主机通过该功能访问外部网络的资源。但它也有一些缺点：

- 由于需要对数据报文进行 IP 地址的转换，涉及 IP 地址的数据报的报头不能被加密。在应用协议中，如果报文中有地址或端口需要转换，则报文不能被加密。例如，不能使用加密的 FTP 连接，否则 FTP 的 port 命令不能被正确转换。
- 网络调试变得更加困难。比如，某一台内部网络的主机试图攻击其它网络，则很难指出究竟哪一台机器是恶意的，因为主机的 IP 地址被屏蔽了。

1.1.2 NAT 实现的功能

1. 多对多地址转换及地址转换的控制

从 图 1-1的地址转换过程可见，当内部网络访问外部网络时，地址转换将会选择一个合适的外部地址，来替代内部网络数据报文的源地址。在 图 1-1中是选择NAT网关出接口的IP地址（公有地址）。这样所有内部网络的主机访问外部网络时，只能拥有一个外部的IP地址，因此，这种情况同时只允许最多有一台内部主机访问外部网络，这称为“一对一地址转换”。当内部网络的多台主机并发的要求访问外部网络时，“一对一地址转换”仅能够实现其中一台主机的访问请求。

NAT 也可实现对并发性请求的响应，允许 NAT 网关拥有多个公有 IP 地址。当第一个内部主机访问外网时，NAT 选择一个公有地址 IP1，在地址转换表中添加记录并发送数据报；当另一内部主机访问外网时，NAT 选择另一个公有地址 IP2，以此类推，从而满足了多台内部主机访问外网的请求。这称为“多对多地址转换”。



说明

NAT 网关拥有的公有 IP 地址数目要远少于内部网络的主机数目，因为所有内部主机并不会同时访问外网。公有 IP 地址数目的确定，应根据网络高峰期可能访问外网的内部主机数目的统计值来确定。

在实际应用中，我们可能希望某些内部的主机可以访问外部网络，而某些主机不允许访问，即当 NAT 网关查看数据报报头内容时，如果发现源 IP 地址属于禁止访问外部网络的内部主机，它将不进行 NAT 转换。这就是对地址转换进行控制的问题。

设备可以通过定义地址池来实现多对多地址转换，同时利用访问控制列表来对地址转换进行控制。

- 利用访问控制列表限制地址转换：可以有效地控制地址转换的使用范围，只有满足访问控制列表条件的数据报文才可以进行地址转换。
- 地址池：用于地址转换的一些连续的公有 IP 地址的集合。用户应根据自己拥有的合法 IP 地址数目、内部网络主机数目以及实际应用情况，配置恰当的地址池。地址转换的过程中，NAT 网关将会从地址池中挑选一个地址做为转换后的源地址。

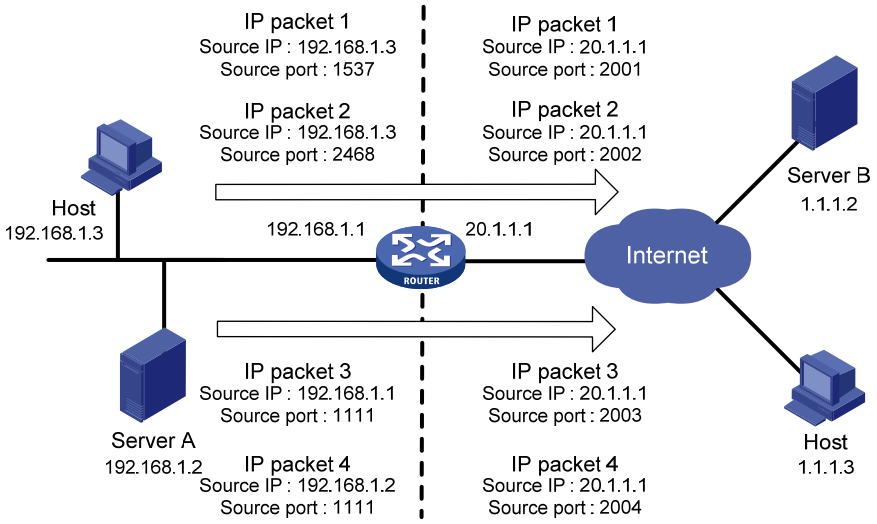
2. NAPT

NAPT（Network Address Port Translation，网络地址端口转换）是 NAT 的一种变形，它允许多个内部地址映射到同一个公有地址上，也可称之为“多对一地址转换”或“地址复用”。

NAPT 同时映射 IP 地址和端口号：来自不同内部地址的数据报的目的地址可以映射到同一外部地址，但它们的端口号被转换为该地址的不同端口号，因而仍然能够共享同一地址，也就是“私有地址＋端口”与“公有地址＋端口”之间的转换。

图 1-2描述了NAPT的基本原理。

图1-2 NAPT 基本原理示意图



如 图 1-2所示，四个带有内部地址的数据报到达NAT网关，其中数据报 1 和 2 来自同一个内部地址但有不同的源端口号，数据报 3 和 4 来自不同的内部地址但具有相同的源端口号。通过NAPT映射，四个数据报的源IP地址都被转换到同一个外部地址，但每个数据报都被赋予了不同的源端口号，因而仍保留了报文之间的区别。当回应报文到达时，NAT网关仍能够根据回应报文的目的地址和端口号来区别该报文应转发到的内部主机。

采用 NAPT 可以更加充分地利用 IP 地址资源，实现更多内部网络主机主机对外部网络的同时访问。

3. 内部服务器

NAT 隐藏了内部网络的结构，具有“屏蔽”内部主机的作用，但是在实际应用中，可能需要给外部网络提供一个访问内部主机的机会，如给外部网络提供一台 Web 服务器，或是一台 FTP 服务器。

使用 NAT 可以灵活地添加内部服务器。例如，可以使用 20.1.1.10 作为 Web 服务器的外部地址；使用 20.1.1.11 作为 FTP 服务器的外部地址；甚至还可以使用 20.1.1.12:8080 这样的地址作为 Web 服务器的外部地址。

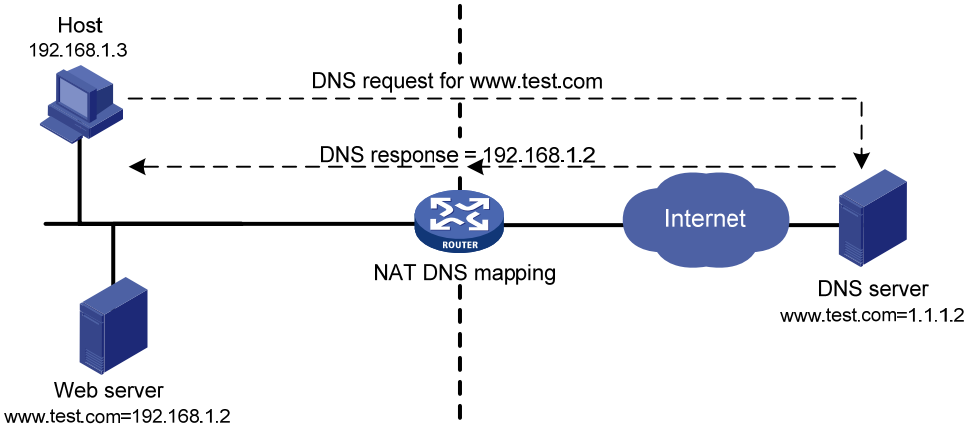
目前设备的 NAT 提供了内部服务器功能供外部网络访问。外部网络的用户访问内部服务器时，NAT 将请求报文内的目的地址转换成内部服务器的私有地址。当内部服务器回应报文时，NAT 要将回应报文的源地址（私有 IP 地址）转换成公有 IP 地址。

4. DNS-mapping

由前面的介绍可知，在 NAT 网关的公网接口上，通过配置内部服务器，可以将公网地址、端口等映射到私网内的服务器上，公网用户可以通过域名或公网地址，访问内部服务器。

在实际应用中，可能会出现这样的情况，DNS 服务器在公网，私网用户希望通过域名访问同一私网内的内部服务器，这就需要配置 DNS-map。DNS-map 通过配置“域名-公网地址-公网端口-协议类型”的映射表，建立域名与内部服务器公网信息的对应关系。在配置了 NAT 的接口上，设备检查接收到的 DNS 响应报文，根据报文中的域名查找用户配置的 DNS-map 映射表，并根据表项内的“公网地址-公网端口-协议类型”查找内部服务器，然后用内部服务器的私网地址替换 DNS 查询结果中的公网地址。这样私网用户收到的 DNS 响应报文中就包含了要访问的私网内部服务器的私网地址，因此就可以使用内部服务器域名访问同一私网内的内部服务器。

图1-3 NAT DNS mapping 工作示意图



在实际应用中，如 图 1-3所示，若DNS服务器在公网，私网用户希望通过域名访问同一私网内的内部服务器。通常，DNS服务器向私网用户响应的报文中包含私网服务器的公网地址，若不在NAT设备上进行相应的处理，会导致收到响应报文的私网用户无法正确利用域名访问私网服务器，该问题可以通过在设备上配置DNS mapping来解决。

DNS mapping 通过配置“域名-公网地址-公网端口-协议类型”的映射表，建立内部服务器域名与内部服务器公网信息的对应关系。在配置了 NAT 的接口上，设备检查接收到的 DNS 响应报文，根据报文中的域名查找用户配置的 DNS mapping 映射表，并根据表项内的“公网地址-公网端口-协议类型”查找内部服务器，然后用内部服务器的私网地址替换 DNS 查询结果中的公网地址。这样私网用户收到的 DNS 响应报文中就包含了要访问的私网内部服务器的私网地址，就可以使用内部服务器域名访问同一私网内的内部服务器。

5. Easy IP

Easy IP 是指进行地址转换时，直接使用接口的公有 IP 地址作为转换后的源地址，能够最大程度的节省 IP 地址资源。它也可以利用访问控制列表控制哪些内部地址可以进行地址转换。

6. NAT 支持的特殊协议

NAT 不仅实现了一般的地址转换功能，同时提供了完善的地址转换 ALG （Application Layer Gateway，应用级网关）机制，使其可以支持一些各种特殊的应用协议，而不需要对 NAT 平台进行任何的修改，具有良好的可扩充性。这些特殊协议的报文载荷里携带了地址或端口信息，该信息也可能需要进行地址转换。

可支持的特殊协议包括：FTP（File Transfer Protocol，文件传输协议）、ICMP（Internet Control Message Protocol，Internet 控制消息协议）、DNS（Domain Name System，域名系统）等。

7. 支持 NAT 多实例

NAT 多实例允许分属于不同 MPLS VPN 的用户通过同一个出口访问外部网络，同时允许分属于不同 MPLS VPN 的用户使用相同的私网地址。当 MPLS VPN 用户访问外部网络时，地址转换将内部网络主机的 IP 地址和端口替换为设备的外部网络地址和端口，同时还记录了用户的 MPLS VPN 信息（如协议类型和路由标识符 RD 等）。回应报文到达时，地址转换将外部网络地址和端口还原为内部网络主机的 IP 地址和端口，同时可得知是哪一个 MPLS VPN 用户的访问。无论是 NAT 方式的地址转换还是 NAPT 方式的地址转换都支持多实例。

同时，地址转换支持内部服务器的多实例，给外部提供访问 MPLS VPN 内主机的机会。例如，MPLS VPN1 内提供 Web 服务的主机地址是 10.110.1.1，可以使用 202.110.10.20 作为 Web 服务器的外部地址，Internet 的用户使用 202.110.10.20 的地址就可以访问到 MPLS VPN1 提供的 Web 服务。

另外，NAT 还可利用外部网络地址所携带的 MPLS VPN 信息，支持多个 MPLS VPN 之间的互访。

1.2 NAT 配置任务简介

表1-1 NAT 配置任务简介

配置任务		说明	详细配置
配置地址池		可选 配置动态地址转换时必配	1.3
配置地址转换	配置静态地址转换	二者必选其一	1.4
	配置动态地址转换		
配置内部服务器		必选	1.5
配置 DNS mapping		可选	1.6
配置基于源地址的连接数限制		可选	1.7

说明

- 接口下的 NAT 相关配置（地址转换或内部服务器配置）改变时，为保证连接的稳定性，建议用户在完成所有 NAT 相关的配置之后，保存配置并重启设备（或通过命令 **reset session** 手工清除与 NAT 相关的表项），以避免可能会产生的问题。这些问题主要包括：NAT 相关的配置删除后，已建立的连接仍然可以进行地址转换处理；在连接过程中进行 NAT 配置，会因为配置顺序的不一致，导致相同的配置产生不同的处理结果。
- 对于分布式设备，必须保证各接口板上引用的地址池所定义的 IP 地址没有重叠。

1.3 配置地址池

动态地址转换的过程中，NAT 网关将会从地址池中挑选一个地址做为转换后的报文源地址。由于动态地址转换的关联配置中需要引用已经定义的地址池，因此需要根据实际网络情况，预先合理规划公网 IP 地址池。

表1-2 配置地址池

操作	命令	说明
进入系统视图	**system-view**	-
定义一个地址池	**nat address-group** *group-number start-address end-address*	必选 对于 Easy IP 功能，不需要配置 NAT 地址池，直接使用接口地址作为转换后的 IP 地址

1.4 配置地址转换

1.4.1 地址转换介绍

通过 NAT 网关设备上静态建立或动态生成的地址映射关系，实现内部网络与外部网络 IP 地址的转换。通常，我们按照地址映射关系的产生方式将地址转换分为动态地址转换和静态地址转换两类：

- 静态地址转换

外部网络和内部网络之间的地址映射关系在配置中确定。适用于内部网络与外部网络之间的少量固定访问需求。

- 动态地址转换

外部网络和内部网络之间的地址映射关系由报文动态决定。通过配置访问控制列表和地址池（或接口地址）的关联，由“具有某些特征的 IP 报文”挑选使用“地址池中地址（或接口地址）”，从而建立动态地址映射关系。适用于内部网络有大量用户需要访问外部网络的需求。这种情况下，关联中指定的地址池资源由内网报文按需从中选择使用，访问外网的会话结束之后该资源便释放给其它用户。

无论静态地址转换还是动态地址转换，都可以支持 NAT 多实例的配置。只要指定地址所属的 *vpn-instance-name*，即可以实现对 MPLS VPN 的支持。

1.4.2 配置静态地址转换

配置静态地址转换时，需要首先在系统视图下配置静态地址转换映射，然后在接口下使该转换生效。

静态地址转换映射支持两种方式：一对一静态转换映射、网段对网段静态转换映射。

目前设备仅支持一对一静态地址转换映射。

1. 配置一对一静态地址转换

表1-3 配置一对一静态地址转换

操作	命令	说明
进入系统视图	**system-view**	-
配置一对一静态地址转换映射	**nat static** *local-ip* [**vpn-instance** *local-name*] *global-ip*	必选
进入接口视图	**interface** *interface-type interface-number*	-
使配置的 NAT 静态转换在接口上生效	**nat outbound static**	必选

1.4.3 配置动态地址转换

一般情况下，通过在接口上配置访问控制列表和地址池（或接口地址）的关联即可实现动态地址转换。另外，设备还支持仅仅通过判断出接口报文的源地址，而不使用访问控制列表的方式来实现出接口报文的动态地址转换。

动态地址转换根据是否使用端口信息分为 NOPAT 和 NAPT 两种方式：NOPAT 为不使用 TCP/UDP 端口信息实现的多对多地址转换；NAPT 为使用 TCP/UDP 端口信息实现的多对一地址转换。若直接使用接口的 IP 地址作为转换后的地址，则可配置 EASY IP 来实现。

地址转换关联一般在 NAT 网关的出接口上配置，但是当某内网主机需要通过多个出接口访问外网时，就需要在多个出接口上配置地址转换关联，过程复杂，因此设备提供了入接口地址关联的配置。这样，当 NAT 设备作为 VPN 间互访的工具时，在出接口较多的情况下，通过在接入各私网的入接口上配置地址转换关联达到简化配置的目的。

- 若配置出接口地址关联，那么从出接口发送的首个数据包会首先由访问控制列表进行判定是否允许进行地址转换，然后根据关联找到与之对应的地址池（或报文源地址）进行源地址转换，并建立地址转换表项，后续数据包直接根据地址转换表项进行转换。
- 若配置入接口地址关联，那么从该接口接收的符合指定访问控制列表的数据包首先会被重定向到 NAT 业务板，然后再做与出接口地址转换类似的源地址转换处理。该方式下地址转换不支持 EASY IP 特性。

目前，设备仅支持配置出接口地址关联。

1. 配置准备

- 配置控制地址转换范围的访问控制列表
- 配置用于地址转换的公网 IP 地址池


说明

- 访问控制列表的配置请参见“安全分册”中的“ACL 配置”。
- 地址池的配置请参见 1.3 配置地址池。

2. 配置 EASY IP

表1-4 配置 Easy IP

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置访问控制列表和接口地址关联，实现 Easy IP 特性	**nat outbound** *acl-number*	必选

3. 配置 NOPAT

表1-5 配置 NOPAT

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在出接口配置访问控制列表和地址池关联，且不使用端口信息，实现 NOPAT	**nat outbound** [*acl-number*] **address-group** *group-number* **no-pat**	必选

4. 配置 NAPT

表1-6 配置 NAPT

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
在出接口配置访问控制列表和地址池关联，并且同时使用 IP 地址和端口信息，实现 NAPT	**nat outbound** *acl-number* **address-group** *group-number*	必选

1.5 配置内部服务器

1.5.1 内部服务器介绍

通过配置内部服务器，可以将相应的外部地址和端口映射到内部服务器的私有地址和端口上，从而使外部网络用户能够访问内部服务器。内部服务器与外部网络的映射表是通过在接口上配置 **nat server** 命令生成的。

配置内部服务器时需要配置的信息包括：外部网络的信息（外部网络地址、外部网络端口）、内部网络的信息（内部网络地址、内部网络端口）以及服务协议类型。根据配置的内外部网络信息的不同，可以将内部服务器分为普通内部服务器和负载分担内部服务器。

内部服务器以及对外公布的公网地址均可以支持 MPLS VPN。当内部服务器位于 MPLS VPN 时，还应指定所属的 *vpn-instance-name*。如果不设置该值，表示内部服务器属于一个普通的私网，不属于某一个 MPLS VPN。

1.5.2 配置普通内部服务器

配置普通的内部服务器是将私网服务器的地址和端口映射为公网地址和端口，允许外部网络访问位于私网的服务器。

表1-7 配置内部服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置普通内部服务器	**nat server protocol** *pro-type* **global** *global-address* [*global-port*] **inside** *local-address* [*local-port*] [**vpn-instance** *local-name*]	二者必选其一
	nat server protocol *pro-type* **global** *global-address global-port1 global-port2* **inside** *local-address1 local-address2 local-port* [**vpn-instance** *local-name*]	

1.6 配置 DNS-map

通过配置 DNS-map，可以实现私网用户通过域名（DNS 服务器在公网）访问位于同一私网内的内部服务器的功能。

表1-8 配置 DNS-map

操作	命令	说明
进入系统视图	**system-view**	-
配置一条域名到内部服务器的映射	**nat dns-map domain** *domain-name* **protocol** *pro-type* **ip** *global-ip* **port** *global-port*	必选

1.7 配置基于 IP 地址的连接数限制

NAT 连接限制功能的配置和具体命令介绍请参见平台“安全分册”中的“连接限制配置”和“连接限制命令”。

1.8 NAT 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示地址转换配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，执行 **reset** 命令可以清除地址转换的统计信息。

表1-9 NAT 显示和维护

操作	命令
显示 NAT 地址池的信息	**display nat address-group** [*group-number*]
显示所有 NAT 配置信息	**display nat all**
显示地址转换关联的配置信息	**display nat bound**
显示 DNS-map 的配置信息	**display nat dns-map**
显示内部服务器的信息	**display nat server**
显示静态配置的信息	**display nat static**
显示当前激活的连接信息（集中式设备）	**display nat session** [**vpn-instance** *vpn-instance-name*] [**source** { **global** *global-address* | **inside** *inside-address* }] [**destination** *dst-address*]
显示当前激活的连接信息（分布式设备）	**display nat session slot** *slot-number* [**vpn-instance** *vpn-instance-name*] [**source** { **global** *global-address* | **inside** *inside-address* }] [**destination** *dst-address*]

操作	命令
显示 NAT 的统计信息（集中式设备）	**display nat statistics**
显示 NAT 的统计信息（分布式设备）	**display nat statistics slot** *slot-number*
显示用户当前配置的连接限制策略	**display connection-limit policy** { *policy-number* | **all** }
查看输出到日志服务器的日志的配置和统计信息（集中式设备）	**display userlog export**
查看输出到日志服务器的日志的配置和统计信息（分布式设备）	**display userlog export slot** *slot-number*

1.9 NAT 典型配置举例

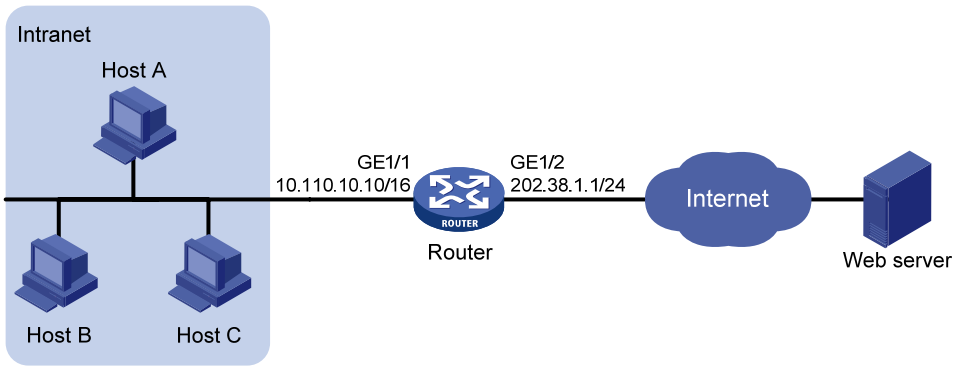
1.9.1 私网访问公网典型配置举例

1. 组网需求

一个公司拥有 202.38.1.1/24 至 202.38.1.3/24 三个公网 IP 地址，内部网址为 10.110.0.0/16。通过配置 NAT 使得仅内部网络中 10.110.10.0/24 网段的用户可以访问 Internet。

2. 组网图

图1-4 私网访问公网典型配置组网



3. 配置步骤

配置 IP 地址池 1，包括两个公网地址 202.38.1.2 和 202.38.1.3。

```
<Router> system-view
[Router] nat address-group 1 202.38.1.2 202.38.1.3
```

配置访问控制列表 2001，仅允许内部网络中 10.110.10.0/24 网段的用户可以访问 Internet。

```
[Router] acl number 2001
[Router-acl-basic-2001] rule permit source 10.110.10.0 0.0.0.255
[Router-acl-basic-2001] rule deny
[Router-acl-basic-2001] quit
```

在出接口 GigabitEthernet1/2 上配置 ACL 2001 与 IP 地址池 1 相关联。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] nat outbound 2001 address-group 1
[Router-GigabitEthernet1/2] quit
```

1.9.2 普通内部服务器典型配置举例

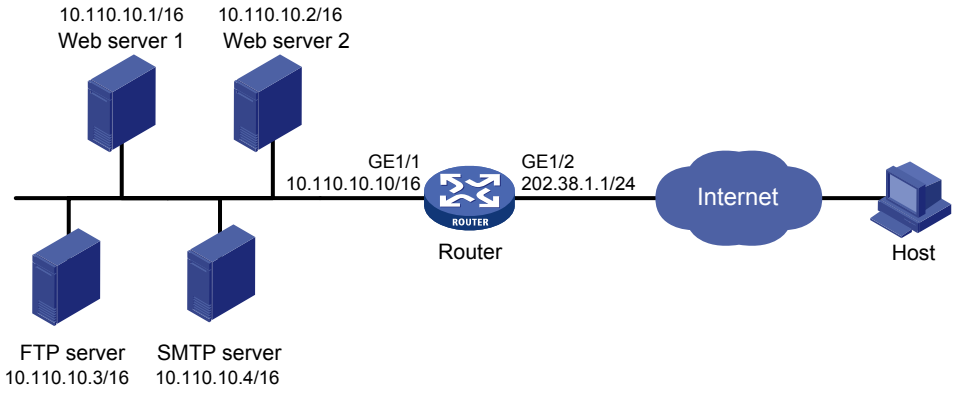
1. 组网需求

某公司内部对外提供 Web、FTP 和 SMTP 服务，而且提供两台 Web 服务器。公司内部网址为 10.110.0.0/16。其中，内部 FTP 服务器地址为 10.110.10.3/16，内部 Web 服务器 1 的 IP 地址为 10.110.10.1/16，内部 Web 服务器 2 的 IP 地址为 10.110.10.2/16，内部 SMTP 服务器 IP 地址为 10.110.10.4/16。公司具有 202.38.1.1/24 至 202.38.1.3/24 三个 IP 地址。需要实现如下功能：

- 外部的主机可以访问内部的服务器。
- 选用 202.38.1.1 作为公司对外提供服务的 IP 地址，Web 服务器 2 对外采用 8080 端口。

2. 组网图

图1-5 普通内部服务器典型配置组网



3. 配置步骤

进入接口 GigabitEthernet1/2。

```
<Router> system-view
[Router] interface gigabitethernet 1/2
```

设置内部 FTP 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.1 21 inside 10.110.10.3
ftp
```

设置内部 Web 服务器 1。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.1 80 inside 10.110.10.1
www
```

设置内部 Web 服务器 2。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.1 8080 inside
10.110.10.2 www
```

设置内部 SMTP 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.1 smtp inside
10.110.10.4 smtp
[Router-GigabitEthernet1/2] quit
```

1.9.3 NAT DNS-mapping 典型配置举例

1. 组网需求

某公司内部对外提供 Web 和 FTP 服务。公司内部网址为 10.110.0.0/16。其中，Web 服务器地址为 10.110.10.1，FTP 服务器的 IP 地址为 10.110.10.2。公司具有 202.38.1.1/24 至 202.38.1.3/24 三个合法的 IP 地址。另外公司在公网有一台 DNS 服务器，IP 地址为 202.38.1.4/24。需要实现如下功能：

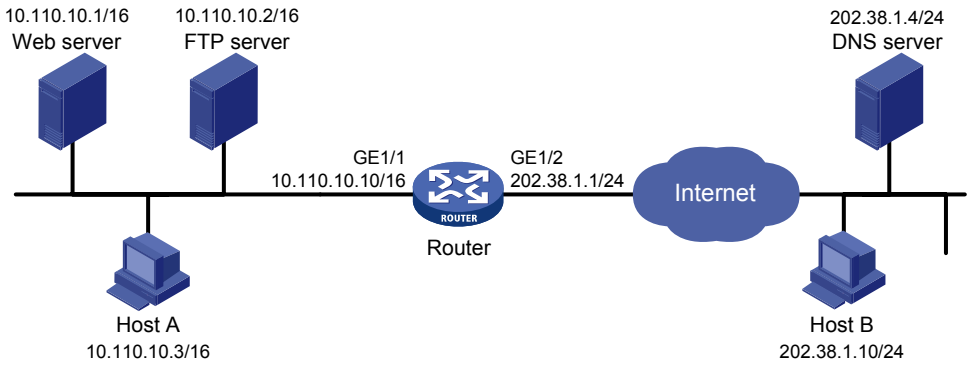
选用 202.38.1.2 作为公司对外提供服务的 IP 地址。

公网用户可以通过域名或 IP 地址访问内部服务器。

私网用户可以通过域名访问内部服务器。

2. 组网图

图1-6 NAT DNS mapping 典型配置组网



3. 配置步骤

```
<Router> system-view
[Router] interface gigabitethernet 1/2
```

配置内部 Web 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.2 inside 10.110.10.1 www
```

配置内部 FTP 服务器。

```
[Router-GigabitEthernet1/2] nat server protocol tcp global 202.38.1.2 inside 10.110.10.2 ftp
[Router-GigabitEthernet1/2] quit
```

配置两条 DNS-mapping 表项：Web 服务器的域名 www.server.com 对应 IP 地址 202.38.1.2；FTP 服务器的域名 ftp.server.com 对应 IP 地址 202.38.1.2。

```
[Router] nat dns-map domain www.server.com protocol tcp ip 202.38.1.2 port www
[Router] nat dns-map domain ftp.server.com protocol tcp ip 202.38.1.2 port ftp
[Router] quit
```

4. 验证配置结果

上述配置完成后，可以用 **display** 命令查看 DNS mapping 的配置情况。

```
<Router> display nat dns-map
NAT DNS mapping information:
  There are currently 2 NAT DNS mapping(s)
  Domain-name: www.server.com
```

```
Global-IP   : 202.38.1.2
Global-port: 80(www)
Protocol    : 6(TCP)

Domain-name: ftp.server.com
Global-IP   : 202.38.1.2
Global-port: 21(ftp)
Protocol    : 6(TCP)
```

私网 Host A 和公网 Host B 均可通过域名 www.server.com 访问私网内的 Web 服务器，以及通过域名 ftp.server.com 访问私网内的 FTP 服务器。

1.10 NAT 常见配置错误举例

1.10.1 故障之一：地址转换不正常。

故障排除：通过打开 NAT 的调试信息开关，根据设备上的调试信息，初步定位错误，然后使用其它命令作进一步的判断。调试时，注意观察地址转换后的源地址，要保证这个地址是希望转换的地址，否则可能会是地址池配置错误。同时要保证目的网络到地址池中地址段的路由可达。注意防火墙以及地址转换本身的访问控制列表对地址转换造成的影响，同时注意路由的配置。

1.10.2 故障之二：内部服务器工作不正常。

故障排除：如果外部主机不能正常访问内部服务器，请检查是否是内部服务器主机的配置有错或路由器上对内部服务器的配置有错，如对内部服务器的 IP 地址指定错误等等。同时也有可能是防火墙禁止了外部主机对内部网络的访问，可以用 **display acl** 命令来查看，详细内容请参见“安全分册”中的“防火墙配置”。

目 录

1 ACL简介 1-1
1.1 IPv4 ACL简介 1-1
1.1.1 IPv4 ACL分类 1-1
1.1.2 IPv4 ACL命名 1-1
1.1.3 IPv4 ACL匹配顺序 1-1
1.1.4 IPv4 ACL步长 1-2
1.1.5 IPv4 ACL生效时间段 1-3
1.1.6 IPv4 ACL对分片报文的处理 1-3
1.2 IPv6 ACL简介 1-3
1.2.1 IPv6 ACL分类 1-3
1.2.2 IPv6 ACL命名 1-3
1.2.3 IPv6 ACL匹配顺序 1-4
1.2.4 IPv6 ACL步长 1-4
1.2.5 IPv6 ACL生效时间段 1-4
2 IPv4 ACL配置 2-1
2.1 配置时间段 2-1
2.1.1 配置时间段 2-1
2.2 配置基本IPv4 ACL 2-2
2.2.1 配置准备 2-2
2.2.2 配置基本IPv4 ACL 2-2
2.3 配置高级IPv4 ACL 2-3
2.3.1 配置准备 2-3
2.3.2 配置高级IPv4 ACL 2-3
2.4 配置二层ACL 2-4
2.4.1 配置准备 2-4
2.4.2 配置二层ACL 2-4
2.5 拷贝IPv4 ACL 2-5
2.5.1 配置准备 2-5
2.5.2 拷贝IPv4 ACL 2-6
2.6 使能ACL加速功能 2-6
2.6.1 配置准备 2-6
2.6.2 使能ACL加速功能 2-6
2.7 IPv4 ACL显示和维护 2-7
2.8 IPv4 ACL典型配置举例 2-7
2.8.1 组网需求 2-7
2.8.2 组网图 2-7
2.8.3 配置步骤 2-8
3 IPv6 ACL配置 3-1
3.1 配置时间段 3-1
3.2 配置基本IPv6 ACL 3-1

3.2.1 配置准备 3-1
3.2.2 配置基本IPv6 ACL 3-1
3.3 配置高级IPv6 ACL 3-2
3.3.1 配置准备 3-2
3.3.2 配置高级IPv6 ACL 3-2
3.4 拷贝IPv6 ACL 3-3
3.4.1 配置准备 3-3
3.4.2 拷贝IPv6 ACL 3-3
3.5 IPv6 ACL显示和维护 3-4
3.6 IPv6 ACL典型配置举例 3-4
3.6.1 组网需求 3-4
3.6.2 配置步骤 3-4

1 ACL 简介

说明

本章所介绍的 ACL 包括 IPv4 ACL 和 IPv6 ACL。

ACL（Access Control List，访问控制列表）是用来实现流识别功能的。网络设备为了过滤报文，需要配置一系列的匹配条件对报文进行分类，这些条件可以是报文的源地址、目的地址、端口号等。

当设备的端口接收到报文后，即根据当前端口上应用的 ACL 规则对报文的字段进行分析，在识别出特定的报文之后，根据预先设定的策略允许或禁止该报文通过。

由 ACL 定义的报文匹配规则，可以被其它需要对流量进行区分的场合引用，如包过滤、QoS 中流分类规则的定义等。

1.1 IPv4 ACL 简介

1.1.1 IPv4 ACL 分类

IPv4 ACL根据ACL序号来区分不同的ACL，可以分为三种类型，如 表 1-1所示。

表1-1 IPv4 ACL 分类

IPv4 ACL 类型	ACL 序号范围	区分报文的依据
基本 IPv4 ACL	2000～2999	只根据报文的源 IP 地址信息制定匹配规则
高级 IPv4 ACL	3000～3999	根据报文的源 IP 地址信息、目的 IP 地址信息、IP 承载的协议类型、协议的特性等三、四层信息制定匹配规则
二层 ACL	4000～4999	根据报文的源 MAC 地址、目的 MAC 地址、802.1p 优先级、二层协议类型等二层信息制定匹配规则

1.1.2 IPv4 ACL 命名

用户在创建 IPv4 ACL 时，可以为 ACL 指定一个名称。每个 IPv4 ACL 最多只能有一个名称。命名的 ACL 使用户可以通过名称唯一地确定一个 IPv4 ACL，并对其进行相应的操作。

在创建 ACL 时，用户可以选择是否配置名称。ACL 创建后，不允许用户修改或者删除 ACL 名称，也不允许为未命名的 ACL 添加名称。

说明

IPv4 ACL 的名称对于 IPv4 ACL 全局唯一，但允许与 IPv6 ACL 使用相同的名称。

1.1.3 IPv4 ACL 匹配顺序

一个 ACL 中可以包含多个规则，而每个规则都指定不同的报文匹配选项，这些规则可能存在重复或矛盾的地方，在将一个报文和 ACL 的规则进行匹配的时候，到底采用哪些规则呢？就需要确定规则的匹配顺序。

IPv4 ACL 支持两种匹配顺序：

- 配置顺序：按照用户配置规则的先后顺序进行规则匹配。
- 自动排序：按照“深度优先”的顺序进行规则匹配。

1. 基本 IPv4 ACL 的“深度优先”顺序判断原则如下

(1) 先看规则中是否带 VPN 实例，带 VPN 实例的规则优先；

(2) 再比较源 IP 地址范围，源 IP 地址范围小（反掩码中“0”位的数量多）的规则优先；

(3) 如果源 IP 地址范围相同，则先配置的规则优先。

2. 高级 IPv4 ACL 的“深度优先”顺序判断原则如下

(1) 先看规则中是否带 VPN 实例，带 VPN 实例的规则优先；

(2) 再比较协议范围，指定了 IP 协议承载的协议类型的规则优先；

(3) 如果协议范围相同，则比较源 IP 地址范围，源 IP 地址范围小（反掩码中“0”位的数量多）的规则优先；

(4) 如果协议范围、源 IP 地址范围相同，则比较目的 IP 地址范围，目的 IP 地址范围小（反掩码中“0”位的数量多）的规则优先；

(5) 如果协议范围、源 IP 地址范围、目的 IP 地址范围相同，则比较四层端口号（TCP/UDP 端口号）范围，四层端口号范围小的规则优先；

(6) 如果上述范围都相同，则先配置的规则优先。

3. 二层 ACL 的“深度优先”顺序判断原则如下

(1) 先比较源 MAC 地址范围，源 MAC 地址范围小（掩码中“1”位的数量多）的规则优先；

(2) 如果源 MAC 地址范围相同，则比较目的 MAC 地址范围，目的 MAC 地址范围小（掩码中“1”位的数量多）的规则优先；

(3) 如果源 MAC 地址范围、目的 MAC 地址范围相同，则先配置的规则优先。

在报文匹配规则时，会按照匹配顺序去匹配定义的规则，一旦有一条规则被匹配，报文就不再继续匹配其它规则了，设备将对该报文执行第一次匹配的规则指定的动作。

1.1.4 IPv4 ACL 步长

1. 步长的含义

步长的含义是：设备自动为 ACL 规则分配编号的时候，每个相邻规则编号之间的差值。例如，如果将步长设定为 5，规则编号分配是按照 0、5、10、15…这样的规律分配的。缺省情况下，步长为 5。

当步长改变后，ACL 中的规则编号会自动从 0 开始重新排列。例如，原来规则编号为 5、10、15、20，当通过命令把步长改为 2 后，则规则编号变成 0、2、4、6。

当使用命令将步长恢复为缺省值后，设备将立刻按照缺省步长调整 ACL 规则的编号。例如：ACL 3001，步长为 2，下面有 4 个规则，编号为 0、2、4、6。如果此时使用命令将步长恢复为缺省值，则 ACL 规则编号变成 0、5、10、15，步长为 5。

2. 步长的作用

使用步长设定的好处是用户可以方便地在规则之间插入新的规则。例如配置好了 4 个规则，规则编号为：0、5、10、15。此时如果用户希望能在第一条规则之后插入一条规则，则可以使用命令在 0

和 5 之间插入一条编号为 1 的规则。

另外，在定义一条 ACL 规则的时候，用户可以不指定规则编号，这时，系统会从 0 开始，按照步长，自动为规则分配一个大于现有最大编号的最小编号。假设现有规则的最大编号是 28，步长是 5，那么系统分配给新定义的规则的编号将是 30。

1.1.5 IPv4 ACL 生效时间段

时间段用于描述一个特殊的时间范围。用户可能有这样的需求：一些 ACL 规则需要在某个或某些特定时间内生效，而在其他时间段则不利用它们进行报文过滤，即通常所说的按时间段过滤。这时，用户就可以先配置一个或多个时间段，然后在相应的规则下通过时间段名称引用该时间段，这条规则只在该指定的时间段内生效，从而实现基于时间段的 ACL 过滤。

如果规则引用的时间段未配置，则系统给出提示信息，并允许这样的规则创建成功，但是规则不能立即生效，直到用户配置了引用的时间段，并且系统时间在指定时间段范围内 ACL 规则才能生效。

1.1.6 IPv4 ACL 对分片报文的处理

传统的报文过滤并不处理所有 IP 报文分片，而是只对首片（第一片）分片报文进行匹配处理，对后续分片不进行匹配处理。这样，网络攻击者可能构造后续的分片报文进行流量攻击，就带来了安全隐患。

目前，设备提供的对分片报文过滤的功能如下：

- 对所有的分片报文进行三层（IP 层）的匹配过滤。
- 对于包含高级信息的 ACL 规则项（例如包含 TCP/UDP 端口号，ICMP 类型），提供标准匹配和精确匹配两种匹配方式，缺省的匹配方式为标准匹配。

说明

标准匹配和精确匹配的含义如下：

- 标准匹配：只匹配三层信息，而三层以外的信息将被忽略。
- 精确匹配：对 ACL 定义的所有的规则项进行匹配。

这两种匹配方式只有防火墙支持，具体介绍请参见“安全分册”中的“防火墙配置”。

1.2 IPv6 ACL 简介

1.2.1 IPv6 ACL 分类

IPv6 ACL根据ACL序号来区分不同的ACL，可以分为两种类型，如 表 1-2所示。

表1-2 IPv6 ACL 分类

IPv6 ACL 类型	ACL 序号范围	区分报文的依据
基本 IPv6 ACL	2000～2999	只根据源 IPv6 地址信息制定匹配规则
高级 IPv6 ACL	3000～3999	根据报文的源 IPv6 地址信息、目的 IPv6 地址信息、IPv6 承载的协议类型、协议的特性等三层、四层信息来制定匹配规则

1.2.2 IPv6 ACL 命名

用户在创建 IPv6 ACL 时，可以为 ACL 指定一个名称。每个 IPv6 ACL 最多只能有一个名称。命名

的 ACL 使用户可以通过名称唯一地确定一个 IPv6 ACL，并对其进行相应的操作。

在创建 ACL 时，用户可以选择是否配置名称。ACL 创建后，不允许用户修改或者删除 ACL 名称，也不允许为未命名的 ACL 添加名称。



说明

IPv6 ACL 的名称对于 IPv6 ACL 全局唯一，但允许与 IPv4 ACL 使用相同的名称。

1.2.3 IPv6 ACL 匹配顺序

一个 ACL 中可以包含多个规则，而每个规则都指定不同的报文匹配选项，这些规则可能存在重复或矛盾的地方，在将一个报文和 ACL 的规则进行匹配的时候，到底采用哪些规则呢？就需要确定规则的匹配顺序。

IPv6 ACL 支持两种匹配顺序：

- 配置顺序：按照用户配置规则的先后顺序进行规则匹配。
- 自动排序：按照“深度优先”的顺序进行规则匹配。

1. 基本 IPv6 ACL 的“深度优先”顺序判断原则如下

(1) 先比较源 IPv6 地址范围，源 IPv6 地址范围小（前缀长）的规则优先；

(2) 如果源 IPv6 地址范围相同，则先配置的规则优先。

2. 高级 IPv6 ACL 的“深度优先”顺序判断原则如下

(1) 先比较协议范围，指定了 IPv6 协议承载的协议类型的规则优先；

(2) 如果协议范围相同，则比较源 IPv6 地址范围，源 IPv6 地址范围小（前缀长）的规则优先；

(3) 如果协议范围、源 IPv6 地址范围相同，则比较目的 IPv6 地址范围，目的 IPv6 地址范围小（前缀长）的规则优先；

(4) 如果协议范围、源 IPv6 地址范围、目的 IPv6 地址范围相同，则比较四层端口号（TCP/UDP 端口号）范围，四层端口号范围小的规则优先；

(5) 如果上述范围都相同，则先配置的规则优先。

在报文匹配规则时，会按照匹配顺序去匹配定义的规则，一旦有一条规则被匹配，报文就不再继续匹配其它规则了，设备将对该报文执行第一次匹配的规则指定的动作。

1.2.4 IPv6 ACL 步长

关于步长的介绍请参见“1.1.4 IPv4 ACL步长”。

1.2.5 IPv6 ACL 生效时间段

关于生效时间段的介绍请参见“1.1.5 IPv4 ACL生效时间段”。

2 IPv4 ACL 配置

2.1 配置时间段

对时间段的配置有如下两种情况：

- 配置周期时间段：采用每周的周几的形式；
- 配置绝对时间段：采用从起始时间到结束时间的形式。

2.1.1 配置时间段

表2-1 配置时间段

操作	命令	说明
进入系统视图	**system-view**	-
创建一个时间段	**time-range** *time-range-name* { *start-time* **to** *end-time days* [**from** *time1 date1*] [**to** *time2 date2*] | **from** *time1 date1* [**to** *time2 date2*] | **to** *time2 date2* }	必选
显示时间段的配置和状态	**display time-range** { *time-range-name* | **all** }	可选 **display** 命令可以在任意视图下执行

需要注意的是：

- 如果用户通过命令 **time-range** *time-range-name start-time* **to** *end-time days* 定义了一个周期时间段，则只有系统时钟在该周期时间段内，该时间段才进入激活状态。
- 如果用户通过命令 **time-range** *time-range-name* { **from** *time1 date1* [**to** *time2 date2*] \| **to** *time2 date2* }定义了一个绝对时间段，则只有系统时钟在该绝对时间段内，该时间段才进入激活状态。
- 如果用户通过命令 **time-range** *time-range-name start-time* **to** *end-time days* { **from** *time1 date1* [**to** *time2 date2*] \| **to** *time2 date2* }同时定义了绝对时间段和周期时间段，则只有系统时钟同时满足绝对时间段和周期时间段的定义时，该时间段才进入激活状态。例如，一个时间段定义了绝对时间段：从 2004 年 1 月 1 日 0 点 0 分到 2004 年 12 月 31 日 24 点 0 分，同时定义了周期时间段：每周三的 12:00 到 14:00。该时间段只有在 2004 年内每周三的 12:00 到 14:00 才进入激活状态。
- 在同一个名字下可以配置多个时间段，来共同描述一个特殊时间，通过名字来引用该时间。在同一个名字下配置的多个周期时间段之间是“或”的关系，多个绝对时间段之间是“或”的关系，而周期时间段和绝对时间段之间是“与”的关系。
- 如果不配置开始日期，时间段就是从系统可表示的最早时间（即 1970 年 1 月 1 日 0 点 0 分）起到结束日期为止。如果不配置结束日期，时间段就是从配置生效之日起到系统可以表示的最晚时间（即 2100 年 12 月 31 日 24 点 0 分）为止。
- 最多可以定义 256 个时间段。

2.2 配置基本 IPv4 ACL

基本 IPv4 ACL 只根据源 IP 地址信息制定匹配规则，对报文进行相应的分析处理。

基本 IPv4 ACL 的序号取值范围为 2000～2999。

2.2.1 配置准备

如果要配置带有时间段参数的规则，则需要定义相应的时间段。

2.2.2 配置基本 IPv4 ACL

表2-2 配置基本 IPv4 ACL

操作	命令	说明
进入系统视图	**system-view**	-
创建基本 IPv4 ACL 并进入基本 IPv4 ACL 视图	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]	必选 缺省情况下，匹配顺序为 **config** 如果用户在创建 IPv4 ACL 时指定了名称，则之后可以通过 **acl name** *acl-name* 命令进入指定名称的 IPv4 ACL 视图
定义规则	**rule** [*rule-id*] { **deny** | **permit** } [**fragment** | **logging** | **source** { *sour-addr sour-wildcard* | **any** } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *	必选 可以重复本步骤创建多条规则 **logging** 参数需要使用该 ACL 的模块支持日志记录功能才能生效，例如防火墙
定义步长	**step** *step-value*	可选 缺省情况下，步长为 5
定义基本 IPv4 ACL 的描述信息	**description** *text*	可选 缺省情况下，基本 IPv4 ACL 没有描述信息
定义规则的描述信息	**rule** *rule-id* **comment** *text*	可选 缺省情况下，规则没有描述信息

需要注意的是：

- 当 ACL 的匹配顺序为 **config** 时，用户可以修改该 ACL 中的任何一条已经存在的规则，在修改 ACL 中的某条规则时，该规则中没有修改到的部分仍旧保持原来的状态；当 ACL 的匹配顺序为 **auto** 时，用户不能修改该 ACL 中的任何一条已经存在的规则，否则系统会提示错误信息。
- 新创建或修改后的规则不能和已经存在的规则内容相同，否则会导致创建或修改不成功，系统会提示该规则已经存在。
- 当 ACL 的匹配顺序为 **auto** 时，新创建的规则将按照“深度优先”的原则插入到已有的规则中，但是所有规则对应的编号不会改变。

注意

- 用户可以通过命令 **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** }修改 IPv4 ACL 的匹配顺序为 **auto** 或者 **config**，但必须在 IPv4 ACL 中没有规则的时候修改，对已经有规则的 IPv4 ACL 是无法修改其匹配顺序的。
- 在使用 **rule comment** 命令为规则定义描述信息时，该规则必须存在。

2.3 配置高级 IPv4 ACL

高级 IPv4 ACL 可以使用报文的源 IP 地址信息、目的 IP 地址信息、IP 承载的协议类型、协议的特性（例如 TCP 或 UDP 的源端口、目的端口，TCP 标记，ICMP 协议的消息类型、消息码等）等信息来制定匹配规则。

高级 IPv4 ACL 支持对三种报文优先级的分析处理：

- ToS（Type of Service，服务类型）优先级；
- IP 优先级；
- DSCP（Differentiated Services Codepoint，差分服务编码点）优先级。

用户可以利用高级 IPv4 ACL 定义比基本 IPv4 ACL 更准确、更丰富、更灵活的匹配规则。

高级 IPv4 ACL 的序号取值范围为 3000～3999。

2.3.1 配置准备

如果要配置带有时间段参数的规则，则需要定义相应的时间段。

2.3.2 配置高级 IPv4 ACL

表2-3 配置高级 IPv4 ACL

操作	命令	说明
进入系统视图	**system-view**	-
创建高级 IPv4 ACL 并进入高级 IPv4 ACL 视图	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]	必选 缺省情况下，匹配顺序为 **config** 如果用户在创建 IPv4 ACL 时指定了名称，则之后可以通过 **acl name** *acl-name* 命令进入指定名称的 IPv4 ACL 视图
定义规则	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [{ **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **destination** { *dest-addr dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **fragment** | **icmp-type** { *icmp-type icmp-code* | *icmp-message* } | **logging** | **precedence** *precedence* | **reflective** | **source** { *sour-addr sour-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **tos** *tos* | **vpn-instance** *vpn-instance-name*] *	必选 可以重复本步骤创建多条规则
定义步长	**step** *step-value*	可选 缺省情况下，步长为 5

操作	命令	说明
定义高级 IPv4 ACL 的描述信息	**description** *text*	可选 缺省情况下，高级 IPv4 ACL 没有描述信息
定义规则的描述信息	**rule** *rule-id* **comment** *text*	可选 缺省情况下，规则没有描述信息

需要注意的是：

- 当 ACL 的匹配顺序为 **config** 时，用户可以修改该 ACL 中的任何一条已经存在的规则，在修改 ACL 中的某条规则时，该规则中没有修改到的部分仍旧保持原来的状态；当 ACL 的匹配顺序为 **auto** 时，用户不能修改该 ACL 中的任何一条已经存在的规则，否则系统会提示错误信息。
- 新创建或修改后的规则不能和已经存在的规则内容相同，否则会导致创建或修改不成功，系统会提示该规则已经存在。
- 当 ACL 的匹配顺序为 **auto** 时，新创建的规则将按照“深度优先”的原则插入到已有的规则中，但是所有规则对应的编号不会改变。

注意

- 用户可以通过命令 **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** }修改 IPv4 ACL 的匹配顺序为 **auto** 或者 **config**，但必须在 IPv4 ACL 中没有规则的时候修改，对已经有规则的 IPv4 ACL 是无法修改其匹配顺序的。
- 在使用 **rule comment** 命令为规则定义描述信息时，该规则必须存在。

2.4 配置二层 ACL

二层 ACL 根据报文的源 MAC 地址、目的 MAC 地址、802.1p 优先级、二层协议类型等二层信息制定匹配规则，对报文进行相应的分析处理。

二层 ACL 的序号取值范围为 4000～4999。

2.4.1 配置准备

如果要配置带有时间段参数的规则，则需要定义相应的时间段。

2.4.2 配置二层 ACL

表2-4 配置二层 ACL

操作	命令	说明
进入系统视图	**system-view**	-
创建二层 ACL 并进入二层 ACL 视图	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto**	**config** }]	必选 缺省情况下，匹配顺序为 **config** 如果用户在创建 IPv4 ACL 时指定了名称，则之后可以通过 **acl name** *acl-name* 命令进入指定名称的 IPv4 ACL 视图

操作	命令	说明
定义规则	**rule** [*rule-id*] { **deny** | **permit** } [**cos** *vlan-pri* **| dest-mac** *dest-addr dest-mask* **| lsap** *lsap-code lsap-wildcard* **| source-mac** *sour-addr source-mask* | **time-range** *time-range-name* **| type** *type-code type-wildcard*] *	必选 可以重复本步骤创建多条规则 **logging** 参数需要使用该 ACL 的模块支持日志记录功能才能生效，例如防火墙
定义步长	**step** *step-value*	可选 缺省情况下，步长为 5
定义二层 IPv4 ACL 的描述信息	**description** *text*	可选 缺省情况下，二层 IPv4 ACL 没有描述信息
定义规则的描述信息	**rule** *rule-id* **comment** *text*	可选 缺省情况下，规则没有描述信息

需要注意的是：

- 当 ACL 的匹配顺序为 **config** 时，用户可以修改该 ACL 中的任何一条已经存在的规则，在修改 ACL 中的某条规则时，该规则中没有修改到的部分仍旧保持原来的状态；当 ACL 的匹配顺序为 **auto** 时，用户不能修改该 ACL 中的任何一条已经存在的规则，否则系统会提示错误信息。
- 新创建或修改后的规则不能和已经存在的规则内容相同，否则会导致创建或修改不成功，系统会提示该规则已经存在。
- 当 ACL 的匹配顺序为 **auto** 时，新创建的规则将按照“深度优先”的原则插入到已有的规则中，但是所有规则对应的编号不会改变。

注意

- 用户可以通过命令 **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** \| **config** }修改 IPv4 ACL 的匹配顺序为 **auto** 或者 **config**，但必须在 IPv4 ACL 中没有规则的时候修改，对已经有规则的 IPv4 ACL 是无法修改其匹配顺序的。
- 在使用 **rule comment** 命令为规则定义描述信息时，该规则必须存在。

2.5 拷贝 IPv4 ACL

拷贝 IPv4 ACL 功能使用户可以通过拷贝一个已经存在的 IPv4 ACL，生成一个新的同类型的 IPv4 ACL。生成的新的 IPv4 ACL 的匹配顺序、包含的匹配规则、步长以及描述信息都和源 IPv4 ACL 相同。

2.5.1 配置准备

源 IPv4 ACL 必须存在，目的 IPv4 ACL 必须不存在。

2.5.2 拷贝 IPv4 ACL

表2-5 拷贝 IPv4 ACL

操作	命令	说明
进入系统视图	**system-view**	-
拷贝生成一个新的同类型的 IPv4 ACL	**acl copy** { *source-acl-number* | **name** *source-acl-name* } **to** { *dest-acl-number* | **name** *dest-acl-name* }	必选

注意

- 源 IPv4 ACL 和目的 IPv4 ACL 的类型要相同。
- 源 IPv4 ACL 的名称不会拷贝到目的 IPv4 ACL。

2.6 使能 ACL 加速功能

基于会话的业务处理一般采用首报文根据策略匹配、后续报文根据会话附加信息直接处理的方式来加快报文的处理速度，但是这种方式不能解决首报文匹配效率问题。当有大量用户同时与设备新建连接，而每个新建连接都需要进行 ACL 规则查找，当 ACL 中包含的规则数量很大时，此查找过程可能需要很长时间，这将导致用户建立连接时间超长，影响设备新建连接的性能。

ACL 加速功能可以在用于提高配置大量 ACL 规则的情况下，大大提高报文的匹配速度，从而提高设备转发性能和新建连接的性能。

- 在使能 ACL 加速功能前，系统在进行报文匹配时，对配置的所有规则（rule）进行线性查找匹配。当 ACL 中配置了大量规则且最后几条中的一条匹配时，匹配效率就会很低。
- 使能 ACL 的加速功能后，系统会将线性排列的大量 ACL 规则数据，采用四级 hash 表的方式重新进行组织和存储，我们将这个新的数据存储结构称之为快速查找库。有了这个快速查找库，系统的匹配速度会大大加快。

由于快速查找库会占用系统内存，建议用户仅在配置了大量 ACL 规则（比如规则数大于 1000）的情况下才有必要启用 ACL 的加速功能。如果用户在配置了少量 ACL 规则的情况下启用 ACL 的加速功能，匹配速度提升效果不大，反而会占用大量内存，得不偿失。

2.6.1 配置准备

将要启用 ACL 加速功能的 ACL 已经存在。

2.6.2 使能 ACL 加速功能

表2-6 使能 ACL 加速功能

操作	命令	说明
进入系统视图	**system-view**	-
使能指定 ACL 的加速功能	**acl accelerate number** *acl-number*	必选

注意

- 只有基本 IPv4 ACL 和高级 IPv4 ACL 支持 ACL 加速功能。
- 如果用户在使能 ACL 加速功能后又修改了 ACL 的配置，ACL 加速功能仍按修改前的配置进行报文匹配，建议在 ACL 启用加速功能后不再对 ACL 的配置做修改，否则无法保证报文正确匹配。

2.7 IPv4 ACL 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IPv4 ACL 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv4 ACL 统计信息。

表2-7 IPv4 ACL 显示和维护

配置	命令
显示配置的 IPv4 ACL 信息	**display acl** { *acl-number* | **all** | **name** *acl-name* }
显示 IPv4 ACL 加速功能的相关信息	**display acl accelerate** { *acl-number* | **all** }
显示时间段的配置和状态	**display time-range** { *time-range-name* | **all** }
清除 IPv4 ACL 统计信息	**reset acl counter** { *acl-number* | **all** | **name** *acl-name* }

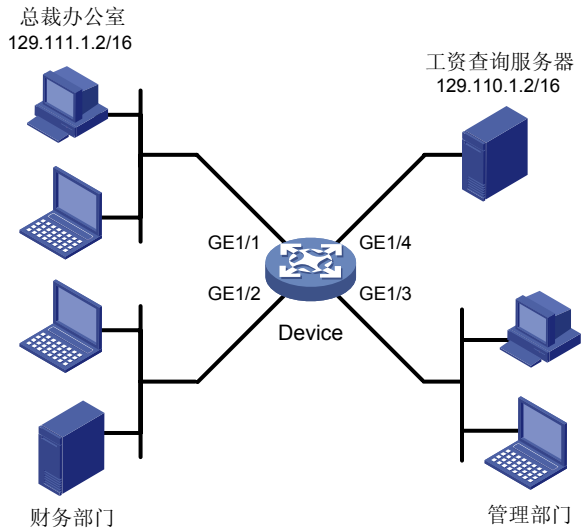
2.8 IPv4 ACL 典型配置举例

2.8.1 组网需求

- 公司企业网通过 Router 实现各部门之间的互连。
- 要求正确配置 ACL，禁止其它部门在上班时间（8:00 至 18:00）访问工资查询服务器（IP 地址为 129.110.1.2），而总裁办公室（IP 地址为 129.111.1.2）不受限制，可以随时访问。

2.8.2 组网图

图2-1 配置 ACL 组网图



2.8.3 配置步骤

(1) 定义上班时间段

定义 8:00 至 18:00 的周期时间段。

```
<Sysname> system-view
[Sysname] time-range trname 8:00 to 18:00 working-day
```

(2) 定义到工资服务器的 ACL

进入高级 IPv4 ACL 视图，编号为 3000。

```
[Sysname] acl number 3000
```

定义总裁办公室到工资服务器的访问规则。

```
[Sysname-acl-adv-3000] rule 1 permit ip source 129.111.1.2 0.0.0.0 destination 129.110.1.2
0.0.0.0
```

定义其它部门到工资服务器的访问规则。

```
[Sysname-acl-adv-3000] rule 2 deny ip source any destination 129.110.1.2 0.0.0.0 time-range
trname
[Sysname-acl-adv-3000] quit
```

(3) 应用 ACL

将 ACL 3000 用于 GigabitEthernet1/4 出方向的包过滤。

```
[Sysname] firewall enable
[Sysname] interface gigabitethernet 1/4
[Sysname-GigabitEthernet1/4] firewall packet-filter 3000 outbound
```

3 IPv6 ACL 配置

3.1 配置时间段

时间段的配置请参见“2.1 配置时间段”。

3.2 配置基本 IPv6 ACL

基本 IPv6 ACL 只根据源 IPv6 地址信息制定匹配规则，对报文进行相应的分析处理。

基本 IPv6 ACL 的序号取值范围为 2000～2999。

3.2.1 配置准备

如果要配置带有时间段参数的规则，则需要定义相应的时间段。

3.2.2 配置基本 IPv6 ACL

表3-1 配置基本 IPv6 ACL

操作	命令	说明
进入系统视图	**system-view**	-
创建基本 IPv6 ACL 并进入基本 IPv6 ACL 视图	**acl ipv6 number** *acl6-number* [**name** *acl6-name*] [**match-order** { **auto** | **config** }]	必选 缺省情况下，匹配顺序为 **config** 如果用户在创建 IPv6 ACL 时指定了名称，则之后可以通过 **acl ipv6 name** *acl6-name* 命令进入指定名称的 IPv6 ACL 视图
定义规则	**rule** [*rule-id*] { **deny** | **permit** } [**fragment** | **logging** | **source** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* | **any** } | **time-range** *time-range-name*] *	必选 可以重复本步骤创建多条规则 **logging** 参数需要使用该 ACL 的模块支持日志记录功能才能生效，例如防火墙
定义步长	**step** *step-value*	可选 缺省情况下，步长为 5
定义基本 IPv6 ACL 的描述信息	**description** *text*	可选 缺省情况下，基本 IPv6 ACL 没有描述信息
定义规则的描述信息	**rule** *rule-id* **comment** *text*	可选 缺省情况下，规则没有描述信息

需要注意的是：

- 当 ACL 的匹配顺序为 **config** 时，用户可以修改该 ACL 中的任何一条已经存在的规则，在修改 ACL 中的某条规则时，该规则中没有修改到的部分仍旧保持原来的状态；当 ACL 的匹配顺序为 **auto** 时，用户不能修改该 ACL 中的任何一条已经存在的规则，否则系统会提示错误信息。
- 新创建或修改后的规则不能和已经存在的规则内容相同，否则会导致创建或修改不成功，系统会提示该规则已经存在。

- 当 ACL 的匹配顺序为 **auto** 时，新创建的规则将按照“深度优先”的原则插入到已有的规则中，但是所有规则对应的编号不会改变。

注意

- 用户可以通过命令 **acl ipv6 number** *acl6-number* [**name** *acl6-name*] **match-order** { **auto** | **config** }修改 IPv6 ACL 的匹配顺序为 **auto** 或者 **config**，但必须在 IPv6 ACL 中没有规则的时候修改，对已经有规则的 IPv6 ACL 是无法修改其匹配顺序的。
- 在使用 **rule comment** 命令为规则定义描述信息时，该规则必须存在。

3.3 配置高级 IPv6 ACL

高级 IPv6 ACL 可以使用报文的源 IPv6 地址信息、目的 IPv6 地址信息、IPv6 承载的协议类型、协议的特性（例如 TCP 或 UDP 的源端口、目的端口，ICMP 协议的消息类型、消息码等）等信息来制定匹配规则。

用户可以利用高级 IPv6 ACL 定义比基本 IPv6 ACL 更准确、更丰富、更灵活的规则。

高级 IPv6 ACL 的序号取值范围 3000～3999。

3.3.1 配置准备

如果要配置带有时间段参数的规则，则需要定义相应的时间段。

3.3.2 配置高级 IPv6 ACL

表3-2 配置高级 IPv6 ACL

操作	命令	说明
进入系统视图	**system-view**	-
创建高级 IPv6 ACL 并进入高级 IPv6 ACL 视图	**acl ipv6 number** *acl6-number* [**name** *acl6-name*] [**match-order** { **auto** | **config** }]	必选 缺省情况下，匹配顺序为 **config** 如果用户在创建 IPv6 ACL 时指定了名称，则之后可以通过 **acl ipv6 name** *acl6-name* 命令进入指定名称的 IPv6 ACL 视图
定义规则	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [**established** | **counting** | **destination** { *dest dest-prefix* | *dest/dest-prefix* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **flow-label** *flow-label-value* | **fragment** | **icmpv6-type** { *icmpv6-type icmpv6-code* | *icmpv6-message* } | **logging** | **source** { *source source-prefix* | *source/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name*] *	必选 可以重复本步骤创建多条规则 **logging** 参数需要使用该 ACL 的模块支持日志记录功能才能生效，例如防火墙
定义步长	**step** *step-value*	可选 缺省情况下，步长为 5

操作	命令	说明
定义高级 IPv6 ACL 的描述信息	**description** *text*	可选 缺省情况下，高级 IPv6 ACL 没有描述信息
定义规则的描述信息	**rule** *rule-id* **comment** *text*	可选 缺省情况下，规则没有描述信息

需要注意的是：

- 当 ACL 的匹配顺序为 **config** 时，用户可以修改该 ACL 中的任何一条已经存在的规则，在修改 ACL 中的某条规则时，该规则中没有修改到的部分仍旧保持原来的状态；当 ACL 的匹配顺序为 **auto** 时，用户不能修改该 ACL 中的任何一条已经存在的规则，否则系统会提示错误信息。
- 新创建或修改后的规则不能和已经存在的规则内容相同，否则会导致创建或修改不成功，系统会提示该规则已经存在。
- 当 ACL 的匹配顺序为 **auto** 时，新创建的规则将按照“深度优先”的原则插入到已有的规则中，但是所有规则对应的编号不会改变。

注意

- 用户可以通过命令 **acl ipv6 number** *acl6-number* [**name** *acl6-name*] **match-order** { **auto** | **config** }修改 IPv6 ACL 的匹配顺序为 **auto** 或者 **config**，但必须在 IPv6 ACL 中没有规则的时候修改，对已经有规则的 IPv6 ACL 是无法修改其匹配顺序的。
- 在使用 **rule comment** 命令为规则定义描述信息时，该规则必须存在。

3.4 拷贝 IPv6 ACL

拷贝 IPv6 ACL 功能使用户可以通过拷贝一个已经存在的 IPv6 ACL，生成一个新的同类型的 IPv6 ACL。生成的新的 IPv6 ACL 的匹配顺序、包含的匹配规则、步长以及描述信息都和源 IPv6 ACL 相同。

3.4.1 配置准备

源 IPv6 ACL 必须存在，目的 IPv6 ACL 必须不存在。

3.4.2 拷贝 IPv6 ACL

表3-3 拷贝 IPv6 ACL

操作	命令	说明
进入系统视图	**system-view**	-
拷贝生成一个新的同类型的 IPv6 ACL	**acl ipv6 copy** { *source-acl6-number* | **name** *source-acl6-name* } **to** { *dest-acl6-number* | **name** *dest-acl6-name* }	必选

注意

- 源 IPv6 ACL 和目的 IPv6 ACL 的类型要相同。
- 源 IPv6 ACL 的名称不会拷贝到目的 IPv6 ACL。
- 简单型 IPv6 ACL 不支持拷贝操作。

3.5 IPv6 ACL 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示 IPv6 ACL 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv6 ACL 统计信息。

表3-4 IPv6 ACL 显示和维护

操作	命令
显示配置的 IPv6 ACL 信息	**display acl ipv6** { *acl6-number* | **all** | **name** *acl6-name* }
显示时间段的配置和状态	**display time-range** { *time-range-name* | **all** }
清除 IPv6 ACL 统计信息	**reset acl ipv6 counter** { *acl6-number* | **all** | **name** *acl6-name* }

3.6 IPv6 ACL 典型配置举例

3.6.1 组网需求

在接口 GigabitEthernet1/1 配置 IPv6 报文过滤，允许接收源 IPv6 地址为 4050::9000 到 4050::90FF 的报文，禁止接收其他报文。

3.6.2 配置步骤

配置 IPv6 ACL（假定 IPv6 ACL 编号为 2000），规则如下。

```
<Sysname> system-view
[Sysname] acl ipv6 number 2000
[Sysname-acl6-basic-2000] rule permit source 4050::9000/120
[Sysname-acl6-basic-2000] rule deny source any
[Sysname-acl6-basic-2000] quit
```

配置接口 GigabitEthernet1/1 入方向的 IPv6 报文过滤，IPv6 ACL 号为 2000。

```
[Sysname] firewall enable
[Sysname] interface gigabitethernet 1/1
[Sysname-GigabitEthernet1/1] firewall packet-filter ipv6 2000 inbound
```

目 录

1 IPsec配置 1-1
1.1 IPsec简介 1-1
1.1.1 IPsec的协议实现 1-1
1.1.2 IPsec基本概念 1-2
1.1.3 协议规范 1-4
1.2 IPsec配置任务简介 1-4
1.3 配置访问控制列表 1-4
1.4 配置安全提议 1-7
1.5 配置安全策略 1-8
1.5.1 手工配置安全策略 1-8
1.5.2 配置IKE协商安全策略 1-10
1.6 在接口上应用安全策略组 1-12
1.7 使能加密引擎功能 1-13
1.8 使能解封装后IPsec报文的ACL检查 1-13
1.9 配置IPsec抗重放功能 1-13
1.10 IPsec显示和维护 1-14
1.11 IPsec典型配置举例 1-15
1.11.1 采用手工方式建立IPsec安全隧道 1-15
1.11.2 采用IKE方式建立IPSec安全隧道 1-17
2 IKE配置 2-1
2.1 IKE简介 2-1
2.1.1 IKE的安全机制 2-1
2.1.2 IKE的交换过程 2-1
2.1.3 IKE在IPsec中的作用 2-2
2.1.4 IPsec与IKE的关系 2-3
2.1.5 协议规范 2-3
2.2 IKE配置任务简介 2-3
2.3 配置本端安全网关的名字 2-4
2.4 配置IKE安全提议 2-4
2.5 配置IKE对等体 2-5
2.6 配置Keepalive定时器 2-6
2.7 配置NAT Keepalive定时器 2-7
2.8 配置对等体存活检测 2-7
2.9 配置取消对next payload域的检查 2-8
2.10 IKE显示和维护 2-8
2.11 IKE典型配置举例 2-8
2.11.1 IKE典型配置组网应用 2-8
2.11.2 IKE野蛮模式及NAT穿越的组网应用 2-10
2.11.3 ADSL与IPsec/IKE相结合的组网应用 2-12

2.12 常见错误配置举例 2-15
2.12.1 非法用户身份信息 2-16
2.12.2 提议不匹配 2-16
2.12.3 无法建立安全隧道 2-16
2.12.4 ACL配置错误 2-17

1 IPsec 配置

1.1 IPsec 简介

IPsec（IP Security）是 IETF 制定的三层隧道加密协议，它为 Internet 上传输的数据提供了高质量的、可互操作的、基于密码学的安全保证。特定的通信方之间在 IP 层通过加密与数据源认证等方式，提供了以下的安全服务：

- 数据机密性（Confidentiality）：IPsec 发送方在通过网络传输包前对包进行加密。
- 数据完整性（Data Integrity）：IPsec 接收方对发送方发送来的包进行认证，以确保数据在传输过程中没有被篡改。
- 数据来源认证（Data Authentication）：IPsec 在接收端可以认证发送 IPsec 报文的发送端是否合法。
- 防重放（Anti-Replay）：IPsec 接收方可检测并拒绝接收过时或重复的报文。

IPsec 具有以下优点：

- 支持 IKE（Internet Key Exchange，因特网密钥交换），可实现密钥的自动协商功能，减少了密钥协商的开销。可以通过 IKE 建立和维护安全联盟的服务，简化了 IPsec 的使用和管理。
- 所有使用 IP 协议进行数据传输的应用系统和服务都可以使用 IPsec，而不必对这些应用系统和服务本身做任何修改。
- 对数据的加密是以数据包为单位的，而不是以整个数据流为单位，这不仅灵活而且有助于进一步提高 IP 数据包的安全性，可以有效防范网络攻击。

1.1.1 IPsec 的协议实现

IPsec协议不是一个单独的协议，它给出了应用于IP层上网络数据安全的一整套体系结构，包括网络认证协议AH（Authentication Header，认证头）、ESP（Encapsulating Security Payload，封装安全载荷）、IKE（Internet Key Exchange，因特网密钥交换）和用于网络认证及加密的一些算法等。其中，AH协议和ESP协议用于提供安全服务，IKE协议用于密钥交换。关于IKE的详细介绍请参见“2 IKE配置”，本节不做介绍。

IPsec 提供了两种安全机制：认证和加密。认证机制使 IP 通信的数据接收方能够确认数据发送方的真实身份以及数据在传输过程中是否遭篡改。加密机制通过对数据进行加密运算来保证数据的机密性，以防数据在传输过程中被窃听。

IPsec 协议中的 AH 协议定义了认证的应用方法，提供数据源认证和完整性保证；ESP 协议定义了加密和可选认证的应用方法，提供数据可靠性保证。

- AH 协议（IP 协议号为 51）提供数据源认证、数据完整性校验和防报文重放功能，它能保护通信免受篡改，但不能防止窃听，适合用于传输非机密数据。AH 的工作原理是在每一个数据包上添加一个身份验证报文头，此报文头插在标准 IP 包头后面，对数据提供完整性保护。可选择的认证算法有 MD5（Message Digest）、SHA-1（Secure Hash Algorithm）等。
- ESP 协议（IP 协议号为 50）提供加密、数据源认证、数据完整性校验和防报文重放功能。ESP 的工作原理是在每一个数据包的标准 IP 包头后面添加一个 ESP 报文头，并在数据包后面追加

一个 ESP 尾。与 AH 协议不同的是，ESP 将需要保护的用户数据进行加密后再封装到 IP 包中，以保证数据的机密性。常见的加密算法有 DES、3DES、AES 等。同时，作为可选项，用户可以选择 MD5、SHA-1 算法保证报文的完整性和真实性。

在实际进行 IP 通信时，可以根据实际安全需求同时使用这两种协议或选择使用其中的一种。AH 和 ESP 都可以提供认证服务，不过，AH 提供的认证服务要强于 ESP。同时使用 AH 和 ESP 时，设备支持的 AH 和 ESP 联合使用的方式为：先对报文进行 ESP 封装，再对报文进行 AH 封装，封装之后的报文从内到外依次是原始 IP 报文、ESP 头、AH 头和外部 IP 头。

1.1.2 IPsec 基本概念

1. 安全联盟（Security Association，SA）

IPsec 在两个端点之间提供安全通信，端点被称为 IPsec 对等体。

SA 是 IPsec 的基础，也是 IPsec 的本质。SA 是通信对等体间对某些要素的约定，例如，使用哪种协议（AH、ESP 还是两者结合使用）、协议的封装模式（传输模式和隧道模式）、加密算法（DES、3DES 和 AES）、特定流中保护数据的共享密钥以及密钥的生存周期等。建立 SA 的方式有手工配置和 IKE 自动协商两种。

SA 是单向的，在两个对等体之间的双向通信，最少需要两个 SA 来分别对两个方向的数据流进行安全保护。同时，如果两个对等体希望同时使用 AH 和 ESP 来进行安全通信，则每个对等体都会针对每一种协议来构建一个独立的 SA。

SA 由一个三元组来唯一标识，这个三元组包括 SPI（Security Parameter Index，安全参数索引）、目的 IP 地址、安全协议号（AH 或 ESP）。

SPI 是为唯一标识 SA 而生成的一个 32 比特的数值，它在 AH 和 ESP 头中传输。在手工配置安全联盟时，需要手工指定 SPI 的取值。使用 IKE 协商产生安全联盟时，SPI 将随机生成。

SA 是具有生存周期的，且只对通过 IKE 协商建立的 SA 有效，手工方式建立的 SA 永不老化。IKE 协商建立的 SA 的生存周期有两种定义方式：

- 基于时间的生存周期，定义了一个 SA 从建立到失效的时间；
- 基于流量的生存周期，定义了一个 SA 允许处理的最大流量。

生存周期到达指定的时间或指定的流量，SA 就会失效。SA 失效前，IKE 将为 IPsec 协商建立新的 SA，这样，在旧的 SA 失效前新的 SA 就已经准备好。在新的 SA 开始协商而没有协商好之前，继续使用旧的 SA 保护通信。在新的 SA 协商好之后，则立即采用新的 SA 保护通信。

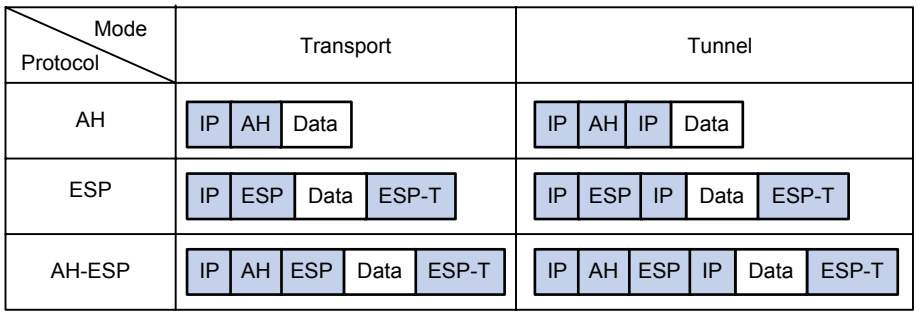
2. 封装模式

IPsec 有如下两种工作模式：

- 隧道（tunnel）模式：用户的整个 IP 数据包被用来计算 AH 或 ESP 头，AH 或 ESP 头以及 ESP 加密的用户数据被封装在一个新的 IP 数据包中。通常，隧道模式应用在两个安全网关之间的通讯。
- 传输（transport）模式：只是传输层数据被用来计算 AH 或 ESP 头，AH 或 ESP 头以及 ESP 加密的用户数据被放置在原 IP 包头后面。通常，传输模式应用在两台主机之间的通讯，或一台主机和一个安全网关之间的通讯。

不同的安全协议在tunnel和transport模式下的数据封装形式如 图 1-1所示，data为传输层数据。

图1-1 安全协议数据封装格式



3. 认证算法与加密算法

(1) 认证算法

认证算法的实现主要是通过杂凑函数。杂凑函数是一种能够接受任意长的消息输入，并产生固定长度输出的算法，该输出称为消息摘要。IPsec 对等体计算摘要，如果两个摘要是相同的，则表示报文是完整未经篡改的。IPsec 使用两种认证算法：

- MD5：MD5 通过输入任意长度的消息，产生 128bit 的消息摘要。
- SHA-1：SHA-1 通过输入长度小于 2 的 64 次方 bit 的消息，产生 160bit 的消息摘要。

MD5 算法的计算速度比 SHA-1 算法快，而 SHA-1 算法的安全强度比 MD5 算法高。

(2) 加密算法

加密算法实现主要通过对称密钥系统，它使用相同的密钥对数据进行加密和解密。目前设备的 IPsec 实现三种加密算法：

- DES（Data Encryption Standard）：使用 56bit 的密钥对一个 64bit 的明文块进行加密。
- 3DES（Triple DES）：使用三个 56bit 的 DES 密钥（共 168bit 密钥）对明文进行加密。
- AES（Advanced Encryption Standard）：使用 128bit、192bit 或 256bit 密钥长度的 AES 算法对明文进行加密。

这三个加密算法的安全性由高到低依次是：AES、3DES、DES，安全性高的加密算法实现机制复杂，运算速度慢。对于普通的安全要求，DES 算法就可以满足需要。

4. 协商方式

有如下两种协商方式建立 SA：

- 手工方式（manual）配置比较复杂，创建 SA 所需的全部信息都必须手工配置，而且不支持一些高级特性（例如定时更新密钥），但优点是可以不依赖 IKE 而单独实现 IPsec 功能。
- IKE 自动协商（isakmp）方式相对比较简单，只需要配置好 IKE 协商安全策略的信息，由 IKE 自动协商来创建和维护 SA。

当与之进行通信的对等体设备数量较少时，或是在小型静态环境中，手工配置 SA 是可行的。对于中、大型的动态网络环境中，推荐使用 IKE 协商建立 SA。

5. 安全隧道

安全隧道是建立在本端和对端之间可以互通的一个通道，它由一对或多对 SA 组成。

1.1.3 协议规范

与 IPsec 相关的协议规范有：

- RFC2401：Security Architecture for the Internet Protocol
- RFC2402：IP Authentication Header
- RFC2406：IP Encapsulating Security Payload

1.2 IPsec 配置任务简介

目前设备实现了上述所介绍的 IPsec 的全部内容。IPsec 基本配置的实现方式是基于如下思路：

(1) 通过 IPsec，对等体之间（此处是指设备及其对端）能够对不同的数据流实施不同的安全保护（认证、加密或两者同时使用），数据流的区分通过配置 ACL 来进行；

(2) 安全保护所用到的安全协议、认证算法和加密算法、封装模式等通过配置安全提议来进行；

(3) 数据流和安全提议的关联（即定义对何种数据流实施何种保护）、SA 的协商方式、对等体 IP 地址的设置（即保护路径的起/终点）、所需要的密钥和 SA 的生存周期等通过配置安全策略来进行；

(4) 最后在设备接口上实施安全策略即完成了 IPsec 的配置。若通过软件实现，则在接口上应用安全策略。

表1-1 IPsec 配置任务简介

配置任务		说明	详细配置
IPsec 基本配置	配置访问控制列表	必选	1.3
	配置安全提议		1.4
	配置安全策略		1.5
	在接口上应用安全策略组		1.6
使能加密引擎		可选	1.7
使能解封装后 IPsec 报文的 ACL 检查开关		可选	1.8
配置 IPsec 抗重放功能		可选	1.9

注意

通常情况下，由于 IKE 协议采用 UDP 的 500 端口进行通信，IPsec 的 AH 和 ESP 协议分别使用 51 或 50 号协议来工作，因此为保障 IKE 和 IPsec 的正常运行，需要确保应用了 IKE 和 IPsec 配置的接口上没有禁止掉属于以上端口和协议的流量。

1.3 配置访问控制列表

ACL（Access Control List，访问控制列表）是用来实现流识别功能的。网络设备为了过滤报文，需要配置一系列的匹配条件对报文进行分类，当设备的端口接收到报文后，即根据当前端口上应用的 ACL 规则对报文进行分析、识别之后，根据预先设定的策略对报文进行不同的处理。

(1) ACL 规则中关键字的使用

IPsec 通过配置 ACL 来定义需要过滤的数据流。在 IPsec 的应用中，ACL 规则中的 **permit** 关键字表示与之匹配的流量需要被 IPsec 保护，而规则中的 **deny** 关键字则表示与之匹配的那些流量不需要保护。一个 ACL 中可以配置多条规则，首个与数据流匹配上的规则决定了对该数据流的处理方式，如果该规则为 **permit**，则该规则就定义了需要建立 SA 来保护的数据流量的范围。

在 IPsec 策略中定义的 ACL 既可用于过滤接口入方向数据流，也可用于过滤接口出方向数据流，且支持多 VPN 多实例。在出方向上，与 ACL 的 **permit** 规则匹配的报文将被 IPsec 保护，未匹配上任何规则或与 **deny** 规则匹配上的报文将不被 IPsec 保护。在入方向上，与 ACL 的 **permit** 规则匹配上的 IPsec 报文将被接收并处理，与 ACL 的 **permit** 规则匹配上的普通 IP 报文将被丢弃。

需要注意的是：

- 仅对确实需要 IPsec 保护的数据流配置 **permit** 规则，避免盲目地使用关键字 **any**。这是因为，在一个 **permit** 规则中使用 **any** 关键字就代表所有出方向流量都需要被 IPsec 保护，所有入方向上被保护的 IPsec 报文将被接收并处理，入方向上未被保护的 IPsec 报文将被丢弃。这种情况下，一旦入方向收到的某流量是未被 IPsec 保护的，那么该流量就会被丢弃，这会造成一些本不需要 IPsec 处理的流量丢失，影响正常的业务流传输。
- 合理使用 **deny** 规则，尤其是在一个安全策略下有多条优先级不同的子安全策略时，避免本应该与优先级较低的子安全策略的 ACL **permit** 规则匹配而被 IPsec 保护的出方向报文，因为先与优先级较高的子安全策略的 ACL **deny** 规则匹配上，而在接收端被当作未被 IPsec 保护的报文丢弃。

deny 规则的错误配置示例：（以下配置信息仅截取了 ACL 的相关内容，其它步骤省略）

Router A 连接的 1.1.2.0/24 网段到 Router B 连接的 3.3.3.0/24 网段之间的报文，在应用了 IPsec 策略 test 的出接口上，优先与 ACL 3000 的 rule 1 deny ip 匹配，Router A 认为它不需要 IPsec 保护，到达 Router B 后将被丢弃。

Router A 上的配置如下：

```
acl number 3000
 rule 0 permit ip source 1.1.1.0 0.0.0.255 destination 2.2.2.0 0.0.0.255
 rule 1 deny ip
acl number 3001
 rule 0 permit ip source 1.1.2.0 0.0.0.255 destination 3.3.3.0 0.0.0.255
 rule 1 deny ip
#
ipsec policy test 1 isakmp
 security acl 3000
 ike-peer aa
 proposal 1
#
ipsec policy test 2 isakmp
 security acl 3001
 ike-peer bb
 proposal 1
```

Router B 上的配置如下：

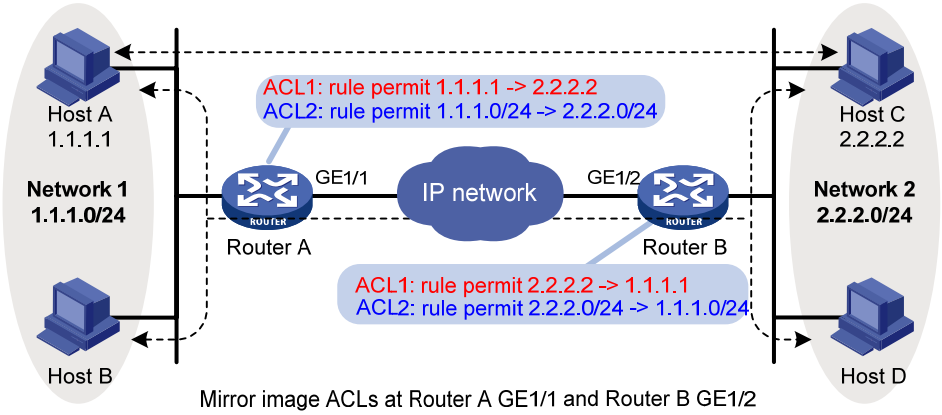
```
acl number 3001
```

```
 rule 0 permit ip source 3.3.3.0 0.0.0.255 destination 1.1.2.0 0.0.0.255
 rule 1 deny ip
#
ipsec policy test 1 isakmp
 security acl 3001
 ike-peer aa
 proposal 1
```

(2) ACL 规则的配置

为保证SA的成功建立，建议将IPsec对等体上的访问控制列表镜像配置，即保证两端要保护的数据流范围是镜像的。例如，图 1-2中Router A和Router B上的ACL配置都是完全镜像对称的，因此用于保护主机Host A与主机Host C之间、子网Network 1 与子网Network 2 之间流量的SA均可成功建立。

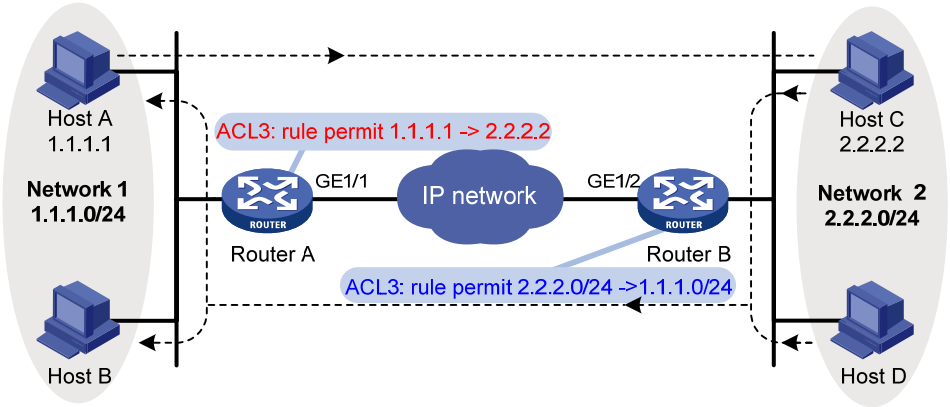
图1-2 镜像 ACL 配置



Mirror image ACLs at Router A GE1/1 and Router B GE1/2

若IPsec对等体上的访问控制列表配置非镜像，那么只有一种情况下，SA的协商是可以建立的。这种情况就是，一端的访问控制列表规则定义的范围是另外一端的子集。如 图 1-3所示，Router A上的访问控制列表允许的范围（Host A->Host C）是Router B上访问控制列表（Network 2->Network 1）的子集。

图1-3 非镜像 ACL 配置



但需要注意的是，在这种 ACL 配置下，并不是任何一端发起的 SA 协商都可以成功，仅当保护范围小（细粒度）的一端向保护范围大（粗粒度）的一端发起的协商才能成功，反之则协商失败。这是因为，协商响应方要求协商发起方发送过来的数据必须在响应方可以接受的范围之内。其结果就是，

从细粒度一端向粗粒度一端发起的协商是可以成功的，例如 Host A->Host C；从粗粒度一方向细粒度一方发起的协商是不能成功的，例如 Host C->Host B、Host D->Host A 等。

(3) 数据流的保护方式

另外，还需要注意的是，仅对确实需要 IPsec 保护的数据流配置 **permit** 规则，避免盲目地使用关键字 **any**。这是因为，在一个 **permit** 规则中使用 **any** 关键字就代表所有出方向流量都需要被 IPsec 保护，所有入方向上被保护的 IPsec 报文将被接收并处理，入方向上未被保护的 IPsec 报文将被丢弃。这种情况下，一旦入方向收到的某流量是未被 IPsec 保护的，那么该流量就会被丢弃，这会造成一些本不需要 IPsec 处理的流量丢失，影响正常的业务流传输。

目前，设备支持的数据流的保护方式包括以下两种：

- 标准方式：一条隧道保护一条数据流。ACL 中的每一个规则对应的数据流都会由一条单独创建的隧道来保护；
- 聚合方式：一条隧道保护 ACL 中定义的所有数据流。ACL 中的所有规则对应的数据流只会由一条创建的隧道来保护。该方式仅在 IKE 协商安全策略的情况下可配。

在本地和远端设备上定义的访问控制列表必须是相对应的，即本端访问控制列表定义的源 IP 地址要与对端的目的 IP 地址保持一致，本端的目的 IP 地址与对端的源 IP 地址一致。

说明

- ACL 的具体配置请参见“安全分册”中的“ACL 配置”。
- 若在接口上同时使能 IPsec 和 QoS，同一个 IPsec 安全联盟的数据流如果被 QoS 分类进入不同队列，会导致部分报文发送乱序。由于 IPsec 具有防重放功能，IPsec 入方向上对于防重放窗口之外的报文会进行丢弃，从而导致丢包现象。因此当 IPsec 与 QoS 结合使用时，必须保证 IPsec 分类与 QoS 分类规则配置保持一致。IPsec 的分类规则完全由引用的 ACL 规则确定，QoS 分类规则的配置请参考“QoS 分册”中的“QoS 配置”。

1.4 配置安全提议

安全提议是安全策略的一个组成部分，它用于保存 IPsec 需要使用的特定安全协议、加密/认证算法以及封装模式，为 IPsec 协商 SA 提供各种安全参数。

表1-2 配置安全提议

操作	命令	说明
进入系统视图	**system-view**	-
创建安全提议，并进入安全提议视图	**ipsec proposal** *proposal-name*	必选 缺省情况下，没有任何安全提议存在
配置安全提议采用的安全协议	**transform** { **ah** | **ah-esp** | **esp** }	可选 缺省情况下，采用 ESP 协议

操作		命令	说明
配置安全算法	配置 ESP 协议采用的加密算法	**esp encryption-algorithm { 3des | aes [*key-length*] | des }**	可选 缺省情况下，ESP 协议采用 DES 加密算法
	配置 ESP 协议采用的认证算法	**esp authentication-algorithm { md5 | sha1 }**	可选 缺省情况下，ESP 协议采用 MD5 认证算法
	配置 AH 协议采用的认证算法	**ah authentication-algorithm { md5 | sha1 }**	可选 缺省情况下，AH 协议采用 MD5 认证算法
配置安全协议对 IP 报文的封装形式		**encapsulation-mode { transport | tunnel }**	可选 缺省情况下，安全协议采用隧道模式对 IP 报文进行封装 传输模式必须应用于数据流的源地址和目的地址与安全隧道两端地址相同的情况下

说明

- 可对安全提议进行修改，但对已协商成功的 SA，新修改的安全提议并不起作用，即 SA 仍然使用原来的安全提议（除非使用 **reset ipsec sa** 命令重置），只有新协商的 SA 将使用新的安全提议。
- 只有选择了相应的安全协议后，该安全协议所需的安全算法才可配置。例如，如果使用 **transform** 命令选择了 **esp**，那么只有 ESP 所需的安全算法才可配置，而 AH 所需的安全算法则不能配置。ESP 协议允许对报文同时进行加密和认证，或只加密，或只认证。
- 最多可以创建 50 个安全提议。

1.5 配置安全策略

安全策略规定了对什么样的数据流采用什么样的安全提议。一条安全策略由“名字”和“顺序号”共同唯一确定。

- 手工配置方式：需要用户手工配置密钥、SPI 等参数，在隧道模式下还需要手工配置安全隧道两个端点的 IP 地址；
- IKE 协商方式：由 IKE 自动协商生成各参数。

1.5.1 手工配置安全策略

1. 配置准备

手工配置安全策略时，除完成该安全策略需要引用的访问控制列表及安全提议的配置之外，为保证 SA 的协商成功，安全隧道两端的配置必须符合以下要求：

手工配置安全策略时，在安全隧道的两端，必须满足以下配置要求：

- 安全策略引用的安全提议应采用相同的安全协议、安全算法和报文封装形式；
- 当前端点的对端地址与对端的本端地址应保持一致；
- SA 的 SPI 及密钥必须是完全匹配的。即，本端的入方向 SA 的 SPI 及密钥必须和对端的出方向 SA 的 SPI 及密钥相同；本端的出方向 SA 的 SPI 及密钥必须和对端的入方向 SA 的 SPI 及密钥相同；

- SA 使用的密钥应当以相同的方式输入。即，一端以字符串方式输入密钥，另一端必须也以字符串方式输入密钥。

2. 手工配置安全策略

表1-3 手工配置安全策略

操作		命令	说明
进入系统视图		**system-view**	-
用手工方式创建一条安全策略，并进入安全策略视图		**ipsec policy** *policy-name seq-number* **manual**	必选 缺省情况下，没有任何安全策略存在
配置安全策略引用的访问控制列表		**security acl** *acl-number*	必选 缺省情况下，安全策略没有指定访问控制列表
配置安全策略所引用的安全提议		**proposal** *proposal-name*	必选 缺省情况下，安全策略没有引用任何安全提议
配置隧道的起点与终点	配置安全隧道的本端地址	**tunnel local** *ip-address*	必选 缺省情况下，没有配置安全隧道的本端地址
	配置安全隧道的对端地址	**tunnel remote** *ip-address*	必选 缺省情况下，没有配置安全隧道的对端地址
配置安全联盟的安全参数索引参数		**sa spi** { **inbound** | **outbound** } { **ah** | **esp** } *spi-number*	必选
配置 SA 使用的密钥	配置协议的认证密钥（以 16 进制方式输入）	**sa authentication-hex** { **inbound** | **outbound** } { **ah** | **esp** } *hex-key*	二者必选其一 对于 ESP 协议，配置认证密钥时，系统会自动地同时生成认证算法的密钥和加密算法的密钥
	配置协议的认证密钥（以字符串方式输入）	**sa string-key** { **inbound** | **outbound** } { **ah** | **esp** } *string-key*	
	配置 ESP 协议的加密密钥（以字符串方式输入）	**sa string-key** { **inbound** | **outbound** } **esp** *string-key*	二者必选其一 以字符串方式输入加密密钥时，系统会自动地同时生成认证算法的密钥和加密算法的密钥
	配置 ESP 协议的加密密钥(以 16 进制方式输入）	**sa encryption-hex** { **inbound** | **outbound** } **esp** *hex-key*	

说明

- 一条安全策略只能引用一条访问控制列表，如果设置安全策略引用了多于一个访问控制列表，最后引用的那条访问控制列表才有效。
- 通过手工方式建立 SA，一条安全策略只能引用一个安全提议，并且如果已经引用了安全提议，必须先取消原先的安全提议才能引用新的安全提议。
- 在为系统配置 SA 时，必须分别设置 **inbound** 和 **outbound** 两个方向 SA 的参数。同时，为保证安全联盟的唯一性，不同 SA 必须对应不同的 SPI。
- 如果分别以两种方式输入了密钥，则最后设定的密钥有效。
- 对于手工方式创建的安全策略，不能直接修改它的创建方式，而必须先删除该安全策略然后再重新创建。

1.5.2 配置 IKE 协商安全策略

在采用 IKE 方式配置安全策略时，有以下两种方式：

- 直接配置安全策略，，在安全策略视图中定义需要协商的各参数；
- 引用安全策略模板创建安全策略，在安全策略模板中定义需要协商的各参数。应用了该类安全策略的设备不能发起协商，仅可以响应远端设备的协商请求。由于策略模板中未定义的可选参数由发起方来决定，而响应方会接受发起方的建议，因此这种方式适用于通信对端（例如对端的 IP 地址）未知的情况下，允许这些对端设备向本端设备主动发起协商。

1. 配置准备

在配置 IKE 协商安全策略之前，需要完成以下配置：

- 配置所引用的访问控制列表和安全提议。
- 配置IKE对等体。具体配置请参见“2.5 配置IKE对等体”。

为保证 IKE 协商成功，安全策略中所有配置的参数必须在本端和对端相匹配。

2. 配置 IKE 协商安全策略

(1) 直接配置 IKE 协商安全策略

表1-4 直接配置 IKE 协商安全策略

操作	命令	说明
进入系统视图	**system-view**	-
创建一条安全策略，并进入安全策略视图	**ipsec policy** *policy-name seq-number* **isakmp**	必选 缺省情况下，没有安全策略存在
配置安全策略引用的访问控制列表	**security acl** *acl-number* [**aggregation**]	必选 缺省情况下，安全策略没有指定访问控制列表
配置安全策略所引用的安全提议	**proposal** *proposal-name*&<1-6>	必选 缺省情况下，安全策略没有引用任何提议
在安全策略中引用 IKE 对等体	**ike-peer** *peer-name*	必选
配置使用此安全策略发起协商时使用 PFS 特性	**pfs** { **dh-group1** | **dh-group2** | **dh-group5** | **dh-group14** }	可选 缺省情况下，安全策略发起协商时没有使用 PFS 特性 PFS（Perfect Forward Secrecy，完善的前向安全）特性请参见“2.1.1 IKE的安全机制”
配置 SA 的生存周期	**sa duration** { **time-based** *seconds* | **traffic-based** *kilobytes* }	可选 缺省情况下，安全策略的 SA 生存周期为当前全局的 SA 生存周期值
退回系统视图	**quit**	-
配置全局 SA 的生存周期	**ipsec sa global-duration** { **time-based** *seconds* | **traffic-based** *kilobytes* }	可选 缺省情况下，安全联盟基于时间的生存周期为 3600 秒，基于流量的生存周期为 1843200 千字节

(2) 引用安全策略模板配置 IKE 协商安全策略

安全策略模板可配置的参数与 IKE 方式的 IPsec 安全策略相同，只是很多参数是可选的。

- 必须配置的参数：IPsec 安全提议和 IKE 对等体，
- 可选配的参数：访问控制列表、PFS 特性和生存周期。与直接方式不同的是，用于定义保护对象范围的访问控制列表在这种方式下是可选的，该参数在未配置的情况下，相当于支持最大范围的保护，即接受协商发起端的访问控制列表设置。

表1-5 引用安全策略模板配置 IKE 协商安全策略

操作	命令	说明
进入系统视图	**system-view**	-
创建一个安全策略模板，并进入安全策略模板视图	**ipsec policy-template** *template-name seq-number*	必选 缺省情况下，没有任何安全策略模板存在
配置安全策略引用的访问控制列表	**security acl** *acl-number*	可选 缺省情况下，安全策略没有指定访问控制列表
配置安全策略所引用的安全提议	**proposal** *proposal-name*&<1-6>	必选 缺省情况下，安全策略没有引用任何提议
在安全策略中引用 IKE 对等体	**ike-peer** *peer-name*	必选
配置使用此安全策略发起协商时使用 PFS 特性	**pfs** { **dh-group1** | **dh-group2** | **dh-group5** | **dh-group14** }	可选 缺省情况下，安全策略发起协商时没有使用 PFS 特性 PFS（Perfect Forward Secrecy，完善的前向安全）特性请参见“2.1.1 IKE的安全机制”
配置 SA 的生存周期	**sa duration** { **time-based** *seconds* | **traffic-based** *kilobytes* }	可选 缺省情况下，安全策略的 SA 生存周期为当前全局的 SA 生存周期值
退回系统视图	**quit**	-
配置全局 SA 的生存周期	**ipsec sa global-duration** { **time-based** *seconds* | **traffic-based** *kilobytes* }	可选 缺省情况下，安全联盟基于时间的生存周期为 3600 秒，基于流量的生存周期为 1843200 千字节
引用安全策略模板创建一条安全策略	**ipsec policy** *policy-name seq-number* **isakmp template** *template-name*	必选 缺省情况下，没有安全策略存在

说明

- 引用安全策略模板创建一条安全策略之后，就不能进入该安全策略视图下进行安全策略的配置与修改了，只能进入安全策略模板视图下配置或修改。
- 一条安全策略只能引用一条访问控制列表，如果设置安全策略引用了多于一个访问控制列表，最后配置的那条访问控制列表才有效。
- 通过 IKE 协商建立 SA，一条安全策略最多可以引用六个安全提议，IKE 协商将在安全隧道的两端搜索能够完全匹配的安全提议。如果 IKE 在两端找不到完全匹配的安全提议，则 SA 不能建立，需要被保护的报文将被丢弃。
- IKE 在使用安全策略发起一个协商时，可以进行一个 PFS 交换。如果本端配置了 PFS 特性，则发起协商的对端也必须配置 PFS 特性，而且本端和对端指定的 DH 组必须一致，否则协商会失败。
- 所有在安全策略视图下没有单独配置生存周期的安全联盟，都采用全局生存周期。IKE 为 IPsec 协商建立安全联盟时，采用本地设置的和对端提议的生存周期中较小的一个。
- 对于 IKE（无论直接或引用模板）方式创建的安全策略，不能直接修改它的创建方式，而必须先删除该安全策略然后再重新创建。

1.6 在接口上应用安全策略组

安全策略组是所有具有相同名字、不同顺序号的安全策略的集合。在同一个安全策略组中，顺序号越小的安全策略，优先级越高。

为使定义的 SA 生效，应在每个要加密的数据流和要解密的数据流所在接口（逻辑的或物理的）上应用一个安全策略组，以对数据进行保护。当取消安全策略组在接口上的应用后，此接口便不再具有 IPsec 的安全保护功能。

当从一个接口发送数据时，将按照从小到大的顺序号查找安全策略组中每一条安全策略。如果数据匹配了一条安全策略引用的访问控制列表，则使用这条安全策略对数据进行处理；如果数据没有匹配安全策略引用的访问控制列表，则继续查找下一条安全策略；如果数据与所有安全策略引用的访问控制列表都不匹配，则直接被发送（IPsec 不对数据加以保护）。

IPsec 安全策略除了可以应用到串口、以太网口等实际物理接口上之外，还能够应用到 Tunnel、Virtual Template 等虚接口上。这样就可以根据实际组网要求，在如 GRE、L2TP 等隧道上应用。

表1-6 在接口上应用安全策略组

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
应用指定的安全策略组	**ipsec policy** *policy-name*	必选

说明

一个接口只能应用一个安全策略组。通过 IKE 方式创建的安全策略可以应用到多个接口上，通过手工创建的安全策略只能应用到一个接口上。

1.7 使能加密引擎功能

加密引擎是设备上的一个协处理器，为 IPsec 的处理提供一个加/解密算法接口。

表1-7 使能加密引擎功能

操作	命令	说明
进入系统视图	**system-view**	-
使能加密引擎功能（集中式设备）	**cryptoengine enable**	可选 缺省情况下，加密引擎功能处于使能状态。
使能加密引擎功能（分布式设备）	**cryptoengine enable [slot** *slot-number*]	

1.8 使能解封装后 IPsec 报文的 ACL 检查

在隧道模式下，对于解封装之后的入方向 IPsec 报文，有可能出现报文的内部 IP 头不在当前安全策略配置的 ACL 保护范围内的情况，如网络中恶意构造的攻击报文头可能不在此范围，所以需要重新检查报文内部 IP 头是否在 ACL 保护范围内。使能该功能后可以保证 ACL 检查不通过的报文被丢弃，从而提高网络安全性。

表1-8 使能解封装后 IPsec 报文的 ACL 检查功能

操作	命令	说明
进入系统视图	**system-view**	-
使能解封装后 IPsec 报文的 ACL 检查功能	**ipsec decrypt check**	可选 缺省情况下，解封装后 IPsec 报文的 ACL 检查功能处于使能状态

1.9 配置 IPsec 抗重放功能

通常，重放报文是指已经处理过的报文。IPsec 通过滑动窗口（抗重放窗口）机制检测重放报文。AH 和 ESP 协议报文中带有序列号，如果报文的序列号与已经解封装过的报文序列号相同，或报文的序列号出现得较早，即已经超过了抗重放窗口的范围，则认为该报文为重放报文。

由于对重放报文的解封装无实际作用，并且解封装过程涉及密码学运算，会消耗设备大量的资源，导致业务可用性下降，实际上构成了拒绝服务攻击。通过使能 IPsec 抗重放检测功能，将检测到的重放报文在解封装处理之前丢弃，可以降低设备资源的消耗。

另外，在某些特定环境下，业务数据报文的序列号顺序可能与正常的顺序差别较大，虽然并非有意的重放攻击，但会被抗重放检测认为是重放报文，导致业务数据报文被丢弃，影响业务的正常运行。因此，这种情况下就可以通过关闭 IPsec 抗重放检测功能来避免业务数据报文的错误丢弃，也可以通过适当地增大抗重放窗口的宽度，来适应业务正常运行的需要。

表1-9 配置 IPsec 抗重放功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPsec 抗重放检测功能	**ipsec anti-replay check**	可选 缺省情况下，IPsec 抗重放检测功能处于使能状态 需要注意的是，目前设备暂不支持关闭 IPsec 抗重放检测功能
配置 IPsec 抗重放窗口宽度	**ipsec anti-replay window** *width*	可选 缺省情况下，IPsec 抗重放窗口宽度为 32

注意

使用较大的抗重放窗口宽度会引起系统开销增大，导致系统性能下降，与抗重放检测用于降低系统在接收重放报文时的开销的初衷不符，因此建议在能够满足业务运行需要的情况下，使用较小的抗重放窗口宽度。

说明

按照 IPsec 协议，只有 IKE 协商的 IPsec SA 才能够支持抗重放检测，手工方式生成的 IPsec SA 不支持抗重放检测。因此该功能使能与否与对手工方式生成的 IPsec SA 没有影响。

1.10 IPsec 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPsec 的运行情况，通过查看显示信息认证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPsec 统计信息。

表1-10 IPsec 显示和维护

操作	命令
显示安全策略的信息	**display ipsec policy** [**brief** | **name** *policy-name* [*seq-number*]]
显示安全策略模板的信息	**display ipsec policy-template** [**brief** | **name** *template-name* [*seq-number*]]
显示安全提议的信息	**display ipsec proposal** [*proposal-name*]
显示安全联盟的相关信息	**display ipsec sa** [**brief** | **duration** | **policy** *policy-name* [*seq-number*] | **remote** *ip-address*]
显示 IPsec 处理报文的统计信息	**display ipsec statistics**
显示 IPsec 隧道的信息	**display ipsec tunnel**
清除已经建立的安全联盟	**reset ipsec sa** [**parameters** *dest-address protocol spi* | **policy** *policy-name* [*seq-number*] | **remote** *ip-address*]
清除 IPsec 的报文统计信息	**reset ipsec statistics**

1.11 IPsec 典型配置举例

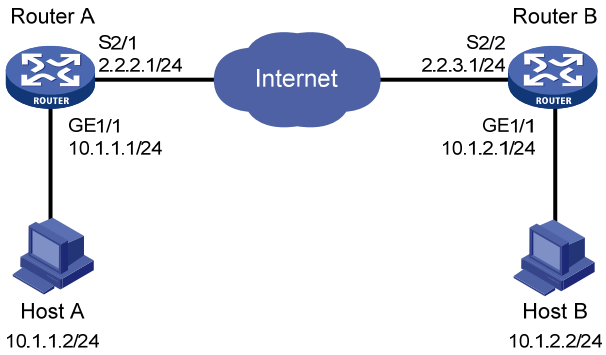
1.11.1 采用手工方式建立 IPsec 安全隧道

1. 组网需求

- 在 Router A 和 Router B 之间建立一个安全隧道，对 Host A 所在的子网（10.1.1.0/24）与 Host B 所在的子网（10.1.2.0/24）之间的数据流进行安全保护。
- 安全协议采用 ESP 协议，加密算法采用 DES，认证算法采用 SHA1-HMAC-96。

2. 组网图

图1-4 IPsec 配置组网图



3. 配置步骤

(1) 配置 Router A

配置一个访问控制列表，定义由子网 10.1.1.0/24 去子网 10.1.2.0/24 的数据流。

```
<RouterA> system-view
[RouterA] acl number 3101
[RouterA-acl-adv-3101] rule permit ip source 10.1.1.0 0.0.0.255 destination 10.1.2.0
0.0.0.255
[RouterA-acl-adv-3101] quit
```

配置到 Host B 的静态路由。

```
[RouterA] ip route-static 10.1.2.0 255.255.255.0 serial 2/1
```

创建名为 tran1 的安全提议。

```
[RouterA] ipsec proposal tran1
```

报文封装形式采用隧道模式。

```
[RouterA-ipsec-proposal-tran1] encapsulation-mode tunnel
```

安全协议采用 ESP 协议。

```
[RouterA-ipsec-proposal-tran1] transform esp
```

选择算法。

```
[RouterA-ipsec-proposal-tran1] esp encryption-algorithm des
[RouterA-ipsec-proposal-tran1] esp authentication-algorithm sha1
[RouterA-ipsec-proposal-tran1] quit
```

创建一条安全策略，协商方式为 manual。

```
[RouterA] ipsec policy map1 10 manual
```

引用访问控制列表。

```
[RouterA-ipsec-policy-manual-map1-10] security acl 3101
```

引用安全提议。

```
[RouterA-ipsec-policy-manual-map1-10] proposal tran1
```

配置对端地址。

```
[RouterA-ipsec-policy-manual-map1-10] tunnel remote 2.2.3.1
```

配置本端地址。

```
[RouterA-ipsec-policy-manual-map1-10] tunnel local 2.2.2.1
```

配置 SPI。

```
[RouterA-ipsec-policy-manual-map1-10] sa spi outbound esp 12345
[RouterA-ipsec-policy-manual-map1-10] sa spi inbound esp 54321
```

配置密钥。

```
[RouterA-ipsec-policy-manual-map1-10] sa string-key outbound esp abcdefg
[RouterA-ipsec-policy-manual-map1-10] sa string-key inbound esp gfedcba
[RouterA-ipsec-policy-manual-map1-10] quit
```

配置串口的 IP 地址。

```
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ip address 2.2.2.1 255.255.255.0
```

在串口上应用安全策略组。

```
[RouterA-Serial2/1] ipsec policy map1
```

(2) 配置 Router B

配置一个访问控制列表，定义由子网 10.1.2.0/24 去子网 10.1.1.0/24 的数据流。

```
<RouterB> system-view
[RouterB] acl number 3101
[RouterB-acl-adv-3101] rule permit ip source 10.1.2.0 0.0.0.255 destination 10.1.1.0
0.0.0.255
[RouterB-acl-adv-3101] quit
```

配置到 HostA 的静态路由。

```
[RouterB] ip route-static 10.1.1.0 255.255.255.0 serial 2/2
```

创建名为 tran1 的安全提议。

```
[RouterB] ipsec proposal tran1
```

报文封装形式采用隧道模式。

```
[RouterB-ipsec-proposal-tran1] encapsulation-mode tunnel
```

安全协议采用 ESP 协议。

```
[RouterB-ipsec-proposal-tran1] transform esp
```

选择算法。

```
[RouterB-ipsec-proposal-tran1] esp encryption-algorithm des
```

```
[RouterB-ipsec-proposal-tran1] esp authentication-algorithm sha1
[RouterB-ipsec-proposal-tran1] quit
```

创建一条安全策略，协商方式为 manual。

```
[RouterB] ipsec policy use1 10 manual
```

引用访问控制列表。

```
[RouterB-ipsec-policy-manual-use1-10] security acl 3101
```

引用安全提议。

```
[RouterB-ipsec-policy-manual-use1-10] proposal tran1
```

配置对端地址。

```
[RouterB-ipsec-policy-manual-use1-10] tunnel remote 2.2.2.1
```

配置本端地址。

```
[RouterB-ipsec-policy-manual-use1-10] tunnel local 2.2.3.1
```

配置 SPI。

```
[RouterB-ipsec-policy-manual-use1-10] sa spi outbound esp 54321
[RouterB-ipsec-policy-manual-use1-10] sa spi inbound esp 12345
```

配置密钥。

```
[RouterB-ipsec-policy-manual-use1-10] sa string-key outbound esp gfedcba
[RouterB-ipsec-policy-manual-use1-10] sa string-key inbound esp abcdefg
[RouterB-ipsec-policy-manual-use1-10] quit
```

配置串口的 IP 地址。

```
[RouterB] interface serial 2/2
[RouterB-Serial2/2] ip address 2.2.3.1 255.255.255.0
```

在串口上应用安全策略组。

```
[RouterB-Serial2/2] ipsec policy use1
```

以上配置完成后，Router A 和 Router B 之间的安全隧道就建立好了，子网 10.1.1.0/24 与子网 10.1.2.0/24 之间的数据流将被加密传输。

1.11.2 采用 IKE 方式建立 IPSec 安全隧道

1. 组网需求

- 如 图 1-4所示，在Router A和Router B之间建立一个安全隧道，对Host A所在的子网（10.1.1.0/24）与Host B所在的子网（10.1.2.0/24）之间的数据流进行安全保护。
- 安全协议采用 ESP 协议，加密算法采用 DES，认证算法采用 SHA1-HMAC-96。

2. 组网图

见 图 1-4。

3. 配置步骤

(1) 配置 Router A

配置一个访问控制列表，定义由子网 10.1.1.0/24 去子网 10.1.2.0/24 的数据流。

```
<RouterA> system-view
```

```
[RouterA] acl number 3101
[RouterA-acl-adv-3101] rule permit ip source 10.1.1.0 0.0.0.255 destination 10.1.2.0
0.0.0.255
[RouterA-acl-adv-3101] quit
```

配置到 Host B 的静态路由。

```
[RouterA] ip route-static 10.1.2.0 255.255.255.0 serial 2/1
```

创建名为 tran1 的安全提议。

```
[RouterA] ipsec proposal tran1
```

报文封装形式采用隧道模式。

```
[RouterA-ipsec-proposal-tran1] encapsulation-mode tunnel
```

安全协议采用 ESP 协议。

```
[RouterA-ipsec-proposal-tran1] transform esp
```

选择算法。

```
[RouterA-ipsec-proposal-tran1] esp encryption-algorithm des
[RouterA-ipsec-proposal-tran1] esp authentication-algorithm sha1
[RouterA-ipsec-proposal-tran1] quit
```

配置 IKE 对等体。

```
[RouterA] ike peer peer
[RouterA-ike-peer-peer] pre-share-key abcde
[RouterA-ike-peer-peer] remote-address 2.2.3.1
[RouterA-ike-peer-peer] quit
```

创建一条安全策略，协商方式为 isakmp。

```
[RouterA] ipsec policy map1 10 isakmp
```

引用安全提议。

```
[RouterA-ipsec-policy-isakmp-map1-10] proposal tran1
```

引用访问控制列表。

```
[RouterA-ipsec-policy-isakmp-map1-10] security acl 3101
```

引用 IKE 对等体。

```
[RouterA-ipsec-policy-isakmp-map1-10] ike-peer peer
[RouterA-ipsec-policy-isakmp-map1-10] quit
```

配置串口的 IP 地址。

```
[RouterA] interface serial 2/1
[RouterA-Serial2/1] ip address 2.2.2.1 255.255.255.0
```

在串口上应用安全策略组。

```
[RouterA-Serial2/1] ipsec policy map1
```

(2) 配置 Router B

配置一个访问控制列表，定义由子网 10.1.2.0/24 去子网 10.1.1.0/24 的数据流。

```
<RouterB> system-view
[RouterB] acl number 3101
```

```
[RouterB-acl-adv-3101] rule permit ip source 10.1.2.0 0.0.0.255 destination 10.1.1.0
0.0.0.255
[RouterB-acl-adv-3101] quit
```

配置到 Host A 的静态路由。

```
[RouterB] ip route-static 10.1.1.0 255.255.255.0 serial 2/2
```

创建名为 tran1 的安全提议。

```
[RouterB] ipsec proposal tran1
```

报文封装形式采用隧道模式。

```
[RouterB-ipsec-proposal-tran1] encapsulation-mode tunnel
```

安全协议采用 ESP 协议。

```
[RouterB-ipsec-proposal-tran1] transform esp
```

选择算法。

```
[RouterB-ipsec-proposal-tran1] esp encryption-algorithm des
[RouterB-ipsec-proposal-tran1] esp authentication-algorithm sha1
[RouterB-ipsec-proposal-tran1] quit
```

配置 IKE 对等体。

```
[RouterB] ike peer peer
[RouterB-ike-peer-peer] pre-share-key abcde
[RouterB-ike-peer-peer] remote-address 2.2.2.1
[RouterB-ike-peer-peer] quit
```

创建一条安全策略，协商方式为 isakmp。

```
[RouterB] ipsec policy use1 10 isakmp
```

引用访问控制列表。

```
[RouterB-ipsec-policy-isakmp-use1-10] security acl 3101
```

引用安全提议。

```
[RouterB-ipsec-policy-isakmp-use1-10] proposal tran1
```

引用 IKE 对等体。

```
[RouterB-ipsec-policy-isakmp-use1-10] ike-peer peer
[RouterB-ipsec-policy-isakmp-use1-10] quit
```

配置串口的 IP 地址。

```
[RouterB] interface serial 2/2
[RouterB-Serial2/2] ip address 2.2.3.1 255.255.255.0
```

在串口上应用安全策略组。

```
[RouterB-Serial2/2] ipsec policy use1
```

以上配置完成后，Router A 和 Router B 之间如果有子网 10.1.1.0/24 与子网 10.1.2.0/24 之间的报文通过，将触发 IKE 进行协商建立 SA。IKE 协商成功并创建了 SA 后，子网 10.1.1.0/24 与子网 10.1.2.0/24 之间的数据流将被加密传输。

2 IKE 配置

2.1 IKE 简介

在实施 IPsec 的过程中，可以使用 IKE（Internet Key Exchange，因特网密钥交换）协议来建立 SA，该协议建立在由 ISAKMP（Internet Security Association and Key Management Protocol，互联网安全联盟和密钥管理协议）定义的框架上。IKE 为 IPsec 提供了自动协商交换密钥、建立 SA 的服务，能够简化 IPsec 的使用和管理，大大简化 IPsec 的配置和维护工作。

IKE 不是在网络上直接传输密钥，而是通过一系列数据的交换，最终计算出双方共享的密钥，并且即使第三者截获了双方用于计算密钥的所有交换数据，也不足以计算出真正的密钥。

2.1.1 IKE 的安全机制

IKE 具有一套自保护机制，可以在不安全的网络上安全地认证身份、分发密钥、建立 IPsec SA。

1. 数据认证

数据认证有如下两方面的概念：

- 身份认证：身份认证确认通信双方的身份。支持两种认证方法：预共享密钥（pre-shared-key）认证和基于 PKI 的数字签名（rsa-signature）认证。
- 身份保护：身份数据在密钥产生之后加密传送，实现了对身份数据的保护。

2. DH

DH（Diffie-Hellman，交换及密钥分发）算法是一种公共密钥算法。通信双方在不传输密钥的情况下通过交换一些数据，计算出共享的密钥。即使第三者（如黑客）截获了双方用于计算密钥的所有交换数据，由于其复杂度很高，不足以计算出真正的密钥。所以，DH 交换技术可以保证双方能够安全地获得公有信息。

3. PFS

PFS（Perfect Forward Secrecy，完善的前向安全性）特性是一种安全特性，指一个密钥被破解，并不影响其他密钥的安全性，因为这些密钥间没有派生关系。对于 IPsec，是通过在 IKE 阶段 2 协商中增加一次密钥交换来实现的。PFS 特性是由 DH 算法保障的。

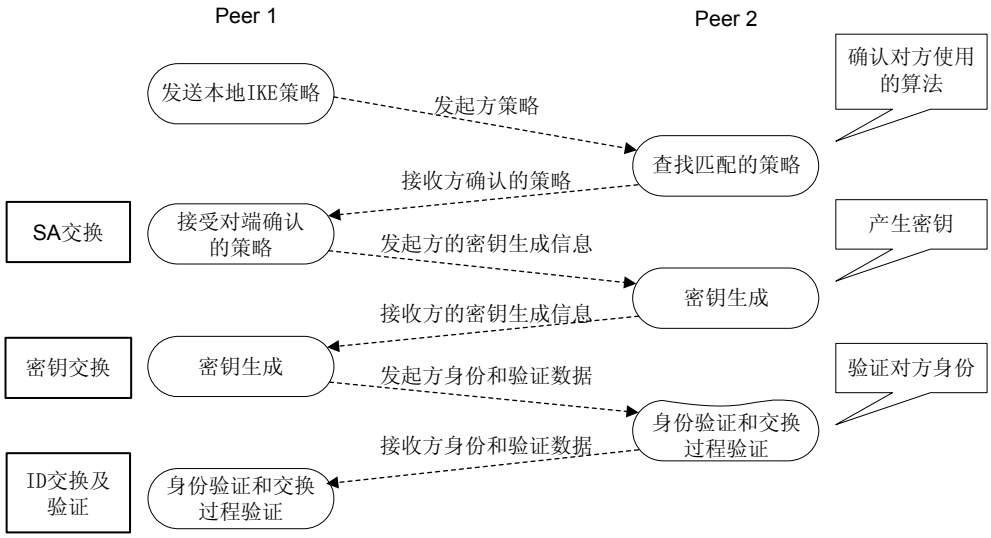
2.1.2 IKE 的交换过程

IKE 使用了两个阶段为 IPsec 进行密钥协商并建立 SA：

(1) 第一阶段，通信各方彼此间建立了一个已通过身份认证和安全保护的通道，即建立一个 ISAKMP SA。第一阶段有主模式（Main Mode）和野蛮模式（Aggressive Mode）两种 IKE 交换方法。

(2) 第二阶段，用在第一阶段建立的安全隧道为 IPsec 协商安全服务，即为 IPsec 协商具体的 SA，建立用于最终的 IP 数据安全传输的 IPsec SA。

图2-1 主模式交换过程



如 图 2-1所示，第一阶段主模式的IKE协商过程中包含三对消息：

- 第一对叫 SA 交换，是协商确认有关安全策略的过程；
- 第二对消息叫密钥交换，交换 Diffie-Hellman 公共值和辅助数据（如：随机数），密钥材料在这个阶段产生；
- 最后一对消息是 ID 信息和认证数据交换，进行身份认证和对整个第一阶段交换内容的认证。

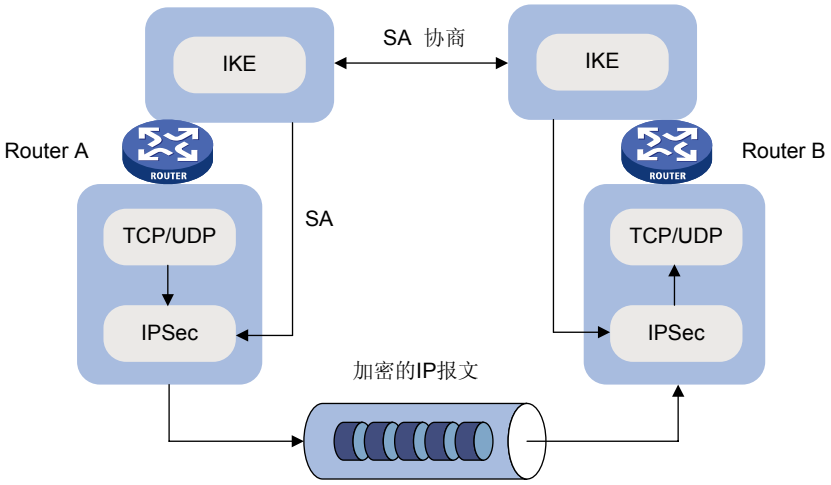
野蛮模式交换与主模式交换的主要差别在于，野蛮模式不提供身份保护，只交换 3 条消息。在对身份保护要求不高的场合，使用交换报文较少的野蛮模式可以提高协商的速度；在对身份保护要求较高的场合，则应该使用主模式。

2.1.3 IKE 在 IPsec 中的作用

- 因为有了 IKE，IPsec 很多参数（如：密钥）都可以自动建立，降低了手工配置的复杂度。
- IKE 协议中的 DH 交换过程，每次的计算和产生的结果都是不相关的。每次 SA 的建立都运行 DH 交换过程，保证了每个 SA 所使用的密钥互不相关。
- IPsec 使用 AH 或 ESP 报文头中的序列号实现防重放。此序列号是一个 32 比特的值，此数溢出后，为实现防重放，SA 需要重新建立，这个过程需要 IKE 协议的配合。
- 对安全通信的各方身份的认证和管理，将影响到 IPsec 的部署。IPsec 的大规模使用，必须有 CA（Certificate Authority，认证中心）或其他集中管理身份数据的机构的参与。
- IKE 提供端与端之间动态认证。

2.1.4 IPsec 与 IKE 的关系

图2-2 IPsec 与 IKE 的关系图



从 图 2-2中我们可以看出IKE和IPsec的关系：

- IKE 是 UDP 之上的一个应用层协议，是 IPsec 的信令协议；
- IKE 为 IPsec 协商建立 SA，并把建立的参数及生成的密钥交给 IPsec；
- IPsec 使用 IKE 建立的 SA 对 IP 报文加密或认证处理。

2.1.5 协议规范

与 IKE 相关的协议规范有：

- RFC2408：Internet Security Association and Key Management Protocol (ISAKMP)
- RFC2409：The Internet Key Exchange (IKE)
- RFC2412：The OAKLEY Key Determination Protocol

2.2 IKE 配置任务简介

进行 IKE 配置之前，用户需要确定以下几个因素，以便配置过程的顺利进行。

- 确定 IKE 交换过程中算法的强度，即确定安全保护的强度（包括身份认证方法、加密算法、认证算法、DH 组）：不同的算法的强度不同，算法强度越高，受保护数据越难被破解，但消耗的计算资源越多。一般来说，密钥越长的算法强度越高。
- 确定通信双方预先约定的预共享密钥或所属的 PKI 域。关于 PKI 的配置，请参见“安全分册”中的“PKI 配置”。

表2-1 IKE 配置任务简介

配置任务	说明	详细配置
配置本端安全网关的名字	可选	2.3
配置 IKE 安全提议	可选	2.4
配置 IKE 对等体	必选	2.5
配置 Keepalive 定时器	可选	2.6
配置 NAT Keepalive 定时器	可选	2.7

配置任务	说明	详细配置
配置对等体存活检测	可选	2.8
配置取消对 next payload 域的检查	可选	2.9

2.3 配置本端安全网关的名字

当 IKE 协商的发起端使用安全网关名字进行协商时（即配置了 **id-type name**），本端需要配置命令 **ike local-name**。

表2-2 配置本端安全网关名字

操作	命令	说明
进入系统视图	**system-view**	-
配置本端安全网关的名字	**ike local-name** *name*	可选 缺省情况下，使用设备名作为本端网关的名字

2.4 配置 IKE 安全提议

IKE 安全提议定义了一套属性数据来描述 IKE 协商怎样进行安全通信。用户可以创建多条不同优先级的 IKE 提议，优先级由 IKE 提议的序号表示，数值越小，优先级越高。在进行 IKE 协商的时候，会从序号最小的 IKE 提议进行匹配，如果匹配则直接使用，否则继续查找。协商双方必须至少有一条匹配的 IKE 提议才能协商成功。匹配的原则是：协商双方具有相同的加密算法、认证方法、认证算法和 DH 组标识。存活时间取两端的最小值。匹配的 IKE 提议将被用来建立安全隧道。

缺省情况下，系统提供一条缺省的 IKE 提议。此缺省的 IKE 提议具有最低的优先级，具有缺省的加密算法、认证方法、认证算法、DH 组标识和 SA 生存周期。

表2-3 配置 IKE 安全提议

操作	命令	说明
进入系统视图	**system-view**	-
创建 IKE 提议，并进入 IKE 提议视图	**ike proposal** *proposal-number*	必选
指定一个供 IKE 提议使用的加密算法	**encryption-algorithm** { **3des-cbc** | **aes-cbc** [*key-length*] | **des-cbc** }	可选 缺省情况下，IKE 提议使用 CBC 模式的 56-bit DES 加密算法
指定一个供 IKE 提议使用的认证方法	**authentication-method** { **pre-share** | **rsa-signature** }	可选 缺省情况下，IKE 提议使用预共享密钥的认证方法
指定一个供 IKE 提议使用的认证算法	**authentication-algorithm** { **md5** | **sha** }	可选 缺省情况下，IKE 提议使用 SHA1 认证算法
配置 IKE 阶段 1 密钥协商时所使用的 DH 密钥交换参数	**dh** { **group1** | **group2** | **group5** | **group14** }	可选 缺省情况下，IKE 阶段 1 密钥协商时所使用的 DH 密钥交换参数为 **group1**，即 768-bit 的 Diffie-Hellman 组

操作	命令	说明
指定一个 IKE 提议的 ISAKMP SA 存活时间	**sa duration** *seconds*	可选 缺省情况下，IKE 提议的 ISAKMP SA 存活时间为 86400 秒

说明

如果存活时间超时，ISAKMP SA 将自动更新。因为 IKE 协商需要进行 DH 计算，在低端设备上需要经过较长的时间，为使 ISAKMP SA 的更新不影响安全通信，建议设置存活时间大于 10 分钟。

2.5 配置 IKE 对等体

表2-4 配置 IKE 对等体

操作		命令	说明
进入系统视图		**system-view**	-
创建一个 IKE 对等体，并进入 IKE-Peer 视图		**ike peer** *peer-name*	必选
配置 IKE 第一阶段的协商模式		**exchange-mode** { **aggressive** | **main** }	可选 缺省情况下，IKE 阶段的协商模式使用主模式
配置采用预共享密钥认证时，所使用的预共享密钥		**pre-shared-key** [**cipher** | **simple**] *key*	二者必选其一 根据 IKE 提议使用的认证方法选择其中一个配置
配置采用数字签名认证时，证书所属的 PKI 域		**certificate domain** *domain-name*	
选择 IKE 第一阶段的协商过程中使用 ID 的类型		**id-type** { **ip** | **name** }	可选 缺省情况下，使用 IP 地址作为 IKE 协商过程中使用的 ID
配置本端及对端安全网关的名字	配置本端安全网关的名字	请参见 2.3 配置本端安全网关的名字	可选 当 IKE 协商的发起端使用网关名字进行协商时（即配置了 **id-type name**），发起端会发送自己名字给对端来标识自己的身份，而对端使用 **remote-name** *name* 来认证发起端，故此时 *name* 应与发起端网关上使用 **ike local-name** 命令所配的名字保持一致
	配置对端安全网关的名字	**remote-name** *name*	
配置本端及对端安全网关的 IP 地址	配置本端安全网关的 IP 地址	**local-address** *ip-address*	可选 当 IKE 协商的发起端使用地址进行协商时（即配置了 **id-type ip**），发起端会发送自己的 IP 地址给对端来标识自己的身份，而对端使用 **remote-address** *ip-address* 来认证发起端，故此时 *ip-address* 应与发起端上 **local-address** 命令所配的 IP 地址保持一致 缺省情况下，IKE 协商时的本端网关 IP 地址使用应用安全策略的接口的主 IP 地址
	配置对端安全网关的 IP 地址	**remote-address** *low-ip-address* [*high-ip-address*]	

操作		命令	说明
配置 IKE/IPsec 的 NAT 穿越功能		**nat traversal**	可选 在 IPsec/IKE 组建的 VPN 隧道中，若存在 NAT 安全网关设备，则必须配置 IPsec/IKE 的 NAT 穿越功能 缺省情况下，没有配置 NAT 穿越功能
配置本端及对端安全网关的子网类型	配置本端安全网关子网类型	**local { multi-subnet | single-subnet }**	可选 缺省情况下，为单子网类型。这两条命令仅在与 NETSCREEN 的设备互通时使用
	配置对端安全网关子网类型	**peer { multi-subnet | single-subnet }**	
为 IKE 对等体应用一个 DPD		**dpd** *dpd-name*	可选 缺省情况下，IKE 对等体没有应用 DPD 关于DPD的配置请参见“2.8 配置对等体存活检测”

说明

- 修改 IPsec IKE 的对等体配置之后，要执行命令 **reset ipsec sa**、**reset ike sa** 来清除原有的 IPsec SA 与 IKE SA，否则重新协商 SA 会失败。
- 当安全隧道一端的 IP 地址为动态获取时，必须将 IKE 的协商模式配置为野蛮模式(**aggressive**)。
- 在预共享密钥认证的主模式下，只能使用 IP 地址类型的身份进行 IKE 协商，建立 SA。在 IKE 野蛮模式下，不但可以使用 IP 地址类型的身份进行协商，也可以使用名字类型的身份进行 IKE 协商，建立 SA。
- IKE 对等体中配置的 IKE 协商模式表示本端作为发起方时所使用的协商模式，响应方将自动适配发起方的协商模式。
- 一般情况下本端的 **local-address** 不需要配置，只有当用户需要指定特殊的本端网关地址时（如指定 loopback 接口地址）才需要配置这条命令。而发起方的对端网关名字或对端网关 IP 地址需要配置，它们用于发起方在协商过程中寻找对端。
- 为了节省 IP 地址空间，ISP 经常会在公网中加入 NAT 网关，以便于将私有 IP 地址分配给用户，此时可能会导致 IPsec/IKE 隧道的两端一端为公网地址，另一端为私网地址，所以必须在私网侧配置 NAT 穿越，保证隧道能够正常协商建立。

2.6 配置 Keepalive 定时器

IKE 通过 Keepalive 报文维护 ISAKMP SA 的链路状态。一般在对端配置了等待 Keepalive 报文的超时时间后，必须在本端配置此 Keepalive 报文发送时间间隔。当对端在配置的超时时间内未收到此 Keepalive 报文时，如果该 ISAKMP SA 带有 TIMEOUT 标记，则删除该 ISAKMP SA 以及由其协商的 IPsec SA；否则，将其标记为 TIMEOUT。

表2-5 配置 Keepalive 定时器

操作	命令	说明
进入系统视图	**system-view**	-
配置 ISAKMP SA 向对端发送 Keepalive 报文的时间间隔	**ike sa keepalive-timer interval** *seconds*	必选 缺省情况下，ISAKMP SA 不向对端发送 Keepalive 报文

操作	命令	说明
配置 ISAKMP SA 等待对端发送 Keepalive 报文的超时时间	**ike sa keepalive-timer timeout** *seconds*	必选 缺省情况下，ISAKMP SA 不向对端发送 Keepalive 报文



说明

本端配置的 Keepalive 报文的等待超时时间要大于对端发送的时间间隔。由于网络中一般不会出现超过连续三次的报文丢失，所以，本端的超时时间可以配置为对端配置的 Keepalive 报文发送时间间隔的三倍。

2.7 配置 NAT Keepalive 定时器

在 IPsec/IKE 组建的 VPN 隧道中，若存在 NAT 安全网关设备，需配置 NAT 穿越功能来实现 NAT 穿越，但由于在 NAT 网关上的 NAT 映射会话有一定存活时间，因此一旦安全隧道建立后如果长时间没有报文穿越，NAT 会话表项会被删除，这样将导致在 NAT 网关外侧的隧道无法继续传输数据。为防止 NAT 表项老化，NAT 网关内网侧的 ISAKMP SA 会以一定的时间间隔向对端发送 NAT Keepalive 报文，以维持 NAT 会话的存活。

表2-6 配置 NAT Keepalive 定时器

操作	命令	说明
进入系统视图	**system-view**	-
配置 ISAKMP SA 向对端发送 NAT Keepalive 报文的时间间隔	**ike sa nat-keepalive-timer interval** *seconds*	必选 缺省情况下，ISAKMP SA 向对端发送 NAT Keepalive 报文的时间间隔为 20 秒

2.8 配置对等体存活检测

DPD（Dead Peer Detection，对等体存活检测）用于 IPsec 邻居状态的检测。启动 DPD 功能后，当接收端在触发 DPD 的时间间隔内收不到对端的 IPsec 加密报文时，能够触发 DPD 查询，主动向对端发送请求报文，对 IKE 对等体是否存在进行检测。

DPD 和 Keepalive 的区别：

- Keepalive 定期发送查询；
- DPD 只在要发送加密报文前并且定时器到期时发送查询。

表2-7 配置对等体存活检测

操作	命令	说明
进入系统视图	**system-view**	-
创建一个 DPD，并进入 DPD 视图	**ike dpd** *dpd-name*	必选
配置触发 DPD 的时间间隔	**interval-time** *interval-time*	可选 缺省情况下，触发 DPD 的时间间隔为 10 秒

操作	命令	说明
配置 DPD 报文的重传时间间隔	**time-out** *time-out*	可选 缺省情况下，DPD 报文的重传时间间隔为 5 秒

2.9 配置取消对 next payload 域的检查

next payload 域是在 IKE 协商报文（由几个 payload 组装而成）的最后一个 payload 的通用头中的一个域。按协议规定如果当前载荷处于消息的最后，该域必须为 0，但某些公司的设备会将该域赋其它值，为增强设备的互通性，可以通过下面的配置取消 IPsec 协商过程对该域的检查。

表2-8 配置取消对 next payload 域的检查

操作	命令	说明
进入系统视图	**system-view**	-
配置在 IPsec 协商过程中取消对 payload 域的检查	**ike next-payload check disabled**	必选 缺省情况下，在 IPsec 协商过程中对 next payload 域进行检查

2.10 IKE 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IKE 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以删除 IKE 建立的安全隧道。

表2-9 IKE 显示和维护

操作	命令
显示 DPD 配置的参数	**display ike dpd** [*dpd-name*]
显示 IKE 对等体配置的参数	**display ike peer** [*peer-name*]
显示当前 IKE SA 的信息	**display ike sa** [**verbose** [**connection-id** *connection-id* | **remote-address** *remote-address*]]
显示每个 IKE 提议配置的参数	**display ike proposal**
清除 IKE 建立的安全隧道	**reset ike sa** [*connection-id*]

2.11 IKE 典型配置举例

2.11.1 IKE 典型配置组网应用

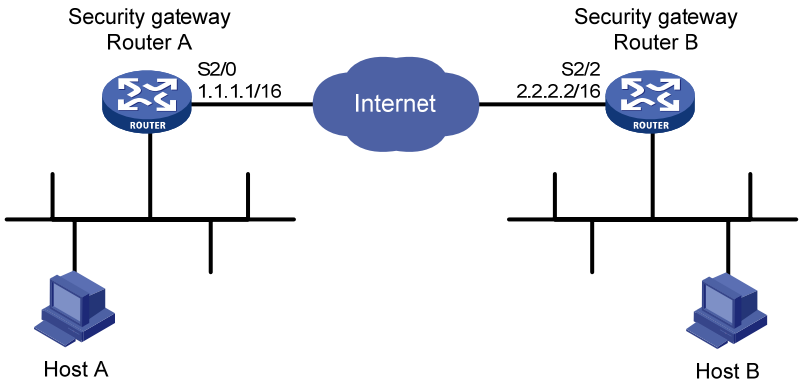
1. 组网需求

- Host A 与 Host B 之间进行安全通信，在安全网关 Router A 与安全网关 Router B 之间使用 IKE 自动协商建立安全隧道。
- 在 Router A 上配置一条 IKE 提议，其提议号为 10，使用的认证算法为 MD5。Router B 使用缺省的 IKE 提议。

- 为使用预共享密钥认证方法的提议配置预共享密钥。

2. 组网图

图2-3 IKE 配置举例组网图



3. 配置步骤

(1) 配置安全网关 Router A

配置 IKE 对等体。

```
<RouterA> system-view
[RouterA] ike peer peer
[RouterA-ike-peer-peer] pre-shared-key abcde
[RouterA-ike-peer-peer] remote-address 2.2.2.2
[RouterA-ike-peer-peer] quit
```

创建一条 IKE 提议 10。

```
[RouterA] ike proposal 10
```

指定 IKE 提议使用的认证算法为 MD5。

```
[RouterA-ike-proposal-10] authentication-algorithm md5
```

使用预共享密钥认证方法。

```
[RouterA-ike-proposal-10] authentication-method pre-share
```

配置 ISAKMP SA 的存活时间为 5000 秒。

```
[RouterA-ike-proposal-10] sa duration 5000
```

(2) 配置安全网关 Router B

配置 IKE 对等体。

```
<RouterB> system-view
[RouterB] ike peer peer
[RouterB-ike-peer-peer] pre-shared-key abcde
[RouterB-ike-peer-peer] remote-address 1.1.1.1
```

以上配置可以保证网关 Router A 和保证网关 Router B 之间进行 IKE 协商。由于 Router A 配置的提议是 proposal 10，使用 authentication-algorithm md5，但 Router B 上只有一条缺省的 IKE 提议，默认是 authentication-algorithm sha。在进行提议匹配的时候，从优先级最高的提议开始，因为 Router B 上没有和 Router A 上提议 10 相匹配的提议，所以双方能够匹配的只有缺省的 IKE 提议。另外，在进行提议匹配的时候，存活时间是不用进行匹配的，它由 IKE 协商双方决定。

2.11.2 IKE 野蛮模式及 NAT 穿越的组网应用

1. 组网需求

- 分公司 LAN 通过专线接入总公司内部网，Router A 的 Serial2/0 接口为固定 IP 地址，Router B 动态获取 IP 地址。
- 分公司自动获得的 IP 地址为私有 IP 地址，Router A 的 Serial2/0 接口的 IP 地址为公网地址，故 Router B 上需要配置 NAT 穿越功能。
- 为了保证信息安全采用 IPsec/IKE 方式创建安全隧道。

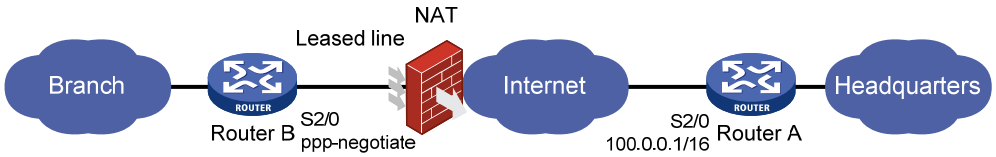
说明

为突出 IKE 野蛮模式及 NAT 穿越的配置，本例中的路由器采用串口通过 Internet 互联，并将一端配置为自动获取 IP 地址。实际应用中的拨号方式及宽带接入方式，均可以参考此例。

---

2. 组网图

图2-4 IKE 野蛮模式及 NAT 穿越的典型配置案例



3. 配置步骤

(1) 配置 Router A

配置本端安全网关设备的名字。

```
<RouterA> system-view
[RouterA] ike local-name routera
```

配置 acl。

```
[RouterA] acl number 3101 match-order auto
[RouterA-acl-adv-3101] rule permit ip source any destination any
[RouterA-acl-adv-3101] quit
```

配置地址池。

```
[RouterA] ip pool 1 10.0.0.2 10.0.0.10
```

配置 IKE 对等体 peer。

```
[RouterA] ike peer peer
[RouterA-ike-peer-peer] exchange-mode aggressive
[RouterA-ike-peer-peer] pre-shared-key abc
[RouterA-ike-peer-peer] id-type name
[RouterA-ike-peer-peer] remote-name routerb
[RouterA-ike-peer-peer] quit
```

创建 IPsec 安全提议 prop。

```
[RouterA] ipsec proposal prop
```

```
[RouterA-ipsec-proposal-prop] encapsulation-mode tunnel
[RouterA-ipsec-proposal-prop] transform esp
[RouterA-ipsec-proposal-prop] esp encryption-algorithm des
[RouterA-ipsec-proposal-prop] esp authentication-algorithm sha1
[RouterA-ipsec-proposal-prop] quit
```

创建安全策略 policy 并指定通过 IKE 协商建立 SA。

```
[RouterA] ipsec policy policy 10 isakmp
```

配置安全策略 policy 引用 IKE 对等体 peer。

```
[RouterA-ipsec-policy-isakmp-policy-10] ike-peer peer
```

配置安全策略 policy 引用访问控制列表 3101。

```
[RouterA-ipsec-policy-isakmp-policy-10] security acl 3101
```

配置安全策略 policy 引用 IPsec 安全提议 prop。

```
[RouterA-ipsec-policy-isakmp-policy-10] proposal prop
[RouterA-ipsec-policy-isakmp-policy-10] quit
```

进入串口 Serial2/0 并配置 IP 地址。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 100.0.0.1 255.255.0.0
```

配置串口 Serial2/0 引用安全策略组 policy。

```
[RouterA-Serial2/0] ipsec policy policy
[RouterA-Serial2/0] remote address pool 1
```

(2) 配置 Router B

配置本端安全网关设备的名字。

```
<RouterB> system-view
[RouterB] ike local-name routerb
```

配置 acl。

```
[RouterB] acl number 3101 match-order auto
[RouterB-acl-adv-3101] rule permit ip source any destination any
[RouterB-acl-adv-3101] quit
```

配置 IKE 对等体 peer。

```
[RouterB] ike peer peer
[RouterB-ike-peer-peer] exchange-mode aggressive
[RouterB-ike-peer-peer] pre-shared-key abc
[RouterB-ike-peer-peer] id-type name
[RouterB-ike-peer-peer] remote-name routera
[RouterB-ike-peer-peer] remote-ip 100.0.0.1
[RouterB-ike-peer-peer] nat traversal
[RouterB-ike-peer-peer] quit
```

创建 IPsec 安全提议 prop。

```
[RouterB] ipsec proposal prop
[RouterB-ipsec-proposal-prop] encapsulation-mode tunnel
[RouterB-ipsec-proposal-prop] transform esp
[RouterB-ipsec-proposal-prop] esp encryption-algorithm des
```

```
[RouterB-ipsec-proposal-prop] esp authentication-algorithm sha1
[RouterB-ipsec-proposal-prop] quit
```

创建安全策略 policy 并指定通过 IKE 协商建立 SA。

```
[RouterB] ipsec policy policy 10 isakmp
```

配置安全策略 policy 引用 IKE 对等体 peer。

```
[RouterB-ipsec-policy-isakmp-policy-10] ike-peer peer
```

配置安全策略 policy 引用访问控制列表 3101。

```
[RouterB-ipsec-policy-isakmp-policy-10] security acl 3101
```

配置安全策略 policy 引用 IPsec 安全提议 prop。

```
[RouterB-ipsec-policy-isakmp-policy-10] proposal prop
[RouterB-ipsec-policy-isakmp-policy-10] quit
```

进入串口 Serial2/0 并配置接口动态协商 IP 地址。

```
[RouterB] interface serial 2/0
[RouterB-Serial2/0] ip address ppp-negotiate
```

配置串口 Serial2/0 引用安全策略组 policy。

```
[RouterB-Serial2/0] ipsec policy policy
```

说明

有关配置 PPP 协商 IP 地址的详细介绍请参考“接入分册”中的“PPP 配置”。

2.11.3 ADSL 与 IPsec/IKE 相结合的组网应用

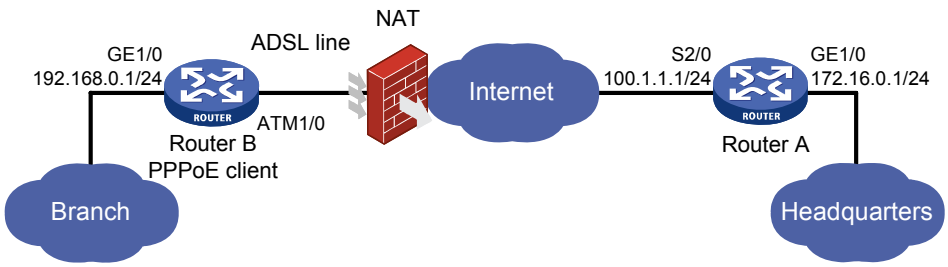
1. 组网需求

本例将 IPsec 和 ADSL 相结合，是目前实际中广泛应用的典型案例。

- Router B 通过 ADSL 卡直接连接公网的 DSLAM 接入端，作为 PPPoE 的 client 端。RouterB 从 ISP 动态获得的 IP 地址为私网地址，故 Router A、Router B 都需要配置 NAT 穿越。
- 总公司局域网通过 Router A 接入到 ATM 网络。
- 为了保证信息安全采用 IPsec/IKE 方式创建安全隧道。

2. 组网图

图2-5 ADSL 与 IPsec/IKE 相结合的组网应用



3. 配置步骤

(1) 配置 Router A

配置本端安全网关设备名称。

```
<RouterA> system-view
[RouterA] ike local-name routera
```

配置 ACL。

```
[RouterA] acl number 3101
[RouterA-acl-adv-3101] rule 0 permit ip source 172.16.0.0 0.0.0.255 destination 192.168.0.0
0.0.0.255
[RouterA-acl-adv-3101] quit
```

配置 IKE 安全提议。

```
[RouterA] ike proposal 1
[RouterA-ike-proposal-1] authentication-algorithm sha
[RouterA-ike-proposal-1] authentication-method pre-share
[RouterA-ike-proposal-1] encryption-algorithm 3des-cbc
[RouterA-ike-proposal-1] dh group2
```

配置 IKE 对等体 peer。

```
[RouterA] ike peer peer
[RouterA-ike-peer-peer] exchange-mode aggressive
[RouterA-ike-peer-peer] pre-shared-key abc
[RouterA-ike-peer-peer] id-type name
[RouterA-ike-peer-peer] remote-name routerb
[RouterA-ike-peer-peer] nat traversal
[RouterA-ike-peer-peer] quit
```

创建 IPsec 安全提议 prop。

```
[RouterA] ipsec proposal prop
[RouterA-ipsec-proposal-prop] encapsulation-mode tunnel
[RouterA-ipsec-proposal-prop] transform esp
[RouterA-ipsec-proposal-prop] esp encryption-algorithm 3des
[RouterA-ipsec-proposal-prop] esp authentication-algorithm sha1
[RouterA-ipsec-proposal-prop] quit
```

创建安全策略 policy 并指定通过 IKE 协商建立 SA。

```
[RouterA] ipsec policy policy 10 isakmp
```

配置安全策略 policy 引用 IKE 对等体 peer。

```
[RouterA-ipsec-policy-isakmp-policy-10] ike-peer peer
```

配置安全策略 policy 引用访问控制列表 3101。

```
[RouterA-ipsec-policy-isakmp-policy-10] security acl 3101
```

配置安全策略 policy 引用 IPsec 安全提议 prop。

```
[RouterA-ipsec-policy-isakmp-policy-10] proposal prop
[RouterA-ipsec-policy-isakmp-policy-10] quit
```

配置 IP 地址。

```
[RouterA] interface serial 2/0
[RouterA-Serial2/0] ip address 100.1.1.1 255.255.255.0
[RouterA-Serial2/0] ipsec policy policy
[RouterA-Serial2/0] quit
```

配置以太网口。

```
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ip address 172.16.0.1 255.255.255.0
[RouterA-GigabitEthernet1/0] quit
```

配置到分公司局域网的静态路由。

```
[RouterA] ip route-static 192.168.0.0 255.255.255.0 serial 2/0
```

(2) 配置 Router B

配置本端安全网关的名称。

```
<RouterB> system-view
[RouterB] ike local-name routerb
```

配置 ACL。

```
[RouterB] acl number 3101
[RouterB-acl-adv-3101] rule 0 permit ip source 192.168.0.0 0.0.0.255 destination 172.16.0.0
0.0.0.255
[RouterB-acl-adv-3101] quit
```

配置 IKE 安全提议。

```
[RouterB] ike proposal 1
[RouterB-ike-proposal-1] authentication-algorithm sha
[RouterB-ike-proposal-1] authentication-method pre-share
[RouterB-ike-proposal-1] encryption-algorithm 3des-cbc
[RouterB-ike-proposal-1] dh group2
```

配置 IKE 对等体 peer。

```
[RouterB] ike peer peer
[RouterB-ike-peer-peer] exchange-mode aggressive
[RouterB-ike-peer-peer] pre-shared-key abc
[RouterB-ike-peer-peer] id-type name
[RouterB-ike-peer-peer] remote-name routera
[RouterB-ike-peer-peer] remote-address 100.1.1.1
[RouterB-ike-peer-peer] nat traversal
[RouterB-ike-peer-peer] quit
```

创建 IPsec 安全提议 prop。

```
[RouterB] ipsec proposal prop
[RouterB-ipsec-proposal-prop] encapsulation-mode tunnel
[RouterB-ipsec-proposal-prop] transform esp
[RouterB-ipsec-proposal-prop] esp encryption-algorithm 3des
[RouterB-ipsec-proposal-prop] esp authentication-algorithm sha1
[RouterB-ipsec-proposal-prop] quit
```

创建安全策略 policy 并指定通过 IKE 协商建立 SA。

```
[RouterB] ipsec policy policy 10 isakmp
```

配置安全策略 policy 引用 IKE 对等体 peer。

```
[RouterB-ipsec-policy-isakmp-policy-10] ike-peer peer
```

配置安全策略 policy 引用访问控制列表 3101。

```
[RouterB-ipsec-policy-isakmp-policy-10] security acl 3101
```

配置安全策略 policy 引用 IPsec 安全提议 prop。

```
[RouterB-ipsec-policy-isakmp-policy-10] proposal prop
[RouterB-ipsec-policy-isakmp-policy-10] quit
```

配置拨号访问控制列表。

```
[RouterB] dialer-rule 1 ip permit
```

创建 Dialer0，使用由 ISP 分配的用户名和密码进行拨号和 PPP 认证的相关配置，并配置 MTU。

```
[RouterB] interface dialer 0
[RouterB-Dialer0] link-protocol ppp
[RouterB-Dialer0] ppp pap local-user test password simple 123456
[RouterB-Dialer0] ip address ppp-negotiate
[RouterB-Dialer0] dialer user 1
[RouterB-Dialer0] dialer-group 1
[RouterB-Dialer0] dialer bundle 1
[RouterB-Dialer0] ipsec policy policy
[RouterB-Dialer0] mtu 1492
[RouterB-Dialer0] quit
```

配置到总公司局域网的静态路由。

```
[RouterB] ip route-static 172.16.0.0 255.255.255.0 dialer 0
```

配置以太网口。

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] tcp mss 1450
[RouterB-GigabitEthernet1/0] ip address 192.168.0.1 255.255.255.0
[RouterB-GigabitEthernet1/0] quit
```

对 ADSL 卡的 ATM 口进行配置。

```
[RouterB] interface atm 1/0
[RouterB-Atm1/0] pvc 0/100
[RouterB-atm-pvc-Atm1/0-0/100] map bridge virtual-ethernet 0
[RouterB-atm-pvc-Atm1/0-0/100] quit
```

配置 VE 口。

```
[RouterB] interface virtual-ethernet 0
[RouterB-Virtual-Ethernet0] pppoe-client dial-bundle-number 1
[RouterB-Virtual-Ethernet0] mac-address 0011-0022-0012
```

2.12 常见错误配置举例

配置参数建立 IPsec 安全隧道时，可以打开 IKE 的 Error 调试开关，帮助我们查找配置问题。其命令是：

```
<Router> debugging ike error
```

2.12.1 非法用户身份信息

1. 故障现象

非法用户身份信息

2. 故障分析

用户身份信息是发起 IPsec 通信的用户用来标识自己的数据。在实际应用中我们可以通过用户身份标识实现对不同的数据流建立不同的安全隧道进行保护。目前我们是通过用户的 IP 地址和名字来标识用户。

可以看到调试信息：

```
got NOTIFY of type INVALID_ID_INFORMATION
```

或者

```
drop message from A.B.C.D due to notification type INVALID_ID_INFORMATION
```

3. 处理过程

检查协商两端接口上配置的安全策略中的 ACL 内容是否相容。建议用户将两端的 ACL 配置成互为镜像的。ACL 镜像的含义请参考 IPsec 配置中“配置访问控制列表”内容。

2.12.2 提议不匹配

1. 故障现象

提议不匹配

2. 故障分析

可以看到调试信息：

```
got NOTIFY of type NO_PROPOSAL_CHOSEN
```

或者：

```
drop message from A.B.C.D due to notification type NO_PROPOSAL_CHOSEN
```

协商双方没有可以匹配的提议。

3. 处理过程

对于阶段 1，检查 IKE proposal 是否有与对方匹配的。对于阶段 2 协商，检查双方接口上应用的 IPsec 安全策略的参数是否匹配，引用的 IPsec 安全提议的协议、加密算法和认证算法是否有匹配的。

2.12.3 无法建立安全隧道

1. 故障现象

无法建立安全隧道

2. 故障分析

实际应用中有时会发现在不稳定的网络状态下，安全隧道无法建立或者存在安全隧道却无法通信，而且检查双方的 ACL 的配置正确，也有匹配的提议。

这种情况一般是安全隧道建立好以后，有一方的设备重启造成的。

3. 处理过程

- 使用 **display ike sa** 命令检查双方是否都已建立阶段 1 的 SA。
- 使用 **display ipsec sa policy** 命令查看接口上的安全策略是否已建立了 IPsec SA。
- 根据以上两步的结果查看，如果有一方存在的 SA 在另一方不存在的情况，使用 **reset ike sa** 命令清除错误存在的 SA，重新发起协商。

2.12.4 ACL 配置错误

1. 故障现象

ACL 配置错误，导致协商成功之后数据流不通

2. 故障分析

多台设备之间先后建立不同的安全隧道，出现同一设备有不同对端的情况。若此设备不配置 ACL 规则，则分别由对端发起报文来与之建立保护粒度不同的安全隧道。由于安全隧道的优先级由它们创建的顺序决定，当这一设备的出方向报文首先匹配到较粗粒度的安全隧道时，将导致此设备无法与其它较细粒度对端互通。

3. 处理过程

为避免这种情况发生，当同一设备有不同对端时，建议用户在此设备上配置 ACL 来区别数据流，且与不同对端尽量避免配置有重复范围的 ACL 子规则。若需要有重复范围的子规则，应该将细粒度的子规则配置为较高的优先级。

目 录

1 SSH2.0 配置 1-1
1.1 SSH2.0 简介 1-1
1.1.1 概述 1-1
1.1.2 SSH工作过程 1-1
1.2 配置设备作为SSH服务器 1-4
1.2.1 SSH服务器配置任务简介 1-4
1.2.2 生成RSA或DSA密钥 1-4
1.2.3 使能SSH服务器功能 1-5
1.2.4 配置SSH客户端登录时的用户界面 1-5
1.2.5 配置客户端的公钥 1-5
1.2.6 配置SSH用户并指定服务类型和认证方式 1-6
1.2.7 配置服务器上的SSH管理功能 1-7
1.3 配置设备作为SSH客户端 1-8
1.3.1 SSH客户端配置任务简介 1-8
1.3.2 为SSH客户端指定源IP地址或源接口 1-8
1.3.3 配置SSH客户端是否支持首次认证 1-9
1.3.4 建立SSH客户端和服务器端的连接 1-10
1.4 SSH协议显示和维护 1-10
1.5 设备作为SSH服务器配置举例 1-11
1.5.1 password认证配置举例 1-11
1.5.2 publickey认证配置举例 1-12
1.6 设备作为SSH客户端配置举例 1-17
1.6.1 password认证配置举例 1-17
1.6.2 publickey认证配置举例 1-19
2 SFTP服务 2-1
2.1 SFTP简介 2-1
2.2 配置SFTP服务器 2-1
2.2.1 配置准备 2-1
2.2.2 启动SFTP服务器 2-1
2.2.3 配置SFTP用户连接的空闲超时时间 2-1
2.3 配置SFTP客户端 2-2
2.3.1 为SFTP客户端指定源IP地址或源接口 2-2
2.3.2 建立与SFTP服务器的连接 2-2
2.3.3 SFTP目录操作 2-2
2.3.4 SFTP文件操作 2-3
2.3.5 显示帮助信息 2-4
2.3.6 终止与远程SFTP服务器的连接 2-4
2.4 设备作为SFTP客户端配置举例 2-4
2.5 设备作为SFTP服务器配置举例 2-8

1 SSH2.0 配置

1.1 SSH2.0 简介

1.1.1 概述

SSH 是 Secure Shell（安全外壳）的简称。用户通过一个不能保证安全的网络环境远程登录到设备时，SSH 可以利用加密和强大的认证功能提供安全保障，保护设备不受诸如 IP 地址欺诈、明文密码截取等攻击。

设备支持 SSH 服务器功能，可以接受多个 SSH 客户端的连接。同时，设备还支持 SSH 客户端功能，允许用户与支持 SSH 服务器功能的设备建立 SSH 连接，从而实现从本地设备通过 SSH 登录到远程设备上。

注意

- 目前，设备作为 SSH 服务器时，支持 SSH2 和 SSH1 两个版本；设备作为 SSH 客户端时，只支持 SSH2 版本。
- 如无特殊说明，文中的 SSH 均指 SSH2。

1.1.2 SSH 工作过程

在整个通讯过程中，为实现 SSH 的安全连接，服务器端与客户端要经历如下五个阶段：

表1-1 SSH 服务器端与客户端建立连接的五个阶段

过程	说明	详细内容
版本号协商阶段	SSH 目前包括 SSH1 和 SSH2 两个版本，双方通过版本协商确定使用的版本	1.1.2 1.
密钥和算法协商阶段	SSH 支持多种加密算法，双方根据本端和对端支持的算法，协商出最终使用的算法	1.1.2 2.
认证阶段	SSH 客户端向服务器端发起认证请求，服务器端对客户端进行认证	1.1.2 3.
会话请求阶段	认证通过后，SSH 客户端向服务器端发送会话请求	1.1.2 4.
交互会话阶段	会话请求通过后，SSH 服务器端和客户端进行信息的交互	1.1.2 5.

1. 版本号协商阶段

具体步骤如下：

(1) 服务器打开端口 22，等待客户端连接。

(2) 客户端向服务器端发起 TCP 初始连接请求，TCP 连接建立后，服务器向客户端发送第一个报文，包括版本标志字符串，格式为“SSH－<主协议版本号>.<次协议版本号>－<软件版本号>”，协议版本号由主版本号和次版本号组成，软件版本号主要是为调试使用。

(3) 客户端收到报文后，解析该数据包，如果服务器端的协议版本号比自己的低，且客户端能支持服务器端的低版本，就使用服务器端的低版本协议号，否则使用自己的协议版本号。

(4) 客户端回应服务器一个报文，包含了客户端决定使用的协议版本号。服务器比较客户端发来的版本号，如果服务器支持该版本，则版本协商结果为使用该版本，否则，版本协商失败。。

(5) 如果协商成功，则进入密钥和算法协商阶段，否则服务器端断开 TCP 连接。



说明

上述报文都是采用明文方式传输的。

2. 密钥和算法协商阶段

具体步骤如下：

- 服务器端和客户端分别发送算法协商报文给对端，报文中包含自己支持的公钥算法列表、加密算法列表、MAC（Message Authentication Code，消息验证码）算法列表、压缩算法列表等。
- 服务器端和客户端根据对端和本端支持的算法列表得出最终使用的算法。任何一种算法协商失败，都会导致服务器端和客户端的算法协商过程失败，服务器将断开与客户端的连接。
- 服务器端和客户端利用 DH 交换（Diffie-Hellman Exchange）算法、主机密钥对等参数，生成会话密钥和会话 ID，并完成客户端对服务器身份的验证。

通过以上步骤，服务器端和客户端就取得了相同的会话密钥和会话 ID。对于后续传输的数据，两端都会使用会话密钥进行加密和解密，保证了数据传送的安全。会话 ID 用来标识一个 SSH 连接，在认证阶段，两端会使用会话 ID 还会用于两端的认证过程。



注意

在协商阶段之前，服务器端已经需要生成 RSA 或 DSADSA 或 RSA 密钥对，他们不仅用于主要用于参与生成会话密钥的生成会话 ID，还用于客户端验证服务器身份。DSARSA 和 DSARSA 密钥对的详细介绍请参见“安全分册”中的“公钥管理配置”。

3. 认证阶段

SSH 提供两种认证方法：

- password 认证：利用 AAA（Authentication、Authorization、Accounting，认证、授权和计费）对用户进行认证。客户端向服务器发出 password 认证请求，将用户名和密码加密后发送给服务器；服务器将该信息解密后得到用户名和密码的明文，通过本地认证或远程认证验证用户名和密码的合法性，并返回认证成功或失败的消息。
- publickey 认证：采用数字签名的方法来认证客户端。目前，设备上可以利用 DSA 和 RSA 两种公共密钥算法实现数字签名。客户端发送包含用户名、公共密钥和公共密钥算法的 publickey 认证请求给服务器端。服务器对公钥进行合法性检查，如果不合法，则直接发送失败消息；否则，服务器利用数字签名对客户端进行认证，并返回认证成功或失败的消息。

认证阶段的具体步骤如下：

- 客户端向服务器端发送认证请求，认证请求中包含用户名、认证方法（password 认证或 publickey 认证）、与该认证方法相关的内容（如：password 认证时，内容为密码）。
- 服务器端对客户端进行认证，如果认证失败，则向客户端发送认证失败消息，其中包含可以再次认证的方法列表。
- 客户端从认证方法列表中选取一种认证方法再次进行认证。
- 该过程反复进行，直到认证成功或者认证次数达到上限，服务器关闭连接为止。

SSH 提供两种认证方法：

- password 认证：客户端向服务器发出 password 认证请求，将用户名和密码加密后发送给服务器；服务器将该信息解密后得到用户名和密码的明文，与设备上保存的用户名和密码进行比较，并返回认证成功或失败的消息。
- publickey 认证：采用数字签名的方法来认证客户端。目前，设备上可以利用 RSA 和 DSA 两种公共密钥算法实现数字签名。客户端发送包含用户名、公共密钥和公共密钥算法的 publickey 认证请求给服务器端。服务器对公钥进行合法性检查，如果不合法，则直接发送失败消息；否则，服务器利用数字签名对客户端进行认证，并返回认证成功或失败的消息。

说明

除了 password 认证和 publickey 认证，SSH2.0 还提供了 **password-publickey** 认证和 **any** 认证。

- **password-publickey** 认证：指定客户端版本为 SSH2 的用户认证方式为必须同时进行 password 和 publickey 两种认证；客户端版本为 SSH1 的用户认证方式为只要进行其中一种认证即可。指定该用户的认证方式为 password 和 publickey 认证同时满足。客户端版本为 SSH1 的用户只要通过其中一种认证即可登录；客户端版本为 SSH2 的用户必须两种认证都通过才能登录。
- **any** 认证：不指定用户的认证方式，用户指定该用户的认证方式既可以是采用 password 认证，也可以是采用 publickey 认证。。

---

4. 会话请求阶段

认证通过后，客户端向服务器发送会话请求。服务器等待并处理客户端的请求。在这个阶段，请求被成功处理后，服务器会向客户端回应 SSH_SMSG_SUCCESS 包，SSH 进入交互会话阶段；否则回应 SSH_SMSG_FAILURE 包，表示服务器处理请求失败或者不能识别请求。

5. 交互会话阶段

会话请求成功后，连接进入交互会话阶段。在这个模式下，数据被双向传送。客户端将要执行的命令加密后传给服务器，服务器接收到报文，解密后执行该命令，将执行的结果加密发送给客户端，客户端将接收到的结果解密后显示到终端上。

说明

- 在交互会话阶段，用户在客户端可以通过粘贴文本会话的方式执行命令，但文本会话不能超过 2000 字节，且粘贴的命令最好是同一视图下的命令，否则服务器可能无法正确执行该命令。
- 如果粘贴的文本会话超过 2000 字节，可以采用将配置文件通过 SFTP(Secure FTP，安全的 FTP) 方式上传到服务器，利用新的配置文件重新启动的方式执行这些命令。

1.2 配置设备作为 SSH 服务器

1.2.1 SSH 服务器配置任务简介

表1-2 SSH 服务器配置任务简介

配置任务	说明	详细配置
生成 RSA 或 DSA 密钥	必选	1.2.2
使能 SSH 服务器功能	必选	1.2.3
配置 SSH 客户端登录时的用户界面	必选	1.2.4
配置客户端的公钥	对于采用 publickey 认证的 SSH 用户为必选 对于采用 password 认证的 SSH 用户为可选	1.2.5
配置 SSH 用户并指定服务类型和认证方式	可选	1.2.6
配置服务器上的 SSH 管理功能	可选	1.2.7

1.2.2 生成 RSA 或 DSA 密钥

服务器端生成的服务器端的 RSA 或 DSADSA 或 RSA 密钥对，用于在密钥和算法协商阶段生成会话 ID，以及客户端认证服务器是完成 SSH 登录的必要操作。

表1-3 生成 RSA 或 DSA 密钥

操作	命令	说明
进入系统视图	**system-view**	-
生成 RSA 或 DSA 密钥对	**public-key local create { dsa | rsa }**	必选 缺省情况下，没有生成 RSA 和 DSA 密钥对

说明

- **public-key local create** 命令的详细介绍请参见“安全分册”中的“公钥管理命令”。
- 虽然一个客户端只会采用 DSA 和 RSA 公共密钥算法中的一种来认证服务器，但是由于不同客户端支持的公共密钥算法不同，为了确保客户端能够成功登录服务器，建议在服务器上生成 DSA 和 RSA 两种密钥对。
- 通过 **public-key local create rsa** 命令生成 RSA 密钥对时，将同时生成两个密钥对——服务器密钥对和主机密钥对，二者分别包括一个公钥和一个私钥。SSH1 中利用 SSH 服务器端的服务器公钥加密会话密钥，以保证会话密钥传输的安全；SSH2 中通过 DH 算法在 SSH 服务器和 SSH 客户端上生成会话密钥，不需要传输会话密钥，因此 SSH2 中没有利用服务器密钥对。
- RSA 服务器密钥和主机密钥模数的最小长度为 512 位，最大长度为 2048 位。在 SSH2 中，有些客户端要求服务器端生成的密钥长度必须大于或等于 768 位。
- 通过 **public-key local create dsa** 命令生成 DSA 密钥对时，只生成一个主机密钥对。SSH1 中不支持 DSA 算法。
- DSA 主机密钥模数的最小长度为 512 位，最大长度为 2048 位。在 SSH2 中，有些客户端要求服务器端生成的密钥长度必须大于或等于 768 位。

1.2.3 使能 SSH 服务器功能

表1-4 使能 SSH 服务器功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 SSH 服务器功能	**ssh server enable**	必选 缺省情况下，SSH 服务器功能处于关闭状态

1.2.4 配置 SSH 客户端登录时的用户界面

SSH 客户端通过 VTY 用户界面访问设备。因此，需要配置 SSH 客户端登录时采用的 VTY 用户界面，使其支持 SSH 远程登录协议。配置结果在客户端下次请求登录时生效。

表1-5 配置 SSH 客户端登录时的用户界面

操作	命令	说明
进入系统视图	**system-view**	-
进入一个或多个 VTY 用户界面视图	**user-interface vty** *number* [*ending-number*]	-
配置登录用户界面的认证方式为 **scheme** 方式	**authentication-mode scheme** [**command-authorization**]	必选 缺省情况下，用户界面认证为 **password** 方式
配置所在用户界面支持 SSH 协议	**protocol inbound** { **all** | **ssh** | **telnet** }	可选 缺省情况下，系统支持所有的协议，即支持 Telnet、PAD 和 SSH

注意

- **authentication-mode** 命令和 **protocol inbound** 命令的详细介绍，请参见“系统分册”中的“用户界面命令”。
- 如果在该用户界面上配置支持的协议是 SSH，为确保配置成功，请务必配置登录用户界面的认证方式为 **authentication-mode scheme**（采用 AAA 认证）。
- 如果某用户界面已经配置成支持 SSH 协议，则无法更改用户界面的认证方式，除非先取消用户界面支持 SSH 协议。。

1.2.5 配置客户端的公钥

说明

本配置适用于采用 publickey 认证的 SSH 用户，采用 password 认证的 SSH 用户，不必进行本配置。

SSH 用户采用 publickey 认证方式时，需要在服务器端配置客户端的 DSA 或 RSA 主机公钥，并在客户端为该 SSH 用户指定与主机公钥对应的 DSA 或 RSA 私钥，以便当客户端登录服务器端时，对客户端进行验证。

可以通过手工配置和从公钥文件中导入两种方式来配置客户端公钥：

- 手工配置客户端的公钥时，可以采用手工输入或拷贝粘贴的方式将客户端的公钥配置到本地。这种方式要求手工输入或拷贝粘贴的主机公钥必须是未经转换的 DER（Distinguished Encoding Rules，特异编码规则）公钥编码格式。
- 从公钥文件中导入客户端的公钥时，系统会自动将客户端的公钥转换为 PKCS（Public Key Cryptography Standards，公共密钥加密标准）编码形式，并实现客户端公钥的配置。这种方式需要客户端事先将公钥文件通过 FTP/TFTP 以二进制（binary）方式上传到本地。

注意

- 建议选用从公钥文件导入的方式配置客户端公钥。
- SSH 服务器上最多可以配置 20 个客户端的公钥。

表1-6 手工配置客户端的公钥

操作	命令	说明
进入系统视图	**system-view**	-
进入公共密钥视图	**public-key peer** *keyname*	-
进入公共密钥编辑视图	**public-key-code begin**	-
配置客户端的公钥	直接输入公钥内容	必选 在输入公钥内容时，字符之间可以有空格，也可以按回车键继续输入数据
退回公共密钥视图	**public-key-code end**	- 退出视图时，系统自动保存配置的公钥密钥
退回系统视图	**peer-public-key end**	-

表1-7 从公钥文件中导入客户端的公钥

操作	命令	说明
进入系统视图	**system-view**	-
从公钥文件中导入 SSH 用户的公钥	**public-key peer** *keyname* **import sshkey** *filename*	必选

说明

客户端公钥配置方法及相关命令的详细介绍，请参见“安全分册”中的“公钥管理配置”。

1.2.6 配置 SSH 用户并指定服务类型和认证方式

通过本配置，可以在创建 SSH 用户的同时，指定 SSH 用户的服务类型和认证方式。

表1-8 配置 SSH 用户并指定服务类型和认证方式

操作		命令	说明
进入系统视图		**system-view**	-
创建 SSH 用户，并指定 SSH 用户的服务类型和认证方式	SSH用户的服务类型为 **stelnet**	**ssh user** *username* **service-type stelnet authentication-type** { **password** | { **any** | **password-publickey** | **publickey** } **assign publickey** *keyname* }	二者必选其一
	SSH用户的服务类型为 **all** 或 **sftp**	**ssh user** *username* **service-type** { **all** | **sftp** } **authentication-type** { **password** | { **any** | **password-publickey** | **publickey** } **assign publickey** *keyname* **work-directory** *directory-name* }	

注意

- 对于 AAA 用户，即使没有创建对应的 SSH 用户，只要能够通过 AAA 认证，且设置的服务类型为 SSH，则该用户仍然可以通过 password 认证方式，使用 Stelnet 或 SFTP 登录服务器。
- SSH 服务器上最多可以创建 1024 个 SSH 用户。
- SSH用户具有两种服务类型：Stelnet和SFTP。Stelnet即Secure Telnet，安全的Stelnet是指传统的SSH服务，详细内容请参见“1.1 SSH2.0 简介”。SFTP即Secure FTP，详细内容请参见“2.1 SFTP简介”。
- 进行 SFTP 登录时，必须设置用户的服务类型为 **sftp** 或者 **all**。
- SSH1 不支持服务类型 **sftp**。如果客户端采用 SSH1 版本登录服务器，则必须在服务器上配置服务类型为 **stelnet** 或者 **all**。否则，客户端无法登录服务器。
- SFTP 用户登录时使用的工作目录与用户使用的认证方式有关。只采用 publickey 认证方式的用户，使用的工作目录为通过 **ssh user** 命令为该用户设置的工作目录；只采用 password 认证方式的用户，使用的工作目录为通过 AAA 授权的工作目录；同时采用 publickey 和 password 两种认证方式的用户，使用的工作目录为通过 **ssh user** 命令为该用户设置的工作目录。
- 配置的认证方式和用户公钥，对于已经登录的 SSH 用户不会生效，只在 SSH 用户下次登录时生效。

说明

对于使用 publickey 认证方式的用户：

- 必须在设备上（交换机或者路由器）配置相应的用户及其公钥。
- 登录服务器后，用户可以访问的命令级别为用户界面上通过 **user privilege level** 命令配置的级别。

对于使用 password 认证方式的用户：

- 用户的账号信息可以配置在设备（交换机或者路由器）或者远程认证服务器（如 RADIUS 认证服务器）上。
- 登录服务器后，用户可以访问的命令级别由 AAA 来授权。

1.2.7 配置服务器上的 SSH 管理功能

SSH 的管理功能包括：

- 设置 SSH 服务器是否兼容 SSH1 版本的客户端
- 设置 RSA 服务器密钥的更新时间，此配置仅对 SSH 客户端版本为 SSH1 的用户有效

- 设置 SSH 用户认证的超时时间
- 设置 SSH 用户请求连接的认证尝试次数

通过定时更新服务器密钥以及对用户认证时间、认证次数的限制，可以防止恶意地对密钥和用户名的猜测和破解，从而提高了 SSH 连接的安全性。

表1-9 配置服务器上的 SSH 管理功能

操作	命令	说明
进入系统视图	**system-view**	-
设置 SSH 服务器是否兼容 SSH1 版本的客户端	**ssh server compatible-ssh1x enable**	可选 缺省情况下，SSH 服务器兼容 SSH1 版本的客户端
设置 RSA 服务器密钥对的更新时间	**ssh server rekey-interval** *hours*	可选 缺省情况下，系统不更新 RSA 服务器密钥对
设置 SSH 用户的认证超时时间	**ssh server authentication-timeout** *time-out-value*	可选 缺省情况下，SSH 用户的认证超时时间为 60 秒
设置 SSH 认证尝试的最大次数	**ssh server authentication-retries** *times*	可选 缺省情况下，SSH 连接认证尝试的最大次数为 3 次

说明

SSH 客户端通过 publickey 和 password 两种方式进行认证尝试的次数总和，不能超过 **ssh server authentication-retries** 命令配置的 SSH 连接认证尝试次数，否则，客户端认证失败，无法登录 SSH 服务器。

1.3 配置设备作为 SSH 客户端

1.3.1 SSH 客户端配置任务简介

表1-10 SSH 客户端配置任务简介

配置任务	说明	详细配置
为 SSH 客户端指定源 IP 地址或源接口	可选	1.3.2
配置 SSH 客户端是否支持首次认证	可选	1.3.3
建立 SSH 客户端和服务器端的连接	必选	1.3.4

1.3.2 为 SSH 客户端指定源 IP 地址或源接口

用户可以通过以下配置，为 SSH 客户端指定源 IP 地址或者源接口，这样客户端可以用指定的 IP 地址或接口地址访问 SSH 服务器，增加了业务的可管理性。

表1-11 为 SSH 客户端指定源 IP 地址或源接口

操作		命令	说明
进入系统视图		**system-view**	-
为 SSH 客户端指定源 IP 地址或源接口	为 SSH 客户端指定源 IPv4 地址或源接口	**ssh client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	必选 缺省情况下，客户端用设备路由指定的接口地址访问 SSH 服务器
	为 SSH 客户端指定源 IPv6 地址或源接口	**ssh client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	

1.3.3 配置 SSH 客户端是否支持首次认证

当设备作为 SSH 客户端和服务器端连接时，可以设置 SSH 客户端对访问的 SSH 服务器是否进行首次认证。

- 如果支持首次认证，则当 SSH 客户端首次访问服务器，而客户端没有配置服务器端的主机公钥时，用户可以选择继续访问该服务器，并在客户端保存该主机公钥；当用户下次访问该服务器时，就以保存的主机公钥来认证该服务器。
- 如果不支持首次认证，则当客户端没有配置服务器端的主机公钥时，客户端将拒绝访问该服务器。用户必须事先将要访问的服务器端的主机公钥配置在本地，同时指定要连接的服务器端的主机公钥名称，以便客户端对连接的服务器进行认证。

1. 配置 SSH 客户端支持首次认证

表1-12 配置 SSH 客户端支持首次认证

操作	命令	说明
进入系统视图	**system-view**	-
设置 SSH 客户端对访问的 SSH 服务器进行首次认证	**ssh client first-time enable**	可选 缺省情况下，客户端进行首次认证

2. 配置 SSH 客户端不支持首次认证

如果配置 SSH 客户端不支持首次认证，则需要在客户端配置服务器端的主机公钥，并为要连接的服务器指定主机公钥名称。

表1-13 配置 SSH 客户端不支持首次认证

操作	命令	说明
进入系统视图	**system-view**	-
设置 SSH 客户端对访问的 SSH 服务器不进行首次认证	**undo ssh client first-time**	必选 缺省情况下，客户端进行首次认证
配置服务器端的公钥	请参见“1.2.5 配置客户端的公钥”	必选 在客户端配置服务器端公钥的方法，与在服务器端配置客户端公钥的方法相同
在客户端上指定要连接的服务器端的主机公钥名称	**ssh client authentication server** *server* **assign publickey** *keyname*	必选

1.3.4 建立 SSH 客户端和服务器端的连接

表1-14 建立 SSH 客户端和服务器端的连接

操作		命令	说明
建立SSH客户端和服务器端的连接，并指定公共密钥算法、客户端和服务器的首选加密算法、首选 HMAC 算法首选密钥交换算法、首选加密算法和首选密钥交换算法	建立 SSH 客户端和 IPv4 服务器端的连接，并指定公共密钥算法、客户端和服务器的首选加密算法、首选 HMAC 算法和首选密钥交换算法	**ssh2** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	二者必选其一 必须在用户视图下执行本命令
	建立 SSH 客户端和 IPv6 服务器端的连接，并指定公共密钥算法、客户端和服务器的首选加密算法、首选 HMAC 算法和首选密钥交换算法	**ssh2 ipv6** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	

1.4 SSH 协议显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令，可以显示配置后 SSH 的运行情况，通过查看显示信息，验证配置的效果。

表1-15 SSH 协议显示和维护

操作	命令
显示当前为 SFTP 客户端设置的源 IP 地址或者源接口	**display sftp client source**
显示当前为 SSH 客户端设置的源 IP 地址或者源接口	**display ssh client source**
在 SSH 服务器端显示该服务器的状态信息或会话信息	**display ssh server** { **status** | **session** }
在 SSH 客户端显示客户端保存的服务器端的主机公钥和服务器的对应关系	**display ssh server-info**
在 SSH 服务器端上显示 SSH 用户信息	**display ssh user-information** [*username*]
显示本地密钥对中的公钥部分	**display public-key local** { **dsa** | **rsa** } **public**
显示保存在本地的远端主机的公钥信息	**display public-key peer** [**brief** | **name** *publickey-name*]



说明

display public-key local 和 **display public-key peer** 命令的详细介绍请参见“安全分册”中的“公钥配置命令”。

1.5 设备作为 SSH 服务器配置举例

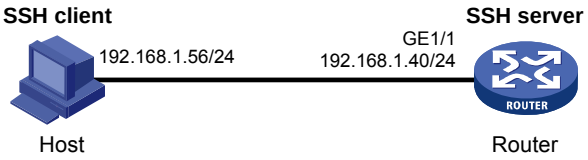
1.5.1 password 认证配置举例

1. 组网需求

- Host 与 Router 通过以太网口直接相连；
- 在 Host 上运行 SSH2.0 的客户端软件可以安全地登录 Router 进行配置管理；
- SSH 用户采用的认证方式为 password 认证，用户名和密码保存在本地设备上。。

2. 组网图

图1-1 SSH 本地配置组网图



3. 配置步骤

(1) 配置 SSH 服务器 Router

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<Router> system-view
[Router] public-key local create rsa
[Router] public-key local create dsa
[Router] ssh server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.1.40 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

设置 SSH 客户端登录用户界面的认证方式为 AAA 认证。

```
[Router] user-interface vty 0 4
[Router-ui-vty0-4] authentication-mode scheme
```

设置 Router 上远程用户登录协议为 SSH。

```
[Router-ui-vty0-4] protocol inbound ssh
[Router-ui-vty0-4] quit
```

创建本地用户 client001，并设置用户访问的命令级别为 3。

```
[Router] local-user client001
[Router-luser-client001] password simple aabbcc
[Router-luser-client001] service-type ssh
[Router-luser-client001] authorization-attribute level 3
[Router-luser-client001] quit
```

配置 SSH 用户 client001 的服务器类型为 Stelnet，认证方式为 password 认证。（此步骤可以不配置）

```
[Router] ssh user client001 service-type stelnet authentication-type password
```

(2) 配置 SSH 客户端 Host

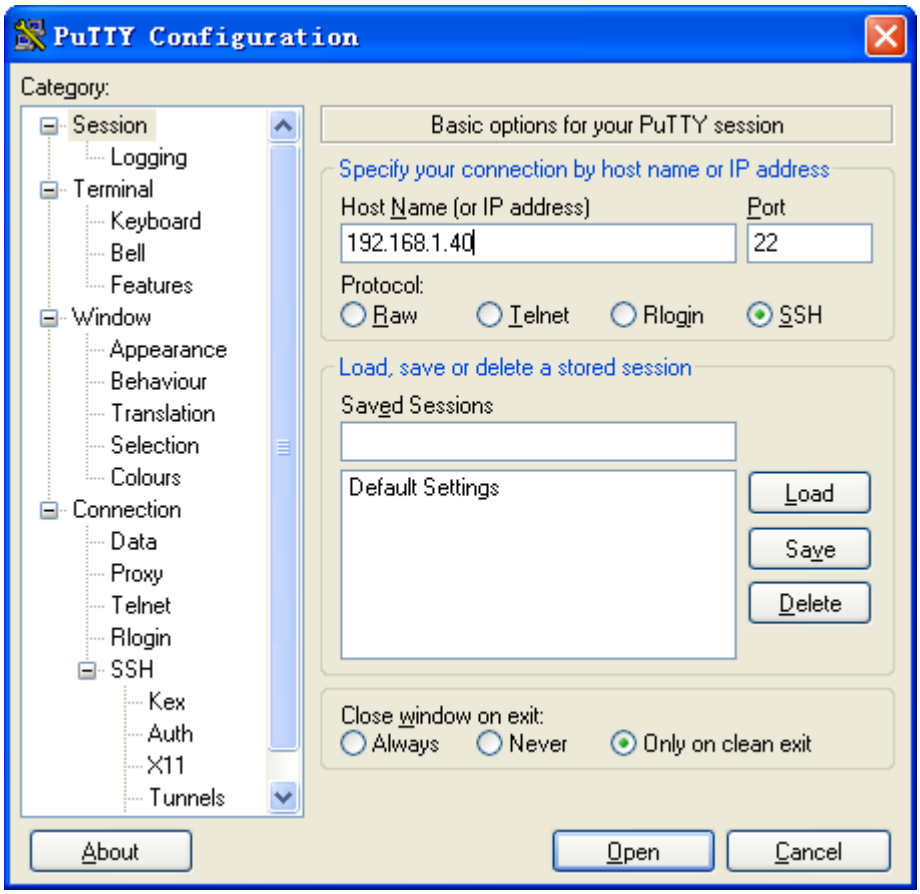
说明

SSH 客户端软件有很多，例如 PuTTY、OpenSSH 等。本文中仅以客户端软件 PuTTY0.58 为例，说明 SSH 客户端的配置方法。

建立与 SSH 服务器的连接。

打开PuTTY.exe程序，出现如 图 1-2所示的客户端配置界面。在“Host Name（or IP address）”文本框中输入SSH服务器的IP地址为 192.168.1.40。

图1-2 SSH 客户端配置界面



在 图 1-2中，单击<Open>按钮。按提示输入用户名client001 及密码aabbcc，即可进入Router的配置界面。

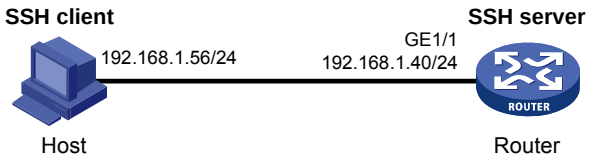
1.5.2 publickey 认证配置举例

1. 组网需求

- Host 与 Router 通过以太网口直接相连；
- 在 Host 上运行 SSH2.0 的客户端软件可以安全地登录 Router 进行配置管理；
- SSH 用户采用的认证方式为 publickey 认证，公共密钥算法为 RSA。

2. 组网图

图1-3 SSH 本地配置组网图



3. 配置步骤

(1) 配置 SSH 服务器 Router

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<Router> system-view
[Router] public-key local create rsa
[Router] public-key local create dsa
[Router] ssh server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.1.40 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

设置用户接口上认证模式为 AAA 认证。

```
[Router] user-interface vty 0 4
[Router-ui-vty0-4] authentication-mode scheme
```

设置 Router 上远程用户登录协议为 SSH。

```
[Router-ui-vty0-4] protocol inbound ssh
```

设置用户能访问的命令级别为 3。

```
[Router-ui-vty0-4] user privilege level 3
[Router-ui-vty0-4] quit
```

说明

这时需要先在 SSH 客户端使用 SSH 客户端软件生成 RSA 密钥对，将生成的 RSA 公钥保存到指定文件中，并将此公钥文件通过 FTP/TFTP 方式上传到服务器端，文件名为 key.pub。相关配置请参见客户端的配置。

从文件 key.pub 中导入远端的公钥，并命名为 Router001。

```
[Router] public-key peer Router001 import sshkey key.pub
```

设置 SSH 用户 client002 的认证方式为 publickey，并指定公钥为 Router001。

```
[Router] ssh user client002 service-type stelnet authentication-type publickey assign publickey Router001
```

(2) 配置 SSH 客户端 Host

生成 RSA 密钥对。

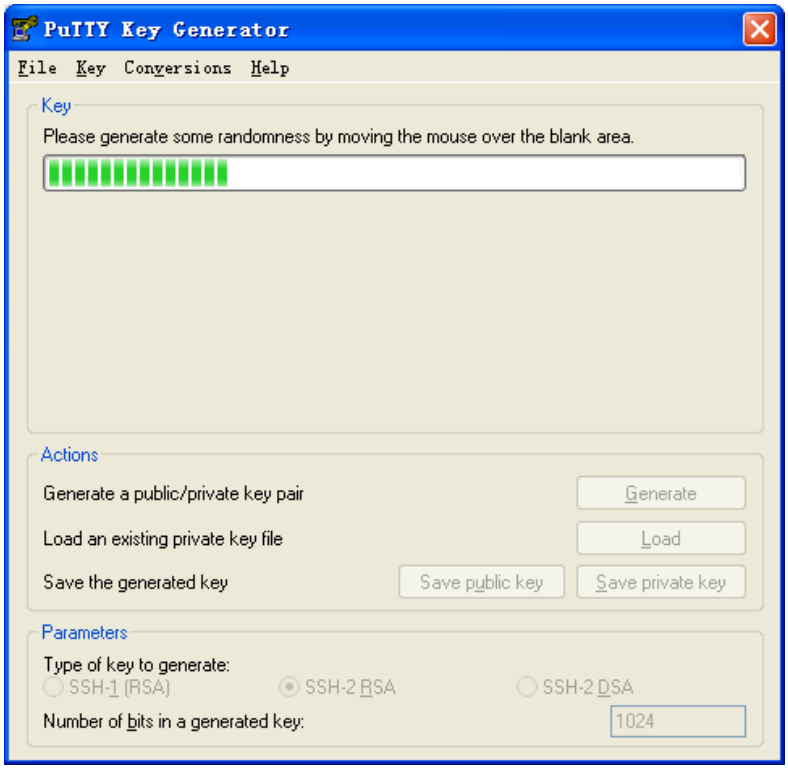
运行 PuTTYGen.exe，在参数栏中选择“SSH-2 RSA”，点击<Generate>，产生客户端密钥对。

图1-4 生成客户端密钥（1）



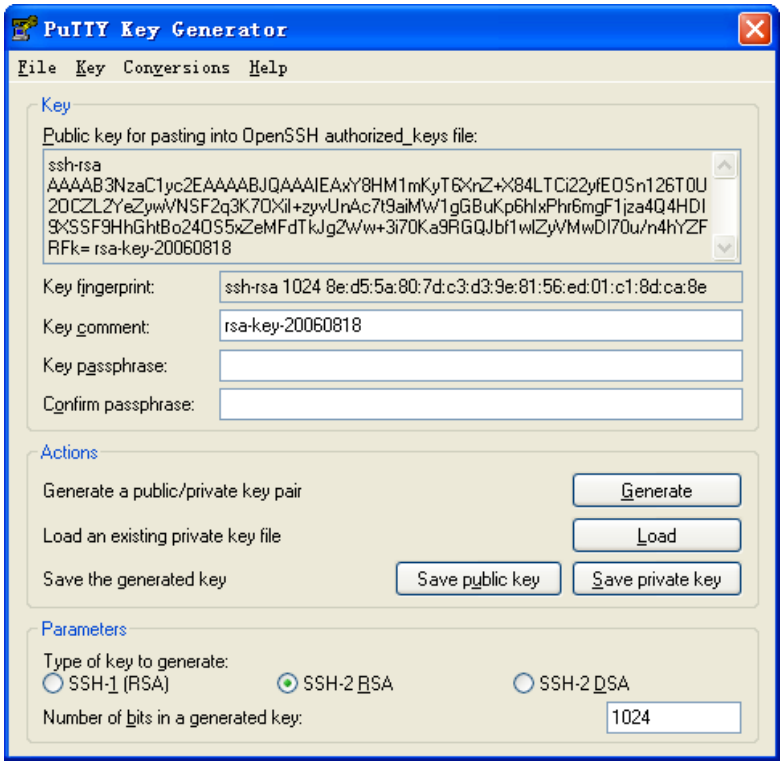
在产生密钥对的过程中需不停的移动鼠标，鼠标移动仅限于下图蓝色框中除绿色标记进程条外的地方，否则进程条的显示会不动，密钥对将停止产生，见 图 1-5。

图1-5 生成客户端密钥（2）



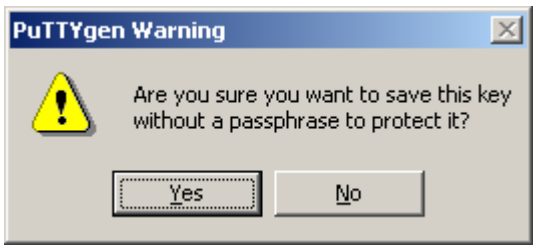
密钥对产生后，点击<Save public key>，输入存储公钥的文件名 key.pub，点击<保存>按钮。

图1-6 生成客户端密钥（3）



点击<Save private key>存储私钥，弹出警告框，提醒是否保存没做任何保护措施的私钥，点击<Yes>，输入私钥文件名为 private，点击保存。

图1-7 生成客户端密钥（4）



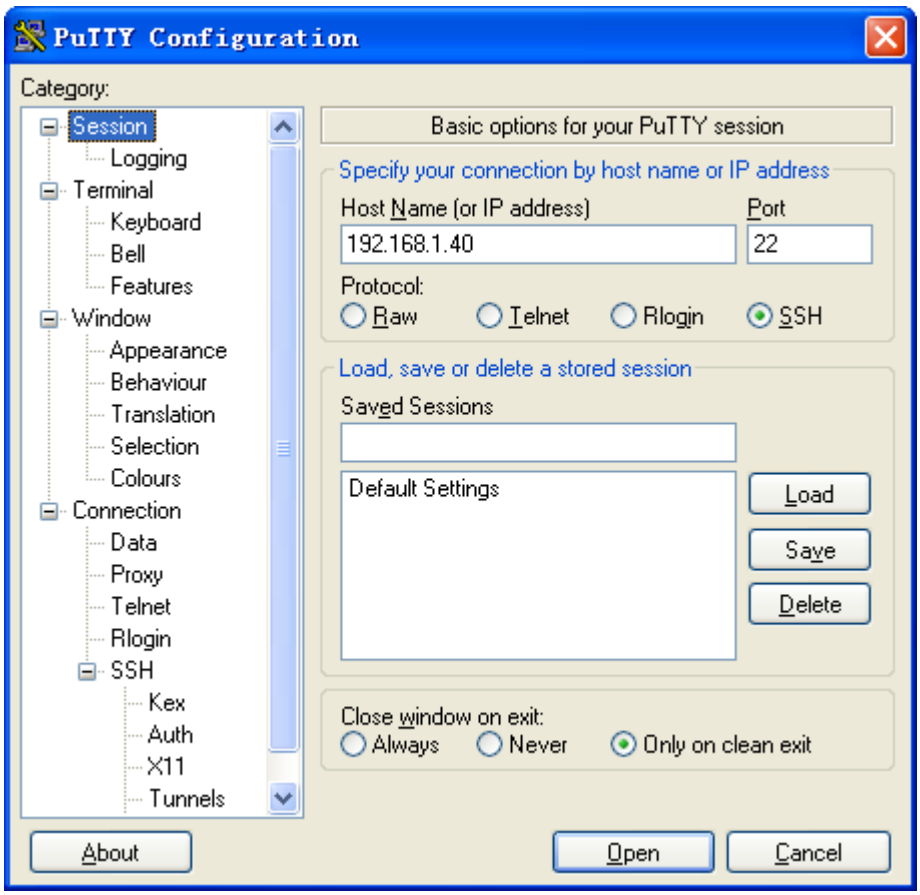
说明

客户端生成密钥对后，需要将保存的公钥文件通过 FTP/TFTP 方式上传到服务器，并完成服务器的配置后，才可继续客户端的配置。

指定私钥文件，并建立与 SSH 服务器的连接。

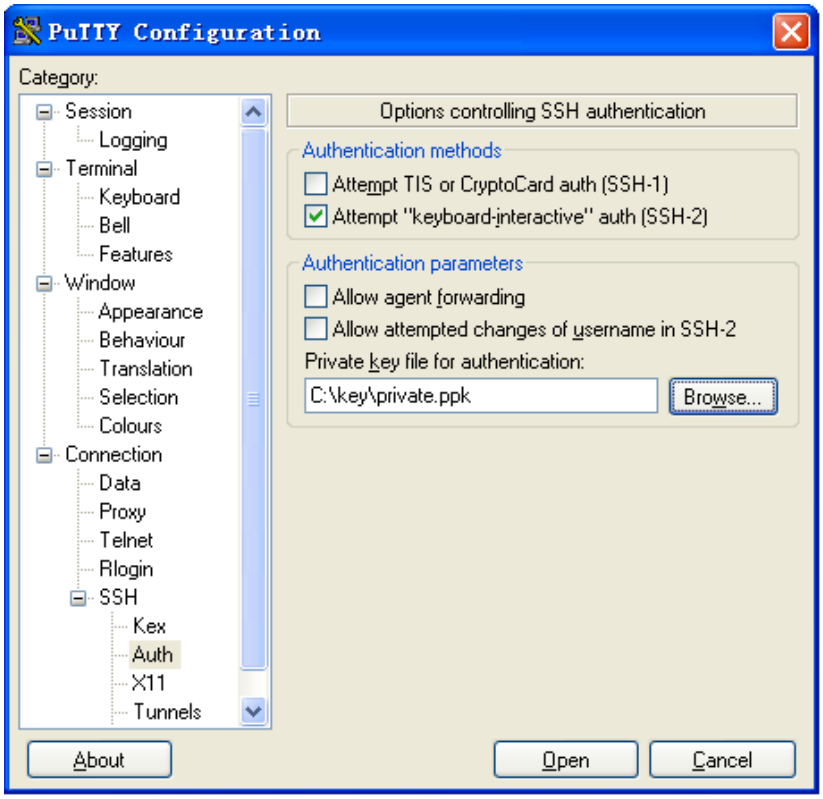
打开PuTTY.exe程序，出现如 图 1-8所示的客户端配置界面。在“Host Name（or IP address）”文本框中输入SSH服务器的IP地址为 192.168.1.40。

图1-8 SSH 客户端配置界面（1）



单击左侧导航树“SSH”下面的“Auth”（认证），出现如 图 1-9的界面。单击<Browse…>按钮，弹出文件选择窗口。选择与配置到服务器端的公钥对应的私钥文件private。

图1-9 SSH 客户端配置界面（2）



如 图 1-9，单击<Open>按钮。按提示输入用户名client002，即可进入Router的配置界面。

1.6 设备作为 SSH 客户端配置举例

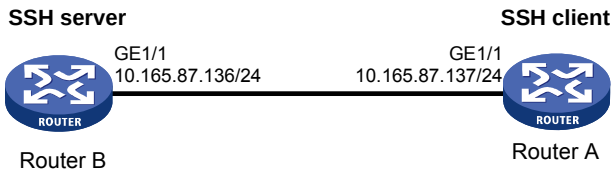
1.6.1 password 认证配置举例

1. 组网需求

如 图 1-10所示，配置Router A作为客户端，采用SSH协议登录到Router B上。SSH用户采用的认证方式为password认证，用户名为client001，密码为aabbcc用户名和密码保存在本地设备上。

2. 组网图

图1-10 设备作为 SSH 客户端配置组网图



3. 配置步骤

(1) 配置 SSH 服务器 Router B

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<RouterB> system-view
[RouterB] public-key local create rsa
[RouterB] public-key local create dsa
```

```
[RouterB] ssh server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ip address 10.165.87.136 255.255.255.0
[RouterB-GigabitEthernet1/1] quit
```

设置 SSH 客户端登录用户界面的认证方式为 AAA 认证。

```
[RouterB] user-interface vty 0 4
[RouterB-ui-vty0-4] authentication-mode scheme
```

设置 Router B 上远程用户登录协议为 SSH。

```
[RouterB-ui-vty0-4] protocol inbound ssh
[RouterB-ui-vty0-4] quit
```

创建本地用户 client001。

```
[RouterB] local-user client001
[RouterB-luser-client001] password simple aabbcc
[RouterB-luser-client001] service-type ssh
RouterB-luser-client001] authorization-attribute level 3
[RouterB-luser-client001] quit
```

配置 SSH 用户 client001 的服务类型为 Stelnet，认证方式为 password 认证（此步骤可以不配置）。

```
[RouterB] ssh user client001 service-type stelnet authentication-type password
```

(2) 配置 SSH 客户端 Router A

配置接口 GigabitEthernet1/1 的 IP 地址。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 10.165.87.137 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
[RouterA] quit
```

- 如果客户端支持首次认证，则可以直接与服务器建立连接。

建立到服务器 10.165.87.136 的 SSH 连接。

```
<RouterA> ssh2 10.165.87.136
Username: client001
Trying 10.165.87.136 ...
Press CTRL+K to abort
Connected to 10.165.87.136 ...
The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
Enter password:
```

输入正确的密码之后，即可成功登录到 Router B 上。

- 如果客户端不支持首次认证，则需要进行如下配置。

配置客户端对服务器不进行首次认证。

```
[RouterA] undo ssh client first-time
```

在客户端配置 SSH 服务器端的主机公钥。在公共密钥编辑视图输入服务器端的主机公钥，即在服务器端通过 **display public-key local dsa public** 命令显示的公钥内容。。

```
[RouterA] public-key peer key1
[RouterA-pkey-public-key] public-key-code begin
[RouterA-pkey-key-code]308201B73082012C06072A8648CE3804013082011F0281810
0D757262C4584C44C211F18BD96E5F0
[RouterA-pkey-key-code]61C4F0A423F7FE6B6B85B34CEF72CE14A0D3A5222FE08CECE
65BE6C265854889DC1EDBD13EC8B274
[RouterA-pkey-key-code]DA9F75BA26CCB987723602787E922BA84421F22C3C89CB9B0
6FD60FE01941DDD77FE6B12893DA76E
[RouterA-pkey-key-code]EBC1D128D97F0678D7722B5341C8506F358214B16A2FAC4B3
68950387811C7DA33021500C773218C
[RouterA-pkey-key-code]737EC8EE993B4F2DED30F48EDACE915F0281810082269009E
14EC474BAF2932E69D3B1F18517AD95
[RouterA-pkey-key-code]94184CCDFCEAE96EC4D5EF93133E84B47093C52B20CD35D02
492B3959EC6499625BC4FA5082E22C5
[RouterA-pkey-key-code]B374E16DD00132CE71B020217091AC717B612391C76C1FB2E
88317C1BD8171D41ECB83E210C03CC9
[RouterA-pkey-key-code]B32E810561C21621C73D6DAAC028F4B1585DA7F42519718CC
9B09EEF0381840002818000AF995917
[RouterA-pkey-key-code]E1E570A3F6B1C2411948B3B4FFA256699B3BF871221CC9C5D
F257523777D033BEE77FC378145F2AD
[RouterA-pkey-key-code]D716D7DB9FCABB4ADBF6FB4FDB0CA25C761B308EF53009F71
01F7C62621216D5A572C379A32AC290
[RouterA-pkey-key-code]E55B394A217DA38B65B77F0185C8DB8095522D1EF044B465E
8716261214A5A3B493E866991113B2D
[RouterA-pkey-key-code]485348
[RouterA-pkey-key-code] public-key-code end
[RouterA-pkey-public-key] peer-public-key end
```

指定服务器 10.165.87.136 对应的主机公钥名称为 key1。

```
[RouterA] ssh client authentication server 10.165.87.136 assign publickey key1
[RouterA] quit
```

建立到服务器 10.165.87.136 的 SSH 连接。

```
<RouterA> ssh2 10.165.87.136
Username: client001
Trying 10.165.87.136
Press CTRL+K to abort
Connected to 10.165.87.136...
Enter password:
```

输入正确的密码之后，即可成功登录到 Router B 上。

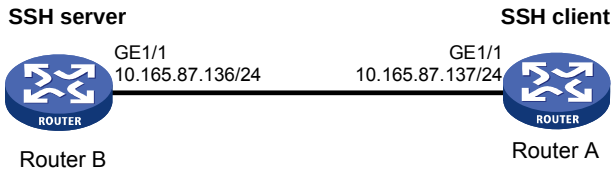
1.6.2 publickey 认证配置举例

1. 组网需求

如 图 1-11所示，配置Router A作为客户端，采用SSH协议登录到Router B上。SSH用户采用的认证方式为publickey认证，公共密钥算法为DSA。

2. 组网图

图1-11 设备作为 SSH 客户端配置组网图



3. 配置步骤

(1) 配置 SSH 服务器 Router B

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<RouterB> system-view
[RouterB] public-key local create rsa
[RouterB] public-key local create dsa
[RouterB] ssh server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ip address 10.165.87.136 255.255.255.0
[RouterB-GigabitEthernet1/1] quit
```

设置 SSH 客户端登录用户界面的认证方式为 AAA 认证。

```
[RouterB] user-interface vty 0 4
[RouterB-ui-vty0-4] authentication-mode scheme
```

设置 Router B 上远程用户登录协议为 SSH。

```
[RouterB-ui-vty0-4] protocol inbound ssh
```

设置用户能访问的命令级别为 3。

```
[RouterB-ui-vty0-4] user privilege level 3
[RouterB-ui-vty0-4] quit
```

说明

这时需要先在 SSH 客户端生成 DSA 密钥对，将生成的 DSA 公钥保存到指定文件中，并将此公钥文件通过 FTP/TFTP 方式上传到服务器端，文件名为 key.pub。相关配置请参见客户端的配置。

从文件 key.pub 中导入远端的公钥。

```
[RouterB] public-key peer Router001 import sshkey key.pub
```

设置 SSH 用户 client002 的认证方式为 publickey，并指定公钥为 Router001。

```
[RouterB] ssh user client002 service-type stelnet authentication-type publickey assign
publickey Router001
```

(2) 配置 SSH 客户端 Router A

配置接口 GigabitEthernet1/1 的 IP 地址。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/1
```

```
[RouterA-GigabitEthernet1/1] ip address 10.165.87.137 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
```

生成 DSA 密钥对。

```
[RouterA] public-key local create dsa
```

将生成的 DSA 主机公钥导出到指定文件 key.pub 中。

```
[RouterA] public-key local export dsa ssh2 key.pub
[RouterA] quit
```



说明

客户端生成密钥对后，需要将保存的公钥文件通过 FTP/TFTP 方式上传到服务器，并完成服务器的配置后，才可继续客户端的配置。

建立到服务器 10.165.87.136 的 SSH 连接。

```
<RouterA> ssh2 10.165.87.136
Username: client002
Trying 10.165.87.136 ...
Press CTRL+K to abort
Connected to 10.165.87.136 ...

The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
```

输入正确的密码之后，即可成功登录到 Router B 上。

2 SFTP 服务

2.1 SFTP 简介

SFTP 是 Secure FTP 的简称，是 SSH 2.0 中新增的功能。

SFTP 建立在 SSH 连接的基础之上，它使得远程用户可以安全地登录设备，进行文件管理和文件传送等操作，为数据传输提供了更高的安全保障。同时，由于设备支持作为客户端的功能，用户可以从本地设备安全登录到远程设备上，进行文件的安全传输。

2.2 配置 SFTP 服务器

2.2.1 配置准备

- 已经完成SSH服务器的配置，具体配置步骤请参见“1.2 配置设备作为SSH服务器”。
- SSH用户的服务类型设置为**sftp**或者**all**。具体配置步骤请参见“1.2.6 配置SSH用户并指定服务类型和认证方式”。

2.2.2 启动 SFTP 服务器

该配置任务用于启动 SFTP 服务，使客户端能用 SFTP 的方式登录到服务器。

表2-1 启动 SFTP 服务器

操作	命令	说明
进入系统视图	**system-view**	-
启动 SFTP 服务器	**sftp server enable**	必选 缺省情况下，SFTP 服务器处于关闭状态



说明

设备作为 SFTP 服务器时，同时只能有一个用户访问 SFTP 服务器。如果以 WinSCP 作为 SFTP 客户端，则无法直接编辑服务器上的文件，需要采用将服务器上的文件下载到本地，在本地修改后上传到服务器的方式实现对文件的编辑。

2.2.3 配置 SFTP 用户连接的空闲超时时间

当 SFTP 用户连接的空闲时间超过设定的阈值后，系统会自动断开此用户的连接，从而有效避免用户长期占用连接而不进行任何操作。

表2-2 配置 SFTP 用户连接的空闲超时时间

操作	命令	说明
进入系统视图	**system-view**	-
设置 SFTP 用户连接的空闲超时时间	**sftp server idle-timeout** *time-out-value*	可选 缺省情况下，SFTP 用户连接的空闲超时时间为 10 分钟

2.3 配置 SFTP 客户端

2.3.1 为 SFTP 客户端指定源 IP 地址或源接口

用户可以通过以下配置，使 SFTP 客户端只能使用指定接口的 IP 地址或指定 IP 地址作为源地址建立与 SFTP 服务器的连接，增加了业务的可管理性。

表2-3 为 SFTP 客户端指定 IP 地址或源接口

操作		命令	说明
进入系统视图		**system-view**	-
为 SFTP 客户端指定源 IP 地址或源接口	为 SFTP 客户端指定源 IPv4 地址或源接口	**sftp client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	二者必选其一 缺省情况下，客户端用设备路由指定的接口地址访问 SFTP 服务器
	为 SFTP 客户端指定源 IPv6 地址或源接口	**sftp client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	

2.3.2 建立与 SFTP 服务器的连接

该配置任务用来启动 SFTP 客户端程序，与远程 SFTP 服务器建立连接，并进入到 SFTP 客户端视图。

表2-4 建立与 SFTP 服务器的连接

操作		命令	说明
与远程 SFTP 服务器建立连接，并进入 SFTP 客户端视图	与远程 IPv4 SFTP 服务器建立连接，并进入 SFTP 客户端视图	**sftp** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	必选其一 请在用户视图下执行此命令
	与远程 IPv6 SFTP 服务器建立连接，并进入 SFTP 客户端视图	**sftp ipv6** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	

2.3.3 SFTP 目录操作

SFTP 目录操作包括：改变或显示当前的工作目录、显示指定目录下的文件或目录信息、改变服务器上指定的文件夹的名字、创建或删除目录。

表2-5 SFTP 目录操作

操作	命令	说明
进入 SFTP 客户端视图	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	必选 请在用户视图下执行此命令
改变远程 SFTP 服务器上的工作路径	**cd** [*remote-path*]	可选
返回到上一级目录	**cdup**	可选
显示远程 SFTP 服务器上的当前工作目录	**pwd**	可选
显示指定目录下的文件列表	**dir** [**-a** | **-l**] [*remote-path*]	可选 **dir** 和 **ls** 两条命令的作用相同
	ls [**-a** | **-l**] [*remote-path*]	
改变 SFTP 服务器上指定的目录的名字	**rename** *oldname newname*	可选
在远程 SFTP 服务器上创建新的目录	**mkdir** *remote-path*	可选
删除 SFTP 服务器上指定的目录	**rmdir** *remote-path*&<1-10>	可选

2.3.4 SFTP 文件操作

SFTP 文件操作包括：改变文件名、下载文件、上传文件、显示文件列表、删除文件。

表2-6 SFTP 文件操作

操作	命令	说明
进入 SFTP 客户端视图	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	必选 请在用户视图下执行此命令
改变 SFTP 服务器上指定的文件或目录的名字	**rename** *old-name new-name*	可选
从远程服务器上下载文件并存储在本地	**get** *remote-file* [*local-file*]	可选
将本地的文件上传到远程 SFTP 服务器	**put** *local-file* [*remote-file*]	可选
查询指定目录下的文件	**dir** [**-a** | **-l**] [*remote-path*]	可选 **dir** 和 **ls** 两条命令的作用相同
	ls [**-a** | **-l**] [*remote-path*]	
删除 SFTP 服务器上指定的文件	**delete** *remote-file*&<1-10>	可选 **delete** 和 **remove** 两条命令的功能相同
	remove *remote-file*&<1-10>	

2.3.5 显示帮助信息

本配置用于显示命令的帮助信息，如命令格式、参数配置等。

表2-7 显示客户端命令的帮助信息

操作	命令	说明
进入 SFTP 客户端视图	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	必选 请在用户视图下执行此命令
显示 SFTP 客户端命令的帮助信息	**help** [**all** | *command-name*]	必选

2.3.6 终止与远程 SFTP 服务器的连接

表2-8 终止与远程 SFTP 服务器的连接

操作	命令	说明
进入 SFTP 客户端视图	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **3des** | **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **3des** | **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	必选 请在用户视图下执行此命令
终止与远程 SFTP 服务器的连接，并退回用户视图	**bye**	必选其一 **bye**，**exit** 和 **quit** 三条命令的功能相同
	exit	
	quit	

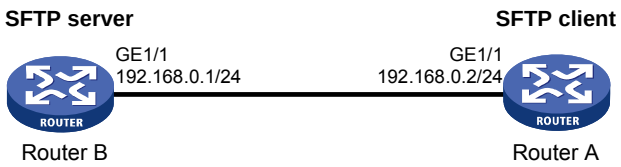
2.4 设备作为 SFTP 客户端配置举例

1. 组网需求

如 图 2-1，Router A和Router B之间建立SSH连接，Router A作为SFTP客户端登录到Router B，进行文件管理和文件传送等操作。SSH用户采用的认证方式为publickey认证，公共密钥算法为RSA。

2. 组网图

图2-1 设备作为 SFTP 客户端组网图



3. 配置步骤

(1) 配置服务器端 Router B

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<RouterB> system-view
[RouterB] public-key local create rsa
[RouterB] public-key local create dsa
[RouterB] ssh server enable
```

启动 SFTP 服务器。

```
[RouterB] sftp server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ip address 192.168.0.1 255.255.255.0
[RouterB-GigabitEthernet1/1] quit
```

设置 SSH 客户端登录用户界面的认证方式为 AAA 认证。

```
[RouterB] user-interface vty 0 4
[RouterB-ui-vty0-4] authentication-mode scheme
```

设置 Router B 上远程用户登录协议为 SSH。

```
[RouterB-ui-vty0-4] protocol inbound ssh
[RouterB-ui-vty0-4] quit
```



说明

这时需要先在 SSH 客户端生成 RSA 密钥对，将生成的 RSA 公钥保存到指定文件中，并将此公钥文件通过 FTP/TFTP 方式上传到服务器端，文件名为 pubkey。相关配置请参见客户端的配置。

从文件 pubkey 中导入远端的公钥。

```
[RouterB] public-key peer Router001 import sshkey pubkey
```

设置 SSH 用户 client001 的服务类型为 SFTP，认证方式为 publickey，并指定公钥为 Router001，工作目录为 cf:/。

```
[RouterB] ssh user client001 service-type sftp authentication-type publickey assign publickey Router001 work-directory cf:/
```

(2) 配置客户端 Router A

配置接口 GigabitEthernet1/1 的 IP 地址。

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.2 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
```

生成 RSA 密钥对。

```
[RouterA] public-key local create rsa
```

将生成的 RSA 主机公钥导出到指定文件 pubkey 中。

```
[RouterA] public-key local export rsa ssh2 pubkey
[RouterA] quit
```

说明

客户端生成密钥对后，需要将保存的公钥文件通过 FTP/TFTP 方式上传到服务器，并完成服务器的配置后，才可继续客户端的配置。

与远程 SFTP 服务器建立连接，进入 SFTP Client 视图。

```
<RouterA> sftp 192.168.0.1 identity-key rsa
Input Username: client001
Trying 192.168.0.1 ...
Press CTRL+K to abort
Connected to 192.168.0.1 ...

The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n

sftp-client>
```

显示服务器的当前目录，删除文件 z，并检查此目录是否删除成功。

```
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
-rwxrwxrwx   1 noone    nogroup         0 Sep 01 08:00 z
End of file
Success

sftp-client> delete z
The following File will be deleted:
/z
Are you sure to delete it? [Y/N]:y
This operation may take a long time.Please wait...

Success
File successfully Removed
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
End of file
Success
```

新增目录 new1，并检查新目录是否创建成功。

```
sftp-client> mkdir new1
Success
```

```
New directory created
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:30 new1
End of file
Success
```

将目录名 new1 更名为 new2，并查看是否更名成功。

```
sftp-client> rename new1 new2
Success
File successfully renamed
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
End of file
Success
```

从服务器上下载文件 pubkey2 到本地，并更名为 public。

```
End of file
Success
sftp-client> get pubkey2 public
Remote  file:/pubkey2 --->  Local file: public
Downloading file successfully ended
```

将本地文件 pu 上传到服务器上，更名为 puk，并查看上传是否成功。

```
sftp-client> put pu puk
Local file:pu --->  Remote file: /puk
Success
Uploading file successfully ended
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey1
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:35 pub
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:36 puk
sftp-client>
```

退出 SFTP。

```
sftp-client> quit
Bye
```

```
Connection closed.
<RouterA>
```

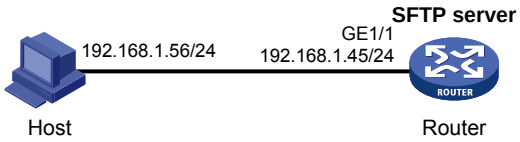
2.5 设备作为 SFTP 服务器配置举例

1. 组网需求

如 图 2-2，Host和Router之间建立SSH连接，Host作为SFTP客户端登录到Router，进行文件管理和文件传送等操作。认证方式为password，用户名为client002，密码为aabbcc用户名和密码保存在本地设备上。

2. 组网图

图2-2 设备作为 SFTP 服务器组网图



3. 配置步骤

(1) 配置 SFTP 服务器 Router

生成 RSA 及 DSA 密钥对，并启动 SSH 服务器。

```
<Router> system-view
[Router] public-key local create rsa
[Router] public-key local create dsa
[Router] ssh server enable
```

启动 SFTP 服务器。

```
[Router] sftp server enable
```

配置接口 GigabitEthernet1/1 的 IP 地址，客户端将通过该地址连接 SSH 服务器。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.1.45 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

设置 SSH 客户端登录用户界面的认证方式为 AAA 认证。

```
[Router] user-interface vty 0 4
[Router-ui-vty0-4] authentication-mode scheme
```

设置 Router 上远程用户登录协议为 SSH。

```
[Router-ui-vty0-4] protocol inbound ssh
[Router-ui-vty0-4] quit
```

创建本地用户 client002，密码为 aabbcc，服务类型为 SSH。

```
[Router] local-user client002
[Router-luser-client002] password simple aabbcc
[Router-luser-client002] service-type ssh
[Router-luser-client002] quit
```

配置 SSH 用户认证方式为 password，服务类型为 SFTP。

```
[Router] ssh user client002 service-type sftp authentication-type password
```

(2) 配置 SFTP 客户端 Host

说明

- SFTP 客户端软件有很多，本文中仅以客户端软件 PuTTY0.58 中的 PSFTP 为例，说明 SFTP 客户端的配置方法。
- PSFTP 只支持 password 认证，不支持 publickey 认证

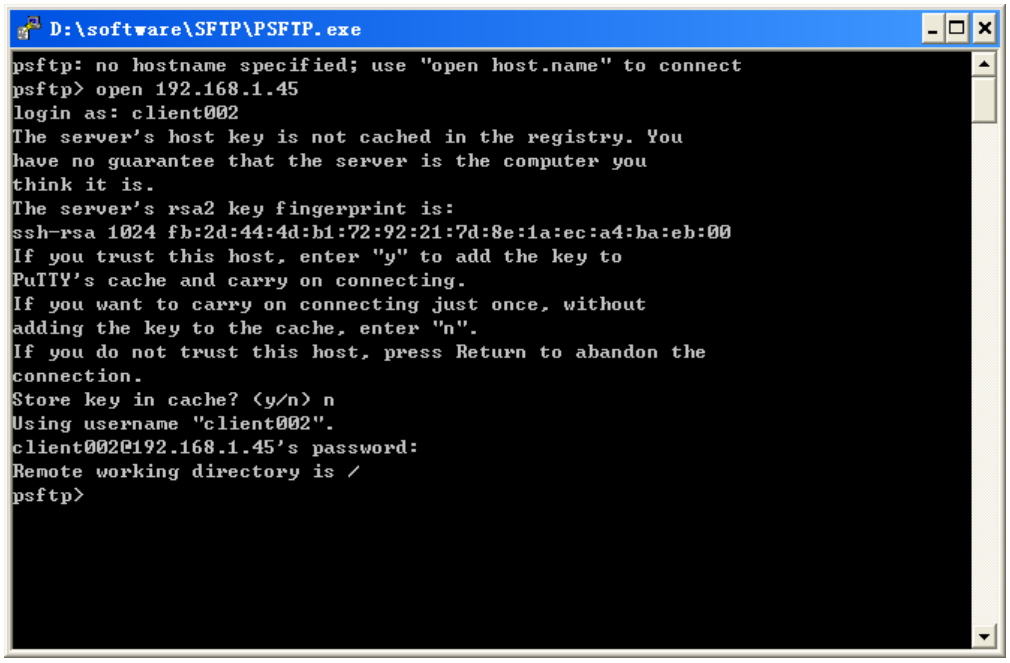
建立与 SFTP 服务器的连接。

打开psftp.exe程序，出现如 图 2-3所示的客户端配置界面。输入如下命令：

```
open 192.168.1.45
```

根据提示输入用户名 client002，密码 aabbcc，即可登录 SFTP 服务器。

图2-3 SFTP 客户端登录界面



目 录

1 Portal配置 1-1
1.1 Portal简介 1-1
1.1.1 Portal概述 1-1
1.1.2 Portal扩展功能 1-1
1.1.3 Portal的系统组成 1-1
1.1.4 Portal的认证方式 1-3
1.1.5 Portal认证过程 1-3
1.2 Portal配置任务简介 1-5
1.3 Portal的基本配置 1-6
1.3.1 配置准备 1-6
1.3.2 配置Portal 1-6
1.4 配置免认证规则 1-7
1.5 配置认证网段 1-8
1.6 配置强制用户下线 1-8
1.7 配置强制认证域 1-8
1.8 配置Portal用户信息同步功能 1-9
1.9 配置Portal用户在线探测功能 1-10
1.10 Portal显示和维护 1-10
1.11 Portal典型配置举例 1-11
1.11.1 Portal直接认证配置举例 1-11
1.11.2 Portal二次地址分配认证配置举例 1-12
1.11.3 三层Portal认证配置举例 1-14
1.11.4 Portal直接认证扩展功能配置举例 1-16
1.11.5 Portal二次地址分配认证扩展功能配置举例 1-18
1.11.6 三层Portal认证扩展功能配置举例 1-20
1.11.7 Portal服务器探测和用户同步功能配置举例 1-22
1.12 常见配置错误举例 1-28
1.12.1 接入设备和Portal服务器上的密钥不一致 1-28
1.12.2 接入设备上服务器端口配置错误 1-28

1 Portal 配置

1.1 Portal 简介

1.1.1 Portal 概述

Portal 在英语中是入口的意思。Portal 认证通常也称为 Web 认证，一般将 Portal 认证网站称为门户网站。

未认证用户上网时，设备强制用户登录到特定站点，用户可以免费访问其中的服务。当用户需要使用互联网中的其它信息时，必须在门户网站进行认证，只有认证通过后才可以使用互联网资源。

用户可以主动访问已知的 Portal 认证网站，输入用户名和密码进行认证，这种开始 Portal 认证的方式称作主动认证。反之，如果用户试图通过 HTTP 访问其他外网，将被强制访问 Portal 认证网站，从而开始 Portal 认证过程，这种方式称作强制认证。

Portal 业务可以为运营商提供方便的管理功能，门户网站可以开展广告、社区服务、个性化的业务等，使宽带运营商、设备提供商和内容服务提供商形成一个产业生态系统。

1.1.2 Portal 扩展功能

Portal 的扩展功能主要是指通过强制接入终端实施补丁和防病毒策略，加强网络终端对病毒攻击的主动防御能力。具体扩展功能如下：

- 在 Portal 身份认证的基础上增加了安全认证机制，可以检测接入终端上是否安装了防病毒软件、是否更新病毒库、是否安装了非法软件、是否更新了操作系统补丁等；
- 用户通过身份认证后仅仅获得访问部分互联网资源（受限资源）的权限，如病毒服务器、操作系统补丁更新服务器等；当用户通过安全认证后便可以访问更多的互联网资源（非受限资源）。

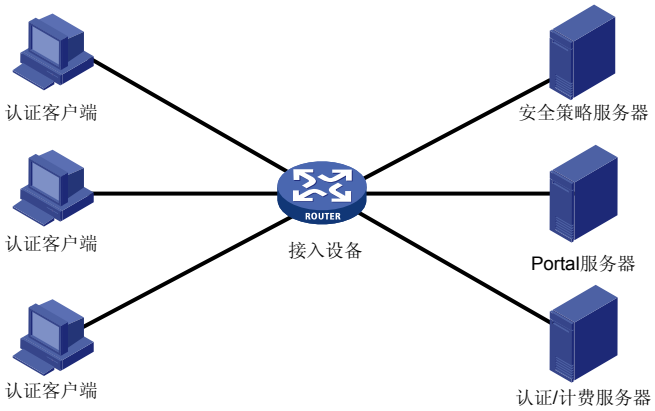
1.1.3 Portal 的系统组成

Portal的典型组网方式如 图 1-1所示，它由五个基本要素组成：认证客户端、接入设备、Portal服务器、认证/计费服务器和安全策略服务器。

说明

由于 Portal 服务器可以是接入设备之外的独立实体，也可以是存在于接入设备之内的内嵌实体，本文称之为“本地 Portal 服务器”，因此下文中除对本地支持的 Portal 服务器做特殊说明之外，其它所有 Portal 服务器均指独立的 Portal 服务器，请勿混淆。

图1-1 Portal 系统组成示意图



1. 认证客户端

安装于用户终端的客户端系统，为运行 HTTP/HTTPS 协议的浏览器或运行 Portal 客户端软件的主机。对接入终端的安全性检测是通过 Portal 客户端和安全策略服务器之间的信息交流完成的。

2. 接入设备

交换机、路由器等宽带接入设备的统称，主要有三方面的作用：

- 在认证之前，将认证网段内用户的所有 HTTP 请求都重定向到 Portal 服务器。
- 在认证过程中，与 Portal 服务器、安全策略服务器、认证/计费服务器交互，完成身份认证/安全认证/计费的功能。
- 在认证通过后，允许用户访问被管理员授权的互联网资源。

3. Portal 服务器

接收 Portal 客户端认证请求的服务器端系统，提供免费门户服务和基于 Web 认证的界面，与接入设备交互认证客户端的认证信息。

4. 认证/计费服务器

与接入设备进行交互，完成对用户的认证和计费。

5. 安全策略服务器

与 Portal 客户端、接入设备进行交互，完成对用户的安全认证，并对用户进行授权操作。

以上五个基本要素的交互过程为：

(1) 未认证用户访问网络时，在 IE 地址栏中输入一个互联网的地址，那么此 HTTP 请求在经过接入设备时会被重定向到 Portal 服务器的 Web 认证主页上；若需要使用 Portal 的扩展认证功能，则用户必须使用 Portal 客户端。

(2) 用户在认证主页/认证对话框中输入认证信息后提交，Portal 服务器会将用户的认证信息传递给接入设备；

(3) 然后接入设备再与认证/计费服务器通信进行认证和计费；

(4) 认证通过后，如果未对用户未采用安全策略，则接入设备会打开用户与互联网的通路，允许用户访问互联网；如果对用户采用了安全策略，则客户端、接入设备与安全策略服务器交互，对用户的安全检测通过之后，安全策略服务器根据用户的安全性授权用户访问非受限资源。

说明

- 由于 Portal 客户端以 IP 地址为身份标识，因此使用 Portal 业务时，在认证客户端、接入设备、Portal 服务器、认证/计费服务器之间不能有配置 NAT（Network Address Translation，网络地址转换）特性的设备，避免地址转换导致认证失败。
- 目前支持 Portal 认证的远端的认证/计费服务器为 RADIUS（Remote Authentication Dial-In User Service，远程认证拨号用户服务）服务器。
- 目前通过访问 IE 页面进行的 Portal 认证不能对用户实施安全策略检查，安全检查功能的实现需要使用 H3C iNode 客户端的配合。

1.1.4 Portal 的认证方式

Portal 的认证方式分为两种：非三层认证方式和三层认证方式。

1. 非三层认证方式

非三层认证方式分为两种：直接认证方式和二次地址分配认证方式。

- 直接认证方式

用户在认证前通过手工配置或 DHCP 直接获取一个公网 IP 地址，只能访问 Portal 服务器，以及设定的免费访问地址；认证通过后即可访问网络资源。认证流程相对二次地址较为简单。

- 二次地址分配认证方式

用户在认证前通过 DHCP 获取一个私网 IP 地址，只能访问 Portal 服务器，以及设定的免费访问地址；认证通过后，用户会申请到一个公网 IP 地址，即可访问网络资源。该认证方式解决了 IP 地址规划和分配问题，对未认证通过的用户不分配公网 IP 地址。例如运营商对于小区宽带用户只在访问小区外部资源时才分配公网 IP。

2. 三层认证方式

和直接认证方式基本相同，但是这种认证方式允许认证客户端和接入设备之间跨越三层转发设备。

3. 两者的区别

- 组网方式不同

三层认证方式的认证客户端和接入设备之间可以跨接三层转发设备；非三层认证方式则要求认证客户端和接入设备之间没有三层转发。

- 用户标识不同

由于三层认证可以跨接三层设备，而接入设备不会学习认证客户端的 MAC 地址信息，所以是以 IP 地址唯一标识用户；而非三层认证方式中的接入设备则可以学习到认证客户端的 MAC 地址，所以是以 IP 和 MAC 地址的组合来唯一标识用户。

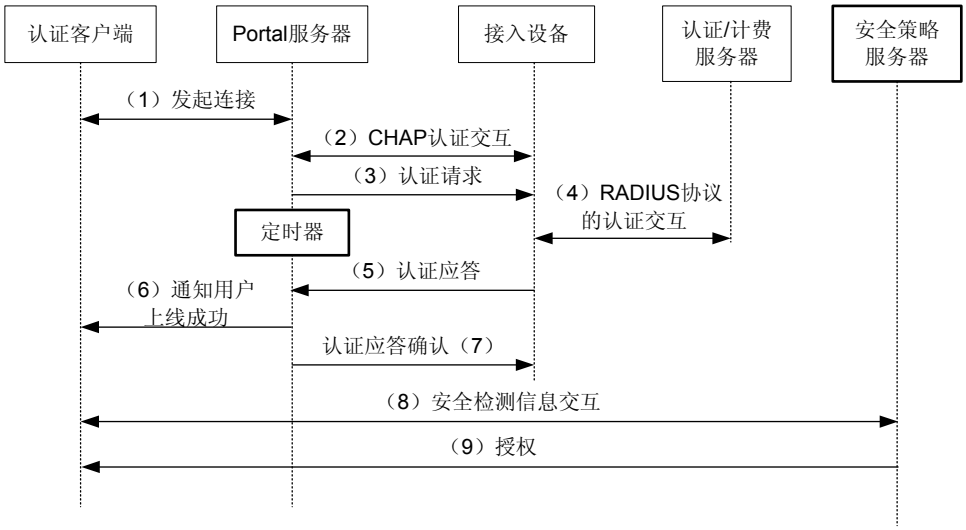
以上不同的组网方式和用户标识特点使得认证客户端的 MAC 地址不变、IP 地址改变时，在三层认证方式下会激发新的 Portal 认证，而在非三层认证方式下不会激发新的 Portal 认证。只有认证客户端的 MAC 地址和 IP 地址同时改变时，非三层认证方式下才会激发新的 Portal 认证。

1.1.5 Portal 认证过程

直接认证和三层 Portal 认证流程相同。二次地址分配认证流程因为有两次地址分配过程，所以其认证流程和另外两种认证方式有所不同。

1. 直接认证和三层 Portal 认证的流程

图1-2 直接认证/三层 Portal 认证流程图



直接认证/三层 Portal 认证流程：

(1) Portal 用户通过 HTTP 协议发起认证请求。HTTP 报文经过接入设备时，对于访问 Portal 服务器或设定的免费访问地址的 HTTP 报文，接入设备允许其通过；对于访问其它地址的 HTTP 报文，接入设备将其重定向到 Portal 服务器。Portal 服务器提供 Web 页面供用户输入用户名和密码来进行认证。

(2) Portal 服务器与接入设备之间进行 CHAP（Challenge Handshake Authentication Protocol，质询握手验证协议）认证交互。若采用 PAP（Password Authentication Protocol，密码验证协议）认证则直接进入下一步骤。

(3) Portal 服务器将用户输入的用户名和密码组装成认证请求报文发往接入设备，同时开启定时器等待认证应答报文。

(4) 接入设备与 RADIUS 服务器之间进行 RADIUS 协议报文的交互。

(5) 接入设备向 Portal 服务器发送认证应答报文。

(6) Portal 服务器向客户端发送认证通过报文，通知客户端认证（上线）成功。

(7) Portal 服务器向接入设备发送认证应答确认。

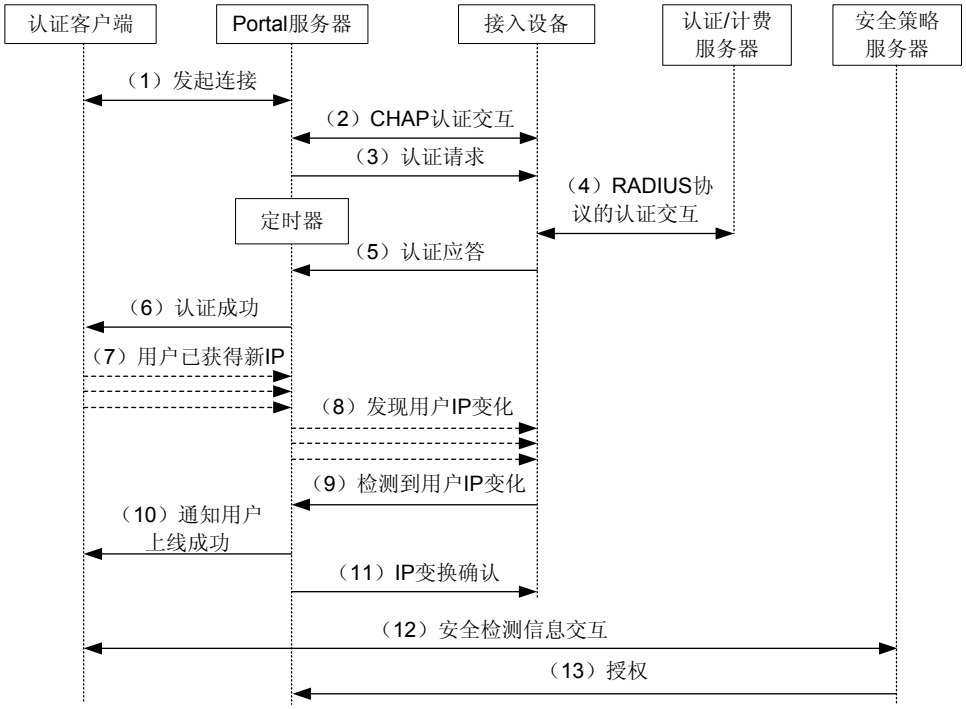
(8) 客户端和安全策略服务器之间进行安全信息交互。安全策略服务器检测接入终端的安全性是否合格，包括是否安装防病毒软件、是否更新病毒库、是否安装了非法软件、是否更新操作系统补丁等。

(9) 安全策略服务器根据用户的安全性授权用户访问非受限资源，授权信息保存到接入设备中，接入设备将使用该信息控制用户的访问。

步骤(8)、(9)为 Portal 认证扩展功能的交互过程。

2. 二次地址分配认证方式的流程

图1-3 二次地址分配认证方式流程图



二次地址分配认证流程：

(1)～(6)同直接/三层 Portal 认证中步骤（1）～（6）。

(7) 客户端收到认证通过报文后，通过 DHCP 请求获得新的公网 IP 地址，并通知 Portal 服务器用户已获得新 IP 地址。

(8) Portal 服务器通知接入设备客户端获得新公网 IP 地址。

(9) 接入设备通过检测 ARP 协议报文发现了用户 IP 变化，并通告 Portal 服务器已检测到用户 IP 变化。

(10) Portal 服务器通知客户端上线成功。

(11) Portal 服务器向接入设备发送 IP 变化确认报文。

(12) 客户端和安全策略服务器之间进行安全信息交互。安全策略服务器检测接入终端的安全性是否合格，包括是否安装防病毒软件、是否更新病毒库、是否安装了非法软件、是否更新操作系统补丁等。

(13) 安全策略服务器根据用户的安全性授权用户访问非受限资源，授权信息保存到接入设备中，接入设备将使用该信息控制用户的访问。

步骤(12)、(13)为 Portal 认证扩展功能的交互过程。

1.2 Portal 配置任务简介

表1-1 Portal 配置任务简介

配置任务	说明	详细配置
Portal 的基本配置	必选	1.3
配置免认证规则	可选	1.4

配置任务	说明	详细配置
配置认证网段	可选	1.5
配置强制用户下线	可选	1.6
配置 Portal 用户信息同步功能	可选	1.8
配置 Portal 用户在线探测功能	可选	1.9

1.3 Portal 的基本配置

1.3.1 配置准备

Portal 提供了一个用户身份认证和安全认证的实现方案，但是仅仅依靠 Portal 不足以实现该方案。接入设备的管理者需选择使用 RADIUS 认证方法，以配合 Portal 完成用户的身份认证。Portal 认证的配置前提：

- 使能 Portal 的接口已配置或者获取了合法的 IP 地址。
- Portal 服务器、RADIUS 服务器已安装并配置成功。
- 若采用二次地址分配认证方式，接入设备需启动 DHCP 中继的安全地址匹配检查功能，另外需要安装并配置好 DHCP 服务器。
- 如果通过远端 RADIUS 服务器进行认证，则需要在 RADIUS 服务器上配置相应的用户名和密码，然后在接入设备端进行 RADIUS 客户端的相关设置。RADIUS 客户端的具体配置请参见“安全分册”中的“AAA 配置”。
- 如果需要支持 Portal 的扩展功能，需要安装并配置安全策略服务器。同时保证在接入设备上的 ACL 配置和安全策略服务器上配置的受限资源 ACL 号、非受限资源 ACL 号对应。接入设备上的安全策略服务器配置请参见“安全分册”中的“AAA 配置”。



说明

- 安全策略服务器的安装和配置请参考《CAMS 综合访问管理服务器 EAD 安全策略组件 用户手册》。
- 受限资源 ACL、非受限资源 ACL 分别对应安全策略服务器中的隔离 ACL 与安全 ACL。
- 如果接入设备上的授权 ACL 配置被修改，则修改后的 ACL 不能对已经在线的 Portal 用户生效，只能对新上线的 Portal 用户有效。

1.3.2 配置 Portal

Portal 的基本配置包括配置 Portal 服务器和在接口上使能 Portal。配置 Portal 服务器时，需要根据实际情况指定 Portal 服务器的 IP 地址。

- 配置 Portal 服务器，需要指定外部 Portal 服务器的 IP 地址；

表1-2 配置 Portal

操作	命令	说明
进入系统视图	**system-view**	-

操作	命令	说明
配置 Portal 服务器	**portal server** *server-name* **ip** *ip-address* [**key** *key-string* | **port** *port-id* | **url** *url-string*] *	必选 缺省情况下，没有配置 Portal 服务器
进入接口视图	**interface** *interface-type interface-number*	-
在接口上使能 Portal	**portal server** *server-name* **method** { **direct** | **layer3** | **redhcp** }	必选 缺省情况下，没有使能 Portal

注意

- 设备向 Portal 服务器主动发送报文时使用的目的端口号必须与远程 Portal 服务器实际使用的端口号保持一致。
- 已配置的 Portal 服务器及其参数仅在该 Portal 服务器未被接口引用时才可以被删除或修改。
- 使能 Portal 的接口上所引用的 Portal 服务器必须已经存在。
- 对于跨三层设备支持 Portal 认证的应用只能配置 **layer3** 方式，但三层 Portal 认证方式不要求接入设备和 Portal 用户之间必需跨越三层设备。
- 在二次地址分配认证方式下，允许用户在未通过 Portal 认证时以公网地址向外发送报文，但相应的回应报文则受限制。

1.4 配置免认证规则

通过配置免认证规则（free-rule）可以让特定的用户不需要通过 Portal 认证即可访问外网特定资源，这是由免认证规则中配置的源信息以及目的信息决定的。

免认证规则的匹配项包括 IP 地址、MAC 地址、所连接设备的接口和 VLAN，只有符合免认证规则的用户报文才不会触发 Portal 认证，因此这些报文所属的用户才可以直接访问网络资源。

表1-3 配置免认证规则

操作	命令	说明
进入系统视图	**system-view**	-
配置 Portal 的免认证规则	**portal free-rule** *rule-number* { **destination** { **any** | **ip** { *ip-address* **mask** { *mask-length* | *netmask* } | **any** } } | **source** { **any** | [**interface** *interface-type interface-number* | **ip** { *ip-address* **mask** { *mask-length* | *mask* } | **any** } | **mac** *mac-address*] * } } *	必选

说明

- 相同内容的免认证规则不能重复配置，否则提示免认证规则已存在或重复。
- 无论接口上是否使能 portal，只能添加或者删除免认证规则，不能修改。

1.5 配置认证网段

通过配置认证网段实现只允许在认证网段范围内的用户报文才能触发 Portal 强制认证。如果非主动认证的用户报文既不满足免认证规则又不在认证网段内，则将被设备丢弃。

表1-4 配置认证网段

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置认证网段	**portal auth-network** *network-address* { *mask-length* | *mask* }	可选 缺省情况下，认证网段为 0.0.0.0/0，表示对任意源 IP 都进行认证



说明

认证网段配置仅对三层 Portal 认证有效。直接认证方式的认证网段为任意源 IP，二次地址分配方式的认证网段为由接口私网 IP 决定的私网网段。

1.6 配置强制用户下线

通过配置强制用户下线可以终止对指定 IP 地址用户的认证过程，或者将已经通过认证的指定 IP 地址的用户删除，对应的用户就被强制下线。

表1-5 配置强制用户下线

操作	命令	说明
进入系统视图	**system-view**	-
强制接入设备上的用户下线	**portal delete-user** { *ip-address* | **all** | **interface** *interface-type interface-number* }	必选

1.7 配置强制认证域

通过在指定接口上配置强制认证域，使得所有从该接口接入的 Portal 用户被强制使用指定的认证域来进行认证、授权和计费。即使 Portal 用户输入的用户名中携带的域名相同，接入设备的管理员也可以通过配置强制认证域对不同接口指定不同的认证域，从而增加了管理员部署 Portal 接入策略的灵活性。

表1-6 配置免认证规则

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置接口上的 Portal 强制认证域	**portal domain** *domain-name*	必选 缺省情况下，未定义 Portal 强制认证域

说明

从指定接口上接入的 Portal 用户将按照如下先后顺序选择认证域：接口上配置的强制 ISP 域-->用户名中指定的 ISP 域-->系统缺省的 ISP 域。关于缺省 ISP 域的相关介绍请参见“安全分册”中的“AAA 配置”。

1.8 配置 Portal 用户信息同步功能

为了解决接入设备与 Portal 服务器通信中断后，两者的 Portal 用户信息不一致问题，设备提供了一种 Portal 用户信息同步功能。该功能利用了 Portal 同步报文的发送及检测机制，具体实现如下：

(1) 由 Portal 服务器周期性地（周期为 Portal 服务器上指定的用户心跳间隔值）将在线用户信息通过用户同步报文发送给接入设备；

(2) 接入设备收到该用户同步报文后，将其中携带的用户信息与自己的用户信息进行对比，如果发现同步报文中有设备上不存在的用户信息，则将这些自己没有的用户信息反馈给 Portal 服务器，Portal 服务器将删除这些用户信息；如果发现接入设备上的某用户信息在连续 N（N 为 **retry** 参数的取值）个周期内，都未在该 Portal 服务器发送过来的用户同步报文中出现过，则认为 Portal 服务器上已不存在该用户，设备将强制该用户下线。

表1-7 配置 Portal 用户同步功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 Portal 用户同步功能	**portal server** *server-name* **user-sync** [**interval** *interval*] [**retry** *retries*]	必选 缺省情况下，未配置 Portal 用户同步功能 本命令中指定的 Portal 服务器必须已经存在 只有在指定的 Portal 服务器已经在接口上使能的情况下，本功能才能立即生效

说明

- 只有在支持 Portal 用户心跳功能（目前仅 iMC 的 Portal 服务器支持）的 Portal 服务器的配合下，本功能才有效。为了实现该功能，还需要在 Portal 服务器上选择支持用户心跳功能，且服务器上配置的用户心跳间隔要小于等于设备上配置的用户同步检测间隔。
- 对于设备上多余的用户信息，即在 N 个周期后被判定为 Portal 服务器上已不存在的用户信息，设备会在第 N + 1 个周期内的某时刻将其删除掉。

1.9 配置 Portal 用户在线探测功能

在接口上配置 Portal 用户在线探测功能后，设备会定期向从该接口上线的 Portal 在线用户发送探测报文（目前支持发送 ARP 请求），来确认该用户是否在线，以便及时发现异常离线用户。

- 若设备在指定的探测次数之内收到了该 Portal 用户的响应报文，则认为此用户在线，并通过继续发送探测报文，来持续确认该用户的在线状态。
- 若设备在指定的探测次数之后仍然未收到该 Portal 用户的响应报文，则认为此用户已经下线，则停止发送探测报文，并删除该用户。

表1-8 配置 Portal 用户在线探测功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
Portal 用户在线探测功能	**access-user detect type arp retransmit** *number* **interval** *interval*	必选 缺省情况下，未配置 Portal 用户在线探测功能


说明

探测报文的发送次数和发送间隔请根据网络的实际情况进行调整。

1.10 Portal 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 Portal 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 Portal 统计信息。

表1-9 Portal 显示和维护

操作	命令
显示接口上 Portal 的 ACL 信息	**display portal acl** { **all** | **dynamic** | **static** } **interface** *interface-type interface-number*
显示接口上 Portal 的连接统计信息	**display portal connection statistics** { **all** | **interface** *interface-type interface-number* }
显示 Portal 的免认证规则信息	**display portal free-rule** [*rule-number*]
显示指定接口的 Portal 配置信息	**display portal interface** *interface-type interface-number*
显示 Portal 服务器信息	**display portal server** [*server-name*]
显示接口上 Portal 服务器的统计信息	**display portal server statistics** { **all** | **interface** *interface-type interface-number* }
显示 TCP 仿冒统计信息	**display portal tcp-cheat statistics**
显示 Portal 用户的信息	**display portal user** { **all** | **interface** *interface-type interface-number* }
清除接口上 Portal 的连接统计信息	**reset portal connection statistics** {**all** | **interface** *interface-type interface-number* }

操作	命令
清除接口上 Portal 服务器的统计信息	**reset portal server statistics** { **all** | **interface** *interface-type interface-number* }
清除 TCP 仿冒统计信息	**reset portal tcp-cheat statistics**

1.11 Portal 典型配置举例

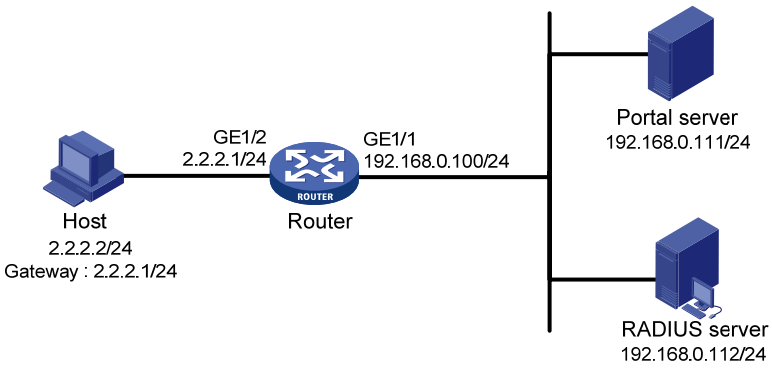
1.11.1 Portal 直接认证配置举例

1. 组网需求

- 用户主机与接入设备 Router 直接相连，采用直接方式的 Portal 认证。用户通过手工配置或 DHCP 获取的一个公网 IP 地址进行认证，在通过 Portal 认证前，只能访问 Portal 服务器；在通过 Portal 认证后，可以使用此 IP 地址访问非受限的互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-4 配置 Portal 直接认证组网图



3. 配置步骤

说明

按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<Router> system-view
[Router] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[Router-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[Router-radius-rs1] primary authentication 192.168.0.112
```

```
[Router-radius-rs1] primary accounting 192.168.0.112
[Router-radius-rs1] key authentication radius
[Router-radius-rs1] key accounting radius
```

配置发送给 RADIUS 服务器的用户名不携带 ISP 域名。

```
[Router-radius-rs1] user-name-format without-domain
[Router-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[Router] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[Router-isp-dm1] authentication portal radius-scheme rs1
[Router-isp-dm1] authorization portal radius-scheme rs1
[Router-isp-dm1] accounting portal radius-scheme rs1
[Router-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[Router] domain default enable dm1
```

(3) 配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
[Router] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

在与用户 Host 相连的接口上使能 Portal 认证。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] ip address 2.2.2.1 255.255.255.0
[Router-GigabitEthernet1/2] portal server newpt method direct
[Router-GigabitEthernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

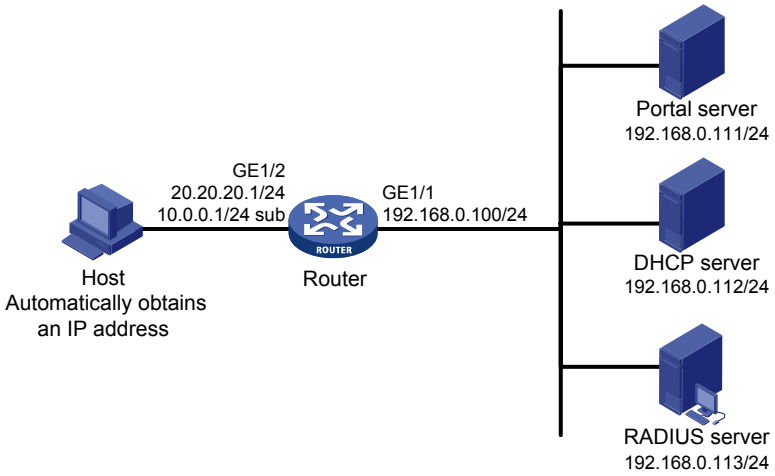
1.11.2 Portal 二次地址分配认证配置举例

1. 组网需求

- 用户主机与接入设备 Router 直接相连，采用二次地址分配方式的 Portal 认证。用户通过 DHCP 服务器获取 IP 地址，Portal 认证前使用分配的一个私网地址；通过 Portal 认证后，用户申请到一个公网地址，才可以访问非受限互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-5 配置 Portal 二次地址分配认证组网图



3. 配置步骤

说明

- Portal 二次地址分配认证方式应用中，DHCP 服务器上需创建公网地址池（20.20.20.0/24）及私网地址池(10.0.0.0/24),具体配置略。关于 DHCP 的详细配置请参见“IP 业务分册”中的“DHCP”。
- Portal 二次地址分配认证方式应用中，接入设备必须配置为 DHCP 中继（不能配置为 Server），且启动 Portal 的接口需要配置主 IP 地址（公网 IP）及从 IP 地址（私网 IP）。
- 按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<Router> system-view
[Router] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[Router-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[Router-radius-rs1] primary authentication 192.168.0.113
[Router-radius-rs1] primary accounting 192.168.0.113
[Router-radius-rs1] key authentication radius
[Router-radius-rs1] key accounting radius
```

配置发送给 RADIUS 服务器的用户名不携带 ISP 域名。

```
[Router-radius-rs1] user-name-format without-domain
[Router-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[Router] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[Router-isp-dm1] authentication portal radius-scheme rs1
[Router-isp-dm1] authorization portal radius-scheme rs1
[Router-isp-dm1] accounting portal radius-scheme rs1
[Router-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[Router] domain default enable dm1
```

(3) 配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
<Router> system-view
[Router] portal server newpt ip 192.168.0.111 key portal port 50100 url
http://192.168.0.111/portal
```

配置 DHCP 中继，并启动 DHCP 中继的安全地址匹配检查功能。

```
[Router] dhcp enable
[Router] dhcp relay server-group 0 ip 192.168.0.112
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] ip address 20.20.20.1 255.255.255.0
[Router-GigabitEthernet1/2] ip address 10.0.0.1 255.255.255.0 sub
[Router-GigabitEthernet1/2] dhcp select relay
[Router-GigabitEthernet1/2] dhcp relay server-select 0
[Router-GigabitEthernet1/2] dhcp relay address-check enable
```

在与用户 Host 相连的接口上使能 Portal 认证。

```
[Router-GigabitEthernet1/2] portal server newpt method redhcp
[Router-GigabitEthernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

1.11.3 三层 Portal 认证配置举例

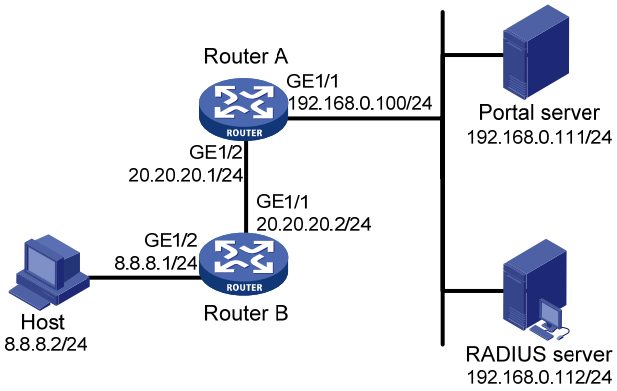
1. 组网需求

Router A 支持 Portal 认证功能。用户 Host 通过 Router B 接入到 Router A。

- 配置 Router A 采用三层 Portal 认证。用户在未通过 Portal 认证前，只能访问 Portal 服务器；用户通过 Portal 认证后，可以访问非受限互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-6 配置三层 Portal 认证组网图



3. 配置步骤

说明

按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router A 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<RouterA> system-view
[RouterA] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[RouterA-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[RouterA-radius-rs1] primary authentication 192.168.0.112
[RouterA-radius-rs1] primary accounting 192.168.0.112
[RouterA-radius-rs1] key authentication radius
[RouterA-radius-rs1] key accounting radius
```

配置发送给 RADIUS 服务器的用户名不携带 ISP 域名。

```
[RouterA-radius-rs1] user-name-format without-domain
[RouterA-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[RouterA] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[RouterA-isp-dm1] authentication portal radius-scheme rs1
[RouterA-isp-dm1] authorization portal radius-scheme rs1
[RouterA-isp-dm1] accounting portal radius-scheme rs1
[RouterA-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[Router] domain default enable dm1
```

(3)　配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
[RouterA] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

在与 Router B 相连的接口上使能 Portal 认证。

```
[RouterA] interface gigabitethernet 1/2
[RouterA-GigabitEthernet1/2] ip address 20.20.20.1 255.255.255.0
[RouterA-GigabitEthernet1/2] portal server newpt method layer3
[RouterA-GigabitEthernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
```

Router B 上需要配置到 192.168.0.0/24 网段的缺省路由，下一跳为 20.20.20.1，具体配置略。

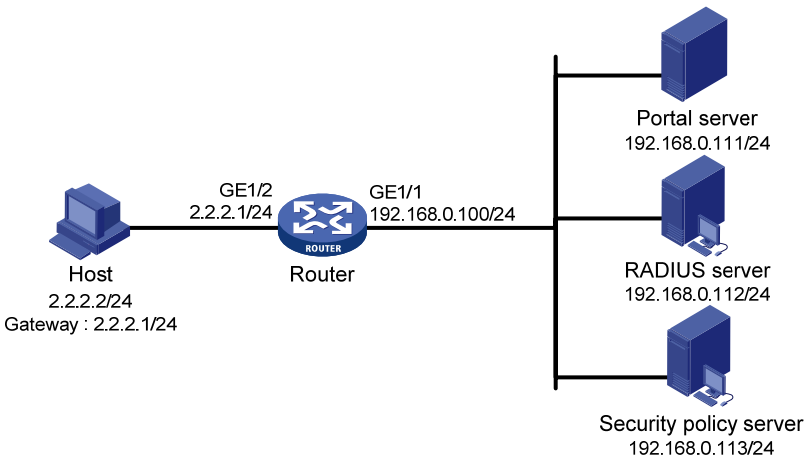
1.11.4 Portal 直接认证扩展功能配置举例

1. 组网需求

- 用户主机与接入设备 Router 直接相连，采用直接方式的 Portal 认证。用户通过手工配置或 DHCP 获取的一个公网 IP 地址进行认证，在通过身份认证而没有通过安全认证时可以访问 192.168.0.0/24 网段；在通过安全认证后，可以使用此 IP 地址访问非受限互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-7 配置 Portal 直接认证扩展功能组网图



3. 配置步骤

说明

按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<Router> system-view
[Router] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[Router-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[Router-radius-rs1] primary authentication 192.168.0.112
[Router-radius-rs1] primary accounting 192.168.0.112
[Router-radius-rs1] key accounting radius
[Router-radius-rs1] key authentication radius
[Router-radius-rs1] user-name-format without-domain
```

配置 RADIUS 方案的安全策略服务器 IP。

```
[Router-radius-rs1] security-policy-server 192.168.0.113
[Router-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[Router] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[Router-isp-dm1] authentication portal radius-scheme rs1
[Router-isp-dm1] authorization portal radius-scheme rs1
[Router-isp-dm1] accounting portal radius-scheme rs1
[Router-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[Router] domain default enable dm1
```

(3) 配置受限资源对应的 ACL 为 3000，非受限资源对应的 ACL 为 3001

说明

安全策略服务器上需要将 ACL 3000 和 ACL 3001 分别指定为隔离 ACL 和安全 ACL。

```
[Router] acl number 3000
[Router-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Router-acl-adv-3000] quit
[Router] acl number 3001
[Router-acl-adv-3001] rule permit ip
```

```
[Router-acl-adv-3001] quit
```

(4) 配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
[Router] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

在与用户 Host 相连的接口上使能 Portal 认证。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] ip address 2.2.2.1 255.255.255.0
[Router-GigabitEthernet1/2] portal server newpt method direct
[Router-GigabitEthernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

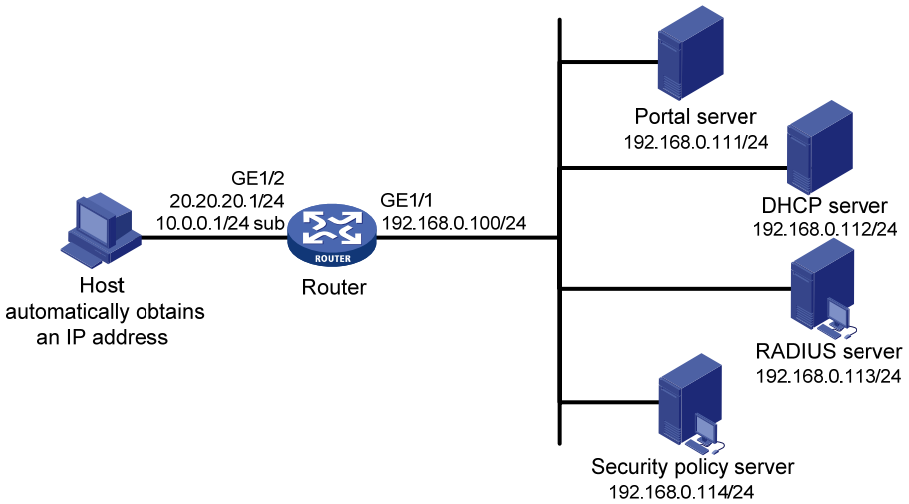
1.11.5 Portal 二次地址分配认证扩展功能配置举例

1. 组网需求

- 用户主机与接入设备 Router 直接相连，采用二次地址分配方式的 Portal 认证。用户通过 DHCP 服务器获取 IP 地址，Portal 认证前使用分配的一个私网地址；通过 Portal 认证后，用户申请到一个公网地址。
- 用户在通过身份认证而没有通过安全认证时可以访问 192.168.0.0/24 网段；在通过安全认证后，可以访问非受限互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-8 配置 Portal 二次地址分配认证扩展功能组网图



3. 配置步骤

说明

- Portal 二次地址分配认证方式应用中，DHCP 服务器上需创建公网地址池（20.20.20.0/24）及私网地址池(10.0.0.0/24),具体配置略。关于 DHCP 的详细配置请参见“IP 业务分册”中的“DHCP”。
- Portal 二次地址分配认证方式应用中，接入设备必须配置为 DHCP 中继（不能配置为 Server），且启动 Portal 的接口需要配置主 IP 地址（公网 IP）及从 IP 地址（私网 IP）。
- 按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<Router> system-view
[Router] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[Router-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[Router-radius-rs1] primary authentication 192.168.0.113
[Router-radius-rs1] primary accounting 192.168.0.113
[Router-radius-rs1] key authentication radius
[Router-radius-rs1] key accounting radius
[Router-radius-rs1] user-name-format without-domain
```

配置 RADIUS 方案的安全策略服务器 IP。

```
[Router-radius-rs1] security-policy-server 192.168.0.114
[Router-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[Router] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[Router-isp-dm1] authentication portal radius-scheme rs1
[Router-isp-dm1] authorization portal radius-scheme rs1
[Router-isp-dm1] accounting portal radius-scheme rs1
[Router-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[Router] domain default enable dm1
```

(3) 配置受限资源对应的 ACL 为 3000，非受限资源对应的 ACL 为 3001

说明

安全策略服务器上需要将 ACL 3000 和 ACL 3001 分别指定为隔离 ACL 和安全 ACL。

```
[Router] acl number 3000
```

```
[Router-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Router-acl-adv-3000] quit
[Router] acl number 3001
[Router-acl-adv-3001] rule permit ip
[Router-acl-adv-3001] quit
```

(4) 配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
<Router> system-view
[Router] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

配置 DHCP 中继，并启动 DHCP 中继的安全地址匹配检查功能。

```
[Router] dhcp enable
[Router] dhcp relay server-group 0 ip 192.168.0.112
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] ip address 20.20.20.1 255.255.255.0
[Router-GigabitEthernet1/2] ip address 10.0.0.1 255.255.255.0 sub
[Router-GigabitEthernet1/2] dhcp select relay
[Router-GigabitEthernet1/2] dhcp relay server-select 0
[Router-GigabitEthernet1/2] dhcp relay address-check enable
```

在与用户 Host 相连的接口上使能 Portal 认证。

```
[Router-GigabitEthernet1/2] portal server newpt method redhcp
[Router-Ethernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[Router] interface gigabitethernet 1/1
[Router-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[Router-GigabitEthernet1/1] quit
```

1.11.6 三层 Portal 认证扩展功能配置举例

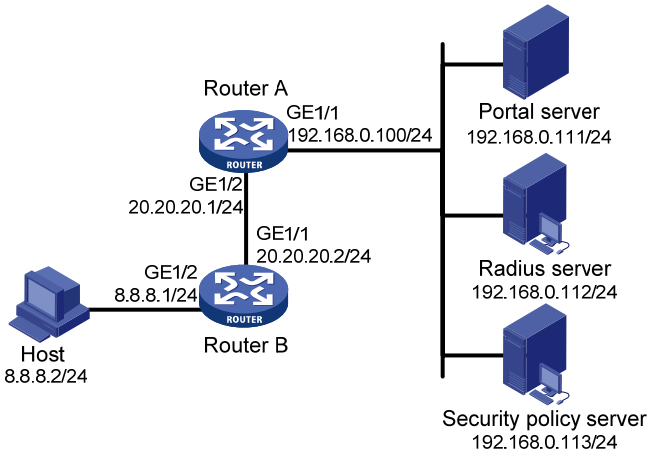
1. 组网需求

Router A 支持 Portal 认证功能。用户 Host 通过 Router B 接入到 Router A。

- 配置 Router A 采用三层 Portal 认证。用户在通过身份认证而没有通过安全认证时可以访问 192.168.0.0/24 网段；在通过安全认证后，可以访问非受限互联网资源。
- 采用 RADIUS 服务器作为认证/计费服务器。

2. 组网图

图1-9 配置三层 Portal 认证扩展功能组网图



3. 配置步骤

说明

按照组网图配置设备各接口的 IP 地址，保证在不启动 Portal 的前提下各设备之间的路由可达。

在 Router A 上进行以下配置。

(1) 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<RouterA> system-view
[RouterA] radius scheme rs1
```

配置 RADIUS 方案的服务器类型为 extended。

```
[RouterA-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[RouterA-radius-rs1] primary authentication 192.168.0.112
[RouterA-radius-rs1] primary accounting 192.168.0.112
[RouterA-radius-rs1] key authentication radius
[RouterA-radius-rs1] key accounting radius
[RouterA-radius-rs1] user-name-format without-domain
```

配置 RADIUS 方案的安全策略服务器 IP。

```
[RouterA-radius-rs1] security-policy-server 192.168.0.113
[RouterA-radius-rs1] quit
```

(2) 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[RouterA] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[RouterA-isp-dm1] authentication portal radius-scheme rs1
[RouterA-isp-dm1] authorization portal radius-scheme rs1
```

```
[RouterA-isp-dm1] accounting portal radius-scheme rs1
[RouterA-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。

```
[RouterA] domain default enable dm1
```

(3) 配置受限资源对应的 ACL 为 3000，非受限资源对应的 ACL 为 3001



说明

安全策略服务器上需要将 ACL 3000 和 ACL 3001 分别指定为隔离 ACL 和安全 ACL。

```
[RouterA] acl number 3000
[RouterA-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[RouterA-acl-adv-3000] quit
[RouterA] acl number 3001
[RouterA-acl-adv-3001] rule permit ip
[RouterA-acl-adv-3001] quit
```

(4) 配置 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111/portal。

```
<RouterA> system-view
[RouterA] portal server newpt ip 192.168.0.111 key portal port 50100 url
http://192.168.0.111/portal
```

在与 Router B 相连的接口上使能 Portal 认证。

```
[RouterA] interface gigabitethernet 1/2
[RouterA-GigabitEthernet1/2] ip address 20.20.20.1 255.255.255.0
[RouterA-GigabitEthernet1/2] portal server newpt method layer3
[RouterA-GigabitEthernet1/2] quit
```

配置与服务器通信的接口 IP 地址。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 192.168.0.100 255.255.255.0
[RouterA-GigabitEthernet1/1] quit
```

Router B 上需要配置到 192.168.0.0/24 网段的缺省路由，下一跳为 20.20.20.1，具体配置略。

1.11.7 Portal 服务器探测和用户同步功能配置举例

1. 组网需求

用户主机与接入设备 Router 直接相连，通过 Portal 认证接入网络，并采用 RADIUS 服务器作为认证/计费服务器。

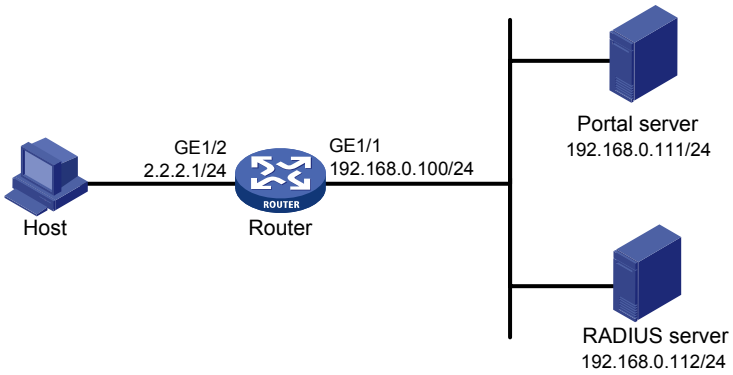
具体要求如下：

- 用户通过手工配置或 DHCP 获取的一个公网 IP 地址进行认证，在通过 Portal 认证前，只能访问 Portal 服务器；在通过 Portal 认证后，可以使用此 IP 地址访问非受限的互联网资源。

- 接入设备能够探测到 Portal 服务器是否可达，并输出可达状态变化的 Trap 信息，在服务器不可达时（例如，网络连接中断、网络设备故障或服务器无法正常提供服务等情况），取消 Portal 认证，使得用户仍然可以正常访问网络。
- 接入设备能够与服务器定期进行用户信息的同步。

2. 组网图

图1-10 Portal 服务器探测和用户同步功能配置组网图



3. 配置思路

(1) 配置 Portal 服务器，并启动逃生心跳功能和用户心跳功能；

(2) 配置 RADIUS 服务器，实现正常的认证及计费功能；

(3) 接入设备通过接口 GigabitEthernet1/2 与用户主机直接相连，在该接口上配置直接方式的 Portal 认证；

(4) 接入设备上配置 Portal 服务器探测功能，在与 Portal 服务器的逃生心跳功能的配合下，对 Portal 服务器的可达状态进行探测；

(5) 接入设备上配置 Portal 用户同步功能，在与 Portal 服务器的用户心跳功能的配合下，与 Portal 服务器上的用户信息进行同步。

4. 配置步骤

说明

- 按照组网图配置设备各接口的 IP 地址，保证启动 Portal 之前各主机、服务器和设备之间的路由可达。
- 完成 RADIUS 服务器上的配置，保证用户的认证/计费功能正常运行。

(1) 配置 Portal server（iMC）

说明

下面以 iMC 为例（使用 iMC 版本为：iMC PLAT 3.20-R2602P06、iMC UAM 3.60-B6202），说明 Portal server 的相关配置。

配置 Portal 服务器。

登录进入 iMC 管理平台，选择“业务”页签，单击导航树中的[Portal 服务器管理/服务器配置]菜单项，进入服务器配置页面。

- 配置逃生心跳间隔时长及用户心跳间隔时长；
- 输入 Portal 认证主页地址，形式为 http://ip:port/portal，其中 ip 和 port 在安装 iMC UAM 的时候确定，port 一般使用缺省值 8080 即可。

图1-11 Portal 服务器配置页面



配置 IP 地址组。

单击导航树中的[Portal 服务器管理/IP 地址组配置]菜单项，进入 Portal IP 地址组配置页面，在该页面中单击<增加>按钮，进入增加 IP 地址组配置页面。

- 填写 IP 地址组名、起始地址和终止地址。用户主机 IP 地址必须包含在该 IP 地址组范围内。

图1-12 增加 IP 地址组配置页面

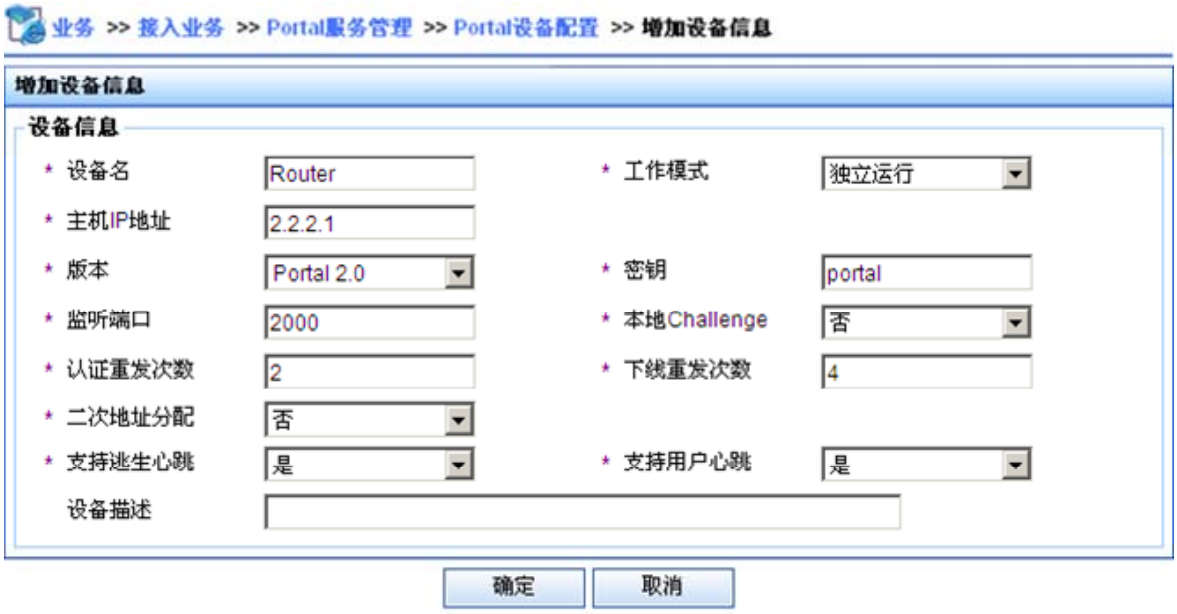


增加 Portal 设备。

单击导航树中的[Portal 服务器管理/设备配置]菜单项，进入 Portal 设备配置页面，在该页面中单击<增加>按钮，进入增加设备信息配置页面。

- 填写设备名；
- 输入 IP 地址为与接入用户相连的设备接口 IP；
- 输入密钥，与接入设备 Router 上的配置保持一致；
- 选择是否进行二次地址分配，本例中为直接认证，因此为否；
- 选择支持逃生心跳供和用户心跳功能。

图1-13 增加设备信息配置页面



Portal 设备关联 IP 地址组

在 Portal 设备配置页面中的设备信息列表中，点击 Router 设备的<端口组信息管理>链接，进入端口组信息配置页面。

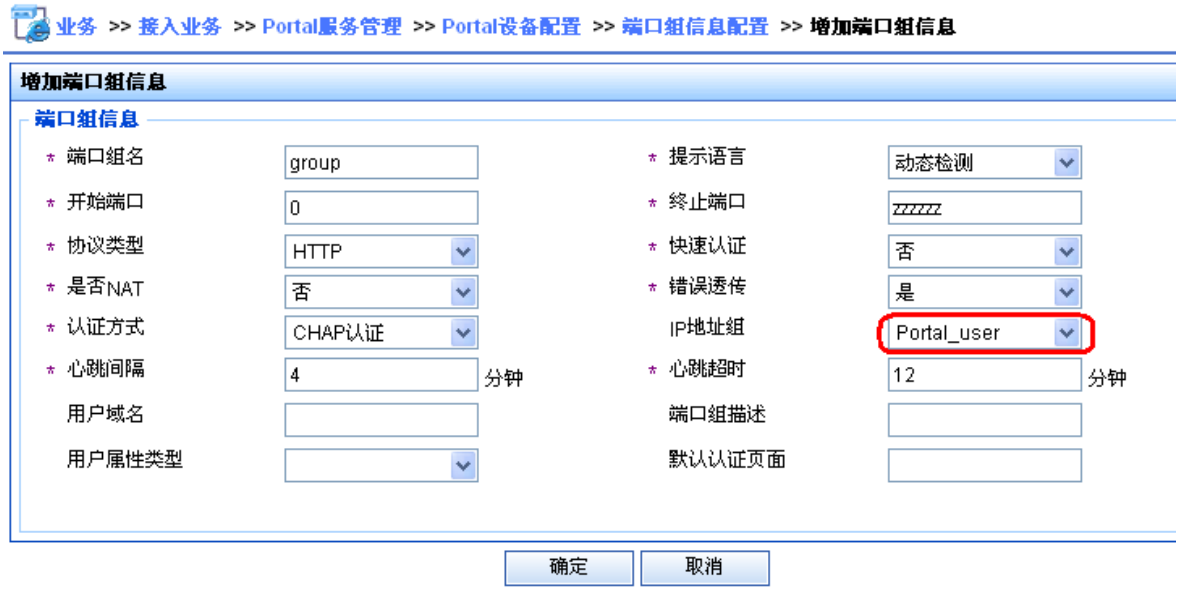
图1-14 设备信息列表



在端口组信息配置页面中点击<增加>按钮，进入增加端口组信息配置页面。

- 填写端口组名；
- 选择 IP 地址组，用户接入网络时使用的 IP 地址必须属于所选的 IP 地址组；
- 其它参数采用缺省值。

图1-15 增加端口组信息配置页面



最后单击导航树中的[业务参数配置/系统配置手工生效]菜单项，使以上 Portal 服务器配置生效。

(2) 配置 Router

- 配置 RADIUS 方案

创建名字为 rs1 的 RADIUS 方案并进入该方案视图。

```
<Router> system-view
[Router] radius scheme rs1
```

配置 RADIUS 方案的服务器类型。使用 iMC 服务器时，RADIUS 服务器类型应选择 **extended**。

```
[Router-radius-rs1] server-type extended
```

配置 RADIUS 方案的主认证和主计费服务器及其通信密钥。

```
[Router-radius-rs1] primary authentication 192.168.0.112
[Router-radius-rs1] primary accounting 192.168.0.112
[Router-radius-rs1] key authentication radius
[Router-radius-rs1] key accounting radius
```

配置发送给 RADIUS 服务器的用户名不携带 ISP 域名。

```
[Router-radius-rs1] user-name-format without-domain
[Router-radius-rs1] quit
```

- 配置认证域

创建并进入名字为 dm1 的 ISP 域。

```
[Router] domain dm1
```

配置 ISP 域的 RADIUS 方案 rs1。

```
[Router-isp-dm1] authentication portal radius-scheme rs1
[Router-isp-dm1] authorization portal radius-scheme rs1
[Router-isp-dm1] accounting portal radius-scheme rs1
[Router-isp-dm1] quit
```

配置系统缺省的 ISP 域 dm1，所有接入用户共用此缺省域的认证和计费方式。若用户登录时输入的用户名未携带 ISP 域名，则使用缺省域下的认证方案。

```
[Router] domain default enable dm1
```

- 使能 Portal 认证

配置 Portal 服务器：名称为 newpt，IP 地址为 192.168.0.111，密钥为 portal，端口为 50100，URL 为 http://192.168.0.111:8080/portal。

```
[Router] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111:8080/portal
```

在与用户 Host 相连的接口上使能 Portal 认证。

```
[Router] interface gigabitethernet 1/2
[Router-GigabitEthernet1/2] portal server newpt method direct
[Router-GigabitEthernet1/2] quit
```

- 配置 Portal 服务器探测功能

配置对 Portal 服务器 pts 的探测功能：探测方式为探测 Portal 心跳报文，每次探测间隔时间为 40 秒，若连续二次探测均失败，则发送服务器不可达的日志信息，并打开网络限制，允许未认证用户访问网络。

```
[Router] portal server newpt server-detect method portal-heartbeat action log permit-all interval 40 retry 2
```

说明

此处 interval 与 retry 的乘积应该大于等于 Portal 服务器的逃生心跳间隔时长，且推荐 interval 取值大于 Portal 服务器的逃生心跳间隔时长。

- 配置 Portal 用户信息同步功能

配置对 Portal 服务器 pts 的 Portal 用户同步功能，检测用户同步报文的时间间隔为 600 秒，如果设备中的某用户信息在连续两个探测周期内都未在该 Portal 服务器发送的同步报文中出现，设备将强制该用户下线。

```
[Router] portal server newpt user-sync interval 600 retry 2
```

说明

此处 interval 与 retry 的乘积应该大于等于 Portal 服务器上的用户心跳间隔时长，且推荐 interval 取值大于 Portal 服务器的逃生心跳间隔时长。

5. 验证配置结果

以上配置完成后，可以通过执行以下命令查看 Portal 服务器的状态。

```
<Router> display portal server newPt
 Portal server:
  1)newPt:
      IP    : 192.168.0.111
      Key   : portal
      Port  : 50100
      URL   : http://192.168.0.111:8080/portal
    Status  : Up
```

1.12 常见配置错误举例

1.12.1 接入设备和 Portal 服务器上的密钥不一致

1. 故障现象

用户被强制去访问 Portal 服务器时没有弹出 Portal 认证页面，也没有错误提示，登录的 Portal 认证服务器 Web 页面为空白。

2. 故障分析

接入设备上配置的 Portal 密钥和 Portal 服务器上配置的密钥不一致，导致 Portal 服务器报文验证出错，Portal 服务器拒绝弹出认证页面。

3. 处理过程

使用 **display portal server** 命令查看接入设备上配置的 Portal 服务器密钥，并在系统视图中使用 **portal server** 命令修改密钥，或者在 Portal 服务器上查看对应接入设备的密钥并修改密钥，直至两者的密钥设置一致。

1.12.2 接入设备上服务器端口配置错误

1. 故障现象

用户通过 Portal 认证后，在接入设备上使用 **portal delete-user** 命令强制用户下线失败，但是使用客户端的“断开”属性可以正常下线。

2. 故障分析

在 Portal 上使用 **portal delete-user** 命令强制用户下线时，由接入设备主动发送下线请求报文到 Portal 服务器，Portal 服务器默认的报文监听端口为 50100，但是因为接入设备上配置的服务器监听端口错误（不是 50100），即其发送的下线请求报文的目的端口和 Portal 服务器真正的监听端口不一致，故 Portal 服务器无法收到下线请求报文，Portal 服务器上的用户无法下线。

当使用客户端的“断开”属性让用户下线时，由 Portal 服务器主动向接入设备发送下线请求，其源端口为 50100，接入设备的下线应答报文的目的端口使用请求报文的源端口，避免了其配置上的错误，使得 Portal 服务器可以收到下线应答报文，从而 Portal 服务器上的用户成功下线。

3. 处理过程

使用 **display portal server** 命令查看接入设备对应服务器的端口，并在系统视图中使用 **portal server** 命令修改服务器的端口，使其和 Portal 服务器上的监听端口一致。

目 录

1 公钥管理配置 1-1
1.1 公共密钥算法简介 1-1
1.1.1 基本概念 1-1
1.1.2 数据加解密算法分类 1-1
1.1.3 非对称密钥算法的应用 1-1
1.2 配置本地的非对称密钥对 1-2
1.2.1 生成非对称密钥对 1-2
1.2.2 导出非对称密钥对的主机公钥 1-2
1.2.3 销毁非对称密钥对 1-3
1.3 配置远端主机的公钥 1-3
1.4 公钥显示和维护 1-4
1.5 公钥管理典型配置举例 1-4
1.5.1 手工配置远端主机的公钥 1-4
1.5.2 从公钥文件中导入远端主机的公钥 1-6

1 公钥管理配置

1.1 公共密钥算法简介

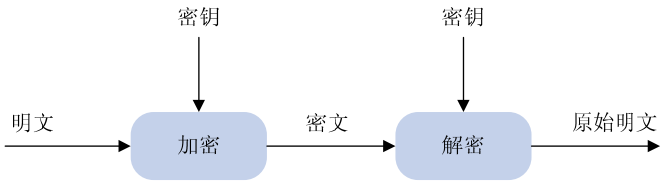
1.1.1 基本概念

- 算法：加密和解密过程中使用的一组变换规则。
- 明文：未加密的信息。
- 密文：加密后的信息。
- 密钥：一组特定的字符串，是控制明文和密文变换的唯一参数，起到“钥匙”的作用。加密和解密都是在密钥的参与下进行的。

1.1.2 数据加解密算法分类

如 图 1-1所示，为了保证数据在网络中安全传输，不被攻击者非法窃听和恶意篡改，发送方在发送数据之前需要对数据进行加密处理，即通过一定的算法，利用密钥将明文数据变换为密文；接收方接收到密文数据后，对其进行解密处理，即通过一定的算法，利用密钥将密文数据恢复为明文，以获得原始数据。

图1-1 加密和解密变换关系



根据加密解密过程中使用的密钥，可以将数据加解密算法分为：

- 对称密钥算法：数据加密和解密时使用相同的密钥和相同的算法。常用的对称密钥算法包括 AES、DES 等。
- 非对称密钥算法：又称为公共密钥算法。数据加密和解密时使用不同的密钥，一个是公开的公钥，一个是由用户秘密保存的私钥，从公钥很难推出私钥。通过公钥（或私钥）加密后的数据只能利用对应的私钥（或公钥）进行解密。

1.1.3 非对称密钥算法的应用

非对称密钥算法可以用于加密和数字签名：

- 加密：发送者利用接收者的公钥对报文进行加密，拥有对应私钥的合法接收者使用私钥对数据进行解密，从而可以保证数据的机密性。
- 数字签名：发送者利用自己的私钥对数据进行加密，接收者利用发送者的公钥对数据进行解密。由于通过私钥加密后的数据只能利用对应的公钥进行解密，因此根据解密是否成功，就可以判断发送者的身份，如同发送者对数据进行了“签名”。例如，用户 1 使用自己的私钥对数据进行签名后，发给用户 2，用户 2 利用用户 1 的公钥验证签名，如果签名是正确的，那么就能够确认该数据来源于用户 1。

非对称密钥算法包括 RSA（Rivest Shamir and Adleman）和 DSA（Digital Signature Algorithm，数字签名算法）。RSA 既可以用于加密，又可以用于签名，而 DSA 只用于签名。



说明

由于非对称密钥算法的计算量较大，速度较慢，因此非对称密钥算法经常用于数字签名，来验证对端的身份。对称密钥算法经常用于加密数据，以保证数据传输的安全性。

1.2 配置本地的非对称密钥对

用户通过命令行可以在本地设备上生成本地非对称密钥对、导出本地非对称密钥对的主机公钥部分和销毁本地非对称密钥对。

1.2.1 生成非对称密钥对

表1-1 生成非对称密钥对

操作	命令	说明
进入系统视图	**system-view**	-
生成本地 DSA 或 RSA 密钥对	**public-key local create { dsa | rsa }**	必选 缺省情况下，没有生成 RSA、DSA 密钥对



说明

- 执行 **public-key local create** 命令后，生成的密钥对将保存在设备中，设备重启后不会丢失只需执行一遍，设备重启后不必再次执行。
- 通过 **public-key local create rsa** 命令生成 RSA 密钥对时，将同时生成两个密钥对——服务器密钥对和主机密钥对，二者分别包括一个公钥和一个私钥。
- RSA 服务器密钥和主机密钥模数的最小长度为 512 位，最大长度为 2048 位。生成 RSA 密钥对时会提示用户输入密钥模数的长度，建议密钥模数的长度大于或等于 768 位，以提高安全性保证。在 SSH2.0 中，有的客户端要求服务器端生成的密钥长度必须大于或等于 768 位。
- 通过 **public-key local create dsa** 命令生成 DSA 密钥对时，只生成一个主机密钥对。
- DSA 主机密钥模数的最小长度为 512 位，最大长度为 2048 位。生成 DSA 密钥对时会提示用户输入密钥模数的长度，建议密钥模数的长度大于或等于 768 位，以提高安全性保证。。

1.2.2 导出非对称密钥对的主机公钥

对于已经生成的 RSA 或 DSA 密钥对，可以根据指定格式在屏幕上显示主机公钥或导出主机公钥到指定文件，从而为在远端配置 RSA 或 DSA 主机公钥作准备。

表1-2 导出 RSA 或 DSA 密钥对的主机公钥

操作	命令	说明
进入系统视图	**system-view**	-
根据指定格式在屏幕上显示本地 RSA 主机公钥或导出本地 RSA 主机公钥到指定文件	**public-key local export rsa** { **openssh** | **ssh1** | **ssh2** } [*filename*]	根据导出的公钥类型选择相应的命令
根据指定格式在屏幕上显示本地 DSA 主机公钥或导出本地 DSA 主机公钥到指定文件	**public-key local export dsa** { **openssh** | **ssh2** } [*filename*]	

1.2.3 销毁非对称密钥对

在非对称密钥对过期或泄漏的情况下，用户需要销毁旧的密钥对，并生成新的密钥对。通过本配置可以实现销毁本地非对称密钥对。

表1-3 销毁非对称密钥对

操作	命令	说明
进入系统视图	**system-view**	-
销毁非对称密钥对	**public-key local destroy** { **dsa** | **rsa** }	必选

1.3 配置远端主机的公钥

为了实现对远端主机的身份验证，需要在本地设备上配置远端主机的 RSA 或 DSA 公钥。

可以通过手工配置和从公钥文件中导入两种方式来配置远端主机的公钥：

- 手工配置远端主机的公钥时，可以采用手工输入或拷贝粘贴的方式将远端主机的公钥配置到本地。这种方式要求手工输入或拷贝粘贴的主机公钥必须是未经转换的 DER（Distinguished Encoding Rules，特异编码规则）公钥编码格式。
- 从公钥文件中导入远端主机的公钥时，系统会自动远端主机的公钥文件转换为 PKCS（Public Key Cryptography Standards，公共密钥加密标准）编码形式，并实现远端主机公钥的配置。这种方式需要远端主机事先将公钥文件通过 FTP/TFTP 以二进制（binary）方式上传到本地。

注意

- 建议选用从公钥文件导入的方式配置远端主机的公钥。
- 设备上最多可以配置 20 个远端主机的公钥。

表1-4 手工配置远端主机的公钥

操作	命令	说明
进入系统视图	**system-view**	-
进入公共密钥视图	**public-key peer** *keyname*	-
进入公共密钥编辑视图	**public-key-code begin**	-

操作	命令	说明
配置远端主机的公钥	直接输入公钥内容	必选 在输入公钥内容时，字符之间可以有空格，也可以按回车键继续输入数据
退回公共密钥视图	**public-key-code end**	- 退出视图时，系统自动保存配置的公钥密钥
退回系统视图	**peer-public-key end**	-

表1-5 从公钥文件中导入远端主机的公钥

操作	命令	说明
进入系统视图	**system-view**	-
从公钥文件中导入远端主机的公钥	**public-key peer** *keyname* **import sshkey** *filename*	必选

1.4 公钥显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后公钥的信息，通过查看显示信息验证配置的效果。

表1-6 公钥显示和维护

操作	命令
显示本地密钥对中的公钥部分	**display public-key local** { **dsa** | **rsa** } **public**
显示保存在本地的远端主机的公钥信息	**display public-key peer** [**brief** | **name** *publickey-name*]

1.5 公钥管理典型配置举例

1.5.1 手工配置远端主机的公钥

1. 组网需求

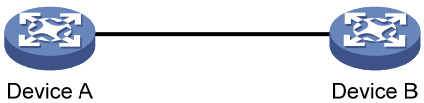
如 图 1-2所示，Device A访问Device B时，Device B要对Device A进行身份验证，为此，需要事先在Device B上配置Device A的公钥。

本例中：

- 采用的非对称密钥算法为 RSA。
- 采用手工方式在 Device B 上配置 Device A 的主机公钥。

2. 组网图

图1-2 手工配置远端主机的公钥组网图



3. 配置步骤

(1) 配置 Device A

在 Device A 上生成 RSA 密钥对。

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits of the modulus[default = 1024]:
Generating Keys...
++++++
++++++
++++++++
++++++++
```

显示生成的 RSA 密钥对的公钥部分。

```
[DeviceA] display public-key local rsa public

=====================================================
Time of Key pair created: 09:50:06  2007/08/07
Key name: HOST_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001

=====================================================
Time of Key pair created: 09:50:07  2007/08/07
Key name: SERVER_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
307C300D06092A864886F70D0101010500036B0030680261009990 89E7AEE9802002D9EB2D0433B87BB6158E
35000AFB3FF310E42F109829D65BF70F7712507BE1A3E0BC5C2C03FAAF00DFDDC63D004B4490DACBA3CFA9E8
4B9151BDC7EECE1C8770D961557D192DE2B36CAF9974B7B293363BB372771C2C1F0203010001
```

(2) 配置 Device B

在 Device B 上配置 Device A 的主机公钥：在公共密钥编辑视图输入 Device A 的主机公钥，即在 Device A 上通过 **display public-key local dsa public** 命令显示的主机公钥 HOST_KEY 内容。

```
<DeviceB> system-view
[DeviceB] public-key peer devicea
Public key view: return to System View with "peer-public-key end".
[DeviceB-pkey-public-key] public-key-code begin
```

```
Public key code view: return to last view with "public-key-code end".
[DeviceB-pkey-key-code]30819F300D06092A864886F70D010101050003818D0030818902818100D90003F
A95F5A44A2A2CD3F814F9854C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A
9AB16C9E766BD995C669A784AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326
470034DC078B2BAA3BC3BCA80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
[DeviceB-pkey-key-code] public-key-code end
[DeviceB-pkey-public-key] peer-public-key end
```

显示 Device B 上保存的 Device A 的主机公钥信息。

```
[DeviceB] display public-key peer name devicea

=====================================
  Key Name  : devicea
  Key Type  : RSA
  Key Module: 1024
=====================================
Key Code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
```

1.5.2 从公钥文件中导入远端主机的公钥

1. 组网需求

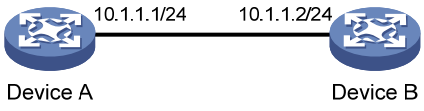
如 图 1-3所示，Device A访问Device B时，Device B要对Device A进行身份验证，为此，需要事先在Device B上配置Device A的公钥。

本例中：

- 采用的非对称密钥算法为 RSA。
- 采用从公钥文件中导入的方式在 Device B 上配置 Device A 的主机公钥。

2. 组网图

图1-3 从公钥文件中导入远端主机的公钥组网图



3. 配置步骤

(1) 在 Device A 上生成密钥对，并导出公钥

在 Device A 上生成 RSA 密钥对。

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
```

```
Press CTRL+C to abort.
Input the bits of the modulus[default = 1024]:
Generating Keys...
++++++
++++++
++++++++
++++++++
```

显示生成的 RSA 密钥对的公钥部分。

```
[DeviceA] display public-key local rsa public

=====================================================
Time of Key pair created: 09:50:06  2007/08/07
Key name: HOST_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001

=====================================================
Time of Key pair created: 09:50:07  2007/08/07
Key name: SERVER_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
307C300D06092A864886F70D0101010500036B0030680261009990 89E7AEE9802002D9EB2D0433B87BB6158E
35000AFB3FF310E42F109829D65BF70F7712507BE1A3E0BC5C2C03FAAF00DFDDC63D004B4490DACBA3CFA9E8
4B9151BDC7EECE1C8770D961557D192DE2B36CAF9974B7B293363BB372771C2C1F0203010001
```

将生成的 RSA 主机公钥导出到指定文件 devicea.pub 中。

```
[DeviceA] public-key local export rsa ssh2 devicea.pub
[DeviceA] quit
```

(2) 在 Device B 上启动 FTP 服务器功能

启动 FTP 服务器功能，并创建 FTP 用户，用户名为 ftp，密码为 123。

```
<DeviceB> system-view
[DeviceB] ftp server enable
[DeviceB] local-user ftp
[DeviceB-luser-ftp] password simple 123
[DeviceB-luser-ftp] service-type ftp
[DeviceB-luser-ftp] authorization-attribute level 3
[DeviceB-luser-ftp] quit
```

(3) Device A 上传公钥文件

将公钥文件 devicea.pub 通过 FTP 方式上传到 Device B。

```
<DeviceA> ftp 10.1.1.2
```

```
Trying 10.1.1.2 ...
Press CTRL+K to abort
Connected to 10.1.1.2.
220 FTP service ready.
User(10.1.1.2:(none)):ftp
331 Password required for ftp.
Password:
230 User logged in.
[ftp] put devicea.pub
227 Entering Passive Mode (10,1,1,2,5,148).
125 ASCII mode data connection already open, transfer starting for /devicea.pub.
226 Transfer complete.
FTP: 299 byte(s) sent in 0.189 second(s), 1.00Kbyte(s)/sec.
```

(4) Device B 从公钥文件中导入公钥

Device A 将公钥文件 devicea.pub 上传到 Device B 后，从公钥文件中导入 Device A 的主机公钥。

```
[DeviceB] public-key peer devicea import sshkey devicea.pub
```

显示 Device B 上保存的 Device A 的主机公钥信息。

```
[DeviceB] display public-key peer name devicea

=====================================
  Key Name  : devicea
  Key Type  : RSA
  Key Module: 1024
=====================================
Key Code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
```

目 录

1 WEB过滤配置 1-1
1.1 WEB过滤简介 1-1
1.1.1 URL网页地址过滤 1-1
1.1.2 URL网页地址过滤对网站IP地址的支持 1-1
1.1.3 URL参数过滤 1-2
1.1.4 Java阻断 1-2
1.1.5 ActiveX阻断 1-3
1.1.6 过滤规则的保存和恢复 1-3
1.2 配置WEB过滤 1-3
1.2.1 配置准备 1-4
1.2.2 配置URL网页地址过滤 1-4
1.2.3 配置URL网页地址过滤对网站IP地址的支持 1-4
1.2.4 配置URL参数过滤 1-5
1.2.5 配置Java阻断 1-5
1.2.6 配置ActiveX阻断 1-6
1.3 WEB过滤显示和维护 1-6
1.4 WEB过滤典型配置举例 1-6
1.4.1 网页地址过滤典型配置举例 1-6
1.4.2 URL参数过滤典型配置举例 1-8
1.4.3 Java阻断典型配置举例 1-9
1.5 常见配置错误举例 1-10
1.5.1 当前关键字或配置条目达到上限，添加失败 1-10
1.5.2 当前配置的参数中出现不允许的字符 1-11
1.5.3 关键字通配符非法使用 1-11
1.5.4 阻断后缀配置非法 1-12
1.5.5 ACL配置错误 1-12
1.5.6 网页地址过滤功能不起作用 1-12
1.5.7 无法访问主机IP地址 1-12

1 WEB 过滤配置

1.1 WEB 过滤简介

在传统的网络安全方案中，对网络攻击的防范主要针对于来自外部的各种攻击。随着网络在各行各业的普及，来自局域网内部的攻击也越来越多，这就要求网络设备能够应对构建安全内部网络的需求，增加内部网络安全特性。

目前设备所支持的 WEB 过滤功能，可以阻止内部用户访问非法的网址，以及对网页内的 Java 或 ActiveX 程序进行阻断。WEB 过滤功能实现了以下功能：

- 网页地址过滤
- URL 参数过滤
- Java 阻断
- ActiveX 阻断
- 配置信息的保存和恢复

下面将对以上五个功能分别进行介绍。

1.1.1 URL 网页地址过滤

1. 特性介绍

使用 URL 网页地址过滤可以阻止内部用户访问非法和不健康的网页，或者只允许用户访问某些特定的网页。

当收到 HTTP 请求报文时，设备会检测报文中的 URL 网页地址。如果该地址是配置为允许通过的，那么该 WEB 请求可以通过；如果该网页地址是配置为不允许通过的，那么该 WEB 请求将被拒绝，同时向发送 WEB 请求的客户端和服务器端发送 TCP reset 报文。

使能 URL 网页地址过滤功能以后，还应当指定 URL 网页地址过滤的默认行为。缺省情况下，URL 网页地址过滤的默认行为是 **deny**，即当 WEB 请求中 URL 网页地址与设备配置的过滤网址条目均不符合的情况下，将按照 URL 网页地址过滤的默认行为操作，拒绝该 WEB 请求通过。

2. 处理过程

(1) 设备接收到 HTTP 请求报文，并从 HTTP 请求报文中获取 URL 网页地址。

(2) 用设备中已配置的网页地址过滤条目与 HTTP 请求报文中的网页地址进行匹配过滤。如果匹配成功，则执行相应的允许或拒绝操作；如果匹配不成功，则按照 URL 网页地址过滤的默认行为操作。

1.1.2 URL 网页地址过滤对网站 IP 地址的支持

1. 特性介绍

URL 网页地址过滤功能启动以后，系统将默认拒绝所有直接以网站 IP 地址访问网站的 WEB 请求。

- 如果希望能够直接以网站 IP 地址访问所有网站，必须将 URL 网页地址过滤对网站 IP 地址的支持配置为 **permit**，则所有以网站 IP 地址访问网站的 WEB 请求允许通过。

- 如果希望部分网站可以通过网站 IP 地址直接访问，必须将 URL 网页地址过滤对网站 IP 地址的支持配置为 **deny**，并配置 ACL 规则允许部分 IP 地址形式的网站 WEB 请求通过。当设备接收到以这部分网站 IP 地址直接访问网站的 WEB 请求时，允许通过。

2. 处理过程

设备接收到以 IP 地址访问网站的请求报文时，做如下处理：

- 如果 URL 网页地址过滤对 IP 地址的支持功能配置为 **permit**，允许报文通过。
- 如果 URL 网页地址过滤对 IP 地址的支持功能配置为 **deny**，则检查 ACL 规则。ACL 允许通过，则报文通过，否则将拒绝该请求。

1.1.3 URL 参数过滤

1. 特性介绍

目前网页一般都是动态的，与数据库相连的，通过 WEB 请求去数据库中查询或修改所需的数据。这使得不法分子可以通过在 WEB 网页中构造特殊的 SQL 语句窃取数据库中机密数据，或不断修改数据库的信息导致数据库瘫痪。这种攻击方式被称为 SQL 注入攻击。

为此，用 SQL 语句中的关键字以及其它可能产生 SQL 语句的字符与 HTTP 请求报文进行匹配。如果匹配成功，则认为是 SQL 注入攻击，禁止其通过，这种过滤方式称为 URL 参数过滤。

WEB 传输参数的方式有很多种，其中最常用的是 get、post 方式。参数传输的方式决定了参数所在的位置，根据参数所在的位置获取参数，然后进行匹配过滤。目前，设备支持对传输方式为 get、post 和 put 的 WEB 请求进行 URL 参数过滤。

SQL 注入攻击一般基于英文 URL 进行，因此该特性不支持中文 URL 参数过滤。

2. 处理过程

设备收到包含 URL 参数的 HTTP 请求报文，根据 WEB 传输参数方式，从报文中获取 URL 参数。

- 非 get、post、put 方式的 HTTP 请求报文不做处理，直接通过。
- 如果是 get、post、put 方式的 HTTP 请求报文，则将其 URL 参数与设备上已配置的过滤参数条目进行匹配过滤，如果匹配成功，则拒绝该请求，否则允许报文通过。

1.1.4 Java 阻断

1. 特性介绍

通过对不可信站点的 Java 阻断功能，可以保护网络不受有害的 Java Applets 的破坏。

Java 阻断功能启动后，所有对 WEB 页面中的 Java Applet 程序的请求将被过滤掉。如果用户仍然希望能够获取部分 WEB 页面的 Java Applet 程序，必须配置 ACL 规则，如果 ACL 规则允许访问，则用户对该 WEB 页面的 Java Applet 程序的请求可以通过。

2. 处理过程

- 使能 Java Applet 阻断功能后，如果没有配置 ACL 规则，则将所有 HTTP 请求报文中的“.class”、“.jar”等文件名后缀都替换为“.block”，并允许该请求报文通过；
- 使能 Java Applet 阻断功能后，如果配置有 ACL 规则，则根据 ACL 过滤规则来决定 HTTP 请求报文中的“.class”、“.jar”是否用“.block”替代。如果 HTTP 请求的目的服务器为 ACL

允许访问的服务器，则不做替代，报文正常通过；否则将后缀替换为“.block”，然后允许该请求报文通过；

- Java 阻断过滤后缀，即 HTTP 请求报文中需要替换的文件名后缀，可通过命令行配置，添加除“.class”、“jar”之外的阻断后缀关键字。

1.1.5 ActiveX 阻断

1. 特性介绍

通过对不可信站点的 ActiveX 阻断功能，可以保护网络不受有害的 ActiveX 插件的破坏。

ActiveX 阻断功能启动后，所有对 WEB 页面中的 ActiveX 插件的请求将被过滤掉。如果用户仍然希望能够获取部分 WEB 页面的 ActiveX 插件，必须配置 ACL 规则，如果 ACL 规则允许访问，则用户可以获取该 WEB 页面的 ActiveX 插件。

2. 处理过程

- 使能 ActiveX 阻断功能后，如果没有配置 ACL 规则，则将所有 HTTP 请求报文中的文件名后缀“.ocx” 都替换为“.block”，然后允许该报文通过；
- 使能 ActiveX 阻断功能后，如果配置有 ACL 规则，则根据 ACL 过滤规则来决定 HTTP 请求报文中的“.ocx”是否用“.block”替代。如果 HTTP 请求的目的服务器为 ACL 允许访问的服务器，则不做替代，报文正常通过；否则将后缀替换为“.block”，然后允许该请求报文通过；
- ActiveX 阻断过滤后缀，即 HTTP 请求报文中需要替换的文件名后缀，可通过命令行配置，添加除“.ocx”之外的阻断后缀关键字。

1.1.6 过滤规则的保存和恢复

1. 过滤规则的保存

系统提供保存过滤规则的功能，可以把用户配置的过滤规则保存在指定的文件里，该功能仅对网页地址过滤、URL 参数过滤有效。

2. 过滤规则的装载

系统为用户提供装载过滤规则的功能，可以在设备启动后将用户指定的过滤规则自动装载到设备上。该功能的配置仅在设备重启后生效，并只对网页地址过滤、URL 参数过滤有效。另外也提供了取消装载过滤规则的命令行，用户可以取消装载已指定的过滤规则。

注意

- 建议用户不要修改装载文件，否则可能造成自动装载功能的失败。装载文件的修改建议通过重新装载新规则实现。
- 由于文件系统不支持热插拔，因此主控板上保存的条目文件并不会在备板插入时同步过来。

1.2 配置 WEB 过滤

只有使能 URL 网页地址过滤功能使能以后，才能配置对网站 IP 地址的支持功能和 URL 参数过滤功能才会生效。对不可信站点的 Java Applet 阻断功能以及 ActiveX 阻断功能均可以单独启用。

1.2.1 配置准备

需要配置 NAT 或 ASPF 特性，WEB 过滤功能才能生效。



说明

NAT 和 ASPF 的相关配置请分别参考“安全分册”中的“NAT 配置”和“防火墙配置”。

1.2.2 配置 URL 网页地址过滤

表1-1 配置 URL 网页地址过滤

操作	命令	说明
进入系统视图	**system-view**	-
使能 URL 网页地址过滤功能	**firewall http url-filter host enable**	必选 缺省情况下，URL 网页地址过滤功能处于关闭状态
配置默认过滤行为	**firewall http url-filter host default { deny | permit }**	可选 缺省情况下，默认的过滤行为为 **deny**
添加网页过滤地址	**firewall http url-filter host url-address { deny | permit }** *url-address*	必选
保存网页地址过滤条目	**firewall http url-filter host save** *file-name*	可选
装载网页地址过滤文件	**firewall http url-filter host load** *file-name*	可选
显示网页地址过滤信息	**display firewall http url-filter host [all | item** *keywords* **| verbose]**	可选

1.2.3 配置 URL 网页地址过滤对网站 IP 地址的支持

表1-2 配置网页地址过滤对网站 IP 地址的支持

操作	命令	说明
进入系统视图	**system-view**	-
使能网页地址过滤功能	**firewall http url-filter host enable**	必选 缺省情况下，网页地址过滤功能处于关闭状态
配置网页地址过滤对网站 IP 地址的支持	**firewall http url-filter host ip-address { deny | permit }**	可选 缺省情况下，网页地址过滤不支持网站 IP 地址
设置网页地址过滤的 ACL 规则	**firewall http url-filter host acl** *acl-number*	可选 缺省情况下，未设置网页地址过滤的 ACL 规则
显示网页地址过滤信息	**display firewall http url-filter host [all | item** *keywords* **| verbose]**	可选

说明

网页地址过滤的 ACL 规则所定义的源地址必须配置为相应网站的 IP 地址。

1.2.4 配置 URL 参数过滤

表1-3 配置 URL 参数过滤

操作	命令	说明
进入系统视图	**system-view**	-
使能 URL 网页地址过滤功能	**firewall http url-filter host enable**	必选 缺省情况下，URL 网页地址过滤功能处于关闭状态
使能 URL 参数过滤功能	**firewall http url-filter parameter enable**	必选 缺省情况下，URL 参数过滤功能处于关闭状态
添加 URL 参数过滤条目	**firewall http url-filter parameter** { **default** | **keywords** *keywords* }	必选
保存 URL 参数过滤条目	**firewall http url-filter parameter save** *file-name*	可选
装载 URL 参数过滤文件	**firewall http url-filter parameter load** *file-name*	可选
显示 URL 参数过滤信息	**display firewall http url-filter parameter** [**all** | **item** *keywords* | **verbose**]	可选

1.2.5 配置 Java 阻断

表1-4 配置 Java 阻断

操作	命令	说明
进入系统视图	**system-view**	-
使能 Java 阻断功能	**firewall http java-blocking enable**	必选 缺省情况下，Java 阻断功能处于关闭状态
添加 Java 阻断后缀关键字	**firewall http java-blocking suffix** *keywords*	可选
设置 Java 阻断的 ACL 规则	**firewall http java-blocking acl** *acl-number*	可选 缺省情况下，未设置 Java 阻断的 ACL 规则
显示 Java 阻断信息	**display firewall http java-blocking** [**all** | **item** *keywords* | **verbose**]	可选

说明

Java 阻断的 ACL 规则中需要将允许访问的 HTTP 服务器地址配置为允许通过（**permit**）的源地址（**source**）。

1.2.6 配置 ActiveX 阻断

表1-5 配置 ActiveX 阻断

操作	命令	说明
进入系统视图	**system-view**	-
使能 ActiveX 阻断功能	**firewall http activex-blocking enable**	必选 缺省情况下，ActiveX 阻断功能处于关闭状态
添加 ActiveX 阻断后缀关键字	**firewall http activex-blocking suffix** *keywords*	可选
设置 ActiveX 阻断的 ACL 规则	**firewall http activex-blocking acl** *acl-number*	可选 缺省情况下，未设置 Java 阻断的 ACL 规则
显示 ActiveX 阻断信息	**display firewall http activex-blocking** [**all** | **item** *keywords* | **verbose**]	可选

说明

ActiveX 阻断的 ACL 规则中需要将允许访问的 HTTP 服务器地址配置为允许通过（**permit**）的源地址（**source**）。

1.3 WEB 过滤显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 WEB 过滤的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 WEB 过滤统计信息。

表1-6 WEB 过滤显示和维护

操作	命令
显示 URL 网页地址过滤信息	**display firewall http url-filter host** [**all** | **item** *keywords* | **verbose**]
显示 URL 参数过滤信息	**display firewall http url-filter parameter** [**all** | **item** *keywords* | **verbose**]
显示 Java 阻断信息	**display firewall http java-blocking** [**all** | **item** *keywords* | **verbose**]
显示 ActiveX 阻断信息	**display firewall http activex-blocking** [**all** | **item** *keywords* | **verbose**]
清除 WEB 过滤统计信息	**reset firewall http** { **activex-blocking** | **java-blocking** | **url-filter host** | **url-filter parameter** } **counter**

1.4 WEB 过滤典型配置举例

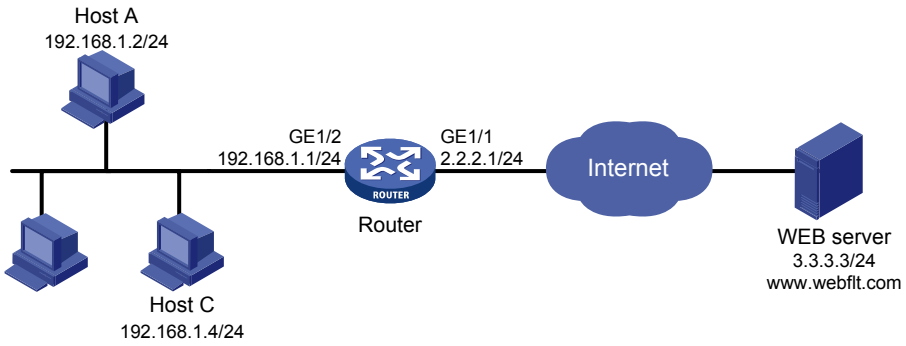
1.4.1 网页地址过滤典型配置举例

1. 组网需求

局域网 192.168.1.0/24 网段内的主机通过设备访问 Internet。启用网页地址过滤功能，只允许该网段内的用户访问地址为 www.webflt.com 的网站，同时允许用户可以使用 IP 地址的方式访问该网站。

2. 组网图

图1-1 网页地址过滤典型组网图



3. 配置步骤

(1) 具体的配置步骤

配置设备各接口的 IP 地址（略）。

配置设备出接口的 NAT 策略。

```
<Router> system-view
[Router] acl number 2200
[Router-acl-basic-2200] rule 0 permit source 192.168.1.0 0.0.0.255
[Router-acl-basic-2200] rule 1 deny source any
[Router-acl-basic-2200] quit
[Router] nat address-group 1 2.2.2.10 2.2.2.11
[Router] interface GigabitEthernet 1/1
[Router-GigabitEthernet1/1] nat outbound 2200 address-group 1
[Router-GigabitEthernet1/1] quit
```

使能 URL 网页地址过滤功能。

```
[Router] firewall http url-filter host enable
```

配置仅允许用户访问 www.webflt.com，并设置默认过滤行为为禁止。

```
[Router] firewall http url-filter host url-address permit www.webflt.com
[Router] firewall http url-filter host default deny
```

配置网页地址过滤的 ACL 规则。

```
[Router] acl number 2000
[Router-acl-basic-2000] rule 0 permit source 3.3.3.3 0.0.0.0
[Router-acl-basic-2000] rule 1 deny source any
[Router-acl-basic-2000] quit
```

配置允许用户以 IP 地址的方式访问该网站。

```
[Router] firewall http url-filter host ip-address deny
[Router] firewall http url-filter host acl 2000
```

(2) 验证配置结果

通过 **display firewall http url-filter host verbose** 命令查看网页地址过滤的详细信息。

```
[Router] display firewall http url-filter host verbose
 URL-filter host is enabled.
```

```
Default method: deny.
The support for IP address: deny.
The configured ACL group is 2000.
No file has been loaded.
There are 1 packet(s) being filtered.
There are 1 packet(s) being passed.
```

通过 **display firewall http url-filter host all** 命令查看网页地址过滤的所有条目信息。

```
[Router] display firewall http url-filter host all
 SN  Match-Times          Keywords
 ------------------------------------
 1   1                    www.webflt.com
```

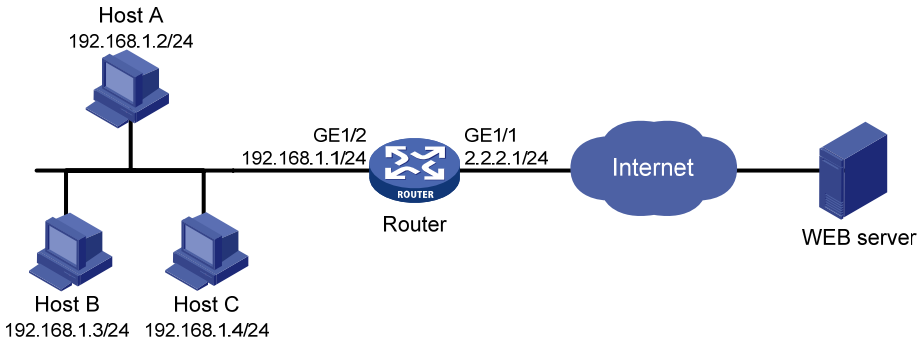
1.4.2 URL 参数过滤典型配置举例

1. 组网需求

局域网 192.168.1.0/24 网段内的主机通过设备访问 Internet。启用 URL 参数过滤功能，使用用户自定义的参数 group 过滤 HTTP 请求报文。

2. 组网图

图1-2 URL 参数过滤典型组网图



3. 配置步骤

(1) 具体的配置步骤

配置设备各接口的 IP 地址（略）。

配置设备出接口的 NAT 策略。

```
<Router> system-view
[Router] acl number 2200
[Router-acl-basic-2200] rule 0 permit source 192.168.1.0 0.0.0.255
[Router-acl-basic-2200] rule 1 deny source any
[Router-acl-basic-2200] quit
[Router] nat address-group 1 2.2.2.10 2.2.2.11
[Router] interface GigabitEthernet 1/1
[Router-GigabitEthernet1/1] nat outbound 2200 address-group 1
[Router-GigabitEthernet1/1] quit
```

使能 URL 参数过滤功能，并添加过滤参数 group。

```
[Router] firewall http url-filter host enable
```

```
[Router] firewall http url-filter parameter enable
[Router] firewall http url-filter parameter keywords group
```

(2) 验证配置结果

通过 **display firewall http url-filter parameter verbose** 命令查看 URL 参数过滤的详细信息。

```
[Router] display firewall http url-filter parameter verbose
 URL-filter parameter is enabled.
 No file has been loaded
 There are 1 packet(s) being filtered.
 There are 2 packet(s) being passed.
```

通过 **display firewall http url-filter parameter all** 命令查看 URL 参数过滤的所有条目信息。

```
[Router] display firewall http url-filter parameter all
 SN   Match-Times          Keywords
 -------------------------------------
 1    1                    group
```

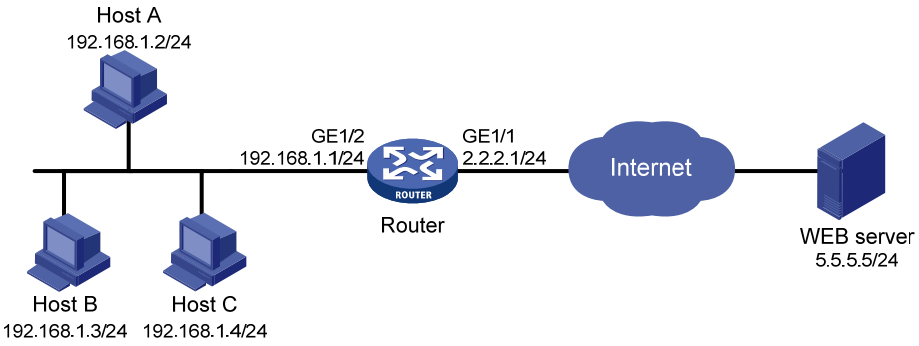
1.4.3 Java 阻断典型配置举例

1. 组网需求

局域网 192.168.1.0/24 网段内的主机通过设备访问 Internet。启用 Java 阻断功能，添加后缀名为“.js”的过滤项，同时只允许网站 IP 地址为 5.5.5.5 的 Java Applet 请求通过。

2. 组网图

图1-3 Java 阻断典型组网图



3. 配置步骤

(1) 具体的配置步骤

配置设备各接口的 IP 地址（略）。

配置设备出接口的 NAT 策略。

```
<Router> system-view
[Router] acl number 2200
[Router-acl-basic-2200] rule 0 permit source 192.168.1.0 0.0.0.255
[Router-acl-basic-2200] rule 1 deny source any
[Router-acl-basic-2200] quit
[Router] nat address-group 1 2.2.2.10 2.2.2.11
[Router] interface GigabitEthernet 1/1
[Router-GigabitEthernet1/1] nat outbound 2200 address-group 1
```

```
[Router-GigabitEthernet1/1] quit
```

配置 Java 阻断的 ACL 规则。

```
[Router] acl number 2100
[Router-acl-basic-2100] rule 0 permit source 5.5.5.5 0.0.0.0
[Router-acl-basic-2100] rule 1 deny source any
[Router-acl-basic-2100] quit
```

使能 Java 阻断功能，添加后缀名为.js 的过滤项，并配置根据 ACL 规则进行阻断。

```
[Router] firewall http java-blocking enable
[Router] firewall http java-blocking suffix .js
[Router] firewall http java-blocking acl 2100
```

(2) 验证配置结果

通过 **display firewall http java-blocking verbose** 命令查看 Java 阻断的详细信息。

```
[Router] display firewall http java-blocking verbose
 Java blocking is enabled.
 The configured ACL group is 2100.
 There are 0 packet(s) being filtered.
 There are 1 packet(s) being passed.
```

通过 **display firewall http java-blocking all** 命令查看 Java 阻断的所有条目信息。

```
[Router] display firewall http java-blocking all
 SN   Match-Times          Keywords
 -----------------------------------
 1    0                    .CLASS
 2    0                    .JAR
 3    1                    .js
```

以上信息显示，Java 阻断配置有 3 个过滤条目，“.CLASS”和“.JAR”是缺省条目，“.js”是用户配置的条目，且仅该条目匹配了一次。

1.5 常见配置错误举例

1.5.1 当前关键字或配置条目达到上限，添加失败

1. 故障现象

- 添加网页地址过滤或 URL 参数过滤条目时，系统提示不能继续添加新条目。
- 添加 Java 阻断或 ActiveX 阻断后缀关键字时，系统提示不能继续添加新后缀关键字。

2. 故障分析

网页地址过滤条目/URL 参数过滤条目/Java 阻断后缀关键字/ActiveX 阻断后缀关键字数目已经达到系统所允许的最大值。

3. 处理过程

如果确实需要添加，请先删除部分已配置的条目或关键字，然后再继续配置。

1.5.2 当前配置的参数中出现不允许的字符

1. 故障现象

配置网页地址过滤或 URL 参数过滤条目时，系统提示输入字符错误。

2. 故障分析

网页地址过滤只允许输入数字、英文字符、“.”、“-”、“_”以及通配符（“^”、“$”、“&”和“*”）；URL 参数的过滤条目只允许输入数字、英文字符、通配符（“^”、“$”、“&”和“*”）以及其它 ASCII 字符（31<ASCII 值<127）。

3. 处理过程

检查输入的命令，保证输入字符合法。

1.5.3 关键字通配符非法使用

1. 故障现象

配置网页地址过滤或 URL 参数过滤条目时，系统提示通配符使用错误。

2. 故障分析

网页地址过滤或 URL 参数过滤条目通配符的使用上分别有如下限制：

表1-7 网页地址过滤条目通配符含义

通配符	含义	使用说明
^	表明是开头匹配	只能出现在过滤关键字的开头，且只能出现一次
$	表明是结尾匹配	只能出现在过滤关键字的结尾，且只能出现一次
&	代替一个字符，不能代替“.”	可出现任意多个，也可连续出现，可位于过滤关键字的任意位置，不能与“*”一起使用
*	可代替任意多个字符，也可代替空格，不能代替“.”	在过滤关键字中只能出现一次，可以位于过滤关键字的开头和中间，不能位于结尾，并且不能和“^”、“$”相邻

表1-8 URL 参数过滤条目通配符含义

通配符	含义	使用说明
^	表明是开头匹配	只能位于过滤关键字的开头，且只能出现一次
$	表明是结尾匹配	只能位于过滤关键字的结尾，且只能出现一次
&	代替一个字符	可出现任意多个，也可连续出现，可位于过滤关键字的任意位置，但不能与“*”相邻，如果出现在开始和结尾的位置，则一定要和“^”或“$”相邻
*	代替不超过 4 个任意字符，可代替空格	只能位于过滤关键字的中间，且只能出现一次

3. 处理过程

根据以上关键字通配符的使用规则正确使用通配符。

1.5.4 阻断后缀配置非法

1. 故障现象

在配置 Java Applet 阻断和 ActiveX 阻断后缀关键字时，系统提示后缀关键字使用错误。

2. 故障分析

阻断后缀要求将“.”作为关键字的一部分进行配置，如果未配置“.”或配置时出现多个“.”将发生此错误。

3. 处理过程

根据阻断后缀的配置要求重新进行配置。

1.5.5 ACL 配置错误

1. 故障现象

ACL 规则配置为允许源地址为指定内网地址，在网页地址过滤、Java Applet 阻断或者 ActiveX 阻断中引用这个 ACL，希望指定内网主机发出的报文能够正常通过，发现配置的 ACL 不起作用。

2. 故障分析

无论是网页地址过滤，还是 Java Applet 阻断和 ActiveX 阻断，都是允许访问某一指定 IP 地址的外网服务器，而不是内网主机。因为内部网络属于信任区域，不需要对内网的 IP 地址进行特殊处理。

3. 处理过程

将外网服务器的 IP 地址作为 ACL 规则中的源地址进行配置。

1.5.6 网页地址过滤功能不起作用

1. 故障现象

在正确使能了网页地址过滤并添加过滤条目以后，发现没有对报文进行过滤。

2. 故障分析

网页地址过滤的前提是该报文被 NAT 或者被 ASPF 处理过，只有被处理过的报文才被认为是需要进行网页过滤处理的。对 URL 参数过滤，Java Applet 阻断，ActiveX 阻断，也是同样的道理。

3. 处理过程

检查是否配置了 NAT 或者 ASPF。

1.5.7 无法访问主机 IP 地址

1. 故障现象

在使能网页地址过滤以后，发现以设备为目的 HTTP 请求都被过滤掉了。

2. 故障分析

网页地址过滤使能以后，缺省情况下是不允许通过 IP 地址进行访问。而我们通过 WEB 对设备进行配置时，都是使用 IP 地址进行访问，因此就被过滤掉了。

3. 处理过程

配置 ACL，允许访问该主机 IP 地址的 HTTP 请求通过。

目 录

1 连接限制配置 1-1
1.1 连接限制简介 1-1
1.2 连接限制配置任务简介 1-1
1.3 创建连接限制策略 1-1
1.4 配置连接限制策略 1-1
1.4.1 配置基于IP地址的连接限制策略 1-1
1.5 应用连接限制策略 1-2
1.6 连接限制的显示和维护 1-2
1.7 连接限制典型配置举例 1-3
1.7.1 连接限制典型配置举例 1-3
1.8 连接限制常见配置错误举例 1-4
1.8.1 不同的连接限制规则中网段存在包含关系时，规则顺序错误 1-4
1.8.2 不同的连接限制规则中协议存在包含关系时，规则顺序错误 1-4

1 连接限制配置

1.1 连接限制简介

内网用户访问外部网络时，如果某一用户在短时间内经过设备向外部网络发起大量连接，将会导致设备系统资源迅速消耗，其它用户无法正常使用网络资源。另外，如果一台内部服务器在短时间内接收到大量的连接请求，将会导致该服务器无法及时进行处理，以至于不能再接受其它客户端的正常连接请求。

因此，为了保护内部网络资源（主机或服务器）以及合理分配设备系统资源，需要制定相应的连接限制策略来对设备上建立的连接进行统计和限制。目前，设备支持的连接限制策略可通过限制用户发起的连接数来实现。

1.2 连接限制配置任务简介

表1-1 连接限制配置任务简介

配置任务		说明	详细配置
创建连接限制策略		必选	1.3
配置连接限制策略	配置基于 IP 地址的连接限制策略	必选	1.4.1
应用连接限制策略		必选	1.5

1.3 创建连接限制策略

连接限制策略用于定义具体的连接限制规则，其中的规则规定了策略生效的范围和实施的参数。

表1-2 创建连接限制策略

操作	命令	说明
进入系统视图	**system-view**	-
创建连接限制策略，并进入连接限制策略视图	**connection-limit policy** *policy-number*	必选

1.4 配置连接限制策略

一个连接限制策略中可定义多条连接限制规则，每条连接限制规则中可指定一个连接限制的对象或范围，与之匹配的用户的连接建立将受到该规则中指定参数的限制。

1.4.1 配置基于 IP 地址的连接限制策略

可以对指定的源 IP 和目的 IP 之间的主机或网段的连接数进行统计和限制。为适应不同的应用需求，设备可支持如下四种类型的连接限制规则：

- 指定源 IP 到指定目的 IP：用于限制特定的内部网络主机或网段向特定的外部网络主机或网段发起的连接数。
- 指定源 IP 到任意 IP：用于限制特定的内部网络主机或网段向外部网络资源发起的连接数。
- 任意源 IP 到指定目的 IP：用于限制外部网络主机或网段向特定的内部服务器发起的连接数。
- 任意源 IP 到任意目的 IP：用于限制经过设备建立的连接总数。

表1-3 配置基于 IP 地址的连接限制策略

操作	命令	说明
进入系统视图	**system-view**	-
进入连接限制策略视图	**connection-limit policy** *policy-number*	-
配置基于 IP 地址的连接限制规则	**limit** *limit-id* { **source ip** { *ip-address mask-length* | **any** } [**source-vpn** *src-vpn-name*] | **destination ip** { *ip-address mask-length* | **any** } [**destination-vpn** *dst-vpn-name*] } * **protocol** { **dns** | **http** | **ip** | **tcp** | **udp** } **max-connections** *max-num* [**per-destination** | **per-source** | **per-source-destination**]	必选

说明

连接限制规则的匹配规则为：

- 按照连接限制规则编号从小到大的顺序进行匹配。
- 如果两条规则的源网段或者目的网段存在包含关系，则按匹配到的第一条规则进行限制。
- 如果两条规则的协议存在包含关系，则按匹配到的第一条规则进行限制。

1.5 应用连接限制策略

将已经配置好的连接限制策略应用到全局或不同的业务模块上，实现对指定网络资源的连接限制。

表1-4 应用连接限制策略

操作	命令	说明
进入系统视图	**system-view**	-
应用连接限制策略	**connection-limit apply policy** *policy-number*	必选

1.6 连接限制的显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示连接限制配置后的运行情况，通过查看显示信息验证配置的效果。

表1-5 NAT 显示和维护

操作	命令
显示连接限制策略的配置信息	**display connection-limit policy** *policy-number*

1.7 连接限制典型配置举例

1.7.1 连接限制典型配置举例

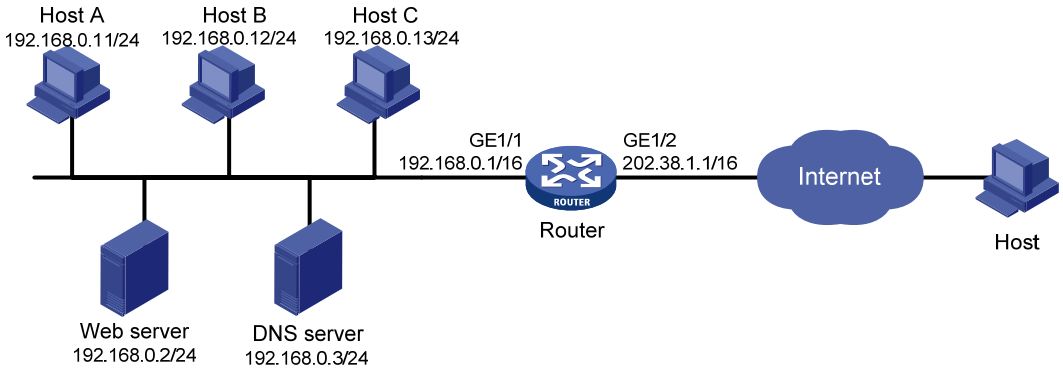
1. 组网需求

某公司拥有 202.38.1.1/24 至 202.38.1.5/24 五个公网 IP 地址，内部网络地址为 192.168.0.0/16。通过配置 NAT 使得内部网络主机可以访问 Internet，并提供两台内部服务器供外网访问。为保护内部网络及网络资源，需要进行以下限制：

- 192.168.0.0/24 网段的每台主机最多只能与外网建立 100 条连接，其他网段的主机不作限制。
- 同一时刻 DNS 服务器只接受 10000 条查询请求。
- 同一时刻 Web 服务器只接受 10000 条连接请求。

2. 组网图

图1-1 连接限制典型配置组网



3. 配置步骤

说明

内网主机访问外部网络的 NAT 配置和内部服务器的配置请参考“安全分册”中的“NAT 配置”，此处不进行介绍，下面仅对连接限制的配置步骤进行详细描述。

创建并进入连接限制策略视图。

```
<Router> system-view
[Router] connection-limit policy 0
```

配置连接限制规则 0，允许 192.168.0.0/24 网段的每台主机最多只能与外网建立 100 条连接。

```
[Router-connection-limit-policy-0] limit 0 source ip 192.168.0.0 24 destination ip any
protocol ip max-connections 100 per-source
```

配置连接限制规则 1，允许同一时刻 DNS 服务器最多接受 10000 条查询请求。

```
[Router-connection-limit-policy-0] limit 1 source ip any destination ip 192.168.0.3 32
protocol dns max-connections 10000
```

配置连接限制规则 2，允许同一时刻 Web 服务器最多接受 10000 条连接请求。

```
[Router-connection-limit-policy-0] limit 2 source ip any destination ip 192.168.0.2 32
protocol http max-connections 10000
```

```
[Router-connection-limit-policy-0] quit
```

应用连接限制策略。

```
[Router] connection-limit apply policy 0
```

4. 验证配置结果

上述配置完成后，执行 **display connection-limit policy** 命令显示连接限制的配置情况，具体内容如下。

```
[Router] display connection-limit policy 0
 Connection-limit policy 0, refcount 1, 3 limits
  limit 0 source ip 192.168.0.0 24 destination ip any protocol ip max-connections 100
per-source
  limit 1 source ip any destination ip 192.168.0.3 32 protocol dns max-connections 10000
  limit 2 source ip any destination ip 192.168.0.2 32 protocol http max-connections 10000
```

1.8 连接限制常见配置错误举例

1.8.1 不同的连接限制规则中网段存在包含关系时，规则顺序错误

1. 故障现象

在 Router 上进行如下配置，希望限制主机 192.168.0.100 最多向外部网络发起 100 条连接请求，以及 192.168.0.0/24 网段的其他主机最多向外部网络发起 10 条连接请求。

```
[Router-connection-limit-policy-0] limit 0 source ip 192.168.0.0 24 destination ip any
protocol ip max-connections 10 per-source
[Router-connection-limit-policy-0] limit 1 source ip 192.168.0.100 32 destination ip any
protocol ip max-connections 100 per-source
```

实际运行过程中，主机 192.168.0.100 最多只能向外部网络发起 10 条连接，后续连接被拒绝。

2. 故障分析

在上述配置中，limit 0 和 limit 1 中指定的源 IP 地址范围存在包含关系，192.168.0.100 发起的连接既符合 limit 0 又符合 limit 1。由于在进行连接限制规则的匹配时，设备以匹配到的第一条有效规则为准，因此对 192.168.0.100 向外部网络发起的连接将只按照 limit 0 进行限制，而不会使用 limit 1 来限制。

3. 处理过程

为实现本需求，需要对 limit 1 与 limit 0 的顺序重新安排，将两个规则的序号进行调换，即将限制粒度更细、限制范围更精确的规则置前。

1.8.2 不同的连接限制规则中协议存在包含关系时，规则顺序错误

1. 故障现象

内部服务器 192.168.0.100 同时对外提供 Web 服务和 FTP 服务，在 Router 上进行如下配置，希望限制该服务器上最多只能建立 1000 条 HTTP 连接和 100 条 TCP 连接。

```
[Router-connection-limit-policy-0] limit 0 source ip any destination ip 192.168.0.100
protocol tcp max-connections 100
[Router-connection-limit-policy-0] limit 1 source ip any destination ip 192.168.0.100
protocol http max-connections 1000
```

实际运行过程中，内部服务器只能接受 100 条 HTTP 连接，后续连接被拒绝。

2. 故障分析

在上述配置中，limit 0 和 limit 1 中指定的协议存在包含关系，即，发往 192.168.0.100 的 HTTP 连接请求既符合 limit 0 又符合 limit 1。由于在进行连接限制规则的匹配时，设备以匹配到的第一条有效规则为准，因此对发往 192.168.0.100 的 HTTP 连接将只按照 limit 0 进行限制，而不会使用 limit 1 来限制。

3. 处理过程

为实现本需求，需要对 limit 1 与 limit 0 的顺序重新安排，将两个规则的序号进行调换，即将限制粒度更细、限制范围更精确的规则置前。

目 录

1 Rsh配置 1-1
1.1 Rsh简介 1-1
1.2 配置Rsh 1-1
1.2.1 配置准备 1-1
1.2.2 配置Rsh终端服务 1-1
1.3 Rsh典型配置举例 1-2

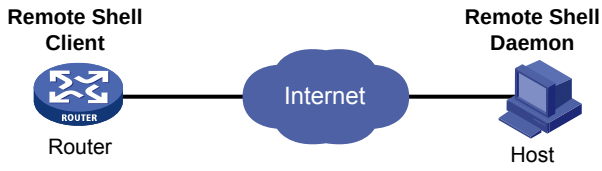
1 Rsh 配置

1.1 Rsh 简介

Rsh（Remote Shell，远程外壳）用来在远程主机上执行特定的命令（即远程主机的操作系统所提供的内部和外部命令）。远程主机需要运行 Rsh 守护程序（Rsh Daemon）以支持 Rsh 服务。Rsh 客户端与远程主机的守护进程之间进行通信。

如 图 1-1所示，设备可实现Rsh客户端功能，用户可以在设备上使用Rsh命令远程执行服务器端主机上的命令。

图1-1 Rsh 典型组网图



Rsh Daemon 提供基于信任主机的特权端口认证的远程命令执行服务。在 Windows NT/2000/XP/2003 上均可以利用服务组件来启动或关闭该服务。

1.2 配置 Rsh

1.2.1 配置准备

- 远程主机需要运行 Rsh Daemon 以支持 Rsh 服务。
- 设备与远程主机之间路由可达。

1.2.2 配置 Rsh 终端服务

设备作为 Rsh 客户端，使用本命令对远程主机进行操作。

表1-1 配置 Rsh 终端服务

操作	命令	说明
执行远程主机命令	**rsh** *host* [**user** *username*] **command** *remote-command*	必选 请在用户视图下执行

说明

如果远程主机启用了认证服务，那么在配置 Rsh 客户端之前需要在远程主机上配置相应的用户名。

1.3 Rsh 典型配置举例

说明

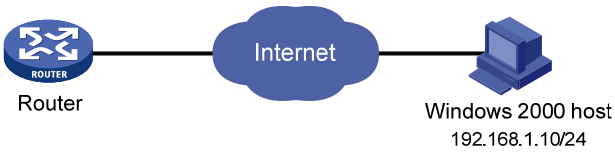
Windows NT/2000/XP/2003 系统不自带 Rsh Daemon，需要安装 Rsh Daemon 后才能启动此项服务。此软件需要用户自行购买或通过其它途径获取。

1. 组网需求

如 图 1-2所示，Router作为Rsh客户端，远程主机的操作系统为Windows 2000，且已安装并启动了Rsh Daemon服务。通过路由器远程设置主机时间。

2. 组网图

图1-2 Rsh 典型配置组网图



3. 配置步骤

在远程主机 Windows 2000 Host 上，检查是否已经安装和启动了 Rsh Daemon。

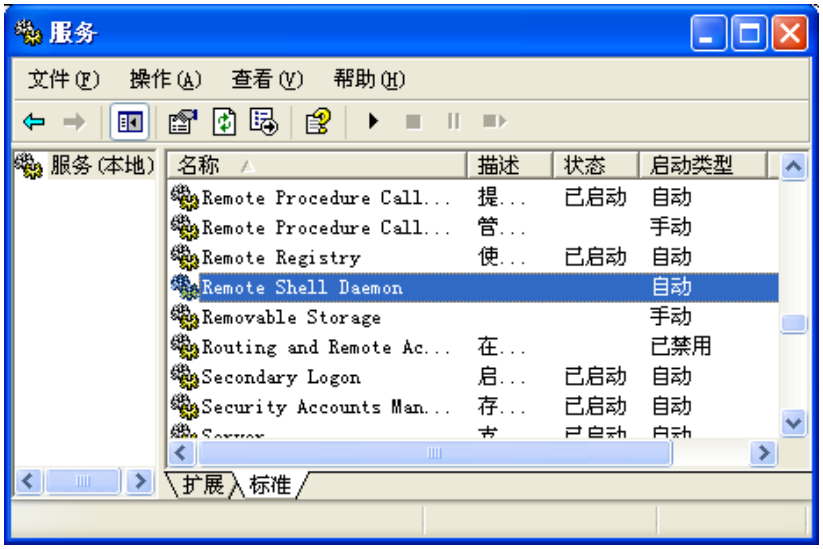
(1) 进入 Windows“控制面板”，打开“管理工具”。（对于 Windows XP 系统，当使用“控制面板”的分类视图时，在“性能和维护”类中选择“管理工具”。）

图1-3 Windows 管理工具文件夹视图



(2) 双击“服务”图标，进入 Windows“服务”管理窗口。

图1-4 Windows“服务”管理窗口



(3) 查找“服务”中是否包含“Remote Shell Daemon”。如果没有，说明尚未安装此服务程序，则需要先安装此程序。

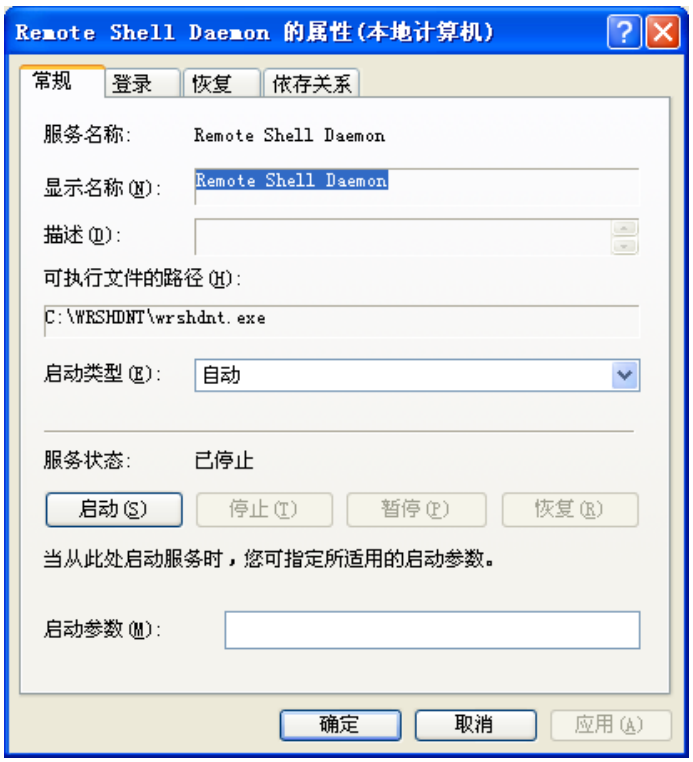
(4) 查看是否已经启动“Remote Shell Daemon”服务。

如 图 1-4，在“Remote Shell Daemon”服务对应状态栏中可以查看该服务是否已经启动。在本例中，该服务尚未启动。

(5) 启动“Remote Shell Daemon”服务。

双击该项服务，进入“Remote Shell Daemon”属性窗口，如 图 1-5所示。单击<启动>按钮，即可启动该项服务。

图1-5 “Remote Shell Daemon”属性窗口



在 Router 上进行如下配置：

配置与主机之间的路由，具体配置过程略。

在 Router 上通过 Rsh 命令设置主机的时间。

```
<Router> rsh 192.168.1.10 command time
Trying 192.168.1.10 ...
Press CTRL+K to abort
当前时间:  6:56:42.57
输入新时间: 12:00
12:00
```

目 录

1 GR概述 1-1
1.1 GR简介 1-1
1.2 GR基本概念 1-1
1.3 GR通讯过程 1-1
1.4 各协议的GR机制 1-3

1 GR 概述

1.1 GR 简介

GR 是 Graceful Restart（平滑重启）的简称，是一种在协议重启时保证转发业务不中断的机制。

GR 机制的核心在于：当某设备进行协议重启时，能够通知其周边设备在一定时间内将到该设备的邻居关系和路由保持稳定。在协议重启完毕后，周边设备协助其进行信息（包括支持 GR 的路由/MPLS 相关协议所维护的各种拓扑、路由和会话信息）同步，在尽量短的时间内使该设备恢复到重启前的状态。在整个协议重启过程中不会产生路由振荡，报文转发路径也没有任何改变，整个系统可以不间断地转发数据。这个过程即称为平滑重启。

1.2 GR 基本概念

配置了 GR 功能的设备称为“具备 GR 能力”的设备。具备 GR 能力的设备在协议重启时，能实现平滑重启，保证转发业务不中断；而不具备 GR 能力的设备在协议重启时，则只能遵循普通的重启过程。GR 中涉及到的基本概念如下：

1. GR Restarter

GR 重启路由器，指由管理员或故障触发而协议重启的设备，它必须具备 GR 能力。

2. GR Helper

即 GR Restarter 的邻居，能协助重启的 GR Restarter 保持路由关系的稳定，它也必须具备 GR 能力。

3. GR Session

GR 会话，是 GR Restarter 和 GR Helper 之间的协商过程。包括协议重启通告，协议重启过程中的信息交互等。通过该会话，GR Restarter 和 GR Helper 可以掌握彼此的 GR 能力。

4. GR Time

GR 时间，是 GR Restarter 和 GR Helper 协商建立一个会话所用的时间。当某 GR 路由器发现邻居路由器处于 down 状态时，将在该时间内仍保留其发出的拓扑或路由信息。

1.3 GR 通讯过程

在网络中配置一个设备为 GR Restarter，该设备与其 GR Helper 必须支持 GR 或具备 GR 能力。这样当 GR Restarter 重启时，其 GR Helper 就可以感知它的重启进程。



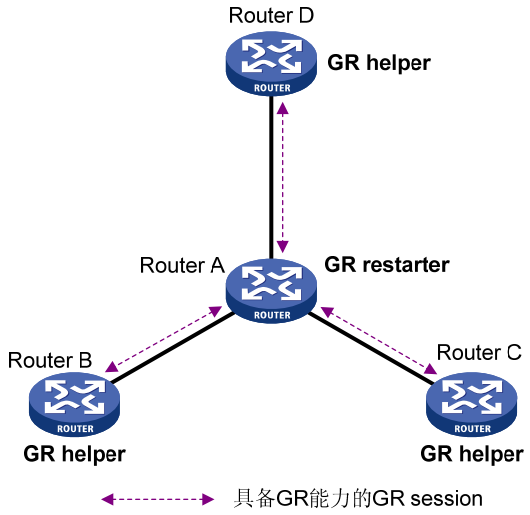
说明

GR Restarter 与 GR Helper 的作用是相互的。在某些情况下，GR Restarter 与 GR Helper 的位置和作用可以互换。

GR Restarter 和 GR Helper 之间的具体通讯过程如下：

1. GR Restarter 和 GR Helper 之间建立 GR Session

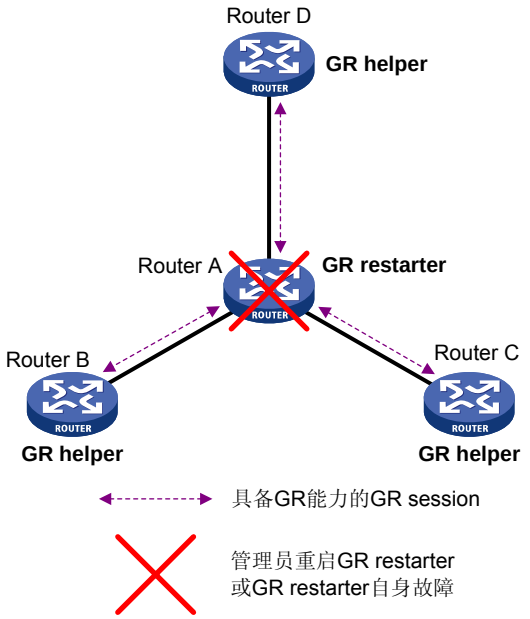
图1-1 在 GR Restarter 与 GR Helper 间建立 GR Session



如 图 1-1所示，Router A承担GR Restarter角色，Router B、Router C和Router D分别是Router A的GR Helper，在GR Restarter和GR Helper之间建立起GR Session。

2. GR Restarter 重新启动

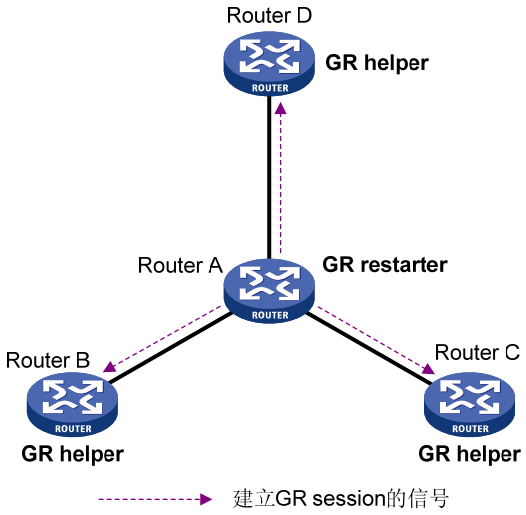
图1-2 GR Restarter 的重启过程



如 图 1-2所示，当各GR Helper发现其对端GR Restarter处于协议重启状态时，不仅继续保持GR Session，而且在GR Time内仍保留来自GR Restarter的拓扑或路由信息。

3. GR Restarter 向 GR Helper 发送信号

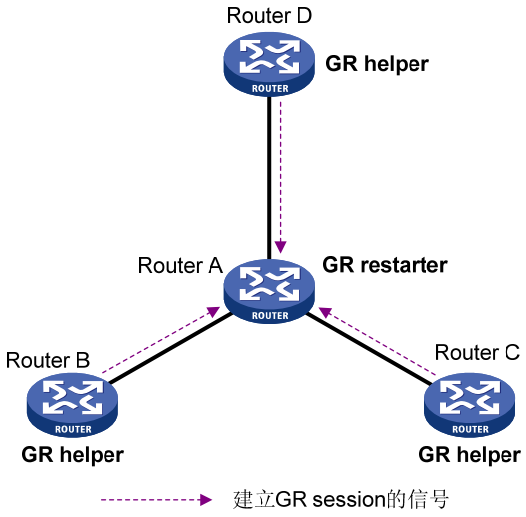
图1-3 GR Restarter 重启完毕后向 GR Helper 发送信号



如 图 1-3所示，GR Restarter的重新启动完成后，会向其每个GR Helper发送信号，从而重新建立GR Session。

4. GR Restarter 从 GR Helper 获取拓扑或路由信息

图1-4 GR Restarter 从 GR Helper 获取拓扑或路由信息



如 图 1-4所示，GR Restarter通过与所有GR Helper建立GR Session，可获得拓扑或路由信息，并以此重新计算自己的路由表。

1.4 各协议的 GR 机制

Comware 支持基于 MPLS LDP、MPLS RSVP-TE、BGP、IS-IS、OSPF 和 OSPF v3 协议的 GR 机制。

有关上述协议各自 GR 机制的实现原理及相关的配置过程，请分别参见“IP 路由分册”中的相关手册，以及“MPLS 分册”中的“MPLS 基本配置”和“MPLS TE 配置”。

目 录

1 VRRP配置 1-1
1.1 VRRP简介 1-1
1.1.1 VRRP概述 1-1
1.1.2 VRRP备份组简介 1-1
1.1.3 VRRP定时器 1-3
1.1.4 VRRP报文格式 1-3
1.1.5 VRRP工作过程 1-5
1.1.6 VRRP监视接口功能 1-5
1.1.7 VRRP应用（以基于IPv4 的VRRP为例） 1-6
1.2 配置基于IPv4 的VRRP 1-7
1.2.1 基于IPv4 的VRRP配置任务简介 1-7
1.2.2 配置虚拟IP地址和MAC地址的对应关系 1-8
1.2.3 创建备份组并配置虚拟IP地址 1-8
1.2.4 配置备份组优先级、抢占方式及监视功能 1-9
1.2.5 配置VRRP报文的相关属性 1-10
1.2.6 开启VRRP的Trap功能 1-11
1.2.7 基于IPv4 的VRRP显示和维护 1-11
1.3 配置基于IPv6 的VRRP 1-11
1.3.1 基于IPv6 的VRRP配置任务简介 1-11
1.3.2 配置虚拟IPv6 地址和MAC地址的对应关系 1-12
1.3.3 创建备份组并配置虚拟IPv6 地址 1-12
1.3.4 配置备份组优先级、抢占方式及监视指定接口 1-13
1.3.5 配置VRRP报文的相关属性 1-14
1.3.6 基于IPv6 的VRRP显示和维护 1-14
1.4 基于IPv4 的VRRP典型配置举例 1-15
1.4.1 VRRP单备份组配置举例 1-15
1.4.2 VRRP监视接口配置举例 1-17
1.4.3 VRRP多备份组配置举例 1-20
1.5 基于IPv6 的VRRP典型配置举例 1-23
1.5.1 VRRP单备份组配置举例 1-23
1.5.2 VRRP监视接口配置举例 1-25
1.5.3 VRRP多备份组配置举例 1-29
1.6 VRRP常见错误配置举例 1-31

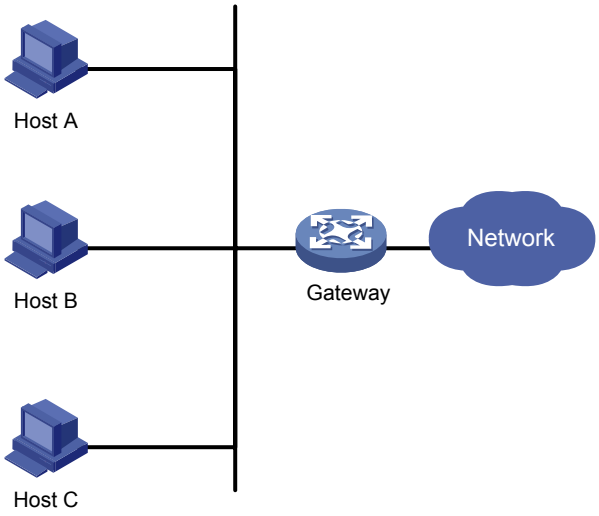
1 VRRP 配置

1.1 VRRP 简介

1.1.1 VRRP 概述

如 图 1-1所示，通常，同一网段内的所有主机都设置一条相同的以网关为下一跳的缺省路由。主机发往其他网段的报文将通过缺省路由发往网关，再由网关进行转发，从而实现主机与外部网络的通信。当网关发生故障时，本网段内所有以网关为缺省路由的主机将无法与外部网络通信。

图1-1 局域网组网方案



缺省路由为用户的配置操作提供了方便，但是对缺省网关设备提出了很高的稳定性要求。增加出口网关是提高系统可靠性的常见方法，此时如何在多个出口之间进行选路就成为需要解决的问题。

VRRP（Virtual Router Redundancy Protocol，虚拟路由器冗余协议）将可以承担网关功能的路由器加入到备份组中，形成一台虚拟路由器，由 VRRP 的选举机制决定哪台路由器承担转发任务，局域网内的主机只需将虚拟路由器配置为缺省网关。VRRP 是一种容错协议，在提高可靠性的同时，简化了主机的配置。在具有多播或广播能力的局域网（如以太网）中，借助 VRRP 能在某台设备出现故障时仍然提供高可靠的缺省链路，有效避免单一链路发生故障后网络中断的问题，而无需修改动态路由协议、路由发现协议等配置信息。

VRRP 协议的实现有 VRRPv2 和 VRRPv3 两个版本。其中，VRRPv2 基于 IPv4，VRRPv3 基于 IPv6。两个版本的 VRRP 在功能实现上并没有区别，只是在 IPv4 设备上和 IPv6 设备上使用的命令不同。

1.1.2 VRRP 备份组简介

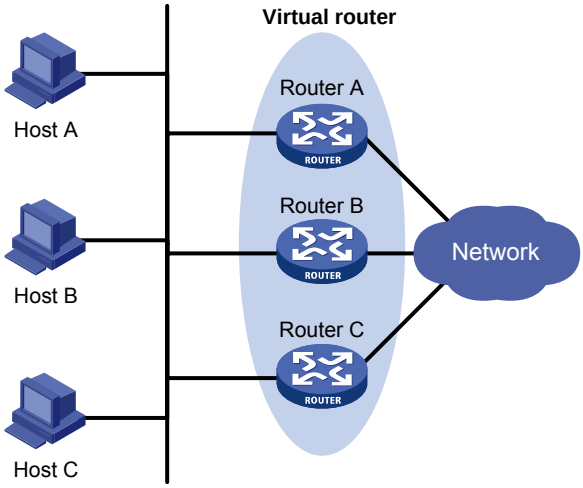
VRRP 将局域网内的一组路由器划分在一起，称为一个备份组。备份组由一个 Master 路由器和多个 Backup 路由器组成，功能上相当于一台虚拟路由器。

VRRP 备份组具有以下特点：

- 虚拟路由器具有 IP 地址。局域网内的主机仅需要知道这个虚拟路由器的 IP 地址，并将其设置为缺省路由的下一跳地址。
- 网络内的主机通过这个虚拟路由器与外部网络进行通信。

- 备份组内的路由器根据优先级，选举出 Master 路由器，承担网关功能。其他路由器作为 Backup 路由器，当 Master 路由器发生故障时，取代 Master 继续履行网关职责，从而保证网络内的主机不间断地与外部网络进行通信。

图1-2 VRRP 组网示意图



如 图 1-2所示，Router A、Router B和Router C组成一个虚拟路由器。此虚拟路由器有自己的IP地址。局域网内的主机将虚拟路由器设置为缺省网关。Router A、Router B和Router C中优先级最高的路由器作为Master路由器，承担网关的功能。其余两台路由器作为Backup路由器。

注意

- 虚拟路由器的 IP 地址可以是备份组所在网段中未被分配的 IP 地址，也可以和备份组内的某个路由器的接口 IP 地址相同。
- 接口 IP 地址与虚拟 IP 地址相同的路由器被称为“IP 地址拥有者”。
- 在同一个 VRRP 备份组中，只允许配置一个 IP 地址拥有者。

1. 备份组中路由器的优先级

VRRP 根据优先级来确定备份组中每台路由器的角色（Master 路由器或 Backup 路由器）。优先级越高，则越有可能成为 Master 路由器。

VRRP 优先级的取值范围为 0 到 255（数值越大表明优先级越高），可配置的范围是 1 到 254，优先级 0 为系统保留给特殊用途来使用，255 则是系统保留给 IP 地址拥有者。当路由器为 IP 地址拥有者时，其优先级始终为 255。因此，当备份组内存在 IP 地址拥有者时，只要其工作正常，则为 Master 路由器。

2. 备份组中路由器的工作方式

备份组中的路由器具有以下两种工作方式：

- 非抢占方式：如果备份组中的路由器工作在非抢占方式下，则只要 Master 路由器没有出现故障，Backup 路由器即使随后被配置了更高的优先级也不会成为 Master 路由器。

- 抢占方式：如果备份组中的路由器工作在抢占方式下，它一旦发现自己的优先级比当前的 Master 路由器的优先级高，就会对外发送 VRRP 通告报文。导致备份组内路由器重新选举 Master 路由器，并最终取代原有的 Master 路由器。相应地，原来的 Master 路由器将会变成 Backup 路由器。

3. 备份组中路由器的认证方式

为了防止非法用户构造报文攻击备份组，VRRP 通过在 VRRP 报文中增加认证字的方式，验证接收到的 VRRP 报文。VRRP 提供了两种认证方式：

- **simple**：简单字符认证。发送 VRRP 报文的路由器将认证字填入到 VRRP 报文中，而收到 VRRP 报文的路由器会将收到的 VRRP 报文中的认证字和本地配置的认证字进行比较。如果认证字相同，则认为接收到的报文是真实、合法的 VRRP 报文；否则认为接收到的报文是一个非法报文。
- **md5**：MD5 认证。发送 VRRP 报文的路由器利用认证字和 MD5 算法对 VRRP 报文进行加密，加密后的报文保存在 Authentication Header（认证头）中。收到 VRRP 报文的路由器会利用认证字解密报文，检查该报文的合法性。

在一个安全的网络中，用户也可以不设置认证方式。

1.1.3 VRRP 定时器

VRRP 定时器分为两种：VRRP 通告报文间隔时间定时器和 VRRP 抢占延迟时间定时器。

1. VRRP 通告报文时间间隔定时器

VRRP 备份组中的 Master 路由器会定时发送 VRRP 通告报文，通知备份组内的路由器自己工作正常。

用户可以通过设置 VRRP 定时器来调整 Master 路由器发送 VRRP 通告报文的时间间隔。如果 Backup 路由器在等待了 3 个间隔时间后，依然没有收到 VRRP 通告报文，则认为自己是 Master 路由器，并对外发送 VRRP 通告报文，重新进行 Master 路由器的选举。

2. VRRP 抢占延迟时间定时器

在性能不够稳定的网络中，Backup 路由器可能因为网络堵塞而无法正常收到 Master 路由器的报文，导致备份组内的成员频繁的进行主备状态转换。用户可以通过设置 VRRP 抢占延迟时间的方法来解决这个问题。

设置了 VRRP 抢占延迟时间后，Backup 路由器会在等待了 3 倍的通告报文时间间隔后，再等待 VRRP 抢占延迟时间。如在此期间还是没有收到 VRRP 通告报文，则此 Backup 路由器将认为自己是 Master 路由器，对外发送 VRRP 通告报文，触发备份组内路由器进行 Master 路由器的选举。

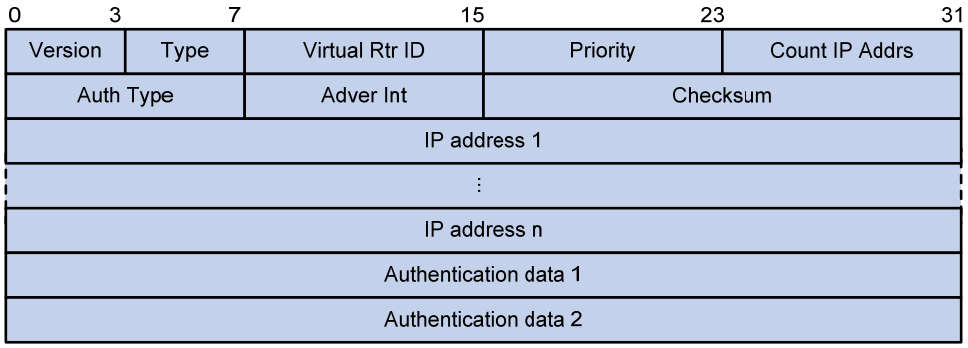
1.1.4 VRRP 报文格式

Master 路由器以组播的方式定时发送 VRRP 报文通告它的存在。这些报文可以用来检测虚拟路由器的各种参数，还可以用于 Master 路由器的选举。

1. VRRPv2 的报文格式

VRRPv2 的报文格式如 图 1-3所示。

图1-3 VRRPv2 的报文格式



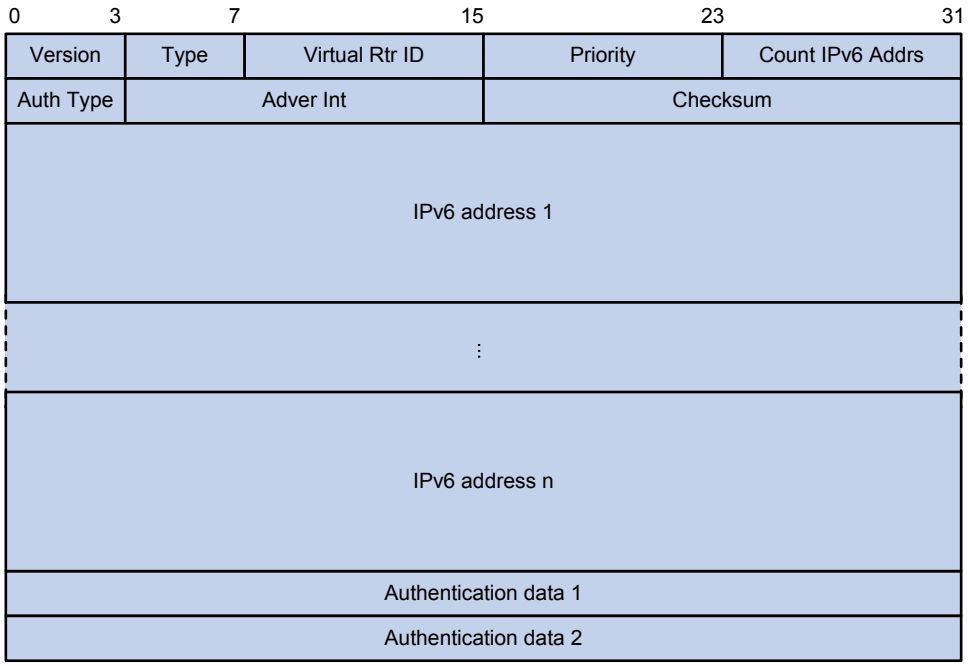
各字段解释如下：

- Version：协议版本号，VRRPv2 对应的版本号为 2。
- Type：VRRP 报文的类型。VRRPv2 报文只有一种类型，即 VRRP 通告报文（Advertisement），该字段取值为 1。
- Virtual Rtr ID（VRID）：虚拟路由器号（即备份组号），取值范围 1～255。
- Priority：路由器在备份组中的优先级，取值范围 0～255，数值越大表明优先级越高。
- Count IP Addrs：备份组虚拟 IP 地址的个数。1 个备份组可对应多个虚拟 IP 地址。
- Auth Type：认证类型。该值为 0 表示无认证，为 1 表示简单字符认证，为 2 表示 MD5 认证。
- Adver Int：发送通告报文的时间间隔，单位为秒，缺省为 1 秒。
- Checksum：16 位校验和，用于检测 VRRP 报文中的数据破坏情况。
- IP Address：备份组虚拟 IP 地址表项。所包含的地址数定义在 Count IP Addrs 字段。
- Authentication Data：验证字，目前只用于简单字符认证，对于其它认证方式一律填 0。

2. VRRPv3 的报文格式

VRRPv3 的报文格式如 图 1-4所示。

图1-4 VRRPv3 的报文格式



各字段解释如下：

- Version：协议版本号，VRRPv3 对应的版本号为 3。
- Type：VRRP 报文的类型。VRRPv3 报文只有一种类型，即 VRRP 通告报文（Advertisement），该字段取值为 1。
- Virtual Rtr ID（VRID）：虚拟路由器号（即备份组号），取值范围 1～255。
- Priority：路由器在备份组中的优先级，取值范围 0～255，数值越大表明优先级越高。
- Count IPv6 Addrs：备份组虚拟 IPv6 地址的个数。1 个备份组可对应多个虚拟 IPv6 地址。
- Auth Type：认证类型。该值为 0 表示无认证，为 1 表示简单字符认证。VRRPv3 不支持 MD5 认证。
- Adver Int：发送通告报文的时间间隔，单位为厘秒，缺省为 100 厘秒。
- Checksum：16 位校验和，用于检测 VRRPv3 报文中的数据破坏情况。
- IPv6 Address：备份组虚拟 IPv6 地址表项。所包含的地址数定义在 Count IPv6 Addrs 字段。

Authentication Data：验证字，目前只用于简单字符认证，对于其它认证方式一律填 0。

1.1.5 VRRP 工作过程

VRRP 的工作过程如下：

(1) 路由器使能 VRRP 功能后，会根据优先级确定自己在备份组中的角色。优先级高的路由器成为 Master 路由器，优先级低的成为 Backup 路由器。Master 路由器定期发送 VRRP 通告报文，通知备份组内的其他设备自己工作正常；Backup 路由器则启动定时器等待通告报文的到来。

(2) 在抢占方式下，当 Backup 路由器收到 VRRP 通告报文后，会将自己的优先级与通告报文中的优先级进行比较。如果大于通告报文中的优先级，则成为 Master 路由器；否则将保持 Backup 状态。

(3) 在非抢占方式下，只要 Master 路由器没有出现故障，备份组中的路由器始终保持 Master 或 Backup 状态，Backup 路由器即使随后被配置了更高的优先级也不会成为 Master 路由器。

(4) 如果 Backup 路由器的定时器超时后仍未收到 Master 路由器发送来的 VRRP 通告报文，则认为 Master 路由器已经无法正常工作，此时 Backup 路由器会认为自己是 Master 路由器，并对外发送 VRRP 通告报文。备份组内的路由器根据优先级选举出 Master 路由器，承担报文的转发功能。

1.1.6 VRRP 监视接口功能

VRRP 的监视接口功能更好地扩充了备份功能：不仅能在备份组中某路由器的接口出现故障时提供备份功能，还能在路由器的其它接口（如连接上行链路的接口）不可用时提供备份功能。

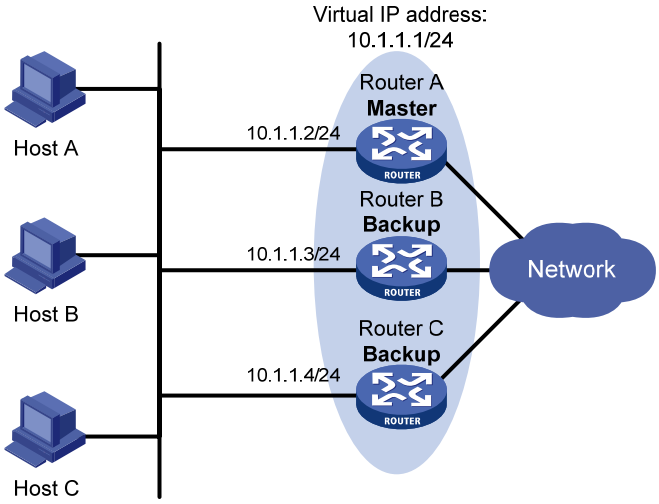
路由器连接上行链路的接口出现故障时，备份组无法感知上行链路的故障，如果该路由器此时处于 Master 状态，将会导致局域网内的主机无法访问外部网络。通过监视指定接口的功能，可以解决该问题。当连接上行链路的接口处于 down 状态时，路由器主动降低自己的优先级，使得备份组内其它路由器的优先级高于这个路由器，以便优先级最高的路由器成为 Master，承担转发任务。

1.1.7 VRRP 应用（以基于 IPv4 的 VRRP 为例）

1. 主备备份

主备备份方式表示业务仅由Master路由器承担。当Master路由器出现故障时，才会从其他Backup路由器选举出一个接替工作。主备备份方式仅需要一个备份组，不同路由器在该备份组中拥有不同优先级，优先级最高的路由器将成为Master路由器，如 图 1-5中所示。

图1-5 主备备份 VRRP



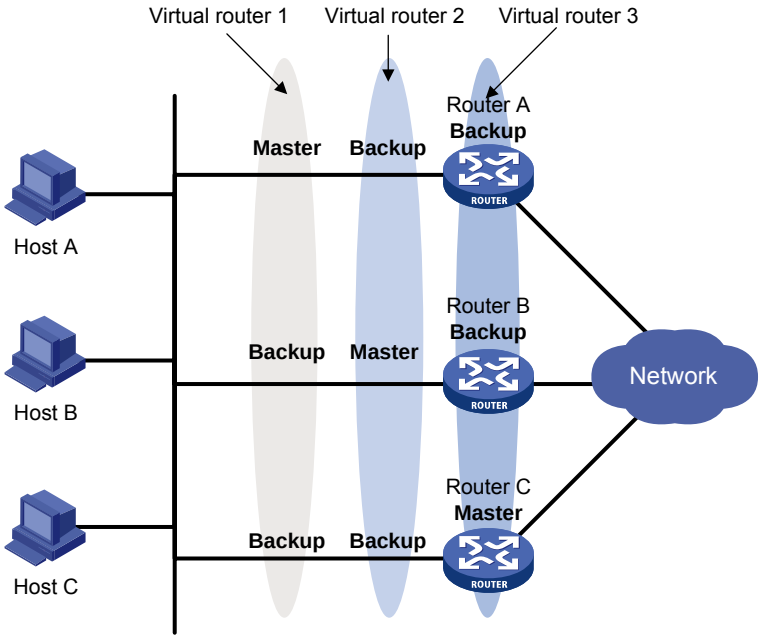
初始情况下，Router A 为 Master 路由器并承担转发任务，Router B 和 Router C 是 Backup 路由器且都处于就绪监听状态。如果 Router A 发生故障，则备份组内处于 Backup 状态的 Router B 和 Router C 路由器将根据优先级选出一个新的 Master 路由器，这个新 Master 路由器继续向网络内的主机提供路由服务。

2. 负载分担

在路由器的一个接口上可以创建多个备份组，使得该路由器可以在一个备份组中作为 Master 路由器，在其他的备份组中作为 Backup 路由器。

负载分担方式是指多台路由器同时承担业务，因此负载分担方式需要两个或者两个以上的备份组，每个备份组都包括一个Master路由器和若干个Backup路由器，各备份组的Master路由器各不相同，如 图 1-6中所示。

图1-6 负载分担 VRRP



同一台路由器同时加入多个 VRRP 备份组，在不同备份组中有不同的优先级。

在 图 1-6中，有三个备份组存在：

- 备份组 1：对应虚拟路由器 1。Router A 作为 Master 路由器，Router B 和 Router C 作为 Backup 路由器。
- 备份组 2：对应虚拟路由器 2。Router B 作为 Master 路由器，Router A 和 Router C 作为 Backup 路由器。
- 备份组 3：对应虚拟路由器 3。Router C 作为 Master 路由器，Router A 和 Router B 作为 Backup 路由器。

为了实现业务流量在 Router A、Router B 和 Router C 之间进行负载分担，需要将局域网内的主机的缺省网关分别设置为虚拟路由器 1、2 和 3。在配置优先级时，需要确保三个备份组中各路由器的 VRRP 优先级形成交叉对应。

1.2 配置基于 IPv4 的 VRRP

1.2.1 基于 IPv4 的 VRRP 配置任务简介

在备份组内的每个路由器上都需进行配置，才能形成一个备份组。

表1-1 VRRP 配置任务简介

配置任务	说明	详细配置
配置虚拟 IP 地址和 MAC 地址的对应关系	可选	1.2.2
创建备份组并配置虚拟 IP 地址	必选	1.2.3
配置备份组优先级、抢占方式及监视功能	可选	1.2.4
配置 VRRP 报文的相关属性	可选	1.2.5
开启 VRRP 的 Trap 功能	可选	1.2.6

1.2.2 配置虚拟 IP 地址和 MAC 地址的对应关系

配置备份组虚拟 IP 地址和 MAC 地址的对应关系后，Master 路由器将配置的 MAC 地址作为发送报文的源 MAC 地址，以便内部网络的主机学习到 IP 地址和 MAC 地址的对应关系，将发往其他网段的报文正确转发给 Master 路由器。

虚拟 IP 地址和 MAC 地址的对应关系有两种：

- 虚拟 IP 地址和虚拟路由器的 MAC 地址对应。缺省情况下，创建备份组后，会自动生成与之对应的虚拟 MAC 地址，虚拟 IP 地址与此虚拟 MAC 地址对应。如果采用这种对应关系，Master 路由器改变时，内部网络的主机不需要更新 IP 地址与 MAC 地址的绑定。
- 虚拟 IP 地址和接口的实际 MAC 地址相对应。当备份组中存在 IP 地址拥有者时，如果配置虚拟 IP 地址和虚拟 MAC 地址对应，会造成一个 IP 地址对应两个 MAC 地址。因此用户可以配置备份组虚拟 IP 地址和实际 MAC 地址对应，主机发送的报文将按照实际 MAC 地址转发给 IP 地址拥有者。

表1-2 配置 MAC 地址和虚拟 IP 地址的对应关系

操作	命令	说明
进入系统视图	**system-view**	-
配置虚拟 IP 地址和 MAC 地址的对应关系	**vrrp method { real-mac | virtual-mac }**	可选 缺省情况下，采用虚拟路由器的 MAC 地址和虚拟 IP 地址对应

注意

本配置需要在备份组创建之前就进行设定。如果路由器上已经创建了备份组，则不允许修改虚拟 IP 地址和 MAC 地址的对应关系。

1.2.3 创建备份组并配置虚拟 IP 地址

创建 VRRP 备份组的同时，需要配置备份组的虚拟 IP 地址。如果接口连接多个子网，则可以为一个备份组配置多个虚拟 IP 地址，以便实现不同子网中路由器的备份。

为备份组指定第一个虚拟 IP 地址时，VRRP 备份组就会自动生成。以后用户再给这个备份组指定虚拟 IP 地址时，VRRP 备份组仅将这个 IP 地址添加到它的备份组虚拟 IP 地址列表中。

注意

建议不要在 Super VLAN 对应的 VLAN 接口下创建 VRRP 备份组，以免对网络性能造成影响。

1. 配置准备

在接口上创建备份组并配置虚拟 IP 地址之前，需要配置接口的 IP 地址，并且保证随后配置的虚拟 IP 地址与接口的 IP 地址在同一网段。

2. 创建备份组并配置虚拟 IP 地址

表1-3 创建备份组并配置虚拟 IP 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
创建备份组，并配置备份组的虚拟 IP 地址	**vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*	必选 缺省情况下，没有创建备份组

注意

- 备份组中所有虚拟 IP 地址都被删除后，该备份组也将同时被删除掉，并且该备份组的所有配置都不再有效。
- 备份组的虚拟 IP 地址可以是备份组所在网段中未被分配的 IP 地址，也可以和备份组内的某个路由器的接口 IP 地址相同。接口 IP 地址与虚拟 IP 地址相同的路由器被称为“IP 地址拥有者”。
- 删除 IP 地址拥有者上的 VRRP 备份组，将导致地址冲突。建议通过修改 IP 地址拥有者的接口 IP 地址的方式解决地址冲突。
- 备份组的虚拟 IP 地址不能为全零地址(0.0.0.0)、广播地址(255.255.255.255)、环回地址、非 A/B/C 类地址和其它非法 IP 地址(如 0.0.0.1)。
- 配置的虚拟 IP 地址和接口 IP 地址在同一网段，且为合法的主机地址时，备份组才能够正常工作；否则，如果配置的虚拟 IP 地址和接口 IP 地址不在同一网段，或为接口 IP 地址所在网段的网络地址或网络广播地址，虽然可以配置成功，但是备份组会始终处于 Initialize 状态，此状态下 VRRP 不起作用。

1.2.4 配置备份组优先级、抢占方式及监视功能

1. 配置准备

在配置备份组优先级、抢占方式及监视功能之前，需要先在接口上创建备份组并配置虚拟 IP 地址。

2. 配置过程

通过优先级、抢占方式和监视指定接口或Track项的配置，可以决定备份组中哪个路由器作为Master路由器。下面这些配置是可选的，可以根据实际需要进行配置。

表1-4 配置备份组优先级、抢占方式及监视功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置路由器在备份组中的优先级	**vrrp vrid** *virtual-router-id* **priority** *priority-value*	可选 缺省情况下，路由器在备份组中的优先级为 100
配置备份组中的路由器工作在抢占方式，并配置抢占延迟时间	**vrrp vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	可选 缺省情况下，备份组中的路由器工作在抢占方式，抢占延迟时间为 0 秒

操作	命令	说明
配置监视指定接口	**vrrp vrid** *virtual-router-id* **track interface** *interface-type interface-number* [**reduced** *priority-reduced*]	可选 缺省情况下，没有指定被监视的接口

注意

- IP 地址拥有者的运行优先级始终为 255，无需用户配置；IP 地址拥有者始终工作在抢占方式。
- 不允许对 IP 地址拥有者进行监视指定接口或 Track 项的配置。
- 被监视的接口可以是三层以太网接口、同步串口、POS 接口、MP-group 接口。目前要求，被监视的同步串口的链路层协议只能为 PPP 协议，且不能加入虚拟模板和 MP-group。
- 被监视接口的状态由 down 变为 up 后，对应设备的优先级数会自动恢复。
- 被监视 Track 项的状态由 negative 变为 positive 后，对应的路由器优先级会自动恢复。

1.2.5 配置 VRRP 报文的相关属性

1. 配置准备

在配置 VRRP 报文的相关属性之前，需要先在接口上创建备份组并配置虚拟 IP 地址。

2. 配置过程

表1-5 配置 VRRP 报文的相关属性

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置备份组发送和接收 VRRP 报文的认证方式和认证字	**vrrp vrid** *virtual-router-id* **authentication-mode** { **md5** | **simple** } *key*	可选 缺省情况下，不进行认证
配置备份组中 Master 路由器发送 VRRP 通告报文的时间间隔	**vrrp vrid** *virtual-router-id* **timer advertise** *adver-interval*	可选 缺省情况下，备份组中 Master 路由器发送 VRRP 通告报文的时间间隔为 1 秒
禁止检查 VRRP 报文的 TTL 域	**vrrp un-check ttl**	可选 缺省情况下，检查 VRRP 报文的 TTL 域 进行此配置之前，不需要创建备份组

说明

- 一个接口上的不同备份组可以设置不同的认证方式和认证字；加入同一备份组的成员需要设置相同的认证方式和认证字。
- 如果网络流量过大或者不同的路由器上的定时器差异等因素，会导致 Backup 路由器的定时器异常超时而发生状态转换。对于这种情况，可以通过将 VRRP 通告报文的发送时间间隔延长和设置抢占延迟时间的办法来解决。

1.2.6 开启 VRRP 的 Trap 功能

开启 VRRP 模块的 Trap 功能后，该模块会生成级别为 errors 的 Trap 报文，用于报告该模块的重要事件。生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-6 开启 VRRP 的 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 VRRP 模块的 Trap 功能	**snmp-agent trap enable vrrp [authfailure | newmaster]**	可选 缺省情况下，VRRP 模块的 Trap 功能处于开启状态



说明

snmp-agent trap enable vrrp 命令的详细介绍请参见“系统分册/SNMP 命令”中的 **snmp-agent trap enable** 命令。

1.2.7 基于 IPv4 的 VRRP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示基于 IPv4 的 VRRP 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除基于 IPv4 的 VRRP 统计信息。

表1-7 VRRP 显示和维护

操作	命令
显示 VRRP 备份组的状态信息	**display vrrp** [**verbose**] [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
显示 VRRP 备份组的统计信息	**display vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
清除 VRRP 备份组的统计信息	**reset vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]

1.3 配置基于 IPv6 的 VRRP

1.3.1 基于 IPv6 的 VRRP 配置任务简介

在备份组内的每个路由器上都需进行配置，才能形成一个备份组。

表1-8 VRRP 配置任务简介

配置任务	说明	详细配置
配置虚拟 IPv6 地址和 MAC 地址的对应关系	可选	1.3.2
创建 VRRP 备份组并配置虚拟 IPv6 地址	必选	1.3.3

配置任务	说明	详细配置
配置备份组优先级、抢占方式及监视指定接口	可选	1.3.4
配置 VRRP 报文的相关属性	可选	1.3.5

1.3.2 配置虚拟 IPv6 地址和 MAC 地址的对应关系

配置备份组虚拟 IPv6 地址和 MAC 地址的对应关系后，Master 路由器将配置的 MAC 地址作为发送报文的源 MAC 地址，以便内部网络的主机学习到 IPv6 地址和 MAC 地址的对应关系，将发往其他网段的报文正确转发给 Master 路由器。

虚拟 IPv6 地址和 MAC 地址的对应关系有两种：

- 虚拟 IPv6 地址和虚拟路由器的 MAC 地址对应。缺省情况下，创建备份组后，会自动生成与之对应的虚拟 MAC 地址，虚拟 IPv6 地址与此虚拟 MAC 地址对应。如果采用这种对应关系，Master 路由器改变时，内部网络的主机不需要更新 IPv6 地址与 MAC 地址的绑定。
- 虚拟 IPv6 地址和接口的实际 MAC 地址相对应。当备份组中存在 IP 地址拥有者时，如果配置虚拟 IPv6 地址和虚拟 MAC 地址对应，会造成一个 IPv6 地址对应两个 MAC 地址。因此用户可以配置备份组虚拟 IPv6 地址和实际 MAC 地址对应，主机发送的报文将按照实际 MAC 地址转发给 IP 地址拥有者。

表1-9 配置 MAC 地址和虚拟 IPv6 地址的对应关系

操作	命令	说明
进入系统视图	**system-view**	-
配置虚拟 IPv6 地址和 MAC 地址的对应关系	**vrrp ipv6 method { real-mac | virtual-mac }**	可选 缺省情况下，采用备份组的虚拟 MAC 地址和虚拟 IPv6 地址对应

注意

本配置需要在备份组创建之前就进行设定。如果路由器上已经创建了备份组，则不允许修改虚拟 IP 地址和 MAC 地址的对应关系。

1.3.3 创建备份组并配置虚拟 IPv6 地址

创建 VRRP 备份组的同时，需要配置备份组的虚拟 IPv6 地址。可以为一个备份组配置多个虚拟 IPv6 地址。

为备份组指定第一个虚拟 IPv6 地址时，VRRP 备份组就会自动生成。以后用户再给这个备份组指定虚拟 IPv6 地址时，VRRP 备份组仅仅将这个 IPv6 地址添加到它的备份组虚拟 IPv6 地址列表中。

注意

建议不要在 Super VLAN 对应的 VLAN 接口下创建 VRRP 备份组，以免对网络性能造成影响。

1. 配置准备

在接口上创建备份组并配置虚拟 IPv6 地址之前，需要配置接口的 IPv6 地址，并且保证随后配置的虚拟 IPv6 地址与接口的 IPv6 地址在同一网段。

2. 创建备份组并配置虚拟 IPv6 地址

表1-10 创建备份组并配置虚拟 IPv6 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
创建备份组，并配置备份组的虚拟 IPv6 地址	**vrrp ipv6 vrid** *virtual-router-id* **virtual-ip** *virtual-address* [**link-local**]	必选 缺省情况下，没有创建备份组 备份组的第一个虚拟 IPv6 地址必须是链路本地地址，并且每个备份组只允许有一个链路本地地址，该地址必须最后一个删除

注意

- 备份组中所有虚拟 IPv6 地址都被删除后，该备份组也将同时被删除掉，并且该备份组的所有配置都不再有效。
- 删除 IP 地址拥有者上的 VRRP 备份组，将导致地址冲突。建议通过修改 IP 地址拥有者的接口 IPv6 地址的方式解决地址冲突。

1.3.4 配置备份组优先级、抢占方式及监视指定接口

1. 配置准备

在配置备份组优先级、抢占方式及监视指定接口之前，需要先在接口上创建备份组并配置虚拟 IPv6 地址。

2. 配置过程

通过优先级、抢占方式和监视指定接口的配置，可以决定备份组中哪个路由器作为 Master 路由器。下面这些配置是可选的，可以根据实际需要进行配置。

表1-11 配置备份组优先级、抢占方式及监视指定接口

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置路由器在备份组中的优先级	**vrrp ipv6 vrid** *virtual-router-id* **priority** *priority-value*	可选 缺省情况下，路由器在备份组中的优先级为 100
配置备份组中的路由器工作在抢占方式，并配置抢占延迟时间	**vrrp ipv6 vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	可选 缺省情况下，备份组中的路由器工作在抢占方式，抢占延迟时间为 0 秒

操作	命令	说明
配置监视指定接口	**vrrp ipv6 vrid** *virtual-router-id* **track interface** *interface-type interface-number* [**reduced** *priority-reduced*]	可选 缺省情况下，没有指定被监视的接口

注意

- IP 地址拥有者的运行优先级始终为 255，无需用户配置；IP 地址拥有者始终工作在抢占方式。
- 不允许对 IP 地址拥有者进行监视指定接口的配置。
- 被监视的接口可以是三层以太网接口、同步串口、POS 接口、MP-group 接口。目前，被监视的同步串口的链路层协议只能为 PPP 协议，且不能加入虚拟模板和 MP-group。
- 被监视接口的状态由 down 变为 up 后，对应设备的优先级数会自动恢复。

1.3.5 配置 VRRP 报文的相关属性

1. 配置准备

在配置 VRRP 报文的相关属性之前，需要先在接口上创建备份组并配置虚拟 IPv6 地址。

2. 配置过程

表1-12 配置 VRRP 报文的相关属性

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置备份组发送和接收 VRRP 报文的认证方式和认证字	**vrrp ipv6 vrid** *virtual-router-id* **authentication-mode simple** *key*	可选 缺省情况下，不进行认证
配置备份组中 Master 路由器发送 VRRP 通告报文的时间间隔	**vrrp ipv6 vrid** *virtual-router-id* **timer advertise** *adver-interval*	可选 缺省情况下，备份组中 Master 路由器发送 VRRP 通告报文的时间间隔为 100 厘秒

说明

- 一个接口上的不同备份组可以设置不同的认证方式和认证字；加入同一备份组的成员需要设置相同的认证方式和认证字。
- 如果网络流量过大或者不同的路由器上的定时器差异等因素，会导致 Backup 路由器的定时器异常超时而发生状态转换。对于这种情况，可以通过将 VRRP 通告报文的发送时间间隔延长和设置抢占延迟时间的办法来解决。

1.3.6 基于 IPv6 的 VRRP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示基于 IPv6 的 VRRP 配置后的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除基于 IPv6 的 VRRP 统计信息。

表1-13 VRRP 显示和维护

操作	命令
显示 VRRP 备份组的状态信息	**display vrrp ipv6** [**verbose**] [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
显示 VRRP 备份组的统计信息	**display vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
清除 VRRP 备份组的统计信息	**reset vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]

1.4 基于 IPv4 的 VRRP 典型配置举例

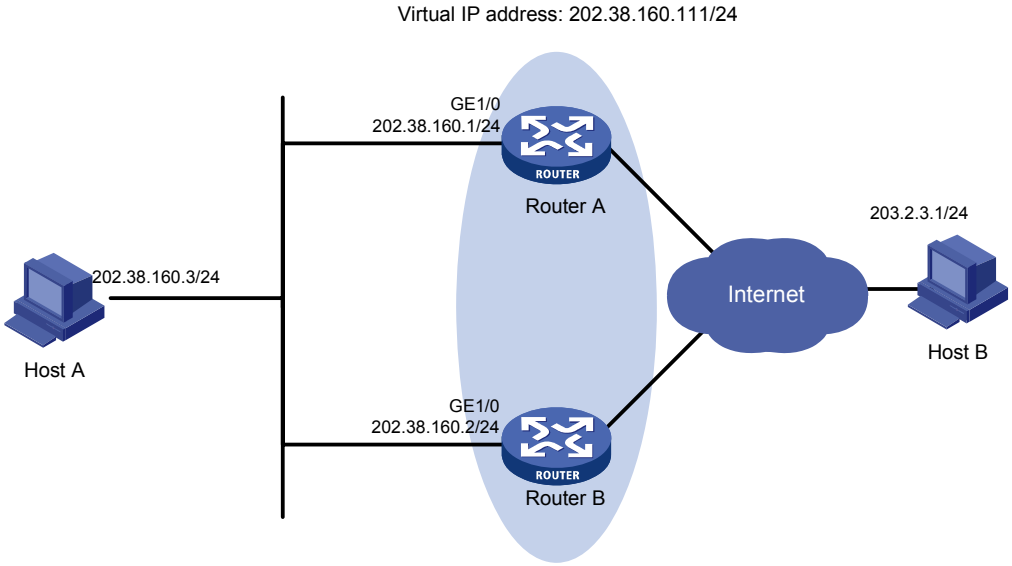
1.4.1 VRRP 单备份组配置举例

1. 组网需求

- Host A 需要访问 Internet 上的 Host B，Host A 的缺省网关为 202.38.160.111/24；
- Router A 和 Router B 属于虚拟 IP 地址为 202.38.160.111/24 的备份组 1；
- 当 Router A 正常工作时，Host A 发送给 Host B 的报文通过 Router A 转发；当 Router A 出现故障时，Host A 发送给 Host B 的报文通过 Router B 转发。

2. 组网图

图1-7 VRRP 单备份组配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ip address 202.38.160.1 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

配置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 priority 110
```

配置 Router A 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 preempt-mode timer delay 5
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface GigabitEthernet1/0
[RouterB-GigabitEthernet1/0] ip address 202.38.160.2 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

配置 Router B 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 preempt-mode timer delay 5
```

(3) 验证配置结果

配置完成后，在 Host A 上可以 ping 通 Host B。通过 **display vrrp verbose** 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers: 1
 Interface       : GigabitEthernet1/0
 VRID            : 1                   Adver. Timer    : 1
 Admin Status    : Up                  State           : Master
 Config Pri      : 110                 Running Pri     : 110
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : None
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1
```

显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers: 1
 Interface       : GigabitEthernet1/0
 VRID            : 1                   Adver. Timer    : 1
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Running Pri     : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : None
 Virtual IP      : 202.38.160.111
 Master IP       : 202.38.160.1
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，Host A 发送给 Host B 的报文通过 Router A 转发。

Router A 出现故障后，在 Host A 上仍然可以 ping 通 Host B。通过 **display vrrp verbose** 命令查看 Router B 上备份组的详细信息。

Router A 出现故障后，显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
 Interface       : GigabitEthernet1/0
 VRID            : 1                    Adver. Timer    : 1
 Admin Status    : Up                   State           : Master
 Config Pri      : 100                  Running Pri     : 100
 Preempt Mode    : Yes                  Delay Time      : 5
 Auth Type       : None
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.2
```

以上显示信息表示 Router A 出现故障后，Router B 成为 Master 路由器，Host A 发送给 Host B 的报文通过 Router B 转发。

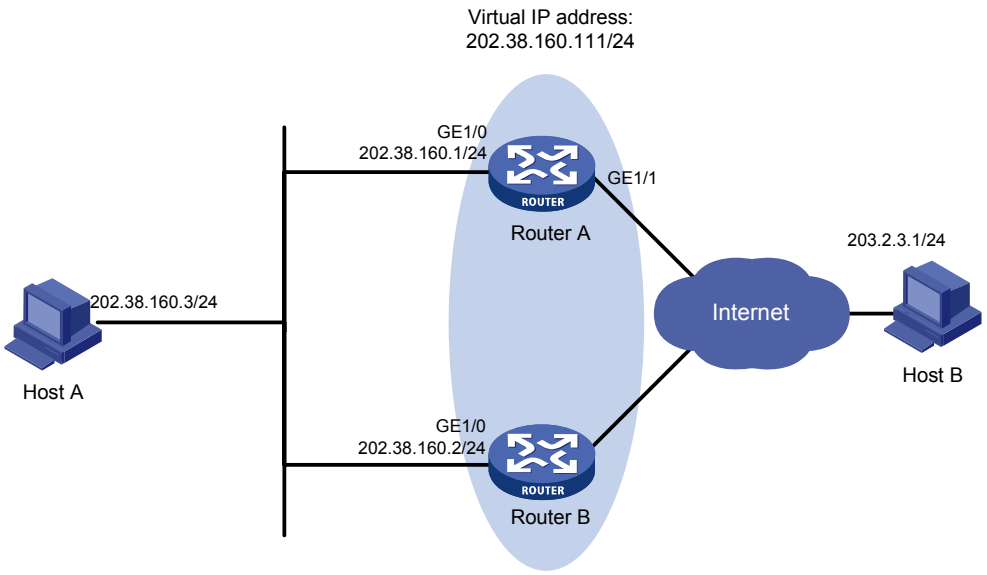
1.4.2 VRRP 监视接口配置举例

1. 组网需求

- Host A 需要访问 Internet 上的 Host B，Host A 的缺省网关为 202.38.160.111/24；
- Router A 和 Router B 属于虚拟 IP 地址为 202.38.160.111/24 的备份组 1；
- 当 Router A 正常工作时，Host A 发送给 Host B 的报文通过 Router A 转发；当 Router A 连接 Internet 的接口 GigabitEthernet1/1 不可用时，Host A 发送给 Host B 的报文通过 Router B 转发。

2. 组网图

图1-8 VRRP 监视接口配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ip address 202.38.160.1 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

设置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 priority 110
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 authentication-mode simple hello
```

设置 Master 路由器发送 VRRP 报文的间隔时间为 5 秒。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 timer advertise 5
```

设置 Router A 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 preempt-mode timer delay 5
```

设置监视接口。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 track interface GigabitEthernet 1/1 reduced 30
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/0
[RouterB-GigabitEthernet1/0] ip address 202.38.160.2 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 authentication-mode simple hello
```

设置 Master 路由器发送 VRRP 报文的间隔时间为 5 秒。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 timer advertise 5
```

设置 Router B 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 preempt-mode timer delay 5
```

(3) 验证配置结果

配置完成后，在 Host A 上可以 ping 通 Host B。通过 **display vrrp verbose** 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers: 1
   Interface       : GigabitEthernet 1/0
     VRID            : 1                   Adver. Timer    : 5
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : Simple              Key             : hello
     Virtual IP      : 202.38.160.111
     Virtual MAC     : 0000-5e00-0101
     Master IP       : 202.38.160.1
   VRRP Track Information:
     Track Interface: GE1/1              State : Up         Pri Reduced : 30
```

显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 5
     Admin Status    : Up                  State           : Backup
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : Simple              Key             : hello
     Virtual IP      : 202.38.160.111
     Master IP       : 202.38.160.1
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，Host A 发送给 Host B 的报文通过 Router A 转发。

Router A 连接 Internet 的接口 GigabitEthernet1/1 不可用时，在 Host A 上仍然可以 ping 通 Host B。通过 **display vrrp verbose** 命令查看备份组的信息。

Router A 的接口 GigabitEthernet1/1 不可用时，显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp verbose
```

```
IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 5
     Admin Status    : Up                  State           : Backup
     Config Pri      : 110                 Running Pri     : 80
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : Simple              Key             : hello
     Virtual IP      : 202.38.160.111
     Master IP       : 202.38.160.2
   VRRP Track Information:
     Track Interface : GE1/1               State : Down       Pri Reduced : 30
```

Router A 的接口 GigabitEthernet1/1 不可用时，显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers: 1
   Interface       : GigabitEthernet 1/0
     VRID            : 1                   Adver. Timer    : 5
     Admin Status    : Up                  State           : Master
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : Simple              Key             : hello
     Virtual IP      : 202.38.160.111
     Virtual MAC     : 0000-5e00-0101
     Master IP       : 202.38.160.2
```

以上显示信息表示 Router A 的接口 GigabitEthernet1/1 不可用时，Router A 的优先级降低为 80，成为 Backup 路由器，Router B 成为 Master 路由器，Host A 发送给 Host B 的报文通过 Router B 转发。

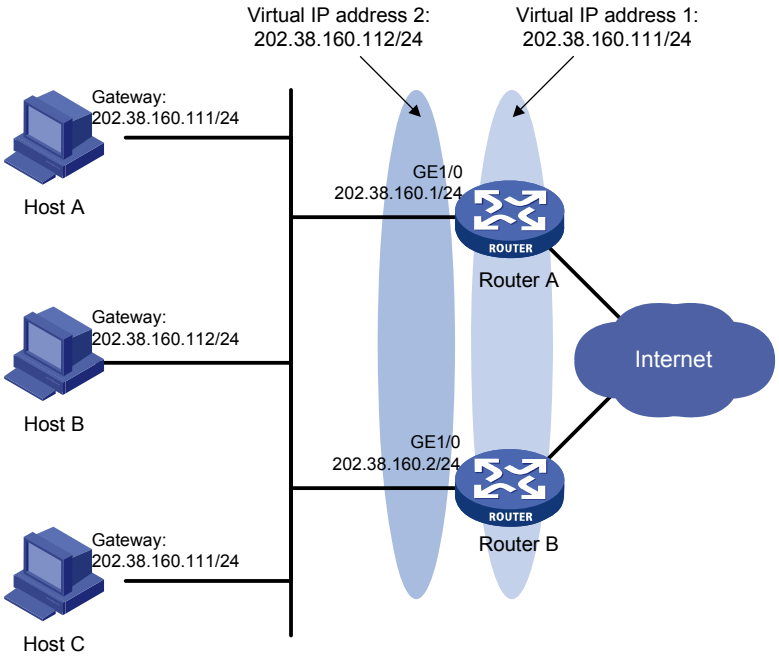
1.4.3 VRRP 多备份组配置举例

1. 组网需求

- 202.38.160.0/24 网段内部分主机的缺省网关为 202.38.160.111/24，部分主机的缺省网关为 202.38.160.112/24；
- 利用 VRRP 备份组实现缺省网关间的负载分担和相互备份。

2. 组网图

图1-9 VRRP 多备份组配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] interface GigabitEthernet1/0
[RouterA-GigabitEthernet1/0] ip address 202.38.160.1 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

设置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 1 priority 110
```

创建备份组 2，并配置备份组 2 的虚拟 IP 地址为 202.38.160.112。

```
[RouterA-GigabitEthernet1/0] vrrp vrid 2 virtual-ip 202.38.160.112
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/0
[RouterB-GigabitEthernet1/0] ip address 202.38.160.2 255.255.255.0
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 202.38.160.111。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 1 virtual-ip 202.38.160.111
```

创建备份组 2，并配置备份组 2 的虚拟 IP 地址为 202.38.160.112。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 2 virtual-ip 202.38.160.112
```

设置 Router B 在备份组 2 中的优先级为 110。

```
[RouterB-GigabitEthernet1/0] vrrp vrid 2 priority 110
```

(3) 验证配置结果

可以通过 **display vrrp verbose** 命令查看配置后的结果。

显示 Router A 上备份组的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp verbose
 IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 2
   Interface       : GigabitEthernet 1/0
     VRID            : 1                   Adver. Timer    : 1
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : 202.38.160.111
     Virtual MAC     : 0000-5e00-0101
     Master IP       : 202.38.160.1

   Interface       : GigabitEthernet1/0
     VRID            : 2                   Adver. Timer    : 1
     Admin Status    : Up                  State           : Backup
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : 202.38.160.112
     Master IP       : 202.38.160.2
```

显示 Router B 上备份组的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp verbose
IPv4 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 2
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 1
     Admin Status    : Up                  State           : Backup
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : 202.38.160.111
     Master IP       : 202.38.160.1

   Interface       : GigabitEthernet1/0
     VRID            : 2                   Adver. Timer    : 1
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : 202.38.160.112
```

```
Virtual MAC      : 0000-5e00-0102
Master IP        : 202.38.160.2
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，缺省网关为 202.38.160.111/24 的主机通过 Router A 访问 Internet；备份组 2 中 Router A 为 Backup 路由器，Router B 为 Master 路由器，缺省网关为 202.38.160.112/24 的主机通过 Router B 访问 Internet。

1.5 基于 IPv6 的 VRRP 典型配置举例

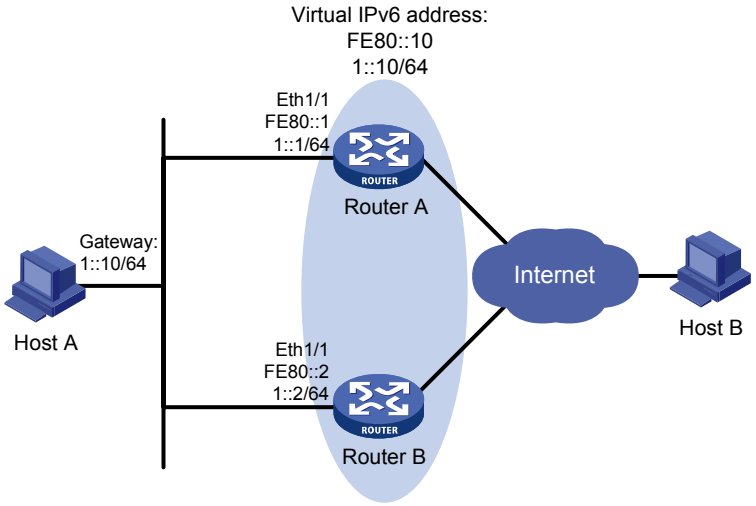
1.5.1 VRRP 单备份组配置举例

1. 组网需求

- Host A 需要访问 Internet 上的 Host B，Host A 的缺省网关为 1::10/64；
- Router A 和 Router B 属于虚拟 IPv6 地址为 1::10/64 和 FE80::10 的备份组 1；
- 当 Router A 正常工作时，Host A 发送给 Host B 的报文通过 Router A 转发；当 Router A 出现故障时，Host A 发送给 Host B 的报文通过 Router B 转发。

2. 组网图

图1-10 VRRP 单备份组配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ipv6 address fe80::1 link-local
[RouterA-GigabitEthernet1/1] ipv6 address 1::1 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterA-Ethernet1/1] vrrp ipv6 vrid 1 virtual-ip 1::10
```

配置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 1 priority 110
```

配置 Router A 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

配置允许发布 RA 消息。

```
[RouterA-GigabitEthernet1/1] undo ipv6 nd ra halt
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ipv6 address fe80::2 link-local
[RouterB-GigabitEthernet1/1] ipv6 address 1::2 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterB-Ethernet1/1] vrrp ipv6 vrid 1 virtual-ip 1::10
```

配置 Router B 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

配置允许发布 RA 消息。

```
[RouterB-GigabitEthernet1/1] undo ipv6 nd ra halt
```

(3) 验证配置结果

配置完成后，在 Host A 上可以 ping 通 Host B。通过 **display vrrp ipv6 verbose** 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/1] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/1
     VRID            : 1                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : None
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::1
```

显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/1] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/1
```

```
    VRID            : 1                   Adver. Timer    : 100
    Admin Status    : Up                  State           : Backup
    Config Pri      : 110                 Running Pri     : 110
    Preempt Mode    : Yes                 Delay Time      : 5
    Auth Type       : None
    Virtual IP      : FE80::10
                      1::10
    Master IP       : FE80::1
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，Host A 发送给 Host B 的报文通过 Router A 转发。

Router A 出现故障后，在 Host A 上仍然可以 ping 通 Host B。通过 **display vrrp ipv6 verbose** 命令查看 Router B 上备份组的信息。

Router A 出现故障后，显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/1] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/1
     VRID            : 1                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Master
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : None
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::2
```

以上显示信息表示 Router A 出现故障后，Router B 成为 Master 路由器，Host A 发送给 Host B 的报文通过 Router B 转发。

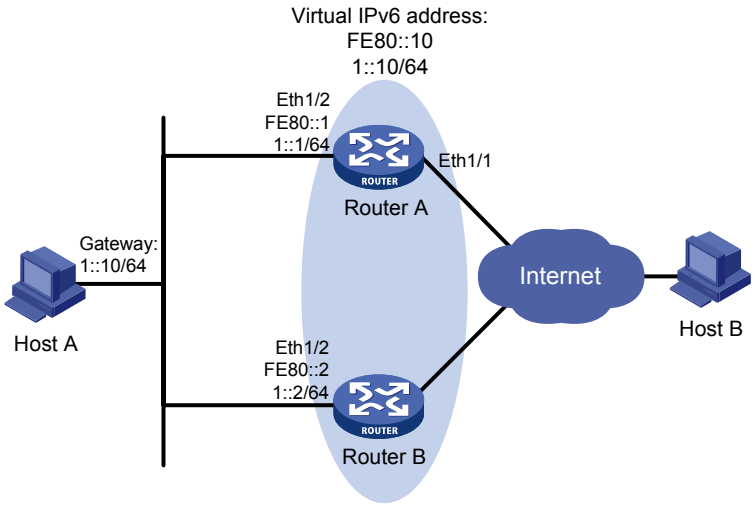
1.5.2 VRRP 监视接口配置举例

1. 组网需求

- Host A 需要访问 Internet 上的 Host B，Host A 的缺省网关为 1::10/64；
- Router A 和 Router B 属于虚拟 IPv6 地址为 1::10/64 和 FE80::10 的备份组 1；
- 当 Router A 正常工作时，Host A 发送给 Host B 的报文通过 Router A 转发；当 Router A 连接 Internet 的接口 GigabitEthernet1/1 不可用时，Host A 发送给 Host B 的报文通过 Router B 转发。

2. 组网图

图1-11 VRRP 监视接口配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ipv6 address fe80::1 link-local
[RouterA-GigabitEthernet1/0] ipv6 address 1::1 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 virtual-ip 1::10
```

设置 Router A 在备份组中的优先级为 110。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 priority 110
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 authentication-mode simple hello
```

设置 VRRP 通告报文发送的间隔时间为 500 厘秒。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 timer advertise 500
```

配置 Router A 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

设置监视接口。

```
[RouterA-GigabitEthernet1/0] vrrp ipv6 vrid 1 track interface gigabitethernet 1/1 reduced
30
```

配置允许发布 RA 消息。

```
[RouterA-GigabitEthernet1/0] undo ipv6 nd ra halt
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] ipv6
```

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] ipv6 address fe80::2 link-local
[RouterB-GigabitEthernet1/0] ipv6 address 1::2 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterB-GigabitEthernet1/0] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterB-GigabitEthernet1/0] vrrp ipv6 vrid 1 virtual-ip 1::10
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterB-GigabitEthernet1/0] vrrp ipv6 vrid 1 authentication-mode simple hello
```

设置 VRRP 通告报文发送的间隔时间为 500 厘秒。

```
[RouterB-GigabitEthernet1/0] vrrp ipv6 vrid 1 timer advertise 500
```

配置 Router B 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterB-GigabitEthernet1/0] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

配置允许发布 RA 消息。

```
[RouterA-GigabitEthernet1/0] undo ipv6 nd ra halt
```

(3) 验证配置结果

配置完成后，在 Host A 上可以 ping 通 Host B。通过 **display vrrp ipv6 verbose** 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 500
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 5
     Auth Type       : Simple              Key             : hello
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::1
   VRRP Track Information:
     Track Interface: GE1/1              State : Up         Pri Reduced : 30
```

显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 500
```

```
    Admin Status    : Up                 State          : Backup
    Config Pri      : 100                Running Pri    : 100
    Preempt Mode    : Yes                Delay Time     : 5
    Auth Type       : Simple             Key            : hello
    Virtual IP      : FE80::10
                      1::10
    Virtual MAC     : 0000-5e00-0201
    Master IP       : FE80::1
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，Host A 发送给 Host B 的报文通过 Router A 转发。

Router A 连接 Internet 的接口 GigabitEthernet1/1 不可用时，在 Host A 上仍然可以 ping 通 Host B。通过 **display vrrp ipv6 verbose** 命令查看备份组的信息。

Router A 的接口 GigabitEthernet1/1 不可用时，显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/0] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                  Adver. Timer   : 500
     Admin Status    : Up                 State          : Backup
     Config Pri      : 110                Running Pri    : 80
     Preempt Mode    : Yes                Delay Time     : 5
     Auth Type       : Simple             Key            : hello
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::2
   VRRP Track Information:
     Track Interface: GE1/1               Pri Reduced    : 30
```

Router A 的接口 GigabitEthernet1/1 不可用时，显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/0] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 1
   Interface       : GigabitEthernet1/0
     VRID            : 1                  Adver. Timer   : 500
     Admin Status    : Up                 State          : Master
     Config Pri      : 100                Running Pri    : 100
     Preempt Mode    : Yes                Delay Time     : 5
     Auth Type       : Simple             Key            : hello
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::2
```

以上显示信息表示 Router A 的接口 GigabitEthernet1/1 不可用时，Router A 的优先级降低为 80，成为 Backup 路由器，Router B 成为 Master 路由器，Host A 发送给 Host B 的报文通过 Router B 转发。

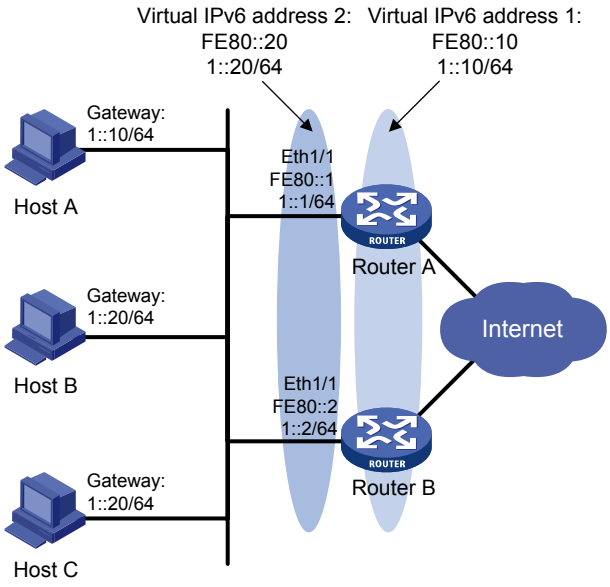
1.5.3 VRRP 多备份组配置举例

1. 组网需求

- 网络内部分主机的缺省网关为 1::10/64，部分主机的缺省网关为 1::20/64；
- 利用 VRRP 备份组实现缺省网关间的负载分担和相互备份。

2. 组网图

图1-12 VRRP 多备份组配置组网图



3. 配置步骤

(1) 配置 Router A

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ipv6 address fe80::1 link-local
[RouterA-GigabitEthernet1/1] ipv6 address 1::1 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterA-Ethernet1/1] vrrp ipv6 vrid 1 virtual-ip 1::10
```

设置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 1 priority 110
```

创建备份组 2，并配置备份组 2 的虚拟 IPv6 地址为 FE80::20 和 1::20。

```
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-local
[RouterA-GigabitEthernet1/1] vrrp ipv6 vrid 2 virtual-ip 1::20
```

配置允许发布 RA 消息。

```
[RouterA-GigabitEthernet1/1] undo ipv6 nd ra halt
```

(2) 配置 Router B

```
<RouterB> system-view
[RouterB] ipv6
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ipv6 address fe80::2 link-local
[RouterB-GigabitEthernet1/1] ipv6 address 1::2 64
```

创建备份组 1，并配置备份组 1 的虚拟 IPv6 地址为 FE80::10 和 1::10。

```
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[RouterB-Ethernet1/1] vrrp ipv6 vrid 1 virtual-ip 1::10
```

创建备份组 2，并配置备份组 2 的虚拟 IPv6 地址为 FE80::20 和 1::20。

```
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-local
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 2 virtual-ip 1::20
```

设置 Router B 在备份组 2 中的优先级为 110。

```
[RouterB-GigabitEthernet1/1] vrrp ipv6 vrid 2 priority 110
```

配置允许发布 RA 消息。

```
[RouterA-GigabitEthernet1/1] undo ipv6 nd ra halt
```

(3) 验证配置结果

可以通过 **display vrrp ipv6 verbose** 命令查看配置后的结果。

显示 Router A 上备份组的详细信息。

```
[RouterA-GigabitEthernet1/1] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 2
   Interface       : GigabitEthernet1/1
     VRID            : 1                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : FE80::10
                       1::10
     Virtual MAC     : 0000-5e00-0201
     Master IP       : FE80::1

   Interface       : GigabitEthernet1/1
     VRID            : 2                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Backup
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : FE80::20
                       1::20
```

```
    Master IP       : FE80::2
```

显示 Router B 上备份组的详细信息。

```
[RouterB-GigabitEthernet1/1] display vrrp ipv6 verbose
 IPv6 Standby Information:
     Run Mode        : Standard
     Run Method      : Virtual-MAC
 Total number of virtual routers : 2
   Interface       : GigabitEthernet1/0
     VRID            : 1                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Backup
     Config Pri      : 100                 Running Pri     : 100
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : FE80::10
                       1::10
     Master IP       : FE80::1

   Interface       : GigabitEthernet1/1
     VRID            : 1                   Adver. Timer    : 100
     Admin Status    : Up                  State           : Master
     Config Pri      : 110                 Running Pri     : 110
     Preempt Mode    : Yes                 Delay Time      : 0
     Auth Type       : None
     Virtual IP      : FE80::20
                       1::20
     Virtual MAC     : 0000-5e00-0202
     Master IP       : FE80::2
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，缺省网关为 1::10/64 的主机通过 Router A 访问 Internet；备份组 2 中 Router A 为 Backup 路由器，Router B 为 Master 路由器，缺省网关为 1::20/64 的主机通过 Router B 访问 Internet。



说明

在实际组网应用中，一般都使用多备份组。IPv6 网络中，主机需要手动配置默认网关才能实现 VRRP 备份组负载分担的功能。

1.6 VRRP 常见错误配置举例

1. 控制台上频频出现配置错误的提示

原因分析：

- 可能是备份组内的另一台路由器配置不一致造成的。
- 可能是有的机器试图发送非法的 VRRP 报文。

解决方法：

- 对于第一种情况，可以通过修改配置来解决。

- 对于第二种情况，则是有些机器有不良企图，应当通过非技术手段来解决。

2. 同一个备份组内出现多个 Master 路由器

原因分析：

- 若短时间内存在多个 Master 路由器，属于正常情况，无需进行人工干预。
- 若多个 Master 路由器长时间共存，这很有可能是由于 Master 路由器之间收不到 VRRP 报文，或者收到的报文不合法造成的。

解决方法：先在多个 Master 路由器之间执行 ping 操作。如果 ping 不通，则检查网络连接是否正确；如果能 ping 通，则检查 VRRP 的配置是否一致。对于同一个 VRRP 备份组的配置，必须要保证虚拟 IP 地址个数、每个虚拟 IP 地址、通告报文的发送时间间隔和认证方式完全一样。

3. VRRP 的状态频繁转换

原因分析：这种情况一般是由于 VRRP 通告报文发送时间间隔太短造成的。

解决方法：增加通告报文的发送时间间隔或者设置抢占延迟都可以解决这种故障。

目 录

1 HA配置 1-1
1.1 HA简介 1-1
1.2 HA配置任务简介 1-1
1.3 忽略对备用板进行版本检查 1-2
1.4 重启备用板 1-2
1.5 手动进行主备倒换 1-2
1.6 HA显示和维护 1-3

1 HA 配置

说明

本特性只有 SR6608 支持。

1.1 HA 简介

HA（High Availability，高可靠性）特性用于实现系统的高可靠性。支持 HA 特性的设备有两块单板，一块为主用板，工作在 Master 模式；另一块为备用板，工作在 Slave 模式。备用板通过同步功能，来保持与主用板当前配置的一致。

HA 主要体现在以下两个方面：

- 当主用板发生故障或者被拔出时，备用板将迅速自动取代主用板成为新的主用板（下面简称该过程为主备倒换），以保证设备的继续运行。
- 当升级设备的时候，可以先升级备用板，再手工执行主备倒换，主备倒换时，原主用板会使用配置的新的下次启动文件重新启动，从而达到升级原主用板的效果。相比整机重启的方式，大大减少了升级过程业务中断的时间。

当使用主备倒换方式升级设备的时候，要求新的软件版本和当前的软件版本能够实现平滑升级（即版本间的差异较小），否则，可能会导致备用板不能启动。这种升级方式的步骤如下：

(1) 将最新的软件版本分别下载到主用板和备用板；

(2) 分别指定主用板和备用板的下次启动文件为新版本的软件；

(3) 忽略对备用板进行版本检查；

(4) 手工重启备用板；

手动进行主备倒换。

注意

用户不能直接对备用板进行命令行操作，需要通过主用板的命令行界面进行配置，然后主用板会将配置同步给备用板。

1.2 HA 配置任务简介

表1-1 HA 配置任务简介

配置任务	说明	详细配置
忽略对备用板进行版本检查	可选	1.3
重启备用板	可选	1.4
手动进行主备倒换	可选	1.5

1.3 忽略对备用板进行版本检查

当备用板上电启动的时候，系统会检测主用板和备用板的版本。当主用板与备用板的版本不一致的时候，备用板将不能启动。用户也可以使用以下配置强制要求系统忽略对备用板进行版本检查。

表1-2 配置 HA

操作	命令	说明
进入系统视图	**system-view**	-
忽略对备用板进行版本检查	**ha slave-ignore-version-check**	必选 缺省情况下，备用板版本检查功能处于使能状态

1.4 重启备用板

当备用板故障或者进行备用板升级时，可以重启备用板，使最新的版本或配置生效。但是在重启前，需要确认主用板和备用板的版本是否一致，如果不一致，需要先配置忽略对备用板进行版本检查。

表1-3 配置 HA

操作	命令	说明
进入系统视图	**system-view**	-
手动重启备用板	**slave restart**	必选

说明

当备用板启动时，主用板会对备用板进行初始同步（即批量备份）。在这个过程中，系统将不会响应用户的输入。初始同步完成后，用户可以在主用板上正常执行命令、进行配置，此时，主用板和备用板将保持实时同步，即用户对主用板的配置将同步到备用板上，以保证主备板当前配置的一致。

1.5 手动进行主备倒换

在进行主用板升级时，需要重启主用板。这时，为了防止业务的中断，可以手动进行主备倒换，让原备用板接替原主用板继续工作。

表1-4 配置 HA

操作	命令	说明
进入系统视图	**system-view**	-
使能手动主备倒换功能	**slave switchover { disable | enable }**	可选 缺省情况下，手动主备倒换功能处于使能状态
手动进行主备倒换	**slave switchover**	必选

注意

主备倒换时，会导致原主用板重启。在进行主备倒换前，请确保主备板的软件版本一致，如果不一致，需要先配置忽略对备用板进行版本检查。

1.6 HA 显示和维护

完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 HA 特性的运行情况，通过查看显示信息验证配置的效果。

表1-5 HA 显示和维护

操作	命令
查看主备倒换的状态	**display switchover state** [**slot** *slot-number*]

目 录

1 设备管理 1-1
1.1 设备管理简介 1-1
1.2 设备管理配置任务简介 1-1
1.3 配置系统异常时的处理方式 1-1
1.4 配置设备重启 1-2
1.5 配置定时自动执行功能 1-3
1.6 升级设备软件 1-4
1.6.1 设备软件简介 1-4
1.6.2 通过命令行升级Bootrom 1-5
1.6.3 通过命令行升级启动文件 1-5
1.6.4 同步升级备用主控板的启动文件 1-6
1.7 单板的热插拔 1-7
1.8 智能电源管理 1-7
1.8.1 配置单板的上电优先级 1-7
1.9 配置单板的温度告警阈值 1-8
1.10 配置接口卡的工作模式 1-8
1.11 清除当前系统中不使用的 16bit接口索引 1-9
1.12 可插拔模块的识别与诊断 1-10
1.12.1 可插拔模块简介 1-10
1.12.2 识别可插拔模块 1-10
1.12.3 诊断可插拔模块 1-10
1.13 配置USB Device端口 1-11
1.13.1 USB Device端口的基本功能 1-11
1.13.2 USB标准的连接器与线缆 1-11
1.13.3 设备与Host的连接 1-12
1.13.4 正确卸载设备的方法 1-13
1.13.5 配置使能USB Device端口 1-14
1.14 设备管理显示和维护 1-14
1.15 设备管理典型配置举例 1-15
1.15.1 远程立即升级配置举例（分布式设备） 1-15
1.15.2 远程定时自动升级配置举例（集中式设备） 1-17
1.15.3 USB Device端口配置典型配置举例 1-19

1 设备管理

1.1 设备管理简介

通过设备管理功能，用户能够查看设备当前的工作状态，配置设备运行的相关参数，实现对设备的日常维护和管理。

目前的设备管理主要提供重启设备、定时重启设备、指定设备下次启动时采用的启动文件和配置单板的温度告警阈值等功能。

1.2 设备管理配置任务简介

表1-1 设备管理配置任务简介

配置任务		说明	详细配置
配置系统异常时的处理方式		可选	1.3
配置设备重启		可选	1.4
配置定时自动执行功能		可选	1.5
升级设备软件	通过命令行升级 Bootrom	可选	1.6.2
	通过命令行升级系统启动文件	可选	1.6.3
	同步升级备用主控板的启动文件	可选	1.6.4
单板的热插拔		可选	1.7
智能电源管理		可选	1.8
配置单板的温度告警阈值		可选	1.9
配置接口卡的工作模式		可选	1.10
清除当前系统中不使用的 16bit 接口索引		可选	1.11
可插拔模块的识别与诊断		可选	1.12
配置 USB Device 端口		可选	1.13

1.3 配置系统异常时的处理方式

当系统检测到软件运行异常时，支持两种处理方式：

- **reboot**：系统通过直接自动重启的方式来进行恢复。
- **maintain**：保持当前状态，系统不会自动采取任何恢复措施。此时需要手工进行恢复操作。某些系统异常可能较难复现，或者异常时打印的一些提示设备重启后会丢失，此时，使用该方式可以保持异常时的状态，以便进行问题定位和修复。但该方式需要手工修复，比如手工重启。

表1-2 配置系统异常时的处理方式（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
配置系统异常时的处理方式	**system-failure** { **maintain** | **reboot** }	可选 缺省情况下，系统异常时的处理方式为 **reboot**

表1-3 配置系统异常时的处理方式（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
配置主用主控板和备用主控板系统异常时的处理方式	**system-failure** { **maintain** | **reboot** }	可选 缺省情况下，主用主控板和备用主控板系统异常时的处理方式为 **reboot**

说明

- 配置本命令后会将主用主控板和备用主控板系统异常时的处理方式设置为相同值。接口板和辅助CPU系统如果出现软件运行异常的情况，系统的处理方式始终是 **reboot**，即重启故障单板。
- 系统异常处理只针对故障单板，不会影响其它单板的功能。（分布式设备）

1.4 配置设备重启

当设备运行出现故障时，用户可以根据实际情况，通过重启设备来排除故障。

重启的方式有三种：

- 通过断电后重新上电立即重启设备（该方式又称为硬件重启或者冷启动）。该方式对设备冲击较大，如果对运行中的设备进行强制断电，可能会造成数据丢失或者硬件损坏。一般情况下，建议不要使用这种方式。
- 通过命令行立即重启设备。
- 通过命令行定时重启设备。该方式下，用户可以设置一个时间，让设备定时自动重启，或者设置一个时延，让设备经过指定时间后自动重启。

后两种方式都属于命令行重启。命令行重启又称为热启动，该操作等效于给设备断电后又上电启动，主要用于远程维护时，可以直接重启设备，而不需要到设备所在地进行硬件重启。

- 集中式设备

表1-4 通过命令行立即重启设备

操作	命令	说明
立即重启设备或者指定子卡	**reboot** [**slot** *slot-number*]	必选 该命令在用户视图下执行

表1-5 通过命令行定时重启设备

操作	命令	说明
开启设备定时重启功能，并指定重启的具体时间和日期	**schedule reboot at** *hh:mm* [*date*]	二者必选其一 缺省情况下，设备定时重启功能处于关闭状态 两命令均在用户视图下执行
开启设备定时重启功能，并指定重启的等待时延	**schedule reboot delay** { *hh:mm* | *mm* }	

- 分布式设备

表1-6 通过命令行立即重启设备

操作	命令	说明
立即重启指定单板、指定子卡或整个设备	**reboot** [**slot** *slot-number* [**subslot** *subslot-number*]]	可选 该命令在用户视图下执行

表1-7 通过命令行定时重启设备

操作	命令	说明
开启整个设备定时重启功能，并指定重启的具体时间和日期	**schedule reboot at** *hh:mm* [*date*]	二者必选其一 缺省情况下，设备定时重启功能处于关闭状态 两命令均在用户视图下执行
开启整个设备的定时重启功能，并设定等待时延	**schedule reboot delay** { *hh:mm* | *mm* }	

注意

- 重新启动会导致业务中断，请谨慎使用。
- 重启前请使用 **save** 命令保存当前配置，以免重启后配置丢失。（**save** 命令的详细介绍请参见“系统分册”中的“文件系统管理命令”）
- 重启前请使用 **display startup** 和 **display boot-loader** 命令分别确认是否设置了合适的下次启动配置文件和下次启动文件。（**display startup** 命令的详细介绍请参见“系统分册”中的“文件系统管理命令”）
- 设备定时器的精度为 1 分钟。在到达用户设定的重启时刻的前一分钟，设备会提示：“REBOOT IN ONE MINUTE（设备将在一分钟后重启）”，并在一分钟后重新启动。
- 对于分布式设备，不能使用该命令重启备用主控板，如需重启备用主控板，请使用 **slave restart** 命令（**slave restart** 命令的详细介绍请参见“系统分册”中的“HA 命令”）；不指定 **slot** 参数时，会重启整个设备（包括主用主控板、备用主控板和接口板）；重启主用主控板而且备用主控板能正常工作时，会触发主备倒换。
- 如果主用启动文件损坏或者不存在，则不能通过 **reboot** 命令重启设备。此时，可以通过指定新的主用启动文件再重启，或者断电后重新上电，系统将自动使用备用启动文件重启。
- 如果设备在准备重启时，用户正在进行文件操作，为了安全起见，系统将不会执行此次重启操作。

1.5 配置定时自动执行功能

定时自动执行功能是指在指定时间、系统自动执行指定视图下的指定命令。该功能主要用来对系统进行定时升级或定时配置。

表1-8 配置定时自动执行功能

操作	命令	说明
在指定时间自动执行指定命令	**schedule job at** *time* [*date*] **view** *view command*	二者必选其一 该命令在用户视图下执行 可选
在指定延时后自动执行指定命令	**schedule job delay** *time* **view** *view command*	

需要注意的是：

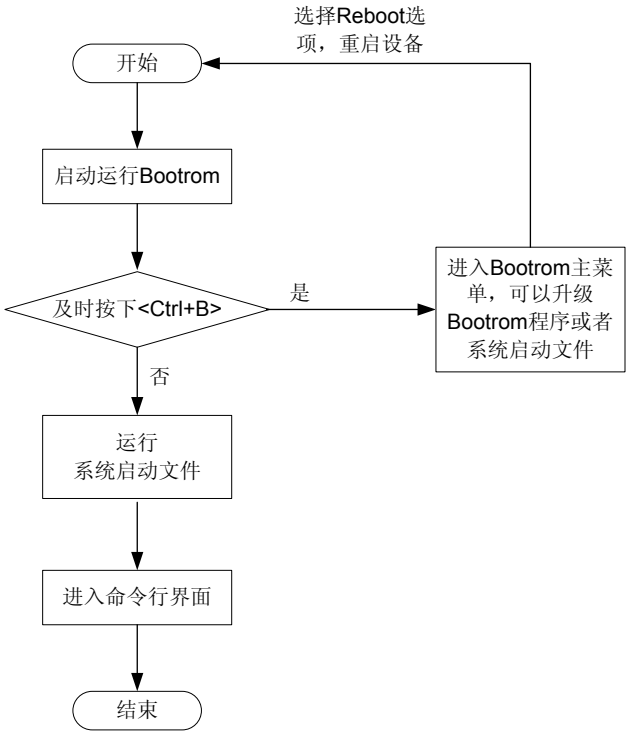
- 当前用户可指定的视图只有用户视图和系统视图。如果要自动执行其他视图下的指定命令或者一次定时自动执行多条命令，可以通过定时自动执行批处理文件来实现（注意在指定批处理文件时必须使用全路径的文件名形式）。
- 系统不对 *view* 和 *command* 参数的值进行检查，请确保 *command* 参数的正确性（包括 *command* 格式正确以及 *command* 与 *view* 的关系正确）。
- 设置的时间点到达时，系统将在后台执行指定命令，不显示任何输出信息（log、trap、debug等系统信息输出除外）。
- 指定命令执行时不进行信息交互。当需要用户交互确认时，系统将自动输入“Y”或“Yes”；当需要用户交互输入字符信息时，系统将自动输入缺省字符串，没有缺省字符串的将自动输入空字符串。
- 对于切换用户操作界面的命令（如 **telnet**、**ftp**、**ssh2** 等）、切换视图的命令（如 **system-view**、**quit**、**interface ethernet** 等）以及修改执行命令用户状态的命令（如 **super** 命令等），自动执行命令后当前用户的操作界面、命令视图和用户状态不变。
- 如果配置定时自动执行功能后，又对系统时间进行了调整，则定时自动执行配置将自动失效。
- 多次执行 **schedule job** 命令时，只有最新的配置生效。
- 对于分布式设备，在主用主控板配置该特性后，**schedule job** 命令不会备份到 Slave 设备，主备倒换后，相当于没有配置该功能。

1.6 升级设备软件

1.6.1 设备软件简介

设备软件主要包括Bootrom程序和系统启动文件。设备开机上电后，Bootrom软件先运行，对DDR等存储器进行尺寸检查和测试，初始化硬件并显示设备的硬件参数，然后引导运行启动文件；启动文件一方面提供对主要部件的硬件驱动和适配功能，另一方面实现了转发、VLAN、SNMP等业务特性。Bootrom程序与系统启动文件是设备启动、运行的必备软件，为整个设备提供支撑、管理、业务等功能，它们的关系如 图 1-1所示。

图1-1 Bootrom 程序与系统启动文件关系示意图



Bootrom 程序和系统启动文件都可以通过 Bootrom 菜单和命令行两种方式来升级，本文只介绍命令行升级方式，通过 Bootrom 菜单升级的具体操作请参见设备的安装手册。

1.6.2 通过命令行升级 Bootrom

请按以下步骤升级 Bootrom：

(1) 使用 FTP 或者 TFTP，将 Bootrom 程序拷贝到设备存储介质的根目录下。

(2) 使用命令行指定下次启动时使用的 Bootrom 程序。

(3) 重启设备，使新的 Bootrom 程序生效。

对于分布式设备，由于不同设备主控板和接口板的 Bootrom 各不相同，用户容易混淆，从而导致 Bootrom 升级错误。通过启动 Bootrom 升级时的合法性检查功能，设备就能够对 Bootrom 升级文件的正确性、版本配套性等进行严格的检查，保证升级成功。

表1-9 升级 Bootrom

操作	命令	说明
进入系统视图	**system-view**	-
读取、恢复、备份或升级设备或子卡的 Bootrom 程序（集中式设备）	**bootrom** { **read** | **restore** | **backup** | **update file** *file-url* } [**slot** *slot-number-list*] [**all** | **part**]	必选 缺省情况下，操作 Bootrom 的全部内容 该命令在用户视图下执行
读取、恢复、备份或升级指定单板或子卡的 Bootrom 程序（分布式设备）	**bootrom** { **read** | **restore** | **backup** | **update file** *file-url* } **slot** *slot-number-list* [**subslot** *subslot-number-list*] [**all** | **part**]	必选 缺省情况下，操作 Bootrom 的全部内容 该命令在用户视图下执行

1.6.3 通过命令行升级启动文件

采用本特性升级启动文件时，请按以下步骤进行（集中式设备）：

(1) 使用 FTP、TFTP 或者其它方式将启动文件下保存到设备存储介质的根目录下。

(2) 使用命令行指定设备下次启动时使用的启动文件（以下简称为下次启动文件）。

(3) 重启设备，使新的启动文件生效。采用本特性升级启动文件时，请按以下步骤进行：

采用本特性升级启动文件时，请按以下步骤进行（分布式设备）：

(1) 使用 FTP、TFTP 或者其它方式将启动文件下保存到主用主控板存储介质的根目录下。

(2) 将新的启动文件拷贝到备用主控板存储介质的根目录下。

(3) 使用命令行分别指定主用主控板和备用主控板下次启动时使用的启动文件（以下简称为下次启动文件）。

(4) 重启设备，使新的启动文件生效。

表1-10 指定下次启动文件（集中式设备）

操作	命令	说明
指定设备的下次启动文件	**boot-loader file** *file-url* **{ main | backup }**	必选 该命令在用户视图下执行

表1-11 指定下次启动文件（分布式设备）

操作	命令	说明
指定主用主控板或者备用主控板的下次启动文件	**boot-loader file** *file-url* **slot** *slot-number* **{ main | backup }**	必选 该命令在用户视图下执行

注意

- 下次启动文件必须存放在设备的根目录下（对于支持存储设备分区的设备，该文件必须存放在第一个分区内）。可使用文件的拷贝或移动操作来调整文件的路径为根目录。
- 对于分布式设备，必须先将下次启动文件保存到相应主控板存储介质的根目录下，**boot-loader** 命令才能执行成功。
- 对于分布式设备，主用主控板和备用主控板的下次启动文件名称可以不同，但是版本必须一致，否则，可能会导致设备不能正常启动。

1.6.4 同步升级备用主控板的启动文件

在使用分布式设备时，如果主用主控板和备用主控板的版本不一致，通常会导致备用主控板不能正常工作。同步升级单板的启动文件功能提供了一种快速升级备用主控板的方法。如果设备支持主用/备用（main/backup）两个下次启动文件，则同步升级的是备用主控板的主用下次启动文件。使用本功能相当于分步执行了以下操作来完成备用主控板的升级：

(1) 将主用主控板当前使用的启动文件拷贝到备用主控板；

(2) 将拷贝后的文件指定为备用主控板的下次启动文件；

(3) 重启备用主控板，使备用主控板的启动文件和主用主控板的一致。

表1-12 同步升级备用主控板的启动文件

操作	命令	说明
同步升级备用主控板的启动文件	**boot-loader update slot** *slot-number*	必选 该命令在用户视图下执行

1.7 单板的热插拔

支持热插拔的单板直接插入插槽中即可完成该单板的安装。

支持热插拔的单板在设备运行的过程中，可以通过以下命令卸载该单板。

表1-13 卸载指定单板或指定子卡

操作		命令	说明
卸载某单板或某单板上某子卡	集中式设备	**remove slot** *slot-number*	可选 该命令在用户视图下执行
	分布式设备	**remove slot** *slot-number* [**subslot** *subslot-number*]	可选 该命令在用户视图下执行

注意

- 只有支持热插拔的单板才可以通过该命令来卸载，其它单板需要在设备断电后才能执行插拔或卸载的操作。
- 该命令会使某单板不可用，从而导致业务中断，请谨慎使用。
- 需要热插拔某单板前请先使用 **remove slot** 命令卸载单板，否则可能会对设备或单板造成损坏。

1.8 智能电源管理

说明

该功能仅 SR6616 支持。

SR6616 支持交流或直流两种供电方式，用户可根据实际需要选用交流或直流电源模块，但是交/直流电源模块不能混插在同一台设备上。设备的电源模块支持在线插拔。

1.8.1 配置单板的上电优先级

SR6616 共有 Slot0～Slot3、Slot6～Slot9 等 8 个灵活接口平台插槽，每个灵活接口平台插槽可以分别配置不同的上电优先级。优先级数字越小则上电优先级越高。

当电源的供电能力恢复时，系统会根据各单板的上电优先级来决定供电的先后顺序，优先级高的先供电，优先级高的后供电。

表1-14 配置单板的电源管理优先级

操作	命令	说明
进入系统视图	**system-view**	-
配置单板的上电优先级	**power-supply policy priority** *priority* **slot** *slot-number*	必选 缺省情况下，单板的电源管理优先级为 2

1.9 配置单板的温度告警阈值

通过以下配置任务，用户可以设置指定单板的温度告警阈值。当单板的温度超出用户设定的阈值时，设备系统会发出告警信号，便于用户及时进行处理。

表1-15 配置单板的温度告警阈值

操作	命令	说明
进入系统视图	**system-view**	-
配置单板的温度告警阈值	**temperature-limit** *slot-number lower-value upper-value*	可选

1.10 配置接口卡的工作模式

设备支持的接口卡类型多样，有些卡是一卡专用，比如 SIC-1FEA（1 端口 10/100M 以太网接口卡，主要用于完成路由器与局域网的通信）；有些卡是具有多功能模块，比如 FIC-2E1（2 端口通道化 E1/PRI 接口卡，它可以配置指定的接口完成 E1、CE1 或者 ISDN PRI 接入）；有些卡是一卡多用，通过配置卡的工作模式，完成整个卡的工作模式的切换。目前可能支持的切换模式如下：

- CPOS 155M（E1/T1）：支持整个接口卡 E1/T1 模式切换。当工作在 E1 模式时，可完成 E1 数据流的收发及处理，提供 CE1 接入，可以实现 ISDN PRI 功能；当工作在 T1 模式时，可完成 T1 数据流的收发及处理，提供 CT1 接入，可以实现 ISDN PRI 功能。
- CPOS 155M（E3/T3）：支持整个接口卡 E3/T3/POS 模式切换。当工作在 E3 模式时，可完成 E3 高速数据流的收发和处理，提供 E3 数据流的接入服务，速率为 34.368Mbps；当工作在 T3 模式时，可完成 T3 高速数据流的收发及处理，提供 T3 数据流的接入服务，速率为 44.736Mbps；当工作在 POS 模式时，整个接口卡的接口都作为 POS 类型接口来工作。
- POS 155M/622M：支持整个接口卡 OC-3c/STM-1c 模式和 OC-12c/STM-4c 模式切换。当工作在 OC-3c/STM-1c 模式时，所有接口的收发速率均为 155Mbps；当工作在 OC-12c/STM-4c 模式时，所有可用接口的收发速率均为 622Mbps（工作模式切换为 OC-12c/STM-4c 模式后，接口板上具有“SPF 622M”丝印标志的接口才可用）。
- E-CPOS：支持整个接口卡切换到 E-CPOS 模式。当工作在 E-CPOS 模式时，整个接口卡的接口都作为 E-CPOS 类型接口来工作。
- IPsec/SSL：支持加密卡 IPsec 和 SSL 加密模式切换。当工作在 IPsec 模式时，加密卡支持 IPsec 协议；当工作在 SSL 模式时，加密卡支持 SSL 协议。

表1-16 配置接口卡的工作模式（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
设置接口卡的工作模式	**card-mode slot** *slot-number* *mode-name*	可选 *mode-name* 的取值可能为 **e1**、**t1**、**e3**、**t3**、**pos**、**e-cpos**、**oc-3**、**oc-12**

表1-17 配置接口卡的工作模式（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
设置接口卡的工作模式	**card-mode slot** *slot-number* **subslot** *subslot-number* *mode-name*	可选 *mode-name* 的取值可能为 **e1**、**t1**、**e3**、**t3**、**pos**、**e-cpos**、**oc-3**、**oc-12**

注意

模式切换后需要重启设备或热插拔接口卡（如果接口卡支持热插拔），新配置的模式才会生效。

1.11 清除当前系统中不使用的 16bit 接口索引

实际组网中，NMS 软件要求设备能够提供统一的 16bit 的接口索引，同时要求接口索引尽量稳定，即同一设备中接口的名字与接口的索引尽量保持一一对应的关系。

为了尽量保证接口索引的稳定性，当系统中的板卡被拔出或者在删除逻辑接口时，系统会保存该接口的 16bit 接口索引，以保证重新创建该接口时其索引值不变。

反复的插拔不同的子卡或接口板，创建或者删除大批量不同类型的逻辑口，都可能会耗尽接口索引，导致创建接口失败。为了避免这种情况，用户可以在用户视图下执行以下操作，清除当前系统中保存的但不使用的 16bit 接口索引。

执行该操作后：

- 对于重新创建的接口，接口的新的索引不能保证与原来的索引一致。
- 对于系统中已经正常存在的接口，不会引起其索引值的改变。

表1-18 清除当前系统中不使用的 16bit 接口索引

操作	命令	说明
清除当前系统中保存的但不使用的 16bit 接口索引	**reset unused porttag**	必选 该命令在用户视图下执行

注意

执行该命令时需要得到用户的确认。若用户在 30 秒之内没有确认操作，或者用户输入字符“N”否定了本次操作，则本命令退出执行。

1.12 可插拔模块的识别与诊断

1.12.1 可插拔模块简介

目前，常用的可插拔模块有四种，这四种模块又可以细分为多种，比如按传输介质类型又可以分为光接口模块和电接口模块，其简介请参见 表 1-19。

表1-19 常用可插拔模块简介

可插拔模块类型	适用环境	能否为光接口模块	能否为电接口模块
SFP（Small Form-factor Pluggable，小封装可热插拔）	一般用于百兆/千兆以太网、POS 155M/622M/2.5G 等环境	是	是
GBIC（Gigabit Interface Converter，千兆以太网接口转换器）	一般用于千兆以太网环境	是	是
XFP（10-Gigabit small Form-factor Pluggable，万兆以太网接口小封装可热插拔）	一般用于 10G 以太网环境	是	否
XENPAK（10-Gigabit Ethernet Transceiver Package，万兆以太网接口收发器集合封装）	一般用于 10G 以太网环境	是	是

1.12.2 识别可插拔模块

因为可插拔模块的类型多样，厂商也不同，通过以下显示命令可以查看可插拔模块的主要特征参数（包括模块型号、连接器类型、发送激光的中心波长、信号的有效传输距离、模块生产或定制厂商等），以便识别可插拔模块。

表1-20 识别可插拔模块信息

操作	命令	说明
显示可插拔模块的主要特征参数	**display transceiver** { **interface** [*interface-type interface-number*] | **controller** [*controller-type controller-number*] }	对可插拔模块均生效
显示 H3C 定制防伪可插拔模块的部分电子标签信息	**display transceiver manuinfo** { **interface** [*interface-type interface-number*] | **controller** [*controller-type controller-number*] }	仅对 H3C 定制防伪可插拔模块生效

- H3C 定制防伪可插拔模块是指 H3C 公司定制的、具有防伪功能的可插拔模块。用户可以通过 **display transceiver** 命令显示信息中 Vendor Name 字段来识别，为“H3C”的可以认定为 H3C 定制防伪可插拔模块。
- 电子标签信息也可以称为永久配置数据或档案信息等，在模块调测（调试、测试）过程中被写入到存储器件中，包括名称、生产序列号、模块生产或定制厂商等信息。

1.12.3 诊断可插拔模块

系统提供故障告警信息标志可插拔模块的故障来源，以便诊断和解决故障。对 H3C 定制防伪光模块系统还提供了数字诊断功能，其原理主要是对影响光模块工作的关键参数进行监控，这些关键参数包括：温度、电压、激光偏置电流、发送光功率和接收光功率等。当这些参数的值异常时，用户可以采取相应的措施，预防故障发生。

表1-21 诊断可插拔模块

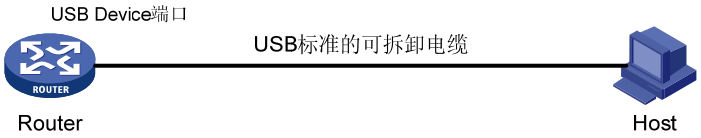
操作	命令	说明
显示可插拔模块的当前故障告警信息	**display transceiver alarm { interface** [*interface-type interface-number*] | **controller** [*controller-type controller-number*] }	对可插拔模块均生效
显示 H3C 定制防伪可插拔光模块的数字诊断参数的当前测量值	**display transceiver diagnosis { interface** [*interface-type interface-number*] | **controller** [*controller-type controller-number*] }	仅对 H3C 定制的防伪可插拔光模块生效

1.13 配置 USB Device 端口

1.13.1 USB Device 端口的基本功能

USB（Universal Serial Bus，通用串行总线）是当今业界最为流行的外设接口总线，它方便易用，可靠性高，性能优越。H3C SR6600 路由器的 USB Device 端口支持 USB2.0 规范，通过 USB 标准的可拆卸电缆可以实现设备与 Host 的 USB 总线连接，并将设备内置的 CF 卡作为 USB Mass Storage Device（USB 大容量存储设备）注册到 Host 的 Windows 操作系统中。路由器在 Windows 操作系统中被识别为 USB Mass Storage Device 之后，就可以利用简单、快捷的 USB 协议以及 Windows 操作系统友好的用户界面，对内置 CF 卡进行快速的文件操作和访问了。

图1-2 设备和 Host 的 USB 连接示意图



1.13.2 USB 标准的连接器与线缆

USB 协议规定了 A 系列和 B 系列两种连接器：

- A 系列连接器：用于 USB 设备直接连接到 Host 或 HUB 的下行端口。
- B 系列连接器：允许设备厂商提供标准的可拆卸电缆，通过该电缆来实现 A 系列到 B 系列两种连接器的转换。

这两种连接器都包含插头和插座，其中：

- A 系列插座与 A 系列插头配对，A 系列插座通常位于 Host 或者 HUB 上，作为 Host 系统或 HUB 系统的输出。A 系列插头则通常连接到 Host 或者 HUB 上与之配对的 A 系列插座上。
- B 系列插座与 B 系列插头配对，B 系列插座通常作为设备的输入，B 系列插头通常连接到设备上与之配对的 B 系列插座上。

图1-3 A 系列插头和插座

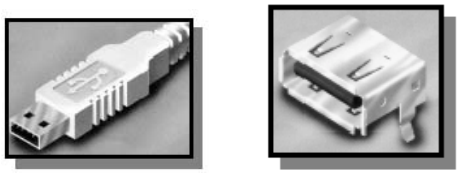
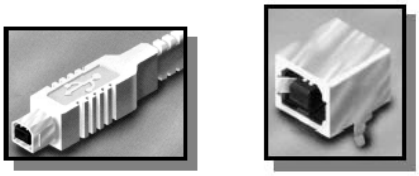


图1-4 B 系列插头和插座



A 系列和 B 系列两种连接器的差异能够保证用户正确地连接 USB。

在实现如 图 1-2所示的组网时，需要采用USB协议规定的标准可拆卸数据线缆。该线缆的一端必须是A系列插头，另一端必须是B系列插头：

- 线缆的 A 系列插头插入 Host 的 A 系列插座上。
- 线缆的 B 系列插头插入 Router 的 B 系列插座上。

1.13.3 设备与 Host 的连接

在实现设备与 Host 的连接时，需要首先使能设备的 USB Device 端口，然后再用 USB 标准的可拆卸线缆把设备和 Host 连接起来。第一次连接时，Windows 操作系统可能会询问是否安装驱动程序，由于 WindowsXP 操作系统提供 USB Mass Storage Device 的支持，因此可以使用 Windows 硬件更新向导的“自动安装软件”功能来搜索并自动安装硬件驱动。安装完毕，打开“我的电脑”后会出现“可移动磁盘”图标，此时设备上的 USB Device 端口指示灯常亮，设备就可以作为 USB Mass Storage Device 来使用了。

图1-5 Windows 操作系统正常加载 USB Mass Storage Device



需要注意的是：

- 由于设备不提供 Host 端的 USB Mass Storage Device 驱动，因此在 Host 端请选用支持 USB Mass Storage Device 驱动的 Windows 操作系统（推荐使用 WindowsXP 系统）。
- 由于设备的文件系统和 Host 端的 Windows 操作系统不能同时操作内置 CF 卡，否则会导致设备发生异常。因此设备与 Host 连接时，设备端的文件系统会先卸载内置 CF 卡，之后 Host 端的 Windows 操作系统才会进行内置 CF 卡的加载操作。
- 当设备处于与 Host 连接的状态时，对于设备的文件系统来说，内置 CF 卡的所有功能将不能使用，直到设备与 Host 的连接断开（此时，Host 端的 Windows 操作系统会卸载内置 CF 卡，并且设备的文件系统将完成内置 CF 卡的重新加载工作）。
- 当设备的 USB Device 端口指示灯还处于闪烁的状态时（此时有数据正在传输），禁止插拔 USB 标准可拆卸线缆，也不能通过配置命令 **undo usb device-port enable** 强行关闭 USB Device 端口，否则可能会导致设备发生异常。

说明

- 设备和 Host 连接成功后，建议不要在 Windows 下对内置 CF 卡进行格式化和删除文件等操作，以免破坏或者删除内置 CF 卡中的应用程序和配置文件等重要文件，导致设备下次无法正常启动。
- 设备的文件系统目前只支持英文，对于在 Windows 下操作的中文文件/文件夹都会显示为乱码，因此，建议不要对中文文件/文件夹进行拷贝、粘贴或者创建等操作。

1.13.4 正确卸载设备的方法

为了保证设备和数据的安全，在 Windows 操作系统下使用完设备后，需要按照 Windows 安全卸载 USB 设备的操作步骤来正确卸载设备：

- 在 Windows 桌面右下角的任务栏中，左键双击“安全删除硬件”图标，在弹出窗口中选择“USB Mass Storage Device”，并点击<停止>按钮。

图1-6 任务栏中的安全卸载图标



图1-7 安全删除硬件对话框



- 在随后出现的[停用硬件设备]窗口中点击<确定>按钮，如果 Windows 提示无法停止，请确认是否有数据正在传输。如果有数据正在传输，则可以等待数据传输完毕或者通过停止数据传输的操作强行结束传输后，重复上述操作。
- 当看到“安全移除硬件”的提示，并且设备上 USB Device 端口指示灯熄灭后，就可以拔出 USB 标准可拆卸线缆了。

1.13.5 配置使能 USB Device 端口

表1-22 配置使能 USB Device 端口

操作		命令	说明
进入系统视图		**system-view**	-
配置使能 USB Device 端口	集中式	**usb device-port enable type storage**	必选 缺省情况下，USB Device 端口处于关闭状态
	分布式	**usb device-port enable type storage slot** *slotnumber*	

注意

在使能 USB Device 端口前，必须确保设备的文件系统没有对内置 CF 卡进行拷贝等操作，否则会导致设备和 Host 连接失败或者拷贝等操作失败。

1.14 设备管理显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后设备的运行情况，通过查看显示信息验证配置的效果。

表1-23 设备管理显示和维护（集中式设备）

操作	命令
显示启动文件信息	**display boot-loader**
显示 CPU 占用率的统计信息	**display cpu-usage** [*entry-number* [*offset*] [**verbose**] [**from-device**]]
以图形方式显示 CPU 占用率统计历史信息	**display cpu-usage history** [**task** *task-id*]
显示设备上单板、子卡、CF 卡、和 USB 等子设备的信息	**display device** [**cf-card** | **usb**] [**slot** *slot-number* | **verbose**]
显示设备的电子标签信息	**display device manuinfo** [**slot** *slot-number*]
显示设备的温度信息	**display environment** [**cpu**]
显示设备内置风扇的工作状态	**display fan** [*fan-id*]
显示设备的内存使用状态	**display memory**
显示设备的电源状态	**display power** [*power-id*]
显示设备的启动方式	**display reboot-type** [**slot** *slot-number*]
显示设备冗余电源系统的状态	**display rps** [*rps-id*]
显示设备的重启时间	**display schedule reboot**

操作	命令
查看定时自动执行功能的具体配置	**display schedule job**
显示异常情况下系统的处理方式	**display system-failure**

表1-24 设备管理显示和维护（分布式设备）

操作	命令
显示启动文件信息	**display boot-loader** [**slot** *slot-number*]
显示 CPU 占用率的统计信息	**display cpu-usage** [**slot** *slot-number* [**cpu** *cpu-number*]] **display cpu-usage** *entry-number* [*offset*] [**verbose**] [**slot** *slot-number*] [**from-device**]
以图形方式显示 CPU 占用率统计历史信息	**display cpu-usage history** [**task** *task-id*] [**slot** *slot-number* [**cpu** *cpu-number*]]
显示设备上单板、子卡、CF 卡、和 USB 等子设备的信息	**display device** [**cf-card** | **usb**] [[**shelf** *shelf-number*] [**frame** *frame-number*] [**slot** *slot-number* [**subslot** *subslot-number*]] | **verbose**]
显示设备的电子标签信息	**display device manuinfo** [**slot** *slot-number* [**subslot** *subslot-number*]]
显示设备的环境信息	**display environment**
显示设备内置风扇的工作状态	**display fan** [*fan-id*]
显示设备的内存使用状态	**display memory** [**slot** *slot-number* [**cpu** *cpu-number*]]
显示设备的电源状态	**display power** [*power-id*]
查看电源的信息	**display power-supply** [**verbose**]
显示设备的启动方式	**display reboot-type** [**slot** *slot-number* [**subslot** *subslot-number*]]
显示设备的重启时间	**display schedule reboot**
查看定时自动执行功能的具体配置	**display schedule job**
显示异常情况下系统的处理方式	**display system-failure**

1.15 设备管理典型配置举例

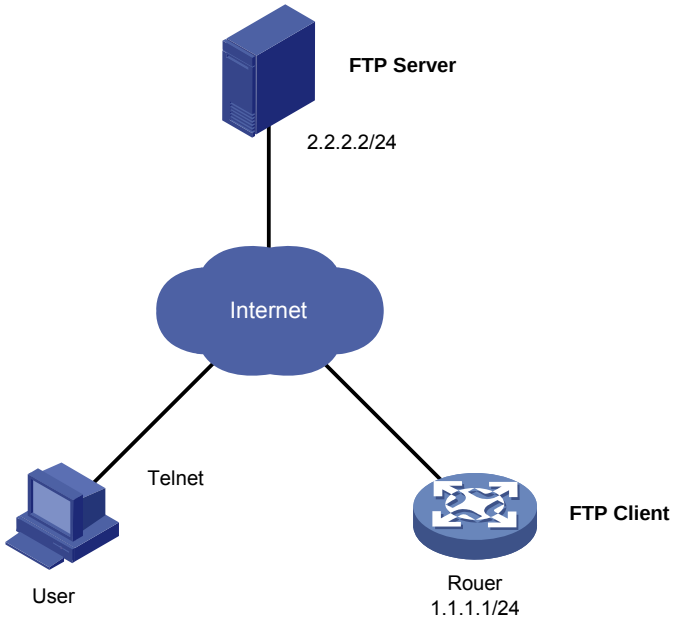
1.15.1 远程立即升级配置举例（分布式设备）

1. 组网需求

- Router 当前的软件版本为 soft-version1、Bootrom 版本为 bootrom-version1，通过远程操作，立即将设备的软件版本升级到 soft-version2、Bootrom 版本升级到 bootrom-version2。
- 最新版本的应用程序 soft-version2.app 和 Bootrom 程序 bootrom-version2.btm 都保存在 FTP Server 的 aaa 目录下；
- Router 的 IP 地址为 1.1.1.1/24，FTP Server 的 IP 地址为 2.2.2.2/24，Router 与 FTP Server 之间路由可达；
- User 通过 Telnet 远程登录到 Router，User 与 Router 之间路由可达。

2. 组网图

图1-8 配置远程升级组网图



3. 配置步骤

(1) FTP Server 的配置（注意：不同的服务器类型配置可能不同）

启动 FTP Server 功能。

```
<FTP-Server> system-view
[FTP-Server] ftp server enable
```

配置一个 FTP 用户名为 aaa，密码为 hello。

```
[FTP-Server] local-user aaa
[FTP-Server-luser-aaa] password cipher hello
```

配置该用户对 aaa 目录具有读写权限。

```
[FTP-Server-luser-aaa] service-type ftp
[FTP-Server-luser-aaa] authorization-attribute work-directory cfa0:/aaa
```

(2) Router 的配置

注意

如果设备 CF 卡的剩余空间不够大，请删除 CF 卡中原有的应用程序后再进行下载。

开始升级前，执行 **save** 命令保存当前配置（配置步骤略）。

#登录 FTP Server（注意：不同的服务器类型显示信息可能不同）。

```
<Router> ftp 2.2.2.2
Trying 2.2.2.2 ...
Press CTRL+K to abort
Connected to 2.2.2.2
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(none): aaa
```

```
331 Give me your password, please
Password:
230 Logged in successfully
[ftp]
```

下载 FTP Server 上的文件 soft-version2.app 和 bootrom-version2.btm。

```
[ftp] binary
[ftp] get soft-version2.app
[ftp] get bootrom-version2.btm
```

中断 FTP 连接，并退回到用户视图下。

```
[ftp] bye
<Router>
```

升级主用主控板的 Bootrom（主用主控板的槽位号为 0）。

```
<Device> bootrom update file bootrom-version2.btm slot 0
```

升级备用主控板的 Bootrom（备用主控板的槽位号为 1）。

```
<Device> copy bootrom-version2.btm slot1#cfa0:/bootrom-version2.btm
<Device> bootrom update file slot1#cfa0:/bootrom-version2.btm slot 1
```

指定主控板下次启动时的应用程序。

```
<Router> boot-loader file soft-version2.app slot 0 main
```

指定备用主控板下次启动时的应用程序。

<Router> copy soft-version2.app slot1#cfa0:/soft-version2.app

```
<Router> boot-loader file slot1#cfa0:/soft-version2.app slot 1 main
```



说明

当指定备用主控板下次启动时的应用程序时，被指定应用程序必须保存在备用主控板的 CF 卡上。

重启设备，实现软件版本的升级。

```
<Router> reboot
```

设备重启后可以使用 **display version** 命令验证升级是否成功。

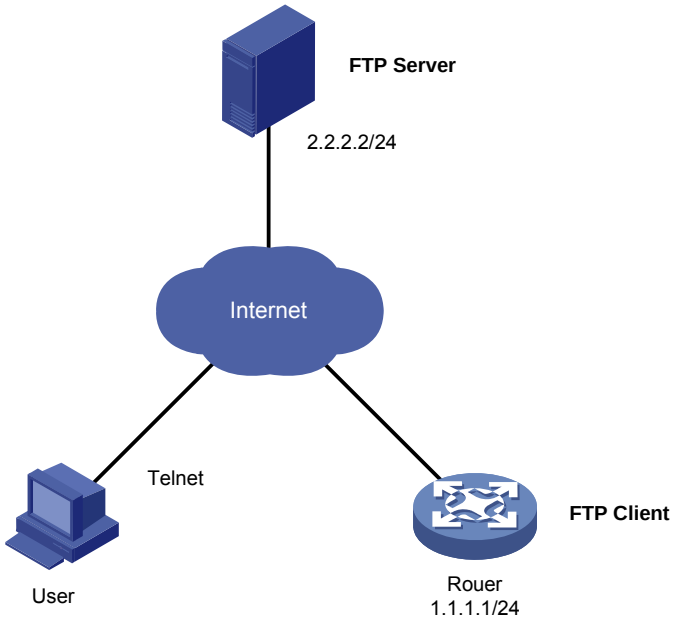
1.15.2 远程定时自动升级配置举例（集中式设备）

1. 组网需求

- Router 当前的软件版本为 soft-version1，通过远程操作，在业务空闲时间（比如凌晨 3 点）将设备的软件版本升级到 soft-version2，配置文件升级到 new-config；
- 最新版本的应用程序 soft-version2.app 和最新配置文件 new-config.cfg 都保存在 FTP Server 的 aaa 目录下；
- Router 的 IP 地址为 1.1.1.1/24，FTP Server 的 IP 地址为 2.2.2.2/24，Router 与 FTP Server 之间路由可达；
- User 通过 Telnet 远程登录到 Device，User 与 Device 之间路由可达。

2. 组网图

图1-9 配置远程定时自动升级组网图



3. 配置步骤

(1) FTP Server 的配置（注意：不同的服务器类型配置可能不同）

- 设置 FTP client 访问参数（包括启动 FTP Server 功能、配置 FTP 用户名为 aaa，密码为 hello、授权路径为 cfa0:/aaa）。

```
<FTP-Server> system-view
[FTP-Server] ftp server enable
[FTP-Server] local-user aaa
[FTP-Server-luser-aaa] password cipher hello
[FTP-Server-luser-aaa] service-type ftp
[FTP-Server-luser-aaa] authorization-attribute work-directory cfa0:/aaa
```

- 在 FTP Server 上使用文本编辑器编辑批处理文档 auto-update.txt。批处理文件的内容如下：

```
return
startup saved-configuration new-config.cfg
boot-loader file soft-version2.app main
reboot
```

(2) Router 的配置

登录 FTP Server（注意：不同的服务器类型显示信息可能不同）。

```
<Device> ftp 2.2.2.2
Trying 2.2.2.2 ...
Press CTRL+K to abort
Connected to 2.2.2.2.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(2.2.2.2:(none)):aaa
331 Give me your password, please
Password:
230 Logged in successfully
```

```
[ftp]
```

下载 FTP Server 上的文件 auto-update.txt。

```
[ftp] ascii
[ftp] get auto-update.txt
```

下载 FTP Server 上的文件 new-config.cfg

```
[ftp]get new-config.cfg
```

下载 FTP Server 上的文件 soft-version2.app。

```
[ftp] binary
[ftp] get soft-version2.app
[ftp] bye
<Device>
```

将 auto-update.txt 文件更名为批处理文件 auto-update.bat。

```
<Device> rename auto-update.txt auto-update.bat
```

为确保文件正确，可以使用 **more** 命令查看确认文件内容。

使用定时自动执行功能实现设备在凌晨 3 点定时自动升级。

```
<Device> schedule job at 03:00 view system execute auto-update.bat
Info: Command execute auto-update.bat in system view will be executed at 03:00 12/11/2007(in
12 hours and  0 minutes).
```

设备重启后可以使用 **display version** 命令验证升级是否成功。

1.15.3 USB Device 端口配置典型配置举例

1. 组网需求

通过设备上的 USB Device 端口，实现和 Host 的连接。

组网图请参见 图 1-2。

2. 配置思路

- 确保设备的文件系统没有对设备内置 CF 卡进行拷贝等操作。
- 使能 USB Device 端口。

3. 配置步骤

使能集中式设备的 USB Device 端口。

```
<sysname> system-view
[sysname] usb device-port enable type storage
```

使能分布式设备 1 号槽位的 USB Device 端口。

```
<sysname> system-view
[sysname] usb device-port enable type storage slot 1
```

目 录

1 NQA配置 1-1
1.1 NQA简介 1-1
1.1.1 NQA概述 1-1
1.1.2 NQA的特点 1-1
1.1.3 NQA的基本概念 1-2
1.1.4 NQA测试操作 1-3
1.2 NQA配置任务简介 1-3
1.3 配置NQA服务器 1-4
1.4 使能NQA客户端功能 1-5
1.5 创建NQA测试组 1-5
1.6 配置NQA测试组 1-5
1.6.1 配置ICMP-echo测试 1-5
1.6.2 配置DHCP测试 1-6
1.6.3 配置DNS测试 1-7
1.6.4 配置FTP测试 1-7
1.6.5 配置HTTP测试 1-8
1.6.6 配置UDP-jitter测试 1-9
1.6.7 配置SNMP测试 1-11
1.6.8 配置TCP测试 1-11
1.6.9 配置UDP-echo测试 1-12
1.6.10 配置Voice测试 1-13
1.6.11 配置DLSw测试 1-15
1.7 配置联动功能 1-15
1.8 配置发送Trap消息 1-16
1.9 配置NQA统计功能 1-17
1.10 配置NQA历史记录功能 1-17
1.11 配置NQA测试组通用可选参数 1-18
1.12 调度NQA测试组 1-19
1.13 NQA显示和维护 1-20
1.14 NQA典型配置举例 1-20
1.14.1 ICMP-echo测试配置举例 1-20
1.14.2 DHCP测试配置举例 1-21
1.14.3 DNS测试配置举例 1-23
1.14.4 FTP测试配置举例 1-24
1.14.5 HTTP测试配置举例 1-25
1.14.6 UDP-jitter测试配置举例 1-26
1.14.7 SNMP测试配置举例 1-29
1.14.8 TCP测试配置举例 1-30
1.14.9 UDP-echo测试配置举例 1-31
1.14.10 Voice测试配置举例 1-33

1.14.11 DLSw测试配置举例 1-35

1 NQA 配置

1.1 NQA 简介

1.1.1 NQA 概述

NQA 是 Network Quality Analyzer（网络质量分析）的简称。NQA 通过发送测试报文，对网络性能、网络提供的服务及服务质量进行分析，并为用户提供网络性能和服务质量的参数，如时延抖动、TCP 连接时延、FTP 连接时延和文件传输速率等。

利用 NQA 的测试结果，用户可以：

(1) 及时了解网络的性能状况，针对不同的网络性能进行相应处理。

(2) 对网络故障进行诊断和定位。

1.1.2 NQA 的特点

1. 支持多种测试类型

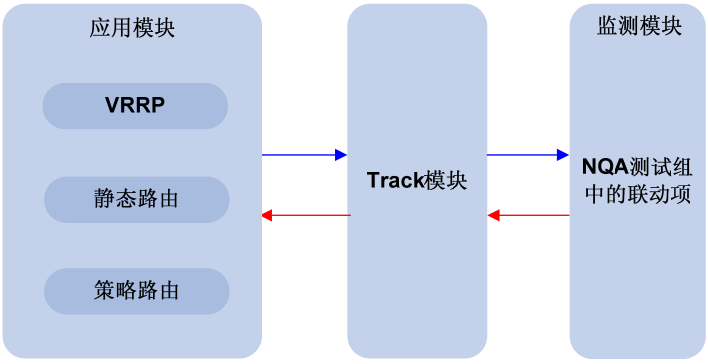
以往的 Ping 功能是使用 ICMP（Internet Control Message Protocol，互联网控制报文协议）测试数据包在本端和指定目的端之间的往返时间。NQA 是对 Ping 功能的扩展和增强，它增加了多种测试类型，提供了更多的功能。

目前 NQA 支持 11 种测试类型：ICMP-echo、DHCP、DNS、FTP、HTTP、UDP-jitter、SNMP、TCP、UDP-echo、Voice 和 DLSw 测试。客户端向对端发送不同类型的测试报文，统计对端是否回应报文以及报文的往返时间等参数，以便用户根据统计结果判断协议的可用性和网络的性能。

2. 支持联动功能

联动功能是指通过建立联动项，对当前所在测试组中的探测结果进行监测，当连续探测失败次数达到一定数目时，就触发其他模块联动。联动功能的实现如 图 1-1所示。

图1-1 联动功能实现示意图



联动功能由监测模块、Track 模块和应用模块三部分组成：

- 监测模块负责对链路状态、网络性能等进行监测，并将探测结果通知给 Track 模块。
- Track 模块接收到监测模块的探测结果后，及时改变 Track 项的状态，并通知应用模块。Track 模块位于应用模块和监测模块之间，主要功能是屏蔽不同监测模块的差异，为应用模块提供统一的接口。

- 应用模块根据 Track 项的状态，进行相应的处理，从而实现联动。

以静态路由为例，用户配置了一条静态路由，下一跳为 192.168.0.88，如果 192.168.0.88 可达，那么该静态路由有效；如果 192.168.0.88 不可达，则该静态路由无效。通过在 NQA、Track 模块和应用模块之间建立联动，可以实现静态路由有效性的实时判断：

(1) 通过 NQA 监测目的地址 192.168.0.88 是否可达；

(2) 如果发现 192.168.0.88 不可达，NQA 将通过 Track 模块通知静态路由模块；

(3) 静态路由模块可以据此判断该静态路由项无效。



说明

Track 模块的详细介绍，请参见“系统分册”中的“Track 配置”。

3. 发送 Trap 消息

NQA 测试过程中可以通过设置 Trap 开关控制是否向网管服务器发送 Trap 消息。在探测失败或测试完成时通知网管服务器，以便网络管理员根据 Trap 消息及时了解测试运行状况和网络性能。

1.1.3 NQA 的基本概念

1. 测试组

进行 NQA 测试前，需要创建 NQA 测试组。在 NQA 测试组中配置 NQA 测试的参数，如测试类型，测试目的地址，测试目的端口等。

每个测试组都有一个管理员名称和一个操作标签，管理员名称和操作标签可以唯一确定一个测试组。

2. 测试和探测

启动 NQA 测试后，每隔一段时间启动一次测试，测试的时间间隔可以由用户来设定。

一次 NQA 测试由若干次连续的探测组成，探测的次数可以由用户来设定。



说明

对于 Voice 测试，一次测试中只能进行一次探测。

不同测试类型中，探测的含义不同：

- 对于 TCP 和 DLSw 测试，一次探测是指一次连接；
- 对于 UDP-jitter 和 Voice 测试，一次探测中连续发送多个探测报文，发送探测报文的个数由用户来设定；
- 对于 FTP、HTTP、DHCP 和 DNS 测试，一次探测是指完成一次相应的功能；
- 对于 ICMP-echo 和 UDP-echo 测试，一次探测发送一个探测报文；
- 对于 SNMP 测试，一次探测发送三个探测报文。

3. NQA 客户端和服务器

NQA 客户端是发起 NQA 测试的设备，NQA 测试组在客户端创建。

NQA服务器负责处理NQA客户端发来的测试报文，如 图 1-2所示。NQA服务器通过监听指定IP地址和端口的报文对客户端发起的测试进行响应。

图1-2 NQA 客户端与服务器之间的关系



在大多数的测试中，只需要配置 NQA 客户端。但在进行 TCP、UDP-echo、UDP-jitter 和 Voice 类型测试时，必须配置 NQA 服务器。

在一个 NQA 服务器上可以创建多个 TCP 或 UDP 监听服务，每个监听服务对应一个目的地址和一个端口号，配置的目的地址和端口号必须与 NQA 客户端的配置一致，且不能与已有的监听服务冲突。

1.1.4 NQA 测试操作

NQA 测试操作的过程为：

(1) NQA 客户端构造指定测试类型的报文，并发送给对端；

(2) 对端收到测试报文后，回复带有时间戳的应答报文；

(3) NQA 客户端根据是否收到应答报文，以及应答报文中的时间戳，计算报文丢失率、往返时间等参数。

要成功进行某项测试，需要在 NQA 客户端上进行如下操作：

(1) 使能 NQA 客户端功能。

(2) 创建测试组，根据测试类型进行相应测试参数的配置。不同测试类型对应的测试参数不同，详见下面具体的配置过程。

(3) 启动 NQA 测试。

(4) 通过显示命令或调试命令查看测试结果。

对于 TCP，UDP-jitter、UDP-echo 和 Voice 测试，还需要在对端设备上配置 NQA 服务器。

1.2 NQA 配置任务简介

对于 TCP，UDP-jitter、UDP-echo 和 Voice 测试，需要在对端设备上配置 NQA 服务器。

表1-1 NQA 服务器配置任务简介

配置任务	说明	详细配置
配置 NQA 服务器	对于 TCP、UDP-echo、UDP-jitter 和 Voice 四种测试类型，为必选	1.3

要成功进行某项测试，NQA 客户端上需要进行如下操作：

(1) 使能 NQA 客户端功能。

(2) 创建测试组，根据测试类型进行相应测试参数的配置。不同测试类型对应的测试参数不同，详见下面具体的配置过程。

(3) 启动 NQA 测试。

测试完成后，可以通过显示命令或调试命令查看测试结果。

表1-2 NQA 客户端配置任务简介

配置任务		说明	详细配置
使能 NQA 客户端功能		必选	1.4
创建 NQA 测试组		必选	1.5
配置 NQA 测试	配置 ICMP-echo 测试	必选其一	1.6.1
	配置 DHCP 测试		1.6.2
	配置 DNS 测试		1.6.3
	配置 FTP 测试		1.6.3
	配置 HTTP 测试		1.6.5
	配置 UDP-jitter 测试		1.6.6
	配置 SNMP 测试		1.6.7
	配置 TCP 测试		1.6.8
	配置 UDP-echo 测试		1.6.9
	配置 Voice 测试		1.6.10
	配置 DLSw 测试		1.6.11
配置联动功能		可选	1.7
配置发送 Trap 消息		可选	1.8
配置 NQA 统计功能		可选	1.9
配置 NQA 历史记录功能		可选	1.10
配置 NQA 测试组通用可选参数		可选	1.11
调度 NQA 测试组		必选	1.12

1.3 配置 NQA 服务器

在进行 TCP、UDP-echo、UDP-jitter 和 Voice 类型测试前，必须在对端设备上配置 NQA 服务器。NQA 服务器通过监听指定 IP 地址和端口的报文，对客户端发起的测试进行响应。

表1-3 配置 NQA 服务器

操作	命令	说明
进入系统视图	**system-view**	-
开启 NQA 服务器功能	**nqa server enable**	必选 缺省情况下，NQA 服务器功能处于关闭状态
在 NQA 服务器上配置监听服务	**nqa server** { **tcp-connect** | **udp-echo** } *ip-address port-number*	必选 配置的 IP 地址和端口号必须与 NQA 客户端的配置一致，且不能与已有的监听服务冲突

1.4 使能 NQA 客户端功能

只有使能 NQA 客户端功能后，NQA 客户端的相关配置才会生效。

表1-4 使能 NQA 客户端功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 NQA 客户端功能	**nqa agent enable**	可选 缺省情况下，NQA 客户端功能处于开启状态

1.5 创建 NQA 测试组

一次测试对应一个测试组，创建测试组并进入测试组视图后，可以配置具体测试类型。

表1-5 创建 NQA 测试组

操作	命令	说明
进入系统视图	**system-view**	-
创建 NQA 测试组，并进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	必选

说明

执行 **nqa entry** 命令进入已经配置测试类型的测试组视图时，将直接进入测试组测试类型视图。

1.6 配置 NQA 测试组

1.6.1 配置 ICMP-echo 测试

ICMP-echo 测试利用 ICMP 协议，根据应答报文或超时信息判断目的主机的可达性。ICMP-echo 测试的功能与 Ping 命令类似，但输出信息更为丰富。网络中出现连通性故障时，可以通过 ICMP-echo 测试对故障进行定位。

表1-6 配置 ICMP-echo 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 ICMP-echo，并进入测试类型视图	**type icmp-echo**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址

操作	命令	说明
配置发送的探测报文的大小	**data-size** *size*	可选 缺省情况下，发送的探测报文为 100 字节
配置发送的探测报文的填充字符串	**data-fill** *string*	可选 缺省情况下，探测报文的填充内容为十六进制数值 00010203040506070809
配置 VPN 实例信息	**vpn-instance** *instance*	可选 缺省情况下，未配置 VPN 实例信息
配置将指定接口的 IP 地址作为 ICMP-echo 探测请求报文的源 IP 地址	**source interface** *interface-type interface-number*	可选 缺省情况下，未配置发送探测请求报文的源接口 如果使用 **source ip** 命令配置了 ICMP-echo 测试的源地址，则此配置无效 该命令指定的接口必须为 up 状态，否则探测将会失败
配置探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 如果没有指定源 IP 地址，但是配置了 **source interface**，则该源接口的 IP 地址将作为探测请求报文中的源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则探测将会失败
配置 IP 报文的下一跳 IP 地址	**next-hop** *ip-address*	可选 缺省情况下，未配置下一跳 IP 地址
配置通用的可选参数	参见 表 1-21	可选

1.6.2 配置 DHCP 测试

DHCP 测试主要用来测试网络上是否有 DHCP 服务器，以及 DHCP 服务器响应客户端请求、为客户端分配 IP 地址所需的时间。

1. 配置准备

在进行 DHCP 测试之前，需要完成 DHCP 服务器的配置。如果 NQA 客户端（即 DHCP 客户端）和 DHCP 服务器在不同的网段，还需要配置 DHCP 中继。DHCP 服务器和 DHCP 中继的配置方法，请参见“IP 业务分册”中的“DHCP 配置”。

2. 配置 DHCP 测试

表1-7 配置 DHCP 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 DHCP，并进入测试类型视图	**type dhcp**	必选

操作	命令	说明
指定进行 DHCP 测试的接口	**operation interface** *interface-type interface-number*	必选 缺省情况下，没有指定进行 DHCP 测试的接口 该命令指定的接口必须为 up 状态，否则会导致测试失败
配置通用的可选参数	参见 表 1-21	可选



- DHCP 测试只是模拟 DHCP 申请地址的过程，进行 DHCP 测试的接口 IP 地址不会改变。
- DHCP 测试完成后，NQA 客户端将主动发送 DHCP-RELEASE 报文释放申请到的 IP 地址。

1.6.3 配置 DNS 测试

DNS 测试主要用来测试 NQA 客户端是否可以通过指定的 DNS 服务器将域名解析为 IP 地址，以及域名解析过程需要的时间。

1. 配置准备

在进行 DNS 测试之前，需要在 DNS 服务器上创建域名和 IP 地址的映射关系。

2. 配置 DNS 测试

表1-8 配置 DNS 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 DNS，并进入测试类型视图	**type dns**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 测试操作的目的 IP 地址，即 DNS 服务器的 IP 地址
配置要解析的域名	**resolve-target** *domain-name*	必选 缺省情况下，没有配置要解析的域名
配置通用的可选参数	参见 表 1-21	可选



DNS 测试只是模拟域名解析的过程，设备上不会保存要解析的域名与 IP 地址的对应关系。

1.6.4 配置 FTP 测试

FTP 测试主要用来测试 NQA 客户端是否可以与指定的 FTP 服务器建立连接，以及与 FTP 服务器之间传送文件的时间。

1. 配置准备

在进行 FTP 测试之前，需要在 FTP 服务器上进行相应的配置，包括 FTP 客户端登录 FTP 服务器的用户名、密码等。FTP 服务器的配置方法，请参见“系统分册”中的“文件系统管理配置”。

2. 配置 FTP 测试

表1-9 配置 FTP 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 FTP，并进入测试类型视图	**type ftp**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 测试操作的目的 IP 地址，即 FTP 服务器的 IP 地址
配置探测请求报文的源 IP 地址	**source ip** *ip-address*	必选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置操作类型	**operation { get | put }**	可选 缺省情况下，FTP 操作方式为 **get** 操作，即从 FTP 服务器获取文件
配置 FTP 登录用户名	**username** *name*	必选 缺省情况下，未配置 FTP 登录用户名
配置 FTP 登录密码	**password** *password*	必选 缺省情况下，未配置 FTP 登录密码
配置 FTP 服务器和客户端传送文件的文件名	**filename** *file-name*	必选 缺省情况下，未配置 FTP 服务器和客户端之间传送文件的文件名
配置通用的可选参数	参见 表 1-21	可选

说明

- 进行 **put** 操作时，服务器上会创建以 *file-name* 为名字的固定大小及内容的文件；进行 **get** 操作时，设备上也不会保存从服务器获取的文件。
- 进行 **get** 操作时，如果 FTP 服务器上没有以 *file-name* 为名字的文件，则测试不会成功。
- 进行 **get** 操作时，请选用较小的文件进行测试，如果文件较大，可能会因为超时而导致测试失败，或由于占用较多的网络带宽而影响其他业务。

1.6.5 配置 HTTP 测试

HTTP 测试主要用来测试 NQA 客户端是否可以与指定的 HTTP 服务器建立连接，以及从 HTTP 服务器获取数据所需的时间，从而判断 HTTP 服务器的连通性及性能。

1. 配置准备

在进行 HTTP 测试之前，需要完成 HTTP 服务器的配置。

2. 配置 HTTP 测试

表1-10 配置 HTTP 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 HTTP，并进入测试类型视图	**type http**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 测试操作的目的地址，即 HTTP 服务器的 IP 地址
配置探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置操作类型	**operation { get | post }**	可选 缺省情况下，HTTP 操作方式为 **get** 操作，即从 HTTP 服务器获取数据
配置 HTTP 测试访问的网址	**url** *url*	必选
配置测试所使用的协议版本	**http-version v1.0**	可选 缺省情况下，HTTP 测试使用的版本为 1.0
配置通用的可选参数	参见 表 1-21	可选



说明

HTTP 测试中，HTTP 服务器的 TCP 端口号只能为 80。如果 HTTP 服务器采用其他 TCP 端口，测试将会失败。

1.6.6 配置 UDP-jitter 测试



说明

建议不要对知名端口，即 1～1023 之间的端口，进行 UDP-jitter 测试，否则可能导致 NQA 测试失败或该知名端口对应的服务不可用。

语音、视频等实时性业务对 Delay jitter（时延抖动）的要求较高。通过 UDP-jitter 测试，可以获得网络的单向和双向时延抖动，从而判断网络是否可以承载实时性业务。

Delay jitter（时延抖动）是指相邻两个报文的接收时间间隔减去这两个报文的发送时间间隔。UDP-jitter 测试的过程如下：

- 源端以一定的时间间隔向目的端发送数据包。

- 目的端每收到一个数据包，就给它打上时间戳，然后再把这个数据包发回到源端。
- 源端收到数据包后通过计算目的端接收数据包时间间隔和源端发送数据包的时间间隔之差，计算出时延抖动，从而清晰地反映出网络状况。

1. 配置准备

UDP-jitter测试需要NQA服务器和客户端配合才能完成。进行UDP-jitter测试之前，必须保证NQA服务器端配置了UDP监听功能，配置方法请参见“1.3 配置NQA服务器”。

2. 配置 UDP-jitter 测试

表1-11 配置 UDP-jitter 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 UDP-jitter，并进入测试类型视图	**type udp-jitter**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 测试操作的目的地址必须与 NQA 服务器上所配置的监听服务的 IP 地址一致
配置测试操作的目的端口	**destination port** *port-number*	必选 缺省情况下，未配置测试操作的目的端口号 测试操作的目的端口号必须与 NQA 服务器上所配置的监听服务的端口号一致
配置请求报文中的源端口号	**source port** *port-number*	可选 缺省情况下，未指定源端口号
配置发送的探测报文的大小	**data-size** *size*	可选 缺省情况下，发送的探测报文为 100 字节
配置发送的探测报文的填充字符串	**data-fill** *string*	可选 缺省情况下，探测报文的填充内容为十六进制数值 00010203040506070809
配置一次 UDP-jitter 探测中发送探测报文的个数	**probe packet-number** *packet-number*	可选 缺省情况下，一次 UDP-jitter 探测中发送 10 个探测报文
配置 UDP-jitter 测试中发送探测报文的时间间隔	**probe packet-interval** *packet-interval*	可选 缺省情况下，UDP-jitter 测试中发送探测报文的时间间隔为 20 毫秒
配置 UDP-jitter 测试中等待响应报文的超时时间	**probe packet-timeout** *packet-timeout*	可选 缺省情况下，UDP-jitter 测试中等待响应报文的超时时间为 3000 毫秒

操作	命令	说明
配置测试操作中探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置通用的可选参数	参见 表 1-21	可选

说明

一次 UDP-jitter 测试中探测的次数取决于 **probe count** 命令的配置，而每次探测所发送的探测包的个数由 **probe packet-number** 命令配置决定。

1.6.7 配置 SNMP 测试

SNMP 查询测试主要用来测试从 NQA 客户端向 SNMP agent 设备发出一个 SNMP 协议查询报文到接收响应报文的时间。

1. 配置准备

在进行 SNMP 测试之前，需要在作为 SNMP agent 的设备上启动 SNMP agent 功能。SNMP agent 的配置方法，请参见“系统分册”中的“SNMP 配置”。

2. 配置 SNMP 测试

表1-12 配置 SNMP 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 SNMP，并进入测试类型视图	**type snmp**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址
配置测试操作中请求报文中的源端口号	**source port** *port-number*	可选 缺省情况下，未指定源端口号
配置测试操作中探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置通用的可选参数	参见 表 1-21	可选

1.6.8 配置 TCP 测试

TCP 测试通过测试客户端和服务器指定端口之间是否能够建立 TCP 连接，以及建立 TCP 连接所需的时间，来判断服务器指定端口上提供的服务是否可用，及服务性能。

1. 配置准备

TCP测试需要NQA服务器和客户端配合才能完成。在TCP测试之前，需要在NQA服务器端配置TCP监听功能，配置方法请参见“1.3 配置NQA服务器”。

2. 配置 TCP 测试

表1-13 配置 TCP 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 TCP，并进入测试类型视图	**type tcp**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 必须与 NQA 服务器上配置的监听服务的 IP 地址一致
配置目的端口	**destination port** *port-number*	必选 缺省情况下，未配置测试操作的目的端口号 必须与 NQA 服务器上配置的监听服务的端口号一致
配置测试操作中探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源IP地址必须是设备上接口的IP地址，且接口为 up 状态，否则测试将会失败
配置通用的可选参数	参见 表 1-21	可选

1.6.9 配置 UDP-echo 测试

UDP-echo 测试可以用来测试客户端和服务器指定 UDP 端口之间的连通性以及 UDP-echo 协议报文的往返时间。

1. 配置准备

UDP-echo测试需要NQA服务器和客户端配合才能完成。在进行UDP-echo测试之前，需要在NQA服务器端配置UDP监听功能，配置方法请参见“1.3 配置NQA服务器”。

2. 配置 UDP-echo 测试

表1-14 配置 UDP-echo 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 UDP-echo，并进入测试类型视图	**type udp-echo**	必选

操作	命令	说明
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 必须与 NQA 服务器上配置的监听服务的 IP 地址一致
配置目的端口	**destination port** *port-number*	必选 缺省情况下，未配置测试操作的目的端口号 必须与 NQA 服务器上配置的监听服务的端口号一致
配置发送的探测报文的大小	**data-size** *size*	可选 缺省情况下，UDP－echo 测试中发送的探测报文为 100 字节
配置发送的探测报文的填充字符串	**data-fill** *string*	可选 缺省情况下，探测报文的填充内容为十六进制数值 00010203040506070809
配置测试操作中探测请求报文中的源端口号	**source port** *port-number*	可选 缺省情况下，未指定源端口号
配置测试操作中探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置通用的可选参数	参见 表 1-21	可选

1.6.10 配置 Voice 测试

说明

建议不要对知名端口，即 1～1023 之间的端口，进行 Voice 测试，否则可能导致 NQA 测试失败或该知名端口对应的服务不可用。

Voice 测试主要用来测试 VoIP（Voice over IP，在 IP 网络上传送语音数据）网络情况，统计 VoIP 网络参数，以便用户根据网络情况进行相应的调整。

Voice 测试的过程如下：

(1) 源端（NQA 客户端）以一定的时间间隔向目的端（NQA 服务器）发送 G.711 A 律、G.711 μ 律或 G.729 A 律编码格式的语音数据包。

(2) 目的端每收到一个数据包，就给它打上时间戳，然后再把这个数据包发回到源端。

(3) 源端收到数据包后通过计算目的端接收两个数据包的时间间隔和源端发送两个数据包的时间间隔之差，计算出时延抖动、单向延迟等网络参数，从而清晰地反映出网络状况。

Voice 测试还可以计算出反映 VoIP 网络状况的语音参数值，包括：

- ICPIF（Calculated Planning Impairment Factor，计算计划损伤元素）：用来量化网络中语音数据的衰减，由单向网络延迟和丢包率等决定。数值越大，表明语音网络质量越差。
- MOS（Mean Opinion Scores，平均意见得分）：用来衡量语音网络的质量。通过计算网络中语音数据的衰减——ICPIF 值，可以估算出 MOS 值，MOS 值的取值范围为 1～15。该值越高，表明语音网络质量越好。

对语音质量的评价具有一定的主观性，不同用户对语音质量的容忍程度不同，因此，衡量语音质量时，需要考虑用户的主观因素。对语音质量容忍程度较强的用户，可以通过 **advantage-factor** 命令配置补偿因子，在计算 ICPIF 值时将减去该补偿因子，修正 ICPIF 和 MOS 值，以便在比较语音质量时综合考虑客观和主观因素。

1. 配置准备

Voice测试需要NQA服务器和客户端配合才能完成。进行Voice测试之前，必须保证NQA服务器端配置了UDP监听功能，配置方法请参见“1.3 配置NQA服务器”。

2. 配置 Voice 测试

表1-15 配置 Voice 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 Voice，并进入测试类型视图	**type voice**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址 测试操作的目的地址必须与 NQA 服务器上所配置的监听服务的 IP 地址一致
配置测试操作的目的端口	**destination port** *port-number*	必选 缺省情况下，未配置测试操作的目的端口号 测试操作的目的端口号必须与 NQA 服务器上所配置的监听服务的端口号一致
配置 Voice 测试的编码格式	**codec-type** { **g711a** | **g711u** | **g729a** }	可选 缺省情况下，语音编码格式为 G.711 A
配置用于计算 MOS 值和 ICPIF 值的补偿因子	**advantage-factor** *factor*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置探测请求报文中的源端口号	**source port** *port-number*	可选 缺省情况下，未指定源端口号
配置发送的探测报文的大小	**data-size** *size*	可选 缺省情况下，发送的探测报文大小与配置的编码格式有关，编码格式为 **g.711a** 和 **g.711u** 时缺省报文大小为 172 字节，**g.729a** 时为 32 字节
配置发送的探测报文的填充字符串	**data-fill** *string*	可选 缺省情况下，探测报文的填充内容为十六进制数值 00010203040506070809
配置一次Voice探测中发送探测报文的个数	**probe packet-number** *packet-number*	可选 缺省情况下，一次 Voice 探测中发送 1000 个探测报文

操作	命令	说明
配置Voice探测中发送探测报文的时间间隔	**probe packet-interval** *packet-interval*	可选 缺省情况下，Voice 探测中发送探测报文的时间间隔为 20 毫秒
配置Voice测试中等待响应报文的超时时间	**probe packet-timeout** *packet-timeout*	可选 缺省情况下，Voice 测试中等待响应报文的超时时间为 5000 毫秒
配置通用的可选参数	参见 表 1-21	可选



说明

一次 Voice 测试只能进行一次探测，而每次探测所发送的探测报文的个数由 **probe packet-number** 命令配置决定。

1.6.11 配置 DLSw 测试

DLSw 测试主要用来测试 DLSw 设备的响应时间。

1. 配置准备

在进行 DLSw 测试之前，需要在对端设备上使能 DLSw 功能。

2. 配置 DLSw 测试

表1-16 配置 DLSw 测试

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
配置测试类型为 DLSw，并进入测试类型视图	**type dlsw**	必选
配置测试操作的目的地址	**destination ip** *ip-address*	必选 缺省情况下，未配置测试操作的目的 IP 地址
配置测试操作中探测请求报文的源 IP 地址	**source ip** *ip-address*	可选 缺省情况下，未指定源 IP 地址 该命令指定的源 IP 地址必须是设备上接口的 IP 地址，且接口为 up 状态，否则测试将会失败
配置通用的可选参数	参见 表 1-21	可选

1.7 配置联动功能

联动功能是通过建立联动项，对当前所在测试组中的探测进行监测，当连续探测失败次数达到阈值时，就触发配置的动作类型。

表1-17 配置联动功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
进入测试组测试类型视图	**type** { **dhcp** | **dlsw** | **dns** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** }	- UDP-jitter 和 Voice 测试不支持联动功能
建立联动项	**reaction** *item-num* **checked-element probe-fail threshold-type consecutive** *occurrences* [**action-type** { **none** | **trigger-only** }]	必选 缺省情况下，未配置联动项
退回系统视图	**quit**	-
创建与 NQA 测试组中指定联动项关联的 Track 项	**track** *entry-number* **nqa entry** *admin-name operation-tag* **reaction** *item-num*	必选 缺省情况下，未配置 Track 项

说明

- 联动项创建后，不能再通过 **reaction** 命令修改该联动项的内容。
- UDP-jitter 和 Voice 测试不支持联动功能。

1.8 配置发送 Trap 消息

NQA 测试中可以通过设置 Trap 开关，控制是否向网管发送 Trap 消息。

1. 配置准备

在配置发送 Trap 消息之前，需要通过 **snmp-agent target-host** 命令配置 Trap 消息的目的地址，正确创建 NQA 测试组并配置相关参数。**snmp-agent target-host** 命令的详细介绍，请参见“系统分册”中的“SNMP 命令”。

2. 配置发送 Trap 消息

表1-18 配置发送 Trap 消息

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
进入测试组测试类型视图	**type** { **dhcp** | **dlsw** | **dns** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	-
配置在指定条件下向网管服务器发送 Trap 消息	**reaction trap** { **probe-failure** *consecutive-probe-failures* | **test-complete** | **test-failure** *cumulate-probe-failures* }	可选 缺省情况下，不向网管服务器发送 Trap 消息

说明

Voice 测试类型只支持 **reaction trap test-complete**，即只支持测试成功时向网管服务器发送 Trap 消息。

1.9 配置 NQA 统计功能

NQA 将在指定时间间隔内完成的 NQA 测试归为一组，计算该组测试结果的统计值，这些统计值构成一个统计组。通过 **display nqa statistics** 命令可以显示该统计组的信息。通过 **statistics interval** 命令可以设置统计的时间间隔。

当保留的统计组数目达到最大值时，如果形成新的统计组，保存时间最久的统计组将被删除。通过 **statistics max-group** 命令可以设置保留的最大统计组个数。

指定时间间隔内最后一次测试结束后，形成一个统计组。统计组具有老化功能，即统计组保存一定时间后，将被删除。通过 **statistics hold-time** 命令可以设置统计组的保留时间。

表1-19 配置 NQA 统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
进入测试组测试类型视图	**type** { **dlsw** | **dns** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	-
配置对测试结果进行统计的时间间隔	**statistics interval** *interval*	可选 缺省情况下，对测试结果进行统计的时间间隔为 60 分钟
配置能够保留的最大统计组个数	**statistics max-group** *number*	可选 缺省情况下，能够保留的最大统计组数为 2 最大统计组个数为 0 时，不进行统计
配置统计组的保留时间	**statistics hold-time** *hold-time*	可选 缺省情况下，统计组的保留时间为 120 分钟

说明

- DHCP 测试不支持配置 NQA 统计功能。
- 如果通过 **frequency** 命令指定连续两次测试开始时间的时间间隔为 0，则不生成统计组信息。

1.10 配置 NQA 历史记录功能

开启 NQA 测试组的历史记录保存功能后，系统将记录 NQA 测试的历史信息，通过 **display nqa history** 命令可以查看该测试组的历史记录信息。

通过本配置任务还可以指定：

- 历史记录的保存时间：历史记录保存时间达到配置的值后，将删除该历史记录。
- 一个测试组中能够保存的最大历史记录个数：如果历史记录个数超过设定的最大数目，则最早的历史记录将会被删除。

表1-20 配置 NQA 历史记录功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
进入测试组测试类型视图	**type** { **dhcp** | **dlsw** | **dns** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	-
开启 NQA 测试组的历史记录保存功能	**history-record enable**	必选 缺省情况下，NQA 测试组的历史记录保存功能处于关闭状态
配置 NQA 测试组中历史记录的保存时间	**history-record keep-time** *keep-time*	可选 缺省情况下，NQA 测试组中历史记录的保存时间为 120 分钟
配置在一个测试组中能够保存的最大历史记录个数	**history-record number** *number*	可选 缺省情况下，一个测试组中能够保存的最大历史记录个数为 50

1.11 配置 NQA 测试组通用可选参数

NQA 测试组的通用可选参数，只对该测试组中的测试有效。

除特别说明外，所有测试类型都可以配置通用可选参数，可以根据实际情况选择配置测试组的参数。

表1-21 配置 NQA 测试组的通用可选参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 NQA 测试组视图	**nqa entry** *admin-name operation-tag*	-
进入测试组测试类型视图	**type** { **dhcp** | **dlsw** | **dns** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	-
配置测试组的描述字符串	**description** *text*	可选 缺省情况下，测试组没有描述字符串
配置测试组连续两次测试开始时间的时间间隔	**frequency** *interval*	可选 缺省情况下，Voice 测试中连续两次测试开始时间的时间间隔为 60000 毫秒；其他类型的测试中连续两次测试开始时间的时间间隔为 0 毫秒，即只进行一次测试
配置一次 NQA 测试中进行探测的次数	**probe count** *times*	可选 缺省情况下，一次测试中的探测次数为 1 次 Voice 测试中探测次数只能为 1，不支持该命令

操作	命令	说明
配置 NQA 探测超时时间	**probe timeout** *timeout*	可选 缺省情况下，探测的超时时间为 3000 毫秒 UDP-jitter 测试不能配置该参数
配置探测报文在网络中可以经过的最大跳数	**ttl** *value*	可选 缺省情况下，探测报文在网络中可以经过的最大跳数为 20 跳 DHCP 测试不能配置该参数
配置 NQA 探测报文 IP 报文头中服务类型域的值	**tos** *value*	可选 缺省情况下，NQA 探测报文 IP 报文头中服务类型域的值为 0 DHCP 测试不能配置该参数
启动路由表旁路功能	**route-option bypass-route**	可选 缺省情况下，路由表旁路功能处于关闭状态 DHCP 测试不能配置该参数

1.12 调度 NQA 测试组

通过本配置，可以设置测试组进行测试的启动时间和持续时间。启动时间取值可以是具体的时间值和 **now**，**now** 表示立即启动测试；持续时间取值可以是具体的时间值和 **forever**，**forever** 表示一直进行测试，直到用户通过 **undo nqa schedule** 命令手动停止测试。

系统时间在启动时间<启动时间>到<启动时间＋持续时间>范围内时，测试组进行测试。执行 **nqa schedule** 命令时，如果系统时间尚未到达启动时间，则到达启动时间后，启动测试；如果系统时间在启动时间～启动时间＋持续时间之间，则立即启动测试；如果系统时间已经超过启动时间＋持续时间，则不会启动测试。通过 **display clock** 命令可以查看系统的当前时间。

1. 配置准备

在调度 NQA 测试组之前，需要保证：

- 测试类型相应的测试参数配置完整；
- 对于需要与 NQA 服务器配合的测试类型，已经完成 NQA 服务器端的配置。

2. 调度 NQA 测试组

表1-22 调度 NQA 测试组

操作	命令	说明
进入系统视图	**system-view**	-
调度 NQA 测试组	**nqa schedule** *admin-name operation-tag* **start-time** { *hh:mm:ss* [*yyyy/mm/dd*] | **now** } **lifetime** { *lifetime* | **forever** }	必选
配置处于测试状态的 NQA 测试的最大个数	**nqa agent max-concurrent** *number*	可选 缺省情况下，处于测试状态的 NQA 测试的最大个数为 40

注意

- 测试组被调度后就不能再进入该测试组视图和测试类型视图。
- 对于已启动的测试组或已经完成测试的测试组，不受系统时间调整的影响，只有等待测试的测试组受系统时间调整的影响。

1.13 NQA 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 NQA 的运行情况，通过查看显示信息验证配置的效果。

表1-23 NQA 显示和维护

操作	命令
显示 NQA 测试操作信息的历史记录	**display nqa history** [*admin-name operation-tag*]
显示最近一次 NQA 测试的结果	**display nqa result** [*admin-name operation-tag*]
显示 NQA 测试的统计信息	**display nqa statistics** [*admin-name operation-tag*]
显示服务器的状态信息	**display nqa server status**

1.14 NQA 典型配置举例

1.14.1 ICMP-echo 测试配置举例

1. 组网需求

使用 NQA 进行 ICMP-echo 测试，测试本端（Router A）发送的报文是否可以到达指定的目的端（Router B），以及报文的往返时间。

2. 组网图

图1-3 ICMP-echo 测试组网图



3. 配置步骤

创建 ICMP-echo 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type icmp-echo
[RouterA-nqa-admin-test-icmp-echo] destination ip 10.2.2.2
```

配置可选参数。

```
[RouterA-nqa-admin-test-icmp-echo] probe count 10
[RouterA-nqa-admin-test-icmp-echo] probe timeout 500
```

```
[RouterA-nqa-admin-test-icmp-echo] frequency 5000
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-icmp-echo] history-record enable
[RouterA-nqa-admin-test-icmp-echo] history-record number 10
[RouterA-nqa-admin-test-icmp-echo] quit
```

启动 ICMP-echo 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 ICMP-echo 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 ICMP-echo 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 10             Receive response times: 10
      Min/Max/Average round trip time: 2/5/3
      Square-Sum of round trip time: 96
      Last succeeded probe time: 2007-08-23 15:00:01.2
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 ICMP-echo 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    370        3            Succeeded        2007-08-23 15:00:01.2
    369        3            Succeeded        2007-08-23 15:00:01.2
    368        3            Succeeded        2007-08-23 15:00:01.2
    367        5            Succeeded        2007-08-23 15:00:01.2
    366        3            Succeeded        2007-08-23 15:00:01.2
    365        3            Succeeded        2007-08-23 15:00:01.2
    364        3            Succeeded        2007-08-23 15:00:01.1
    363        2            Succeeded        2007-08-23 15:00:01.1
    362        3            Succeeded        2007-08-23 15:00:01.1
    361        2            Succeeded        2007-08-23 15:00:01.1
```

1.14.2 DHCP 测试配置举例

1. 组网需求

使用 DHCP 功能测试 Router A 从 DHCP 服务器 Router B 申请到 IP 地址所需的时间。

2. 组网图

图1-4 配置 DHCP 组网图



3. 配置步骤

创建 DHCP 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type dhcp
[RouterA-nqa-admin-test-dhcp] operation interface gigabitethernet 1/1
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-dhcp] history-record enable
[RouterA-nqa-admin-test-dhcp] quit
```

启动 DHCP 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 DHCP 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 DHCP 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 512/512/512
      Square-Sum of round trip time: 262144
      Last succeeded probe time: 2007-11-22 09:54:03.8
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 DHCP 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          512          Succeeded        2007-11-22 09:54:03.8
```

1.14.3 DNS 测试配置举例

1. 组网需求

使用 DNS 功能测试 Router A 是否可以通过指定的 DNS 服务器将域名 host.com 解析为 IP 地址，并测试域名解析所需的时间。

2. 组网图

图1-5 配置 DNS 组网图



3. 配置步骤

创建 DNS 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type dns
[RouterA-nqa-admin-test-dns] destination ip 10.2.2.2
[RouterA-nqa-admin-test-dns] resolve-target host.com
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-dns] history-record enable
[RouterA-nqa-admin-test-dns] quit
```

启动 DNS 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 DNS 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 DNS 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 62/62/62
      Square-Sum of round trip time: 3844
      Last succeeded probe time: 2008-11-10 10:49:37.3
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 DNS 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          62           Succeeded        2008-11-10 10:49:37.3
```

1.14.4 FTP 测试配置举例

1. 组网需求

使用 FTP 功能测试 Router A 是否可以和指定的 FTP 服务器 Router B 建立连接，以及往 FTP 服务器上传一个文件的时间。登录 FTP 服务器的用户名为 admin，密码为 systemtest，要传送到服务器的文件名为 config.txt。

2. 组网图

图1-6 配置 FTP 组网图



3. 配置步骤

创建 FTP 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type ftp
[RouterA-nqa-admin-test-ftp] destination ip 10.2.2.2
[RouterA-nqa-admin-test-ftp] source ip 10.1.1.1
[RouterA-nqa-admin-test-ftp] operation put
[RouterA-nqa-admin-test-ftp] username admin
[RouterA-nqa-admin-test-ftp] password systemtest
[RouterA-nqa-admin-test-ftp] filename config.txt
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-ftp] history-record enable
[RouterA-nqa-admin-test-ftp] quit
```

启动 FTP 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 FTP 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 FTP 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 173/173/173
```

```
    Square-Sum of round trip time: 29929
    Last succeeded probe time: 2007-11-22 10:07:28.6
  Extended results:
    Packet lost in test: 0%
    Failures due to timeout: 0
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to sequence error: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packet(s) arrived late: 0
```

显示 FTP 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          173          Succeeded        2007-11-22 10:07:28.6
```

1.14.5 HTTP 测试配置举例

1. 组网需求

使用 HTTP 功能，测试是否可以和指定的 HTTP 服务器之间建立连接，以及从 HTTP 服务器获取数据的时间。

2. 组网图

图1-7 HTTP 测试组网图



3. 配置步骤

创建 HTTP 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type http
[RouterA-nqa-admin-test-http] destination ip 10.2.2.2
[RouterA-nqa-admin-test-http] operation get
[RouterA-nqa-admin-test-http] url /index.htm
[RouterA-nqa-admin-test-http] http-version v1.0
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-http] history-record enable
[RouterA-nqa-admin-test-http] quit
```

启动 HTTP 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 HTTP 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 HTTP 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 64/64/64
      Square-Sum of round trip time: 4096
      Last succeeded probe time: 2007-11-22 10:12:47.9
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 HTTP 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          64           Succeeded        2007-11-22 10:12:47.9
```

1.14.6 UDP-jitter 测试配置举例

1. 组网需求

使用 UDP-jitter 功能测试本端（Router A）和指定的目的端（Router B）之间传送报文的时延抖动。

2. 组网图

图1-8 UDP-jitter 测试组网图



3. 配置步骤

(1) 配置 Router B

使能 NQA 服务器，配置监听的 IP 地址为 10.2.2.2，端口号为 9000。

```
<RouterB> system-view
[RouterB] nqa server enable
[RouterB] nqa server udp-echo 10.2.2.2 9000
```

(2) 配置 Router A

创建 UDP-jitter 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type udp-jitter
[RouterA-nqa-admin-test-udp-jitter] destination ip 10.2.2.2
[RouterA-nqa-admin-test-udp-jitter] destination port 9000
[RouterA-nqa-admin-test-udp-jitter] frequency 1000
[RouterA-nqa-admin-test-udp-jitter] quit
```

启动 UDP-jitter 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 UDP-jitter 测试操作。

```
[RouterA] undo nqa schedule admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 10             Receive response times: 10
      Min/Max/Average round trip time: 15/32/17
      Square-Sum of round trip time: 3235
      Last succeeded probe time: 2008-05-29 13:56:17.6
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
    UDP-jitter results:
     RTT number: 10
      Min positive SD: 4                   Min positive DS: 1
      Max positive SD: 21                  Max positive DS: 28
      Positive SD number: 5                Positive DS number: 4
      Positive SD sum: 52                  Positive DS sum: 38
      Positive SD average: 10              Positive DS average: 10
      Positive SD square sum: 754          Positive DS square sum: 460
      Min negative SD: 1                   Min negative DS: 6
      Max negative SD: 13                  Max negative DS: 22
      Negative SD number: 4                Negative DS number: 5
      Negative SD sum: 38                  Negative DS sum: 52
      Negative SD average: 10              Negative DS average: 10
      Negative SD square sum: 460          Negative DS square sum: 754
    One way results:
      Max SD delay: 15                     Max DS delay: 16
      Min SD delay: 7                      Min DS delay: 7
      Number of SD delay: 10               Number of DS delay: 10
      Sum of SD delay: 78                  Sum of DS delay: 85
      Square sum of SD delay: 666          Square sum of DS delay: 787
      SD lost packet(s): 0                 DS lost packet(s): 0
```

```
    Lost packet(s) for unknown reason: 0
```

显示 UDP-jitter 测试的统计结果。

```
[RouterA] display nqa statistics admin test
  NQA entry(admin admin, tag test) test statistics:
    NO. : 1
    Destination IP address: 10.2.2.2
      Start time: 2008-05-29 13:56:14.0
      Life time: 47
      Send operation times: 410            Receive response times: 410
      Min/Max/Average round trip time: 1/93/19
      Square-Sum of round trip time: 206176
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
    UDP-jitter results:
     RTT number: 410
      Min positive SD: 3                   Min positive DS: 1
      Max positive SD: 30                  Max positive DS: 79
      Positive SD number: 186              Positive DS number: 158
      Positive SD sum: 2602                Positive DS sum: 1928
      Positive SD average: 13              Positive DS average: 12
      Positive SD square sum: 45304        Positive DS square sum: 31682
      Min negative SD: 1                   Min negative DS: 1
      Max negative SD: 30                  Max negative DS: 78
      Negative SD number: 181              Negative DS number: 209
      Negative SD sum: 181                 Negative DS sum: 209
      Negative SD average: 13              Negative DS average: 14
      Negative SD square sum: 46994        Negative DS square sum: 3030
    One way results:
      Max SD delay: 46                     Max DS delay: 46
      Min SD delay: 7                      Min DS delay: 7
      Number of SD delay: 410              Number of DS delay: 410
      Sum of SD delay: 3705                Sum of DS delay: 3891
      Square sum of SD delay: 45987        Square sum of DS delay: 49393
      SD lost packet(s): 0                 DS lost packet(s): 0
      Lost packet(s) for unknown reason: 0
```

说明

display nqa history 命令的显示信息无法反映 UDP-jitter 测试的结果，如果想了解 UDP-jitter 测试的结果，建议通过 **display nqa result** 命令查看最近一次 NQA 测试的结果，或通过 **display nqa statistics** 命令查看 NQA 测试的统计信息。

1.14.7 SNMP 测试配置举例

1. 组网需求

使用 SNMP 查询测试从 Router A 发出一个 SNMP 协议查询报文到收到 SNMP agent（Router B）响应报文所用的时间。

2. 组网图

图1-9 SNMP 配置测试组网图



3. 配置步骤

(1) 在 SNMP agent 上进行配置

启动 SNMP agent 服务，设置 SNMP 版本为 all、只读团体名为 public、读写团体名为 private。

```
<RouterB> system-view
[RouterB] snmp-agent sys-info version all
[RouterB] snmp-agent community read public
[RouterB] snmp-agent community write private
```

(2) 在 Router A 上进行配置

创建 SNMP 查询类型的测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type snmp
[RouterA-nqa-admin-test-snmp] destination ip 10.2.2.2
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-snmp] history-record enable
[RouterA-nqa-admin-test-snmp] quit
```

启动测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 SNMP 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 SNMP 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
```

```
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 50/50/50
      Square-Sum of round trip time: 2500
      Last succeeded probe time: 2007-11-22 10:24:41.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 SNMP 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          50           Timeout          2007-11-22 10:24:41.1
```

1.14.8 TCP 测试配置举例

1. 组网需求

使用 TCP 功能测试本端（Router A）和指定目的端（Router B）之间建立 TCP 连接所需的时间，使用的端口号为 9000。

2. 组网图

图1-10 TCP 测试组网图



3. 配置步骤

(1) 配置 Router B

使能 NQA 服务器，配置监听 IP 地址 10.2.2.2 和端口号 9000。

```
<RouterB> system-view
[RouterB] nqa server enable
[RouterB] nqa server tcp-connect 10.2.2.2 9000
```

(2) 配置 Router A

创建 TCP 类型的测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type tcp
[RouterA-nqa-admin-test-tcp] destination ip 10.2.2.2
```

```
[RouterA-nqa-admin-test-tcp] destination port 9000
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-tcp] history-record enable
[RouterA-nqa-admin-test-tcp] quit
```

启动测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 TCP 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 TCP 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 13/13/13
      Square-Sum of round trip time: 169
      Last succeeded probe time: 2007-11-22 10:27:25.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 TCP 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          13           Succeeded        2007-11-22 10:27:25.1
```

1.14.9 UDP-echo 测试配置举例

1. 组网需求

使用 UDP-echo 功能测试本端（Router A）和指定目的端（Router B）之间 UDP 协议报文的往返时间，使用的端口号设置为 8000。

2. 组网图

图1-11 UDP-echo 测试组网图



3. 配置步骤

(1) 配置 Router B

使能 NQA 服务器，配置监听的 IP 地址为 10.2.2.2，端口号为 8000。

```
<RouterB> system-view
[RouterB] nqa server enable
[RouterB] nqa server udp-echo 10.2.2.2 8000
```

(2) 配置 Router A

创建 UDP-echo 类型的测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type udp-echo
[RouterA-nqa-admin-test-udp-echo] destination ip 10.2.2.2
[RouterA-nqa-admin-test-udp-echo] destination port 8000
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-udp-echo] history-record enable
[RouterA-nqa-admin-test-udp-echo] quit
```

启动测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 UDP-echo 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 UDP-echo 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 25/25/25
      Square-Sum of round trip time: 625
      Last succeeded probe time: 2007-11-22 10:36:17.9
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 UDP-echo 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          25           Succeeded        2007-11-22 10:36:17.9
```

1.14.10 Voice 测试配置举例

1. 组网需求

使用 Voice 功能测试本端（Router A）和指定的目的端（Router B）之间传送语音报文的时延抖动和网络语音质量参数。

2. 组网图

图1-12 Voice 测试组网图



3. 配置步骤

(1) 配置 Router B

使能 NQA 服务器，配置监听的 IP 地址为 10.2.2.2，端口号为 9000。

```
<RouterB> system-view
[RouterB] nqa server enable
[RouterB] nqa server udp-echo 10.2.2.2 9000
```

(2) 配置 Router A

创建 Voice 类型的 NQA 测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type voice
[RouterA-nqa-admin-test-voice] destination ip 10.2.2.2
[RouterA-nqa-admin-test-voice] destination port 9000
[RouterA-nqa-admin-test-voice] quit
```

启动 Voice 测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 Voice 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 Voice 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1000          Receive response times: 1000
      Min/Max/Average round trip time: 31/1328/33
      Square-Sum of round trip time: 2844813
      Last succeeded probe time: 2008-06-13 09:49:31.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
```

```
    Failures due to no connection: 0
    Failures due to sequence error: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packet(s) arrived late: 0
  Voice results:
   RTT number: 1000
    Min positive SD: 1                    Min positive DS: 1
    Max positive SD: 204                  Max positive DS: 1297
    Positive SD number: 257               Positive DS number: 259
    Positive SD sum: 759                  Positive DS sum: 1797
    Positive SD average: 2                Positive DS average: 6
    Positive SD square sum: 54127         Positive DS square sum: 1691967
    Min negative SD: 1                    Min negative DS: 1
    Max negative SD: 203                  Max negative DS: 1297
    Negative SD number: 255               Negative DS number: 259
    Negative SD sum: 759                  Negative DS sum: 1796
    Negative SD average: 2                Negative DS average: 6
    Negative SD square sum: 53655         Negative DS square sum: 1691776
  One way results:
    Max SD delay: 343                     Max DS delay: 985
    Min SD delay: 343                     Min DS delay: 985
    Number of SD delay: 1                 Number of DS delay: 1
    Sum of SD delay: 343                  Sum of DS delay: 985
    Square sum of SD delay: 117649        Square sum of DS delay: 970225
    SD lost packet(s): 0                  DS lost packet(s): 0
    Lost packet(s) for unknown reason: 0
  Voice scores:
    MOS value: 4.38                       ICPIF value: 0
```

显示 Voice 测试的统计结果。

```
[RouterA] display nqa statistics admin test
  NQA entry(admin admin, tag test) test statistics:
    NO. : 1
    Destination IP address: 10.2.2.2
      Start time: 2008-06-13 09:45:37.8
      Life time: 331
      Send operation times: 4000          Receive response times: 4000
      Min/Max/Average round trip time: 15/1328/32
      Square-Sum of round trip time: 7160528
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

```
Voice results:
 RTT number: 4000
  Min positive SD: 1                    Min positive DS: 1
  Max positive SD: 360                  Max positive DS: 1297
  Positive SD number: 1030              Positive DS number: 1024
  Positive SD sum: 4363                 Positive DS sum: 5423
  Positive SD average: 4                Positive DS average: 5
  Positive SD square sum: 497725        Positive DS square sum: 2254957
  Min negative SD: 1                    Min negative DS: 1
  Max negative SD: 360                  Max negative DS: 1297
  Negative SD number: 1028              Negative DS number: 1022
  Negative SD sum: 1028                 Negative DS sum: 1022
  Negative SD average: 4                Negative DS average: 5
  Negative SD square sum: 495901        Negative DS square sum: 5419
One way results:
  Max SD delay: 359                     Max DS delay: 985
  Min SD delay: 0                       Min DS delay: 0
  Number of SD delay: 4                 Number of DS delay: 4
  Sum of SD delay: 1390                 Sum of DS delay: 1079
  Square sum of SD delay: 483202        Square sum of DS delay: 973651
  SD lost packet(s): 0                  DS lost packet(s): 0
  Lost packet(s) for unknown reason: 0
Voice scores:
  Max MOS value: 4.38                   Min MOS value: 4.38
  Max ICPIF value: 0                    Min ICPIF value: 0
```

说明

display nqa history 命令的显示信息无法反映 Voice 测试的结果，如果想了解 Voice 测试的结果，建议通过 **display nqa result** 命令查看最近一次 NQA 测试的结果，或通过 **display nqa statistics** 命令查看 NQA 测试的统计信息。

1.14.11 DLSw 测试配置举例

1. 组网需求

使用 DLSw 功能测试 DLSw 设备的响应时间。

2. 组网图

图1-13 DLSw 测试组网图



3. 配置步骤

创建 DLSw 类型的测试组并配置相关测试参数。

```
<RouterA> system-view
[RouterA] nqa entry admin test
[RouterA-nqa-admin-test] type dlsw
[RouterA-nqa-admin-test-dlsw] destination ip 10.2.2.2
```

配置 NQA 历史记录功能。

```
[RouterA-nqa-admin-test-dlsw] history-record enable
[RouterA-nqa-admin-test-dlsw] quit
```

启动测试操作。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

测试执行一段时间后，停止 DLSw 测试操作。

```
[RouterA] undo nqa schedule admin test
```

显示 DLSw 测试中最后一次测试的结果。

```
[RouterA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 19/19/19
      Square-Sum of round trip time: 361
      Last succeeded probe time: 2007-11-22 10:40:27.7
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

显示 DLSw 测试的历史记录。

```
[RouterA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          19           Succeeded        2007-11-22 10:40:27.7
```

目 录

1 NetStream配置 1-1
1.1 NetStream简介 1-1
1.1.1 NetStream概述 1-1
1.1.2 NetStream输出报文版本格式 1-2
1.1.3 NetStream聚合功能简介 1-2
1.1.4 NetStream实现 1-3
1.2 NetStream配置任务简介 1-3
1.3 配置NetStream统计功能 1-4
1.3.1 配置NetStream统计功能 1-4
1.4 配置NetStream聚合统计功能 1-4
1.4.1 配置准备 1-4
1.4.2 配置NetStream聚合统计功能 1-4
1.5 配置NetStream统计输出报文的属性 1-5
1.5.1 配置NetStream统计输出报文的属性 1-5
1.5.2 配置NetStream统计输出报文版本 9 模板的刷新率 1-6
1.6 配置NetStream MPLS报文统计功能 1-6
1.6.1 MPLS统计简介 1-6
1.6.2 配置NetStream MPLS报文统计功能 1-7
1.7 配置NetStream的流老化 1-7
1.7.1 流老化简介 1-7
1.7.2 配置NetStream的流老化 1-8
1.8 配置NetStream的采样过滤功能 1-8
1.8.1 采样过滤简介 1-8
1.8.2 配置NetStream的采样过滤 1-8
1.9 NetStream显示和维护 1-9
1.10 NetStream典型配置举例 1-9
1.10.1 配置NetStream分类统计功能 1-9
1.10.2 配置NetStream统计输出功能 1-10

1 NetStream 配置

1.1 NetStream 简介

Internet 的高速发展为用户提供了更高的带宽，支持的业务和应用日渐增多，传统流量统计如SNMP、端口镜像等，由于统计流量方式不灵活或是需要投资专用服务器成本高等原因，无法满足对网络进行更细致的管理，需要一种新技术来更好的支持网络流量统计。

NetStream 技术是一种基于网络流信息的统计技术，可以对网络中的业务流量情况进行统计和分析。在网络的接入层、汇聚层、核心层上，都可以通过部署 NetStream 进行网络规划、网络监控和安全监控等。

1.1.1 NetStream 概述

1. 流定义

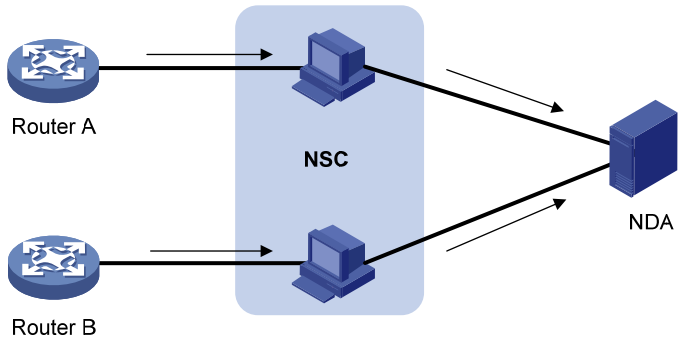
NetStream 提供报文统计功能，它根据 IPv4 报文的目的 IP 地址、源 IP 地址、目的端口号、源端口号、协议号、ToS（Type of Service，服务类型）以及输入或者输出接口来区分流，相同的七元组标识为同一条流。

2. 流老化

在实际网络环境中，可能在很短的时间内产生大量流，这就需要根据一定的规则把当前一部分流从缓冲区中删除，这个过程称为流老化（aging）。

3. NetStream 流量统计方式

图1-1 NetStream 数据采集和分析过程示意图



如 图 1-1所示，NetStream进行数据采集和分析的过程如下：

(1) 配置了 NetStream 功能的设备把采集到的关于流的详细信息定期发送给 NSC（NetStream Collector，网络流数据收集器）；

(2) 信息由 NSC 初步处理后发送给 NDA（NetStream Data Analyzer，网络流数据分析器）；

(3) NDA 对数据进行分析，以用于计费、网络规划等应用。NDA 可以通过 XLog 软件对输出的数据进行分析。

说明

XLog 是可扩展的网络日志审计系统（XLog Network Analysis System）的英文简称，它通过与网络设备（路由器、交换机等）配合组网，为网络管理者提供用户上网日志审计的解决方案。XLog 支持的日志类型有 NAT、FLOW、DIG、NetStream V5 等。

1.1.2 NetStream 输出报文版本格式

目前 NetStream 输出的报文主要有 5、8、9 三个版本。

- 版本 5：根据七元组产生原始的数据流，但报文格式固定，不易扩展。
- 版本 8：支持聚合输出格式，但报文格式固定，不易扩展。
- 版本 9：基于模板方式，使统计信息的输出更为灵活，可以用来灵活输出各种组合格式的数据。

  版本 9 支持对 BGP 下一跳、MPLS 等统计输出。

1.1.3 NetStream 聚合功能简介

NetStream 支持聚合功能。老化的流在输出前先按照一定的规则进行分类，生成聚合的信息后再发送出去。

在目前的实现中，支持 12 种聚合方式，如 表 1-1所示。

表1-1 NetStream 的 5 种聚合方式

聚合方式	分类依据
自治系统聚合	源 AS 号、目的 AS 号、输入接口索引、输出接口索引
协议-端口聚合	协议号、源端口、目的端口
源前缀聚合	源 AS 号、源掩码长度、源前缀、输入接口索引
目的前缀聚合	目的 AS 号、目的掩码长度、目的前缀、输出接口索引
源和目的前缀聚合	源 AS 号、目的 AS 号、源掩码长度、目的掩码长度、源前缀、目的前缀、输入接口索引、输出接口索引
前缀端口聚合	源前缀、目的前缀、源掩码长度、目的掩码长度、ToS、协议号、源端口、目的端口、输入接口索引、输出接口索引
服务类型-自治系统聚合	ToS、源 AS 号、目的 AS 号、输入接口索引、输出接口索引
服务类型-源前缀聚合	ToS、源 AS 号、源前缀、源掩码长度、输入接口索引
服务类型-目的前缀聚合	ToS、目的 AS 号、目的前缀、目的掩码长度、输出接口索引
服务类型-前缀聚合	ToS、源 AS 号、源前缀、源掩码长度、目的 AS 号、目的前缀、目的掩码长度、输入接口索引和输出接口索引
服务类型-协议-端口聚合	ToS、协议类型、源端口、目的端口、输入接口索引、输出接口索引
服务类型-BGP 下一跳聚合	ToS、BGP 下一跳地址、输出接口索引

说明

- 系统根据所选择聚合方式的分类依据，将符合对应聚合方式的多条流合并为一条聚合流，对应一条聚合记录。12 种聚合方式相互独立，可以同时配置。
- 在统计 AS 域号时，如果流量没有按照 BGP 的路由表进行转发则系统无法统计出 AS 域号。
- 在统计 BGP 下一跳地址时如果流量没有按照 BGP 的路由表进行转发则系统无法统计出 BGP 下一跳地址。

1.1.4 NetStream 实现

当设备启用 NetStream 功能后，IPv4 流信息首先被存储在设备的 NetStream 缓冲区中。当设定的 NetStream 流的老化时间超时或者对 NetStream 流执行了强制老化命令后，设备就将流信息通过 NetStream 统计输出报文（即，流信息经过 UDP 封装以后产生的报文）发送给 NSC，并从设备的 NetStream 缓冲区中将该流信息删除。

NetStream 统计输出报文有 3 种版本：版本 5、版本 8 和版本 9。目前用户可以配置版本 5 和版本 9 的 NetStream 统计输出报文的属性。

当用户配置版本 5 的 NetStream 统计输出报文属性时：

- 如果没有配置 NetStream 聚合方式，设备将直接通过版本 5 报文将老化后的 IPv4 流信息发送给 NSC。
- 如果配置了 NetStream 聚合方式，设备将按照一定的规则对 IPv4 流信息进行分类、合并后生成聚合信息，再通过版本 8 报文发送给 NSC。

当用户配置版本 9 的 NetStream 统计输出报文属性时：

- 如果没有配置 NetStream 聚合方式，设备将直接通过版本 9 报文将老化后的 IPv4 流信息发送给 NSC。
- 如果配置了 NetStream 聚合方式，设备不但直接通过版本 9 报文将老化后的 IPv4 流信息发送给 NSC，同时会按照一定的规则对 IPv4 流信息进行分类、合并后生成聚合信息，再通过版本 9 报文发送给 NSC。

MPLS 流信息只能通过版本 9 报文发送，仅当用户配置版本 9 的 NetStream 统计输出报文属性时，带 IPv4 选项信息的 MPLS 流信息才会发送给 NSC。

1.2 NetStream 配置任务简介

表1-2 NetStream 配置任务简介

配置任务	说明	详细配置
配置 NetStream 统计功能	必选	1.3
配置 NetStream 聚合统计功能	可选	1.4
配置 NetStream 统计输出报文的属性	必选	1.5
配置 NetStream MPLS 报文统计功能	可选	1.6
配置 NetStream 的流老化	可选	1.7
配置 NetStream 的采样过滤功能	可选	1.8

1.3 配置 NetStream 统计功能

1.3.1 配置 NetStream 统计功能

表1-3 配置 NetStream 统计功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 NetStream 流缓存区大小，即所能容纳 NetStream 流的最大数量，以及当流达到流缓存区的最大数量时，是进行强制老化，还是禁止建流	**ip netstream max-entry** { *max-entries* | **aging** | **disable-caching** }	可选 缺省情况下，NetStream 流缓存区大小为 20000，在达到最大流表个数时进行强制老化 该命令仅 SR6602 支持
进入接口视图	**interface** *interface-type interface-number*	-
使能接口的 NetStream 统计功能	**ip netstream** { **inbound** | **outbound** }	必选 缺省情况下，接口未启动 NetStream 统计功能

1.4 配置 NetStream 聚合统计功能

1.4.1 配置准备

在配置此统计功能前，必须使能 NetStream 统计功能。

1.4.2 配置 NetStream 聚合统计功能

1. 配置 NetStream 聚合统计功能

表1-4 配置 NetStream 聚合统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 NetStream 聚合视图	**ip netstream aggregation** { **as** | **destination-prefix** | **prefix** | **prefix-port** | **protocol-port** | **source-prefix** | **tos-as** | **tos-destination-prefix** | **tos-prefix** | **tos-protocol-port** | **tos-source-prefix** | **tos-bgp-nexthop** }	必选
使能该类的 NetStream 聚合功能	**enable**	必选 缺省情况下，未使能任何 NetStream 聚合功能

2. 配置 NetStream 聚合统计输出报文的属性

表1-5 配置 NetStream 聚合统计输出报文的属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 NetStream 聚合视图	**ip netstream aggregation** { **as** | **destination-prefix** | **prefix** | **prefix-port** | **protocol-port** | **source-prefix** | **tos-as** | **tos-destination-prefix** | **tos-prefix** | **tos-protocol-port** | **tos-source-prefix** | **tos-bgp-nexthop** }	必选
配置 NetStream 聚合统计输出报文的目的地址	**ip netstream export host** *ip-addrress udp-port* [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，系统视图下没有配置目的 IP 地址和目的端口号，聚合视图下目的 IP 地址和目的端口号为系统视图下所配置的值
配置 NetStream 统计输出报文的源接口	**ip netstream export source interface** *interface-type interface-number*	必选 缺省情况下，采用统计输出报文的出接口作为源接口

1.5 配置 NetStream 统计输出报文的属性

1.5.1 配置 NetStream 统计输出报文的属性

用户可以对 NetStream 统计输出报文的源接口和目的 IP 地址及目的端口号进行设置，其中源接口 IP 地址将作为此报文的源 IP 地址。

NetStream 统计输出报文支持 BGP 的自治系统选项，用户可以选择以起始自治系统号或者以邻接自治系统号作为给定 IP 地址所在的自治系统号。

NetStream 统计输出报文还支持 BGP 下一跳选项。当用户配置版本 9 的 NetStream 统计输出报文属性时，用户可以选择是否在流信息中记录 BGP 下一跳地址。

表1-6 配置 NetStream 统计输出报文的属性

操作	命令	说明
进入系统视图	**system-view**	-
配置 NetStream 统计输出报文的目的 IP 地址、端口号、私网路由	**ip netstream export host** *ip-address udp-port* [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，系统视图下没有配置目的 IP 地址和目的端口号，聚合视图下目的 IP 地址和目的端口号为系统视图下所配置的值
配置 NetStream 统计输出报文的源接口	**ip netstream export source interface** *interface-type interface-number*	可选 缺省情况下，采用统计报文输出的接口作为源接口

操作	命令	说明
配置 NetStream 统计输出报文版本以及其自治系统选项、BGP 下一跳选项	**ip netstream export version 5 [origin-as | peer-as]**	可选 缺省情况下，IPv4 普通流信息通过版本 5 的 NetStream 统计输出报文发送，IPv4 聚合统计流信息通过版本 8 的 NetStream 统计输出报文发送，带 IPv4 选项信息的 MPLS 流信息不输出。自治系统选项使用邻接自治系统号（**peer-as**），流信息中不记录 BGP 下一跳地址
	ip netstream export version 9 [origin-as | peer-as] [bgp-nexthop]	
配置输出速率限制	**ip netstream export rate** *rate*	可选 缺省情况下，NetStream 统计输出报文的输出速率不受限制

注意

- 在聚合视图下，用户配置的 NetStream 统计输出报文的属性，仅对聚合报文生效；而系统视图下的配置对普通报文生效，并且当聚合视图下没有配置以上属性时，也会对聚合报文生效。
- 自治系统相关知识，请参考“IP 路由分册”中的“BGP 配置”。

1.5.2 配置 NetStream 统计输出报文版本 9 模板的刷新率

MPLS 流信息通过版本 9 的 NetStream 统计输出报文发送。

表1-7 配置 NetStream 统计输出报文版本 9 模板的刷新率

操作	命令	说明
进入系统视图	**system-view**	-
配置 NetStream 统计输出报文版本 9 模板的包刷新率	**ip netstream export v9-template refresh-rate packet** *packets*	可选 缺省情况下，每隔 20 个包发送一次版本 9 模板
配置 NetStream 统计输出报文版本 9 模板的时间刷新率	**ip netstream export v9-template refresh-rate time** *minutes*	可选 缺省情况下，每隔 30 分钟发送一次版本 9 模板

注意

- 由于容量问题，XLog 接收到的版本 9 数据模板会被定期老化，因此用户需要配置版本 9 模板的刷新率，及时更新模板。
- 可以同时配置包刷新率和时间刷新率，只要满足任意一个刷新条件，设备就会将符合条件的模板发送给 XLog。

1.6 配置 NetStream MPLS 报文统计功能

1.6.1 MPLS 统计简介

NetStream 采集流（NetStream 七元组以及 MPLS 标签相同的报文被视为同一条流）信息中 MPLS 报文数据，按照标签进行统计并记录标签栈顶标签的类型、标签栈顶标签对应的 IP 地址及掩码长。在缺省情况下还会统计 IP 数据内容。

1.6.2 配置 NetStream MPLS 报文统计功能

表1-8 配置 NetStream MPLS 报文统计功能

操作	命令	说明
进入系统视图	**system-view**	-
使能 MPLS 报文统计功能	**ip netstream mpls** [**label-positions** { *label-position1* [*label-position2*] [*label-position3*] }] [**no-ip-fields**]	必选 缺省情况下，禁用统计和输出 MPLS 格式报文



说明

ip netstream mpls [**label-positions** { *label-position1* [*label-position2*] [*label-position3*] }] [**no-ip-fields**]不仅使能 IPv4 NetStream 对 MPLS 报文的统计功能，同时也使能了 IPv6 NetStream 对 MPLS 报文的统计功能。

1.7 配置 NetStream 的流老化

1.7.1 流老化简介

在实际网络环境中，可能在很短的时间内产生大量流，这就需要根据一定的算法把当前一部分流从缓冲区中删除，称为老化（aging）。

NetStream 的老化机制有三种：

- 按时老化
- 强制老化
- TCP 的 FIN 和 RST 报文触发老化（该形式的老化在 TCP 连接拆除时自动进行）。

1. 按时老化

有两种按时老化方式：

- 不活跃的流在超时后老化：对从最后一个报文流过到当前的时间超过 **inactive timeout** 的流进行老化；
- 活跃的流在超时后老化：对从第一个报文流过到当前的时间超过 **active timeout** 的流进行老化。

2. 强制老化

执行强制老化命令，用户可以将 NetStream 缓冲区中所有流老化，并清除 NetStream 缓冲区的状态信息、NetStream 输出报文信息、NetStream V9 模板的状态信息，输出报文信息及 V9 模板的状态信息从强制老化时刻开始重新统计。

3. TCP 的 FIN 和 RST 报文触发老化

对于 TCP 连接，当有标志为 FIN 或 RST 的报文发送时，表示一次会话结束。因此当一条已经存在的 TCP 协议 NetStream 流中流过一条标志为 FIN 或 RST 的报文时，可以立即老化相应的 NetStream 流。但是假如一条流的第一个报文就是 TCP 的 FIN 或 RST 报文，则会按正常的流程创建一条新流，不进行老化。

1.7.2 配置 NetStream 的流老化

表1-9 配置 NetStream 的流老化

操作	命令	说明
进入系统视图	**system-view**	-
配置 NetStream 活跃老化时间	**ip netstream timeout active** *minutes*	可选 缺省情况下，普通流信息的活跃老化时间为 30 分钟
配置 NetStream 不活跃老化时间	**ip netstream timeout inactive** *seconds*	可选 缺省情况下，普通流信息的不活跃老化时间为 30 秒

1.8 配置 NetStream 的采样过滤功能

1.8.1 采样过滤简介

1. 采样

为了减少 NetStream 给设备带来的性能影响，可以通过与 Sampler 采样器配合来减少进入 NetStream 统计的报文数量。与 Sampler 有关的内容请参见“系统分册”中的“Sampler 配置”。

2. 过滤

通过配置 NetStream 过滤，可以使得 NetStream 只统计符合 ACL 流分类的特定报文，减少进入 NetStream 统计的报文数量，更能满足用户的统计要求。在 NetStream 中通过引用 ACL 的方式来配置过滤。

1.8.2 配置 NetStream 的采样过滤

1. 使能 NetStream 采样功能

表1-10 使能 NetStream 的采样过滤功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 NetStream 采样过滤功能	**ip netstream sampler** *sampler-name* { **inbound** | **outbound** }	可选 缺省情况下，NetStream 统计所有 IPv4 报文，不启用采样功能 本命令也可以在系统视图下配置开启所有接口采样功能

2. 配置根据指定条件对 NetStream 过滤功能

表1-11 配置根据指定条件对 NetStream 过滤功能

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置根据指定条件对NetStream过滤	**ip netstream filter acl** *acl-number* { **inbound** | **outbound** }	可选 缺省情况下，NetStream统计所有IPv4报文，不启用过滤功能

注意

- 在配置采样功能时，采样器引用后不允许删除，不同采样模式或采样率的表项不能进行聚合和不能在一个V5和V8报文中输出。
- 在配置过滤功能时，ACL引用后不允许删除和修改。如果过滤和采样同时配置，报文先进行过滤，而后采样。
- 配置采样时，要求采样器必须存在（即需要**sampler**命令先创建采样器）。与Sampler有关的内容请参见“系统分册”中的“Sampler配置”。
- 配置过滤时，要求ACL必须存在。与ACL有关的内容请参见“安全分册”中的“ACL配置”。

1.9 NetStream显示和维护

在完成上述配置后，在任意视图下执行**display**命令可以显示配置后NetStream的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行**reset**命令可以清除NetStream的统计信息。

表1-12 NetStream显示和维护

操作	命令
查看NetStream流缓存区的配置和状态信息（集中式设备）	**display ip netstream cache** [**verbose**]
查看NetStream流缓存区的配置和状态信息（分布式设备）	**display ip netstream cache** [**slot** *slot-id*] [**verbose**]
查看NetStream统计输出报文信息	**display ip netstream export**
查看NetStream V9模板的配置和状态信息（集中式设备）	**display ip netstream template**
查看NetStream V9模板的配置和状态信息（分布式设备）	**display ip netstream template** [**slot** *slot-id*]
清除NetStream缓冲区的状态信息、NetStream统计输出报文信息、NetStream V9模板状态信息	**reset ip netstream statistics**

1.10 NetStream典型配置举例

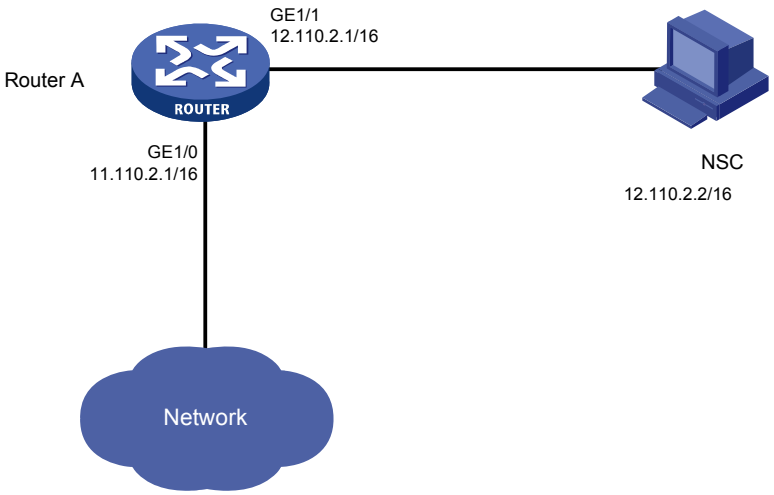
1.10.1 配置NetStream分类统计功能

1. 组网需求

如图1-2所示，在Router A的GE1/0上配置NetStream入统计，GE1/1是配置NetStream出统计，并将统计结果发送到NSC 12.110.2.2/16的5000端口。

2. 组网图

图1-2 配置 NetStream 分类统计功能组网图



3. 配置步骤

配置 GigabitEthernet1/0，在此接口上启动 NetStream 入统计。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ip address 11.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/0] ip netstream inbound
[RouterA-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1，在此接口上启动 NetStream 出统计。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 12.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/1] ip netstream outbound
[RouterA-GigabitEthernet1/1] quit
```

配置 NetStream 统计输出报文的目的 IP 地址和目的端口（即 NSC 的地址以及端口），源接口采用缺省配置。

```
[RouterA] ip netstream export host 12.110.2.2 5000
```

1.10.2 配置 NetStream 统计输出功能

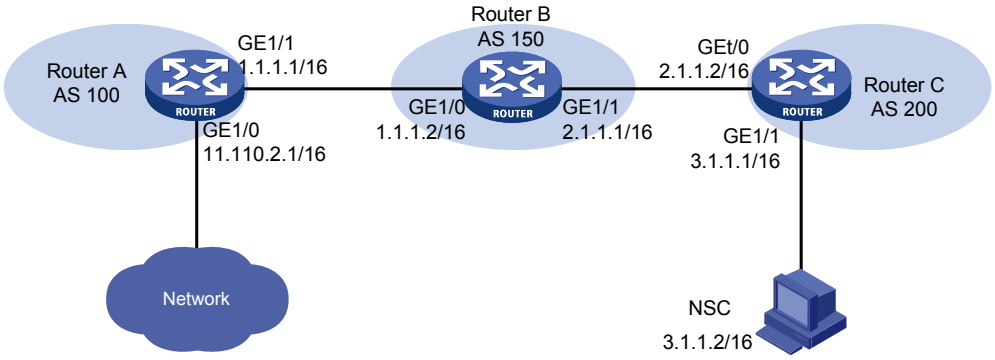
1. 组网需求

在 Router A 上配置 NetStream，普通流统计输出报文使用版本 5 格式输出到 NSC 3.1.1.2/16 的 5000 端口，5 种 NetStream 聚合（**as**、**protocol-port**、**source-prefix**、**destination-prefix** 和 **prefix**）的聚合流统计输出报文使用版本 8 格式分别输出到该 NSC 的 2000、3000、4000、6000 和 7000 端口。

所有路由器均运行 BGP 路由协议。Router A、Router B 和 Router C 之间建立了 EBGP 邻居关系。

2. 组网图

图1-3 配置版本 5 和版本 8 报文输出组网图



3. 配置步骤

(1) 配置 Router A

配置 GigabitEthernet1/0，在出、入方向启动 NetStream 统计。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ip address 11.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/0] ip netstream inbound
[RouterA-GigabitEthernet1/0] ip netstream outbound
[RouterA-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] ip address 1.1.1.1 255.255.0.0
[RouterA-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterA] bgp 100
[RouterA-bgp] peer 1.1.1.2 as-number 150
[RouterA-bgp] network 11.110.0.0 255.255.0.0
[RouterA-bgp] quit
```

配置 NetStream 版本号和自治系统选项。

```
[RouterA] ip netstream export version 5 origin-as
```

配置 NetStream 最大流数。

```
[RouterA] ip netstream max-entry 5000
```

配置版本 5 输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream export host 3.1.1.2 5000
[RouterA] ip netstream export source interface GigabitEthernet 1/0
```

配置自治系统聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream aggregation as
[RouterA-aggregation-as] enable
[RouterA-aggregation-as] ip netstream export host 3.1.1.2 2000
[RouterA-aggregation-as] ip netstream export source interface GigabitEthernet 1/1
```

```
[RouterA-aggregation-as] quit
```

配置协议－端口聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream aggregation protocol-port
[RouterA-aggregation-protport] enable
[RouterA-aggregation-protport] ip netstream export host 3.1.1.2 3000
[RouterA-aggregation-protport] ip netstream export source interface GigabitEthernet 1/1
[RouterA-aggregation-protport] quit
```

配置源前缀聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream aggregation source-prefix
[RouterA-aggregation-srcpre] enable
[RouterA-aggregation-srcpre] ip netstream export host 3.1.1.2 4000
[RouterA-aggregation-srcpre] ip netstream export source interface GigabitEthernet 1/1
[RouterA-aggregation-srcpre] quit
```

配置目的前缀聚合模式，以及该模式下输出的报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream aggregation destination-prefix
[RouterA-aggregation-dstpre] enable
[RouterA-aggregation-dstpre] ip netstream export host 3.1.1.2 6000
[RouterA-aggregation-dstpre] ip netstream export source interface GigabitEthernet 1/1
[Sysname-aggregation-dstpre] quit
```

配置前缀聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ip netstream aggregation prefix
[RouterA-aggregation-prefix] enable
[RouterA-aggregation-prefix] ip netstream export host 3.1.1.2 7000
[RouterA-aggregation-prefix] ip netstream export source interface GigabitEthernet 1/1
[RouterA-aggregation-prefix] quit
```

(2) 配置 Router B

配置 GigabitEthernet1/0。

```
<RouterB> system-view
[RouterB] interface GigabitEthernet 1/0
[RouterB-GigabitEthernet1/0] ip address 1.1.1.2 255.255.0.0
[RouterB-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] ip address 2.1.1.1 255.255.0.0
[RouterB-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterB] bgp 150
[RouterB-bgp] peer 1.1.1.1 as-number 100
[RouterB-bgp] peer 2.1.1.2 as-number 200
```

(3) 配置 Router C

配置 GigabitEthernet1/0。

```
[RouterC] interface GigabitEthernet 1/0
[RouterC-GigabitEthernet1/0] ip address 2.1.1.2 255.255.0.0
```

```
[RouterC-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ip address 3.1.1.1 255.255.0.0
[RouterC-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterC] bgp 200
[RouterC-bgp] peer 2.1.1.1 as-number 150
```

目 录

1 NTP配置 1-1
1.1 NTP简介 1-1
1.1.1 NTP的应用 1-1
1.1.2 NTP工作原理 1-1
1.1.3 NTP的报文格式 1-2
1.1.4 NTP的工作模式 1-3
1.1.5 NTP多实例 1-6
1.2 NTP配置任务简介 1-6
1.3 配置NTP工作模式 1-6
1.3.1 配置NTP客户端/服务器模式 1-7
1.3.2 配置NTP对等体模式 1-7
1.3.3 配置NTP广播模式 1-8
1.3.4 配置NTP组播模式 1-9
1.4 配置本地时钟作为参考时钟 1-9
1.5 配置NTP可选参数 1-10
1.5.1 配置NTP报文的源接口 1-10
1.5.2 配置禁止接口接收NTP报文 1-10
1.5.3 配置允许建立的动态会话数目 1-11
1.6 配置访问控制权限 1-11
1.6.1 配置准备 1-11
1.6.2 配置访问控制权限 1-12
1.7 配置NTP验证功能 1-12
1.7.1 配置准备 1-12
1.7.2 配置过程 1-13
1.8 NTP显示和维护 1-14
1.9 NTP典型配置举例 1-14
1.9.1 配置NTP客户端/服务器模式 1-14
1.9.2 配置NTP对等体模式 1-15
1.9.3 配置NTP广播模式 1-17
1.9.4 配置NTP组播模式 1-18
1.9.5 配置带身份验证的NTP客户端/服务器模式 1-21
1.9.6 配置带身份验证的NTP广播模式 1-22
1.9.7 配置采用客户端/服务器模式实现MPLS VPN网络的时间同步 1-24
1.9.8 配置采用对等体模式实现MPLS VPN网络的时间同步 1-26

1 NTP 配置

1.1 NTP 简介

NTP（Network Time Protocol，网络时间协议）是由 RFC 1305 定义的时间同步协议，用来在分布式时间服务器和客户端之间进行时间同步。NTP 基于 UDP 报文进行传输，使用的 UDP 端口号为 123。

使用 NTP 的目的是对网络内所有具有时钟的设备进行时钟同步，使网络内所有设备的时钟保持一致，从而使设备能够提供基于统一时间的多种应用。

对于运行 NTP 的本地系统，既可以接受来自其他时钟源的同步，又可以作为时钟源同步其他的时钟，并且可以和其他设备互相同步。

1.1.1 NTP 的应用

对于网络中的各台设备来说，如果依靠管理员手工输入命令来修改系统时钟是不可能的，不但工作量巨大，而且也不能保证时钟的精确性。通过 NTP，可以很快将网络中设备的时钟同步，同时也能保证很高的精度。

NTP 主要应用于需要网络中所有设备时钟保持一致的场合，比如：

- 在网络管理中，对于从不同设备采集来的日志信息、调试信息进行分析的时候，需要以时间作为参照依据。
- 计费系统要求所有设备的时钟保持一致。
- 完成某些功能，如定时重启网络中的所有设备，此时要求所有设备的时钟保持一致。
- 多个系统协同处理同一个比较复杂的事件时，为保证正确的执行顺序，多个系统必须参考同一时钟。
- 在备份服务器和客户端之间进行增量备份时，要求备份服务器和所有客户端之间的时钟同步。

NTP 的优势如下：

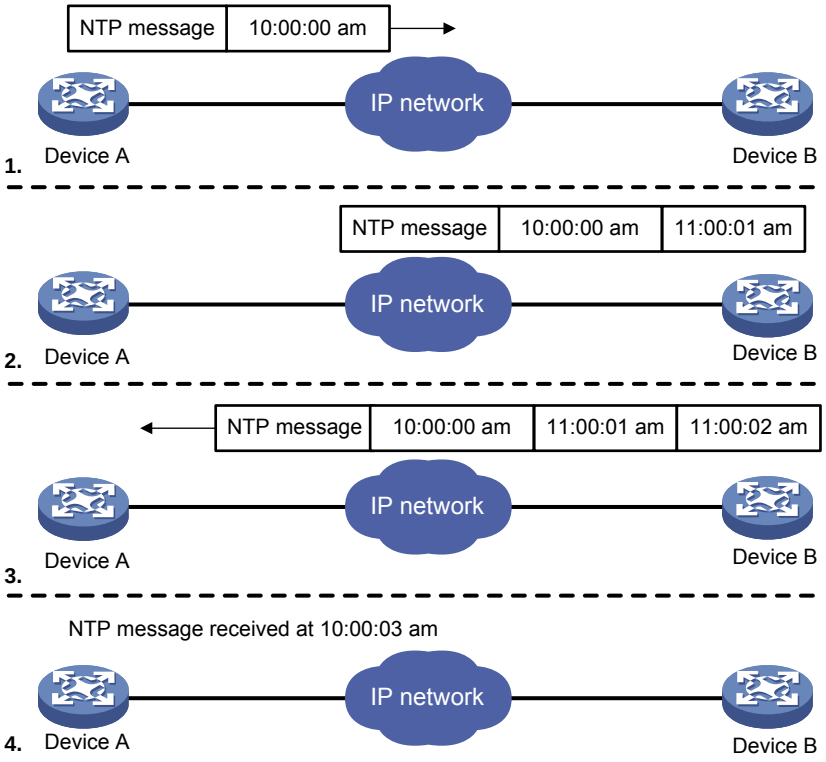
- 采用分层的方法定义时钟的准确性，可以迅速同步网络中各台设备的时间。
- 支持访问控制和 MD5 验证。
- 可以选择采用单播、组播或广播的方式发送协议报文。

1.1.2 NTP 工作原理

NTP的基本工作原理如 图 1-1所示。Device A和Device B通过网络相连，它们都有自己独立的系统时钟，需要通过NTP实现各自系统时钟的自动同步。为便于理解，作如下假设：

- 在 Device A 和 Device B 的系统时钟同步之前，Device A 的时钟设定为 10:00:00am，Device B 的时钟设定为 11:00:00am。
- Device B 作为 NTP 时间服务器，即 Device A 将使自己的时钟与 Device B 的时钟同步。
- NTP 报文在 Device A 和 Device B 之间单向传输所需要的时间为 1 秒。

图1-1 NTP 基本原理图



系统时钟同步的工作过程如下：

- Device A 发送一个 NTP 报文给 Device B，该报文带有它离开 Device A 时的时间戳，该时间戳为 10:00:00am（T_1）。
- 当此 NTP 报文到达 Device B 时，Device B 加上自己的时间戳，该时间戳为 11:00:01am（T_2）。
- 当此 NTP 报文离开 Device B 时，Device B 再加上自己的时间戳，该时间戳为 11:00:02am（T_3）。
- 当 Device A 接收到该响应报文时，Device A 的本地时间为 10:00:03am（T_4）。

至此，Device A 已经拥有足够的信息来计算两个重要的参数：

- NTP 报文的往返时延 Delay=（T_4-T_1）-（T_3-T_2）=2 秒。
- Device A 相对 Device B 的时间差 offset=（（T_2-T_1）+（T_3-T_4））/2=1 小时。

这样，Device A 就能够根据这些信息来设定自己的时钟，使之与 Device B 的时钟同步。

以上内容只是对 NTP 工作原理的一个粗略描述，详细内容请参阅 RFC 1305。

1.1.3 NTP 的报文格式

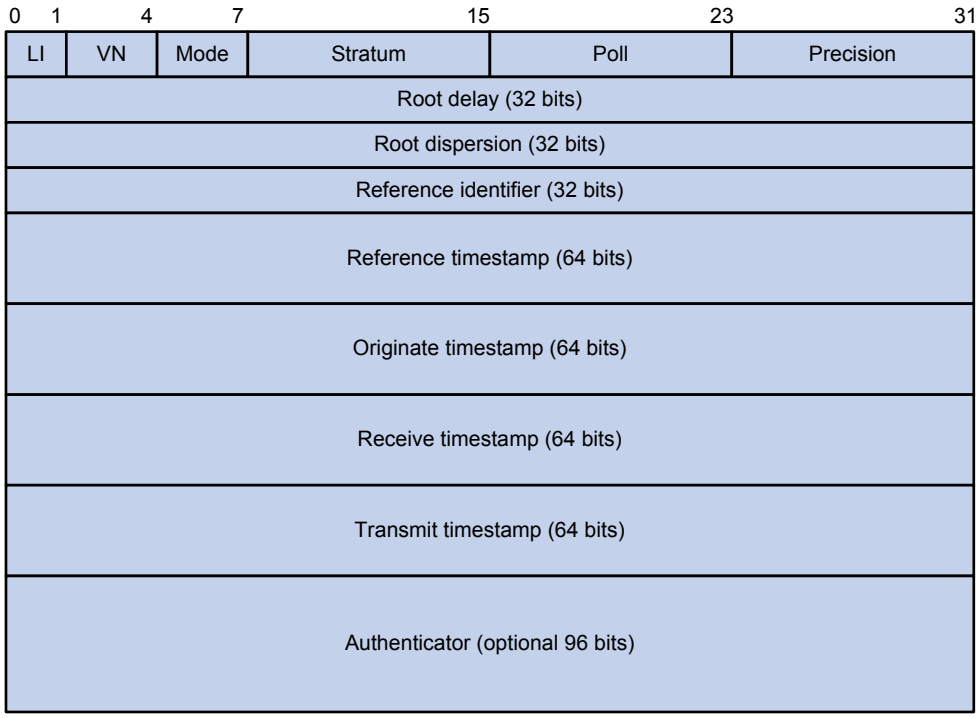
NTP 有两种不同类型的报文，一种是时钟同步报文，另一种是控制报文。控制报文仅用于需要网络管理的场合，它对于时钟同步功能来说并不是必需的，这里不做介绍。

说明

本文中提到的 NTP 报文，均为 NTP 时钟同步报文。

时钟同步报文封装在UDP报文中，其格式如 图 1-2所示。

图1-2 时钟同步报文格式



主要字段的解释如下：

- LI（Leap Indicator）：长度为 2 比特，值为“11”时表示告警状态，时钟未被同步。为其他值时 NTP 本身不做处理。
- VN（Version Number）：长度为 3 比特，表示 NTP 的版本号，目前的最新版本为 3。
- Mode：长度为 3 比特，表示 NTP 的工作模式。不同的值所表示的含义分别是：0 未定义、1 表示主动对等体模式、2 表示被动对等体模式、3 表示客户模式、4 表示服务器模式、5 表示广播模式或组播模式、6 表示此报文为 NTP 控制报文、7 预留给内部使用。
- Stratum：系统时钟的层数，取值范围为 1～16，它定义了时钟的准确度。层数为 1 的时钟准确度最高，准确度从 1 到 16 依次递减，层数为 16 的时钟处于未同步状态，不能作为参考时钟。
- Poll：轮询时间，即两个连续 NTP 报文之间的时间间隔。
- Precision：系统时钟的精度。
- Root Delay：本地到主参考时钟源的往返时间。
- Root Dispersion：系统时钟相对于主参考时钟的最大误差。
- Reference Identifier：参考时钟源的标识。
- Reference Timestamp：系统时钟最后一次被设定或更新的时间。
- Originate Timestamp：NTP 请求报文离开发送端时发送端的本地时间。
- Receive Timestamp：NTP 请求报文到达接收端时接收端的本地时间。
- Transmit Timestamp：应答报文离开应答者时应答者的本地时间。
- Authenticator：验证信息。

1.1.4 NTP 的工作模式

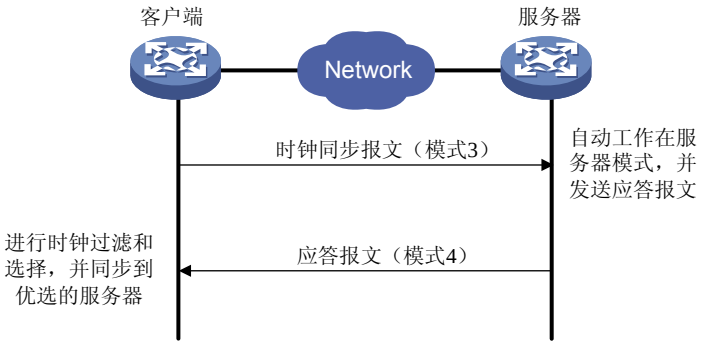
设备可以采用多种 NTP 工作模式进行时间同步：

- 客户端/服务器模式
- 对等体模式
- 广播模式
- 组播模式

用户可以根据需要选择合适的工作模式。在不能确定服务器或对等体 IP 地址、网络中需要同步的设备很多等情况下，可以通过广播或组播模式实现时钟同步；服务器和对等体模式中，设备从指定的服务器或对等体获得时钟同步，增加了时钟的可靠性。

1. 客户端/服务器模式

图1-3 客户端/服务器模式

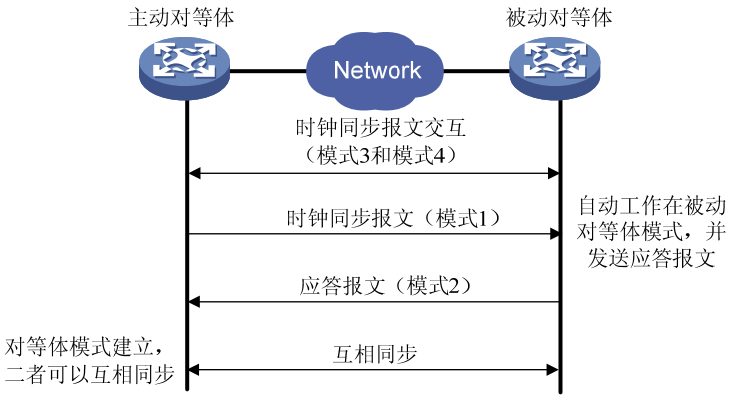


在客户端/服务器模式中，客户端向服务器发送时钟同步报文，报文中的 Mode 字段设置为 3（客户模式）。服务器端收到报文后会自动工作在服务器模式，并发送应答报文，报文中的 Mode 字段设置为 4（服务器模式）。客户端收到应答报文后，进行时钟过滤和选择，并同步到优选的服务器。

在该模式下，客户端能同步到服务器，而服务器无法同步到客户端。

2. 对等体模式

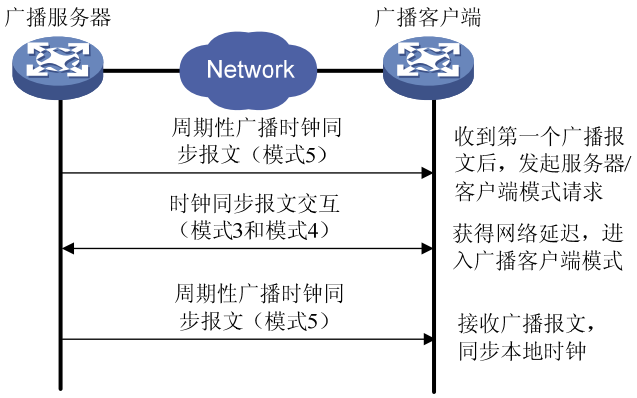
图1-4 对等体模式



在对等体模式中，主动对等体和被动对等体之间首先交互 Mode 字段为 3（客户端模式）和 4（服务器模式）的 NTP 报文。之后，主动对等体向被动对等体发送时钟同步报文，报文中的 Mode 字段设置为 1（主动对等体），被动对等体收到报文后自动工作在被动对等体模式，并发送应答报文，报文中的 Mode 字段设置为 2（被动对等体）。经过报文的交互，对等体模式建立起来。主动对等体和被动对等体可以互相同步。如果双方的时钟都已经同步，则以层数小的时钟为准。

3. 广播模式

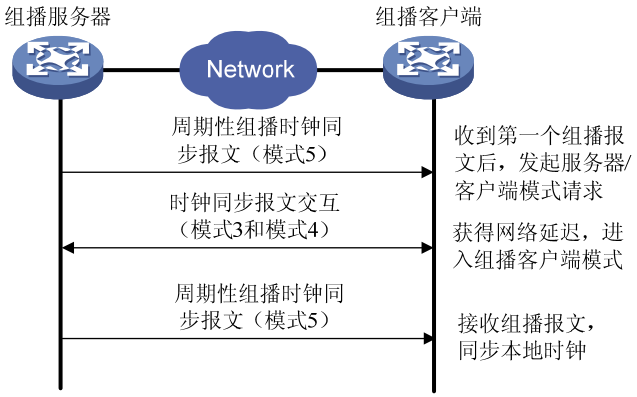
图1-5 广播模式



在广播模式中，服务器端周期性地向广播地址 255.255.255.255 发送时钟同步报文，报文中的 Mode 字段设置为 5（广播模式）。客户端侦听来自服务器的广播报文。当客户端接收到第一个广播报文后，客户端与服务器交互 Mode 字段为 3（客户模式）和 4（服务器模式）的 NTP 报文，以获得客户端与服务器间的网络延迟。之后，客户端就进入广播客户端模式，继续侦听广播报文的到来，根据到来的广播报文对系统时钟进行同步。

4. 组播模式

图1-6 组播模式



在组播模式中，服务器端周期性地向用户配置的组播地址（若用户没有配置组播地址，则使用默认的 NTP 组播地址 224.0.1.1）发送时钟同步报文，报文中的 Mode 字段设置为 5（组播模式）。客户端侦听来自服务器的组播报文。当客户端接收到第一个组播报文后，客户端与服务器交互 Mode 字段为 3（客户模式）和 4（服务器模式）的 NTP 报文，以获得客户端与服务器间的网络延迟。之后，客户端就进入组播客户模式，继续侦听组播报文的到来，根据到来的组播报文对系统时钟进行同步。

说明

在对等体模式、广播模式和组播模式中，客户端（或主动对等体）和服务器（或被动对等体）之间首先要交互 Mode 字段为 3（客户端模式）和 4（服务器模式）的 NTP 报文，之后，才能进入指定的 NTP 工作模式。在此报文交互过程中，可以实现时钟的同步。

1.1.5 NTP 多实例

客户端/服务器模式和对等体模式支持 NTP 多实例，可以实现 MPLS VPN 网络的时间同步，即物理位置不同的网络设备（CE、PE），只要属于同一个 VPN，就可以通过 NTP 来获得时间同步。具体功能如下：

- CE 上的 NTP 客户端可以同步到另一个 CE 上的 NTP 服务器；
- CE 上的 NTP 客户端可以同步到 PE 上的 NTP 服务器；
- PE 上的 NTP 客户端可以通过指定的 VPN 实例同步到 CE 上的 NTP 服务器；
- PE 上的 NTP 客户端可以通过指定的 VPN 实例同步到另一个 PE 上的 NTP 服务器；
- PE 上的 NTP 服务器可以同步不同 VPN 内多个 CE 上的 NTP 客户端。

说明

- CE（Customer Edge）设备：用户网络边缘设备，有接口直接与服务提供商（Service Provider，SP）相连。CE“感知”不到 VPN 的存在。
- PE（Provider Edge）设备：服务提供商网络的边缘设备，与用户的 CE 直接相连。

1.2 NTP 配置任务简介

表1-1 NTP 配置任务简介

配置任务	说明	详细配置
配置 NTP 工作模式	必选	1.3
配置本地时钟作为参考时钟	可选	1.4
配置 NTP 可选参数	可选	1.5
配置访问控制权限	可选	1.6
配置 NTP 验证功能	可选	1.7

1.3 配置 NTP 工作模式

设备可以采用以下 NTP 工作模式进行时钟同步：

- 客户端/服务器模式
- 对等体模式
- 广播模式
- 组播模式

设备采用客户端/服务器模式或对等体模式时，只需要对客户端或主动对等体进行配置；设备采用广播模式或组播模式时，则需要在服务器端和客户端都进行配置。

说明

同一设备同一时间内存在的连接数目最多为 128 个，其中包括静态连接数和动态连接数。静态连接是用户手动配置 NTP 相关命令而建立的连接；动态连接是系统运行过程中建立的临时连接，若系统长期收不到报文就会删除该临时连接。例如，在客户端/服务器模式中，当用户在客户端配置向服务器端同步的命令的时候，系统会在客户端建立一个静态连接，服务器端在收到报文之后只是被动的响应报文，而不会建立连接（包括静态和动态连接）；在对等体模式中，主动对等体端会建立静态连接，被动对等体端会建立动态连接；在组播和广播模式中，服务器端会建立静态连接，而在客户端会建立动态连接。

1.3.1 配置 NTP 客户端/服务器模式

当设备采用客户端/服务器模式时，只需在客户端进行配置，服务器端不需进行配置。

表1-2 配置 NTP 客户端

操作	命令	说明
进入系统视图	**system-view**	-
指定设备的 NTP 服务器	**ntp-service unicast-server** [**vpn-instance** *vpn-instance-name*] { *ip-address* | *server-name* } [**authentication-keyid** *keyid* | **priority** | **source-interface** *interface-type interface-number* | **version** *number*] *	必选 缺省情况下，没有为设备指定 NTP 服务器

说明

- **ntp-service unicast-server** 命令中的 *ip-address* 是一个单播地址，不能为广播地址、组播地址或本地时钟的 IP 地址。
- 通过 **source-interface** 参数指定 NTP 报文的源接口后，NTP 报文的源 IP 地址将被设置为指定接口的主 IP 地址。
- 服务器端只有当其时钟被同步后，才能作为时间服务器去同步其他设备。当服务器端的时钟层数大于或等于客户端的时钟层数时，客户端将不会向其同步。
- 可以通过多次执行 **ntp-service unicast-server** 命令配置多个服务器，客户端依据时钟优选来选择最优的时钟源。

1.3.2 配置 NTP 对等体模式

当设备采用对等体模式时，需要在主动对等体上指定被动对等体。

表1-3 配置主动对等体

操作	命令	说明
进入系统视图	**system-view**	-
指定设备的被动对等体	**ntp-service unicast-peer** [**vpn-instance** *vpn-instance-name*] { *ip-address* | *peer-name* } [**authentication-keyid** *keyid* | **priority** | **source-interface** *interface-type interface-number* | **version** *number*] *	必选 缺省情况下，没有为设备指定被动对等体

说明

- 在对等体模式中，被动对等体上需要执行**ntp-service refclock-master**或“1.3 配置NTP工作模式”中的任何一条NTP配置命令来使能NTP，否则被动对等体不会处理来自主动对等体的NTP报文。
- **ntp-service unicast-peer** 命令中的 *ip-address* 是一个单播地址，不能为广播地址、组播地址或本地时钟的 IP 地址。
- 通过 **source-interface** 参数指定 NTP 报文的源接口后，NTP 报文的源 IP 地址将被设置为指定接口的主 IP 地址。
- 通常，主、被动对等体中至少有一个处于同步状态，否则他们将都无法同步。
- 可以通过多次执行 **ntp-service unicast-peer** 命令配置多个被动对等体。

1.3.3 配置 NTP 广播模式

广播服务器周期性地向广播地址 255.255.255.255 发送 NTP 报文，工作在 NTP 广播客户端模式的设备将回应这个报文，从而开始时钟同步过程。

当设备采用广播模式时，需要在服务器端和客户端都进行配置。由于广播服务器上需要指定一个发送 NTP 广播报文的接口，广播客户端上也需要指定一个接收 NTP 广播报文的接口，所以广播模式的配置只能在具体的接口视图下进行。

1. 配置广播客户端

表1-4 配置广播客户端

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	- 进入要接收 NTP 广播报文的接口
配置设备工作在 NTP 广播客户端模式	**ntp-service broadcast-client**	必选

2. 配置广播服务器

表1-5 配置广播服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	- 进入要发送 NTP 广播报文的接口
配置设备工作在 NTP 广播服务器模式	**ntp-service broadcast-server** [**authentication-keyid** *keyid* | **version** *number*] *	必选

说明

广播服务器只有当其时钟同步后，才能去同步广播客户端。

1.3.4 配置 NTP 组播模式

NTP 组播服务器以组播形式周期性地发送时钟同步报文，工作在 NTP 组播客户端模式的设备将回应这个报文，从而开始时钟同步过程。

设备采用组播模式时，需要在服务器端和客户端都进行配置。组播模式的配置只能在具体的接口视图下进行。

1. 配置组播客户端

表1-6 配置组播客户端

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	- 进入要接收 NTP 组播报文的接口
配置设备工作在 NTP 组播客户端模式	**ntp-service multicast-client** [*ip-address*]	必选

2. 配置组播服务器

表1-7 配置组播服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	- 进入要发送 NTP 组播报文的接口
配置设备工作在 NTP 组播服务器模式	**ntp-service multicast-server** [*ip-address*] [**authentication-keyid** *keyid* | **ttl** *ttl-number* | **version** *number*] *	必选

说明

- 组播服务器只有当其时钟同步后，才能去同步组播客户端。
- 目前最多可以配置 1024 个组播客户端，但同时起作用的最多为 128 个。

1.4 配置本地时钟作为参考时钟

网络中的设备可以通过下面两种方式进行时间同步：

- 与本地时钟进行同步：即采用本地时钟作为参考时钟。
- 与网络中的其他设备进行同步：可以采用前面介绍的四种 NTP 工作模式中的任何一种。

如果同时配置了两种方式，设备将通过时钟优选来选择最优的时钟源。

表1-8 配置本地时钟作为参考时钟

操作	命令	说明
进入系统视图	**system-view**	-
配置本地时钟作为参考时钟	**ntp-service refclock-master** [*ip-address*] [*stratum*]	必选

说明

- 实际网络中，通常将从权威时钟（如原子时钟）获得时钟同步的 NTP 服务器的层数设置为 1，并将其作为主参考时钟源同步网络中其他设备的时钟。网络中的设备与主参考时钟源的 NTP 距离，即 NTP 同步链上 NTP 服务器的数目，决定了设备上时钟的层数。
- 配置本地时钟作为参考时钟后，本地设备可以作为时钟源同步网络中的其他设备。请谨慎使用本配置，以免导致网络中设备的时钟错误。
- 命令中的 *ip-address* 只能为 127.127.1.u。u 的取值范围为 0～3，表示 NTP 的进程号。

1.5 配置 NTP 可选参数

1.5.1 配置 NTP 报文的源接口

如果指定了 NTP 报文的源接口，则设备在主动发送 NTP 报文时，将报文的源 IP 地址设置为指定接口的主 IP 地址。

设备对接收到的 NTP 请求报文进行应答时，应答报文的源 IP 地址始终为接收到 NTP 请求报文的接口的 IP 地址。

表1-9 配置 NTP 报文的源接口

操作	命令	说明
进入系统视图	**system-view**	-
配置 NTP 报文的源接口	**ntp-service source-interface** *interface-type interface-number*	必选 缺省情况下，没有指定 NTP 报文的源接口，即根据路由选择 NTP 报文的源 IP 地址

注意

- 如果在命令 **ntp-service unicast-server** 或 **ntp-service unicast-peer** 中指定了 NTP 报文的源接口，则以 **ntp-service unicast-server** 或 **ntp-service unicast-peer** 指定的为准。
- 如果在接口视图下配置了 **ntp-service broadcast-server** 或 **ntp-service multicast-server**，则 NTP 广播或组播模式报文的源接口为配置了上述命令的接口。

1.5.2 配置禁止接口接收 NTP 报文

使能 NTP 功能后，缺省情况下所有接口都可以接收 NTP 报文。可以通过本配置，禁止从某个接口接收 NTP 报文。

表1-10 配置禁止接口接收 NTP 报文

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置禁止接口接收 NTP 报文	**ntp-service in-interface disable**	必选 缺省情况下，允许接口接收 NTP 报文

1.5.3 配置允许建立的动态会话数目

表1-11 配置允许建立的动态会话数目

操作	命令	说明
进入系统视图	**system-view**	-
配置允许建立的动态 NTP 会话数目	**ntp-service max-dynamic-sessions** *number*	必选 缺省情况下，允许建立的动态 NTP 会话的数目为 100

1.6 配置访问控制权限

用户可以配置对本地设备 NTP 服务的访问控制权限。访问控制权限可以分为四种：

- **query**：允许控制查询权限。该权限只允许对端设备对本地设备的 NTP 服务进行控制查询，但是不能向本地设备同步。所谓的控制查询，就是查询 NTP 的一些状态，比如告警信息，验证状态，时钟源信息等。
- **synchronization**：只允许服务器访问权限。该权限只允许对端设备向本地设备同步，但不能进行控制查询。
- **server**：允许服务器访问与查询权限。该权限允许对端设备向本地设备同步和控制查询，但本地设备不会同步到对端设备。
- **peer**：完全访问权限。该权限既允许对端设备向本地设备同步和控制查询，同时本地设备也可以同步到对端设备。

NTP 服务的访问控制权限从高到低依次为 **peer**、**server**、**synchronization**、**query**。当设备接收到一个 NTP 服务请求时，会按照此顺序进行匹配，以第一个匹配的权限为准。

1.6.1 配置准备

在配置对本地设备 NTP 服务的访问控制权限之前，需要创建并配置与访问权限关联的 ACL。ACL 的配置方法请参见“安全分册”中的“ACL 配置”。

1.6.2 配置访问控制权限

表1-12 配置对本地设备 NTP 服务的访问控制权限

操作	命令	说明
进入系统视图	**system-view**	-
配置对端设备对本地设备 NTP 服务的访问控制权限	**ntp-service access** { **peer** | **query** | **server** | **synchronization** } *acl-number*	必选 缺省情况下，对端设备对本地设备 NTP 服务的访问控制权限为 **peer**



说明

配置对本地设备 NTP 服务的访问控制权限，仅提供了一种最小限度的安全措施，更安全的方法是进行身份验证。

1.7 配置 NTP 验证功能

在一些对安全性要求较高的网络中，运行 NTP 协议时需要启用验证功能。通过客户端和服务器端的密码验证，保证客户端只与通过验证的设备进行同步，提高了网络安全性。

1.7.1 配置准备

配置 NTP 验证功能可以分为配置客户端的 NTP 验证和配置服务器端的 NTP 验证两个部分。

在配置 NTP 验证功能时，应注意以下原则：

- 对于所有同步模式，如果使能了 NTP 验证功能，应同时配置验证密钥并将密钥设为可信密钥，即如果执行了 **ntp-service authentication enable** 命令，则必须同时执行 **ntp-service authentication-keyid** 命令和 **ntp-service reliable authentication-keyid** 命令。否则，无法正常启用 NTP 验证功能。
- 对于客户端/服务器模式和对等体模式，还应在客户端（对等体模式中的主动对等体）将指定密钥与对应的 NTP 服务器（对等体模式的被动对等体）关联；对于广播服务器模式和组播服务器模式，应在广播服务器或组播服务器上将指定密钥与对应的 NTP 服务器关联。否则，无法正常启用 NTP 验证功能。
- 对于客户端/服务器同步模式，如果客户端没有成功启用 NTP 验证功能，不论服务器端是否使能 NTP 验证，客户端均可以与服务器端同步；如果客户端上成功启用了 NTP 验证功能，则客户端只会同步到提供可信密钥的服务器，如果服务器提供的密钥不是可信的密钥，那么客户端不会与其同步。
- 对于所有同步模式，服务器端的配置与客户端的配置应保持一致。

1.7.2 配置过程

1. 配置客户端的 NTP 验证

表1-13 配置客户端的 NTP 验证

操作	命令	说明
进入系统视图	**system-view**	-
使能NTP身份验证功能	**ntp-service authentication enable**	必选 缺省情况下，NTP 身份验证功能处于关闭状态
配置NTP验证密钥	**ntp-service authentication-keyid** *keyid* **authentication-mode md5** *value*	必选 缺省情况下，没有配置 NTP 验证密钥
配置指定密钥为可信密钥	**ntp-service reliable authentication-keyid** *keyid*	必选 缺省情况下，没有指定可信密钥
将指定密钥与对应的NTP服务器关联	客户端/服务器模式： **ntp-service unicast-server** { *ip-address* | *server-name* } **authentication-keyid** *keyid*	必选 可以将不存在的密钥与 NTP 服务器关联。但是若想成功启用 NTP 验证功能，则必须在关联密钥后，配置该密钥，并将其指定为可信密钥
	对等体模式： **ntp-service unicast-peer** { *ip-address* | *peer-name* } **authentication-keyid** *keyid*	

说明

客户端使能 NTP 验证功能后，必须配置与服务器端相同的验证密钥，并且必须声明该密钥是可信的，否则无法与服务器同步。

2. 配置服务器端的 NTP 验证

表1-14 配置服务器端的 NTP 验证

操作	命令	说明
进入系统视图	**system-view**	-
使能NTP验证功能	**ntp-service authentication enable**	必选 缺省情况下，NTP 身份验证功能处于关闭状态
配置NTP验证密钥	**ntp-service authentication-keyid** *keyid* **authentication-mode md5** *value*	必选 缺省情况下，没有配置 NTP 验证密钥
配置指定密钥为可信密钥	**ntp-service reliable authentication-keyid** *keyid*	必选 缺省情况下，没有指定可信密钥
进入接口视图	**interface** *interface-type interface-number*	-
将指定密钥与对应的NTP服务器关联	广播服务器模式： **ntp-service broadcast-server authentication-keyid** *keyid*	必选 可以将不存在的密钥与 NTP 服务器关联。但是若想成功启用 NTP 验证功能，则必须在关联密钥后，配置该密钥，并将其指定为可信密钥
	组播服务器模式： **ntp-service multicast-server authentication-keyid** *keyid*	

说明

服务器端的 NTP 验证配置步骤与客户端的相同，并且两端必须配置相同的密钥。

1.8 NTP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 NTP 的运行情况，通过查看显示信息验证配置的效果。

表1-15 NTP 显示与维护

操作	命令
显示 NTP 服务的状态信息	**display ntp-service status**
显示 NTP 服务维护的会话信息	**display ntp-service sessions [verbose]**
显示从本地设备回溯到主参考时钟源的各个 NTP 时间服务器的简要信息	**display ntp-service trace**

1.9 NTP 典型配置举例

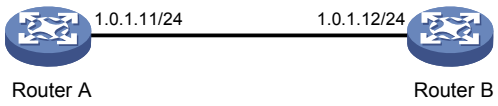
1.9.1 配置 NTP 客户端/服务器模式

1. 组网需求

- Router A 设置本地时钟作为参考时钟，层数为 2；
- Router B 工作在客户端模式，指定 Router A 为 NTP 服务器。

2. 组网图

图1-7 配置 NTP 客户端/服务器模式组网图



3. 配置步骤

(1) 配置 Router A

设置本地时钟作为参考时钟，层数为 2。

```
<RouterA> system-view
[RouterA] ntp-service refclock-master 2
```

(2) 配置 Router B

同步前查看 Router B 的 NTP 状态。

```
<RouterB> display ntp-service status
Clock status: unsynchronized
Clock stratum: 16
Reference clock ID: none
```

```
Nominal frequency: 64.0000 Hz
Actual frequency: 64.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 0.00 ms
Root dispersion: 0.00 ms
Peer dispersion: 0.00 ms
Reference time: 00:00:00.000 UTC Jan 1 1900 (00000000.00000000)
```

设置 Router A 为 Router B 的 NTP 服务器。

```
<RouterB> system-view
[RouterB] ntp-service unicast-server 1.0.1.11
```

以上配置将 Router B 向 Router A 进行时间同步，同步后查看 Router B 的 NTP 状态。

```
[RouterB] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 1.0.1.11
Nominal frequency: 64.0000 Hz
Actual frequency: 64.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 1.05 ms
Peer dispersion: 7.81 ms
Reference time: 14:53:27.371 UTC Sep 19 2005 (C6D94F67.5EF9DB22)
```

此时 Router B 已经与 Router A 同步，层数比 Router A 的层数大 1，为 3。

查看 Router B 的 NTP 会话信息，可以看到 Router B 与 Router A 建立了连接。

```
[RouterB] display ntp-service sessions
      source      reference   stra  reach  poll  now  offset  delay  disper
**************************************************************************
[12345] 1.0.1.11  127.127.1.0    2    63    64    3    -75.5    31.0  16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

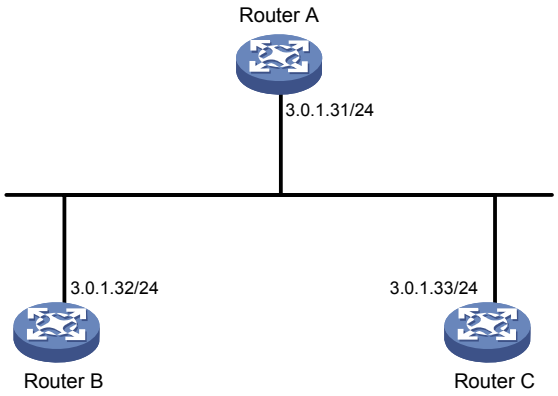
1.9.2 配置 NTP 对等体模式

1. 组网需求

- Router A 设置本地时钟作为参考时钟，层数为 2；
- Router B 工作在客户端模式，指定 Router A 为 NTP 服务器；
- Router C 工作在对等体模式，将 Router B 设为对等体。Router C 为主动对等体，Router B 为被动对等体。

2. 组网图

图1-8 配置 NTP 对等体模式组网图



3. 配置步骤

(1) 配置 Router A

设置本地时钟作为参考时钟，层数为 2。

```
<RouterA> system-view
[RouterA] ntp-service refclock-master 2
```

(2) 配置 Router B

设置 Router A 为 Router B 的 NTP 服务器。

```
<RouterB> system-view
[RouterB] ntp-service unicast-server 3.0.1.31
```

(3) 配置 Router C（Router B 向 Router A 同步后）

设置本地时钟作为参考时钟，层数为 1。

```
<RouterC> system-view
[RouterC] ntp-service refclock-master 1
```

本地同步后，设置 Router B 为对等体。

```
[RouterC] ntp-service unicast-peer 3.0.1.32
```

以上配置将 Router B 和 Router C 配置为对等体，Router C 处于主动对等体模式，Router B 处于被动对等体模式，由于 Router C 系统时钟的层数为 1，而 Router B 的层数为 3，所以 Router B 向 Router C 同步。

同步后查看 Router B 的状态。

```
[RouterB] display ntp-service status
Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 3.0.1.33
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: -21.1982 ms
 Root delay: 15.00 ms
 Root dispersion: 775.15 ms
```

```
 Peer dispersion: 34.29 ms
 Reference time: 15:22:47.083 UTC Sep 19 2005 (C6D95647.153F7CED)
```

此时 Router B 已经与 Router C 同步，层数比 Router C 的层数大 1，为 2。

查看 Router B 的 NTP 会话信息，可以看到 Router B 与 Router C 建立了连接。

```
[RouterB] display ntp-service sessions
       source     reference    stra  reach  poll  now   offset delay  disper
**************************************************************************
[245] 3.0.1.31  127.127.1.0    2    15    64    24   10535.0  19.6   14.5
[1234] 3.0.1.33   LOCL          1    14    64    27    -77.0   16.0   14.8
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  2
```

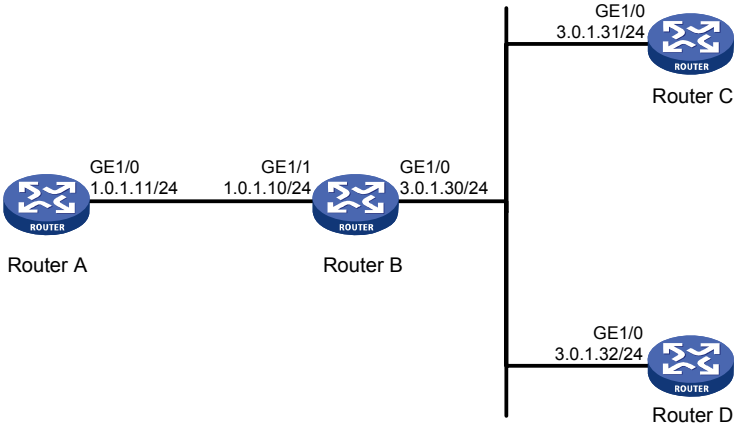
1.9.3 配置 NTP 广播模式

1. 组网需求

- Router C 设置本地时钟作为参考时钟，层数为 2；
- Router C 工作在广播服务器模式，从接口 GigabitEthernet1/0 向外发送广播报文；
- Router A 和 Router D 工作在广播客户端模式，分别从各自的接口 GigabitEthernet1/0 监听广播报文。

2. 组网图

图1-9 配置 NTP 广播模式组网图



3. 配置步骤

(1) 配置 Router C

设置本地时钟作为参考时钟，层数为 2。

```
<RouterC> system-view
[RouterC] ntp-service refclock-master 2
```

设置 Router C 为广播服务器，从接口 GigabitEthernet1/0 发送广播报文。

```
[RouterC] interface GigabitEthernet 1/0
[RouterC-GigabitEthernet1/0] ntp-service broadcast-server
```

(2) 配置 Router D

设置 Router D 为广播客户端，从接口 GigabitEthernet1/0 监听广播报文。

```
<RouterD> system-view
[RouterD] interface GigabitEthernet 1/0
[RouterD-GigabitEthernet1/0] ntp-service broadcast-client
```

(3) 配置 Router A

设置 Router A 为广播客户端，从接口 GigabitEthernet1/0 监听广播报文。

```
<RouterA> system-view
[RouterA] interface GigabitEthernet 1/0
[RouterA-GigabitEthernet1/0] ntp-service broadcast-client
```

由于 Router A 与 Router C 不在同一网段，所以接收不到 Router C 发出的广播报文，而 Router D 接收到 Router C 发出的广播报文后与其同步。

同步后查看 Router D 的状态。

```
[RouterD-GigabitEthernet1/0] display ntp-service status
Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 3.0.1.31
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 8.31 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

此时 Router D 已经与 Router C 同步，层数比 Router C 的层数大 1，为 3。

查看 Router D 的 NTP 会话信息，可以看到 Router D 与 Router C 建立了连接。

```
[RouterD-GigabitEthernet1/0] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset   delay   disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0  2    254     64    62   -16.0    32.0    16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

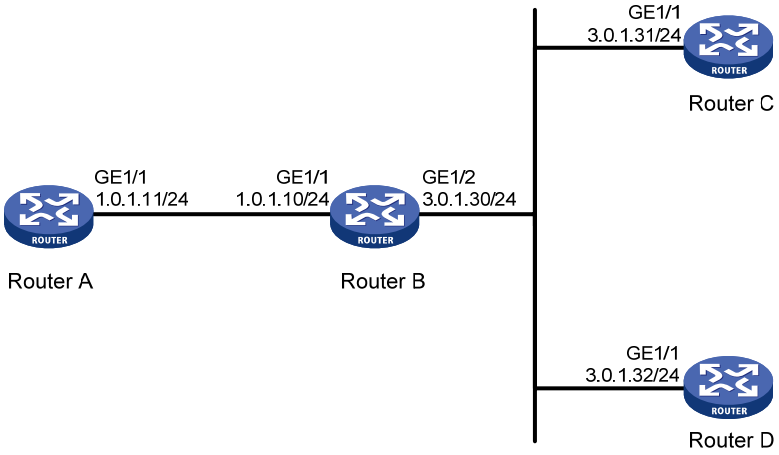
1.9.4 配置 NTP 组播模式

1. 组网需求

- Router C 设置本地时钟作为参考时钟，层数为 2；
- Router C 工作在组播服务器模式，从接口 GigabitEthernet1/1 向外发送组播报文；
- Router A 和 Router D 工作在组播客户端模式，分别从各自的接口 GigabitEthernet1/1 监听组播报文。

2. 组网图

图1-10 配置 NTP 组播模式组网图



3. 配置步骤

(1) 配置 Router C

设置本地时钟作为参考时钟，层数为 2。

```
<RouterC> system-view
[RouterC] ntp-service refclock-master 2
```

设置 Router C 为组播服务器，从接口 GigabitEthernet1/1 发送组播报文。

```
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] ntp-service multicast-server
```

(2) 配置 Router D

设置 Router D 为组播客户端，从接口 GigabitEthernet1/1 监听组播报文。

```
<RouterD> system-view
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] ntp-service multicast-client
```

由于 Router D 和 Router C 在同一个网段，不需要配置组播功能，Router D 就可以收到 Router C 发出的组播报文，并与其同步。

同步后查看 Router D 的状态。

```
[RouterD-GigabitEthernet1/0] display ntp-service status
Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 3.0.1.31
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 8.31 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

此时 Router D 已经与 Router C 同步，层数比 Router C 的层数大 1，为 3。

查看 Router D 的 NTP 会话信息，可以看到 Router D 与 Router C 建立了连接。

```
[RouterD-GigabitEthernet1/1] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset  delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0    2   254     64    62   -16.0    31.0   16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

(3) 配置 Router B

由于 Router A 与 Router C 不在同一网段，所以 Router B 上需要配置组播功能，否则 Router A 收不到 Router C 发出的组播报文。

配置组播功能。

```
<RouterB> system-view
[RouterB] multicast routing-enable
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] igmp enable
[RouterB-GigabitEthernet1/1] igmp static-group 224.0.1.1
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface gigabitethernet 1/2
[RouterB-GigabitEthernet1/2] pim dm
```

(4) 配置 Router A

```
<RouterA> system-view
[RouterA] interface gigabitethernet 1/1
```

设置 Router A 为组播客户端，从接口 GigabitEthernet1/1 监听组播报文。

```
[RouterA-GigabitEthernet1/1] ntp-service multicast-client
```

同步后查看 Router A 的状态。

```
[RouterA-GigabitEthernet1/1] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 3.0.1.31
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 40.00 ms
 Root dispersion: 10.83 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:02:49.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

此时 Router A 已经与 Router C 同步，层数比 Router C 的层数大 1，为 3。

查看 Router A 的 NTP 会话信息，可以看到 Router A 与 Router C 建立了连接。

```
[RouterA-GigabitEthernet1/1] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset  delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0    2   255     64    26   -16.0    40.0   16.6
```

```
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

说明

组播功能的详细介绍请参见“IP 组播分册”中的“IGMP 配置”和“PIM 配置”。

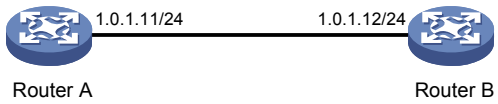
1.9.5 配置带身份验证的 NTP 客户端/服务器模式

1. 组网需求

- Router A 设置本地时钟作为参考时钟，层数为 2；
- Router B 工作在客户端模式，指定 Router A 为 NTP 服务器；
- Router A 和 Router B 上同时配置 NTP 验证。

2. 组网图

图1-11 配置带身份验证的 NTP 客户端/服务器模式组网图



3. 配置步骤

(1) 配置 Router A

设置本地时钟作为参考时钟，层数为 2。

```
<RouterA> system-view
[RouterA] ntp-service refclcok-master 2
```

(2) 配置 Router B

```
<RouterB> system-view
```

在 Router B 上启动身份验证。

```
[RouterB] ntp-service authentication enable
```

设置密钥。

```
[RouterB] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

指定密钥为可信密钥。

```
[RouterB] ntp-service reliable authentication-keyid 42
```

指定 NTP 服务器为 Router A。

```
[RouterB] ntp-service unicast-server 1.0.1.11 authentication-keyid 42
```

以上配置将 Router B 向 Router A 进行时间同步，但由于 Router A 没有使能 NTP 身份验证，所以，Router B 还是无法向 Router A 同步。

现在，向 Router A 增加以下配置：

在 Router A 上启动身份验证。

```
[RouterA] ntp-service authentication enable
```

设置密钥。

```
[RouterA] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

指定密钥为可信密钥。

```
[RouterA] ntp-service reliable authentication-keyid 42
```

此时，Router B 可以向 Router A 同步。

同步后查看 Router B 的状态。

```
[RouterB] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 1.0.1.11
Nominal frequency: 64.0000 Hz
Actual frequency: 64.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 1.05 ms
Peer dispersion: 7.81 ms
Reference time: 14:53:27.371 UTC Sep 19 2005 (C6D94F67.5EF9DB22)
```

可以看出，Router B 已经与 Router A 同步，层数比 Router A 的层数大 1，为 3。

查看 Router B 的 NTP 会话信息，可以看到 Router B 与 Router A 建立了连接。

```
[RouterB] display ntp-service sessions
      source      reference   stra  reach  poll  now  offset  delay  disper
**************************************************************************
[12345] 1.0.1.11  127.127.1.0    2    63    64    3    -75.5    31.0  16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

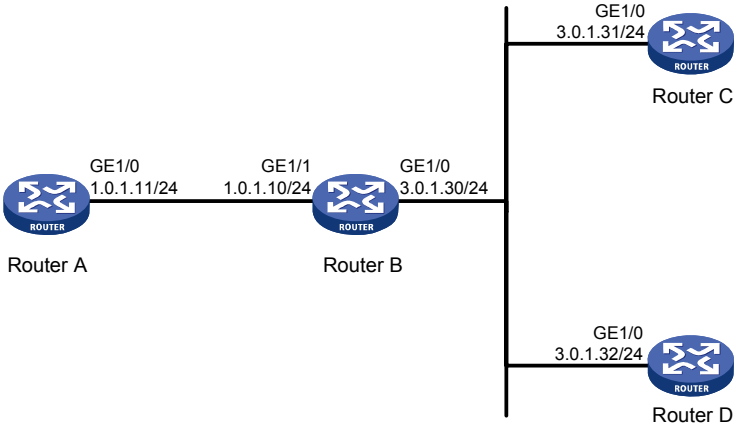
1.9.6 配置带身份验证的 NTP 广播模式

1. 组网需求

- Router C 设置本地时钟作为参考时钟，层数为 3；
- Router C 工作在广播服务器模式，从接口 GigabitEthernet1/0 向外发送广播报文；
- Router D 工作在广播客户端模式，从接口 GigabitEthernet1/0 监听广播报文；
- 同时在 Router C 和 Router D 上配置 NTP 验证。

2. 组网图

图1-12 配置带身份验证的 NTP 广播模式组网图



3. 配置步骤

(1) 配置 Router C

设置本地时钟作为参考时钟，层数为 3。

```
<RouterC> system-view
[RouterC] ntp-service refclock-master 3
```

配置 NTP 验证。

```
[RouterC] ntp-service authentication enable
[RouterC] ntp-service authentication-keyid 88 authentication-mode md5 123456
[RouterC] ntp-service reliable authentication-keyid 88
```

设置 Router C 为 NTP 广播服务器并指定密钥编号。

```
[RouterC] interface GigabitEthernet 1/0
[RouterC-GigabitEthernet1/0] ntp-service broadcast-server authentication-keyid 88
```

(2) 配置 Router D

配置 NTP 验证。

```
<RouterD> system-view
[RouterD] ntp-service authentication enable
[RouterD] ntp-service authentication-keyid 88 authentication-mode md5 123456
[RouterD] ntp-service reliable authentication-keyid 88
```

设置 Router D 为 NTP 广播客户端。

```
[RouterD] interface GigabitEthernet 1/0
[RouterD-GigabitEthernet1/0] ntp-service broadcast-client
```

以上配置将 Router D 配置为从接口 GigabitEthernet1/0 监听广播报文，Router C 从接口 GigabitEthernet1/0 发送广播报文。Router D 接收到 Router C 发出的广播报文后与其同步。

同步后查看 Router D 的状态。

```
[RouterD-GigabitEthernet1/0] display ntp-service status
Clock status: synchronized
 Clock stratum: 4
 Reference clock ID: 3.0.1.31
```

```
Nominal frequency: 64.0000 Hz
Actual frequency: 64.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 8.31 ms
Peer dispersion: 34.30 ms
Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

此时 Router D 已经与 Router C 同步，层数比 Router C 的层数大 1，为 4。

查看 Router D 的 NTP 会话信息，可以看到 Router D 与 Router C 建立了连接。

```
[RouterD-GigabitEthernet1/0] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset   delay   disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0  3    254     64    62   -16.0    32.0    16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

1.9.7 配置采用客户端/服务器模式实现 MPLS VPN 网络的时间同步

1. 组网需求

- PE 1、PE 2 上同时存在两个 VPN： VPN 1 和 VPN 2；
- CE 1 和 CE 3 是 VPN 1 内的设备；CE 2 和 CE 4 是 VPN 2 内的设备；
- CE 1 设置本地时钟作为参考时钟，层数为 1；
- 采用客户端/服务器模式实现 CE 3 向 CE 1 进行时间同步。



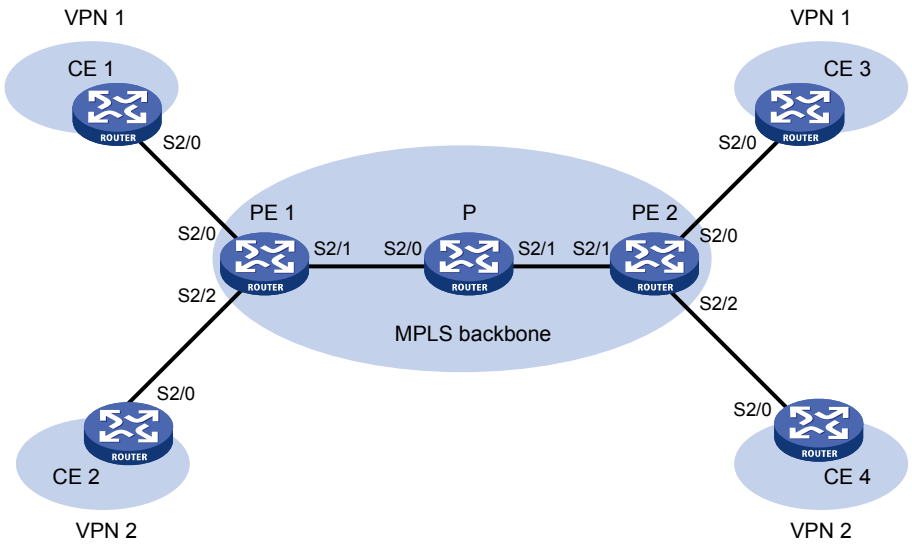
说明

目前只有单播方式（客户端/服务器模式或者对等体模式）的 NTP 时间同步支持 MPLS VPN 功能，组播和广播模式的时间同步暂时不支持 MPLS VPN 功能。

---

2. 组网图

图1-13 配置 MPLS VPN 网络的时间同步组网图



设备	接口	IP 地址	设备	接口	IP 地址
CE 1	S2/0	10.1.1.1/24	PE 1	S2/0	10.1.1.2/24
CE 2	S2/0	10.2.1.1/24		S2/1	172.1.1.1/24
CE 3	S2/0	10.3.1.1/24		S2/2	10.2.1.2/24
CE 4	S2/0	10.4.1.1/24	PE 2	S2/0	10.3.1.2/24
P	S2/0	172.1.1.2/24		S2/1	172.2.1.2/24
	S2/1	172.2.1.1/24		S2/2	10.4.1.2/24

3. 配置步骤

说明

在下面的配置之前，MPLS VPN 的相关配置必须完成，CE 1 和 PE 1 之间、PE 1 和 PE 2 之间、PE 2 和 CE 3 之间都必须有路由可达。MPLS VPN 的配置方法请参见“MPLS 分册”。

(1) 配置 CE 1

设置本地时钟作为参考时钟，层数为 1。

```
<CE1> system-view
[CE1] ntp-service refclcok-master 1
```

(2) 配置 CE 3

设置 VPN1 中的 CE 1 为 CE 3 的 NTP 服务器。

```
<CE3> system-view
[CE3] ntp-service unicast-server 10.1.1.1
```

经过一段时间之后在 CE 3 上可以查看 NTP 的会话信息、状态信息等，可以看出 CE 3 已经同步到 CE 1 了，层数为 2。

```
[CE3] display ntp-service status
 Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 10.1.1.1
 Nominal frequency: 63.9100 Hz
```

```
 Actual frequency: 63.9100 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 47.00 ms
 Root dispersion: 0.18 ms
 Peer dispersion: 34.29 ms
 Reference time: 02:36:23.119 UTC Jan 1 2001(BDFA6BA7.1E76C8B4)
[CE3] display ntp-service sessions
source          reference       stra reach poll  now offset  delay disper
**************************************************************************
[12345]10.1.1.1       LOCL         1    7   64   15    0.0   47.0    7.8
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
[CE2] display ntp-service trace
 server 127.0.0.1,stratum 2, offset -0.013500, synch distance 0.03154
 server 10.1.1.1,stratum 1, offset -0.506500, synch distance 0.03429
 refid 127.127.1.0
```

1.9.8 配置采用对等体模式实现 MPLS VPN 网络的时间同步

1. 组网需求

- PE 1 设置本地时钟作为参考时钟，层数为 1；
- 采用对等体模式实现 PE 2 向 PE 1 进行时间同步。

2. 组网图

如 图 1-13所示。

3. 配置步骤

(1) 配置 PE 1

设置本地时钟作为参考时钟，层数为 1。

```
<PE1> system-view
[PE1] ntp-service refclcok-master 1
```

(2) 配置 PE 2

设置 VPN 1 中的 PE 1 为 PE 2 的被动对等体。

```
<PE2> system-view
[PE2] ntp-service unicast-peer vpn-instance vpn1 10.1.1.2
```

经过一段时间之后在 PE 2 上可以查看 NTP 的会话信息、状态信息等，可以看出 PE 2 已经同步到 PE 1 了，层数为 2。

```
[PE2] display ntp-service status
Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 10.1.1.2
 Nominal frequency: 63.9100 Hz
 Actual frequency: 63.9100 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
```

```
 Root delay: 32.00 ms
 Root dispersion: 0.60 ms
 Peer dispersion: 7.81 ms
 Reference time: 02:44:01.200 UTC Jan 1 2001(BDFA6D71.33333333)
[PE2] display ntp-service sessions
source          reference       stra reach poll  now offset  delay disper
**************************************************************************
[12345]10.1.1.2       LOCL          1    1   64   29    -12.0   32.0    15.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
[PE2] display ntp-service trace
 server 127.0.0.1,stratum 2, offset -0.012000, synch distance 0.02448
 server 10.1.1.2,stratum 1, offset 0.003500, synch distance 0.00781
 refid 127.127.1.0
```

目 录

1 RMON配置 1-1
1.1 RMON简介 1-1
1.1.1 RMON概述 1-1
1.1.2 RMON工作机制 1-1
1.1.3 RMON组 1-1
1.2 配置RMON 1-3
1.2.1 配置准备 1-3
1.2.2 配置过程 1-3
1.3 RMON显示和维护 1-4
1.4 RMON典型配置举例 1-4

1 RMON 配置

1.1 RMON 简介

1.1.1 RMON 概述

RMON（Remote Monitoring，远程网络监视）的实现完全基于 SNMP 体系结构，它与现存的 SNMP 框架相兼容，不需对该协议进行任何修改。RMON 使 SNMP 更有效、更积极主动地监测远程网络设备，为监控子网的运行提供了一种高效的手段。RMON 能够减少网管站同代理间的通讯流量，达到简便而有力地管理大型互连网络的目的。

RMON 包括两部分：NMS（Network Management Station，网络管理站）和运行在各网络设备上的 Agent。

- RMON NMS 对它所在范围内的 Agent 进行管理。
- RMON Agent 在网络监视器或网络探测器上，跟踪统计其端口所连接的网段上的各种流量信息（如某段时间内某网段上的报文总数，或发往某台主机的正确报文总数等）。

1.1.2 RMON 工作机制

RMON 允许有多个监控者，监控者可用两种方法收集数据：

- 第一种方法利用专用的 RMON probe（探测仪）收集数据，NMS 直接从 RMON probe 获取管理信息并控制网络资源。这种方式可以获取 RMON MIB 的全部信息；
- 第二种方法是将 RMON Agent 直接植入网络设备（路由器、交换机、HUB 等），使它们成为带 RMON probe 功能的网络设施。RMON NMS 使用 SNMP 的基本命令与 RMON Agent 交换数据信息，收集网络管理信息，但这种方法受设备资源限制，一般不能获取 RMON MIB 的所有数据，大多数只收集四个组的信息。这四个组是：事件组、告警组、历史组和统计组。

我们的设备采用第二种方法。通过运行在网络监视器上的支持 RMON 的 RMON Agent，网管站可以获得与被管网络设备端口相连的网段上的整体流量、错误统计和性能统计等信息，进而实现对网络的管理。

1.1.3 RMON 组

RMON 规范（RFC1757）定义的 10 个 RMON 组中，H3C 系列设备支持的有事件组、告警组、历史组和统计组。此外，H3C 还自定义和实现了扩展告警组，以增强告警组的功能。下面将对这五个组作简要介绍。

1. 事件组

事件组用来定义事件索引号及事件的处理方式。事件组定义的事件主要用在告警组配置项和扩展告警组配置项中。当监控对象达到告警条件时，必然会触发事件，事件有如下几种处理方式：

- 将事件相关信息记录在事件日志表中
- 向网管站发送 Trap 消息
- 将事件相关信息记录在事件日志表中并向网管站发送 Trap 消息
- 不做任何处理

2. 告警组

RMON 告警管理可对指定的告警变量（如端口的统计数据）进行监视。被监视的告警变量的采样值大于或等于上限阈值时，触发一次上限告警事件；被监视的告警变量的采样值小于或等于下限阈值，触发一次下限告警事件，告警管理将按照事件的定义进行相应的处理。

用户定义了告警表项后，系统对告警表项的处理如下：

(1) 对所定义的告警变量按照定义的时间间隔进行采样。

(2) 将采样值和设定的阈值进行比较，越过阈值就触发相应事件。



说明

当告警变量的采样值在同一方向上连续多次超过阈值时，只会在第一次产生告警事件，后面的几次不会产生告警事件，即上限告警和下限告警是交替产生的，出现了一次上限告警，则下一次必为下限告警。

3. 扩展告警组

扩展告警表项可以对告警变量的采样值进行运算，然后将运算结果和设置的阈值比较，实现更为丰富的告警功能。

用户定义了扩展告警表项后，系统对扩展告警表项的处理如下：

(1) 对定义的扩展告警公式中的告警变量按照定义的时间间隔进行采样。

(2) 将采样值按照定义的运算公式进行计算。

(3) 将计算结果和设定的阈值进行比较，越过阈值就触发相应事件。



说明

当运算结果在同一方向上连续多次超过阈值时，只会在第一次产生告警事件，后面的几次不会产生告警事件，即上限告警和下限告警是交替产生的。

4. 历史组

历史组是按周期对端口的使用情况进行统计，并将统计结果存储在历史记录表中以便以后查看。统计数据包括带宽利用率、错误包数和总包数等。

在指定接口下创建历史表项成功后，统计组就对当前接口的报文数按周期进行统计，它统计的结果是一个周期内端口收发报文的情况。

5. 统计组

利用 RMON 统计管理功能，可以监视端口的使用情况。统计信息包括网络冲突数、CRC 校验错误报文数、过小（或超大）的数据报文数、广播、多播的报文数以及接收字节数、接收报文数等。

在指定接口下创建统计表项成功后，统计组就对当前接口的报文数进行统计，它统计的结果是一个连续的累加值。

1.2 配置 RMON

1.2.1 配置准备

在配置 RMON 功能之前，必须保证 SNMP Agent 已经正确配置。SNMP Agent 的配置请参见“系统分册”中的“SNMP 配置”。

1.2.2 配置过程

表1-1 RMON 配置过程

操作	命令	说明
进入系统视图	**system-view**	-
添加事件表的一个表项	**rmon event** *entry-number* [**description** *string*] { **log** | **log-trap** *log-trapcommunity* | **none** | **trap** *trap-community* } [**owner** *text*]	可选
进入以太网接口视图	**interface** *interface-type interface-number*	-
添加历史控制表的一个表项	**rmon history** *entry-number* **buckets** *number* **interval** *sampling-interval* [**owner** *text*]	可选
添加统计表的一个表项	**rmon statistics** *entry-number* [**owner** *text*]	可选
退出以太网接口视图	**quit**	-
添加告警表的一个表项	**rmon alarm** *entry-number alarm-variable sampling-interval* { **absolute** | **delta** } **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* [**owner** *text*]	可选 在添加告警表项之前，需要通过 **rmon event** 命令定义告警表项中引用的事件 如果告警变量是接口某统计参数，则要使告警表项生效需要配置相应的统计组
添加扩展告警表的一个表项	**rmon prialarm** *entry-number prialarm-formula prialarm-des sampling-interval* { **absolute** | **delta** } **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* **entrytype** { **forever** | **cycle** *cycle-period* } [**owner** *text*]	可选 在添加扩展告警表项之前，需要通过 **rmon event** 命令定义扩展告警表项中引用的事件 如果扩展告警变量是接口某统计参数，则要使扩展告警表项生效需要配置相应的统计组

说明

- 系统不允许创建两个配置完全相同的表项。如果新建表项参数的值和已存在表项对应参数的值完全相同（对于历史表项，只与同一接口下已存在的历史表项进行比较），系统将认为这两个表项的配置相同，创建操作失败。不同表项需要比较的参数请参见 表 1-2。
- 系统对每种类型表项的总数均进行了限制（具体数目请参见 表 1-2），当某种类型表项的总数达到系统允许创建的最大数目时，创建操作失败。
- 在新建历史控制表项时，如果指定的 **buckets** *number* 参数值超出了设备实际支持的历史表容量时，该历史控制表项会被添加，但该表项对应生效的 **buckets** *number* 的值为设备实际支持的历史表容量。

表1-2 RMON 配置约束表

表项名	需要比较的参数	最多可创建的表项数
事件表项	事件描述（**description** *string*）、事件类型（**log**、**trap**、**logtrap** 或 **none**）和团体名（*trap-community* 或 *log-trapcommunity*）	60
历史表项	采样间隔（**interval** *sampling-interval*）	100
统计表项	每个接口下只能定义一个统计表项	100
告警表项	告警变量（*alarm-variable*）、采样间隔（*sampling-interval*）、采样类型（**absolute** 或 **delta**）、上限阈值（*threshold-value1*）和下限阈值（*threshold-value2*）	60
扩展告警表项	告警变量公式（*prialarm-formula*）、采样间隔（*sampling-interval*）、采样类型（**absolute**、**changeratio** 或 **delta**）、上限阈值（*threshold-value1*）和下限阈值（*threshold-value2*）	50

1.3 RMON 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 RMON 的运行情况，通过查看显示信息验证配置的效果。

表1-3 RMON 显示和维护

操作	命令
显示 RMON 统计消息	**display rmon statistics** [*interface-type interface-number*]
显示 RMON 历史控制表及历史采样信息	**display rmon history** [*interface-type interface-number*]
显示 RMON 告警配置信息	**display rmon alarm** [*entry-number*]
显示 RMON 扩展告警配置信息	**display rmon prialarm** [*entry-number*]
显示 RMON 事件配置信息	**display rmon event** [*entry-number*]
显示指定事件的记录	**display rmon eventlog** [*entry-number*]

1.4 RMON 典型配置举例

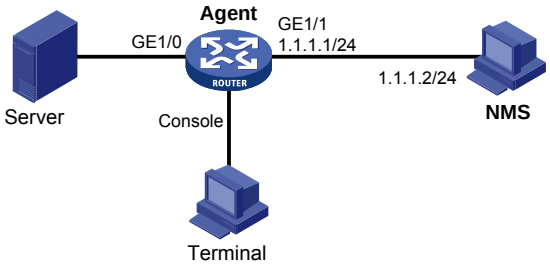
1. 组网需求

Agent 通过 console 口连接配置终端，通过以太网连接 NMS。NMS 使用 Quidview 网管系统，可以通过 NMS 查询路由器的运行状况。

- GigabitEthernet1/0 接口连接 FTP 服务器，现需要对该服务器的流量进行统计。以 5 秒的采样间隔采样，当流量过大或过低时发送相应的 trap 信息通知 NMS；
- 在 Agent 上通过 **display rmon statistics** 命令查看显示结果，同时在 NMS 上查看统计结果。

2. 组网图

图1-1 配置 RMON 组网图



3. 配置步骤

配置 SNMP Agent（注意：SNMP Agent 的读、写团体及版本应与 NMS 侧的配置一致）。

```
<Sysname> system-view
[Sysname] snmp-agent
[Sysname] snmp-agent community read public
[Sysname] snmp-agent community write private
[Sysname] snmp-agent community write user1
[Sysname] snmp-agent sys-info version v1
[Sysname] snmp-agent trap enable
[Sysname] snmp-agent target-host trap address udp-domain 1.1.1.2 params securityname user1
```

配置以太网接口 GigabitEthernet1/1 的 IP 地址。

```
[Sysname] interface GigabitEthernet 1/1
[Sysname-GigabitEthernet1/1] ip address 1.1.1.1 255.255.255.0
```

配置 RMON 对以太网接口 GigabitEthernet1/0 进行统计。

```
[Sysname-GigabitEthernet1/1] quit
[Sysname] interface GigabitEthernet 1/0
[Sysname-GigabitEthernet1/0] rmon statistics 1 owner user1-rmon
[Sysname-GigabitEthernet1/0] quit
```

配置 RMON 告警表项。（当节点 1.3.6.1.2.1.16.1.1.1.4.1 的相对采样值超过 100 时，触发事件 1——发起 trap 信息；当节点的相对采样值低于 50 时，触发事件 2——发起 trap 信息。）

```
[Sysname] rmon event 1 description rising trap router1 owner user1-rmon
[Sysname] rmon event 2 description falling trap router1 owner user1-rmon
[Sysname] rmon alarm 1 1.3.6.1.2.1.16.1.1.1.4.1 5 delta rising-threshold 100 1
falling-threshold 50 2
```

查看 RMON 告警表信息。

```
<Sysname> display rmon alarm 1
Alarm table 1 owned by null is Valid.
  Samples type          : delta
  Variable formula      : 1.3.6.1.2.1.16.1.1.1.4.1<etherStatsOctets.1>
  Sampling interval     : 5(sec)
  Rising threshold      : 100(linked with event 1)
  Falling threshold     : 50(linked with event 2)
  When startup enables  : risingOrFallingAlarm
```

```
  Latest value          : 0
```

查看以太网口的统计信息。

```
<Sysname> display rmon statistics GigabitEthernet 1/0
Statistics entry 1 owned by user1-rmon is VALID.
  Interface : GigabitEthernet1/0<ifIndex.983040>
  etherStatsOctets         : 57329     , etherStatsPkts          : 455
  etherStatsBroadcastPkts  : 53        , etherStatsMulticastPkts : 353
  etherStatsUndersizePkts  : 0         , etherStatsOversizePkts  : 0
  etherStatsFragments      : 0         , etherStatsJabbers       : 0
  etherStatsCRCAlignErrors : 0         , etherStatsCollisions    : 0
  etherStatsDropEvents (insufficient resources): 0
  Packets received according to length:
  64     : 7         ,  65-127  : 413       ,  128-255  : 35
  256-511: 0         ,  512-1023: 0         ,  1024-1518: 0
```

完成以上配置后，当告警事件被触发时，在 NMS 的告警管理部分可以查看相应的记录。在设备侧也会有相应信息，如下所示。

```
[Sysname]
#Aug 27 16:31:34:12 2005 Sysname RMON/2/ALARMFALL:Trap 1.3.6.1.2.1.16.0.2 Alarm table 1
monitors 1.3.6.1.2.1.16.1.1.1.4.1 with sample type 2,has sampled alarm value 0 less than(or
=) 50.
```

目 录

1 SNMP配置 1-1
1.1 SNMP简介 1-1
1.1.1 SNMP的工作机制 1-1
1.1.2 SNMP的协议版本 1-1
1.1.3 MIB简介 1-2
1.2 配置SNMP基本参数 1-2
1.3 配置SNMP日志 1-4
1.3.1 SNMP日志介绍 1-4
1.3.2 开启SNMP日志功能 1-5
1.4 配置SNMP Trap 1-5
1.4.1 开启Trap功能 1-5
1.4.2 配置Trap报文发送参数 1-6
1.5 SNMP显示和维护 1-7
1.6 SNMP典型配置举例 1-8
1.7 SNMP日志典型配置举例 1-9
2 MIB风格配置 2-1
2.1 设置MIB风格 2-1
2.2 MIB显示和维护 2-1

1 SNMP 配置

1.1 SNMP 简介

SNMP（Simple Network Management Protocol，简单网络管理协议）是使用 TCP/IP 协议族对互联网上的设备进行管理的一个框架，它提供一组基本的操作来监视和维护互联网。SNMP 具有以下优势：

- 自动化网络管理。网络管理员可以利用 SNMP 平台在网络上的节点检索信息、修改信息、发现故障、完成故障诊断、进行容量规划和生成报告。
- 屏蔽不同设备的物理差异，实现对不同厂商产品的自动化管理。SNMP 只提供最基本的功能集，使得管理任务分别与被管设备的物理特性和下层的联网技术相对独立，从而实现对不同厂商设备的管理，特别适合在小型、快速和低成本的环境中使用。

1.1.1 SNMP 的工作机制

SNMP 网络元素分为 NMS 和 Agent 两种。

- NMS（Network Management Station，网络管理站）是运行 SNMP 客户端程序的工作站，能够提供非常友好的人机交互界面，方便网络管理员完成绝大多数的网络管理工作。
- Agent 是驻留在设备上的一个进程，负责接收、处理来自 NMS 的请求报文。在一些紧急情况下，如接口状态发生改变等，Agent 也会通知 NMS。

NMS 是 SNMP 网络的管理者，Agent 是 SNMP 网络的被管理者。NMS 和 Agent 之间通过 SNMP 协议来交互管理信息。

SNMP 提供四种基本操作：

- Get 操作：NMS 使用该操作查询 Agent 的一个或多个对象的值。
- Set 操作：NMS 使用该操作重新设置 Agent 数据库（MIB，Management Information Base）中的一个或多个对象的值。
- Trap 操作：Agent 使用该操作向 NMS 发送报警信息。
- Inform 操作：NMS 使用该操作向其他 NMS 发送报警信息。

1.1.2 SNMP 的协议版本

目前，设备的 SNMP Agent 支持 SNMP v3 版本，兼容 SNMP v1 版本和 SNMP v2c 版本。

- SNMP v1 采用团体名（Community Name）认证。团体名用来定义 SNMP NMS 和 SNMP Agent 的关系。如果 SNMP 报文携带的团体名没有得到设备的认可，该报文将被丢弃。团体名起到了类似于密码的作用，用来限制 SNMP NMS 对 SNMP Agent 的访问。
- SNMP v2c 也采用团体名认证。它在兼容 SNMP v1 的同时又扩充了 SNMP v1 的功能：它提供了更多的操作类型（GetBulk 和 InformRequest）；它支持更多的数据类型（Counter64 等）；它提供了更丰富的错误代码，能够更细致地区分错误。

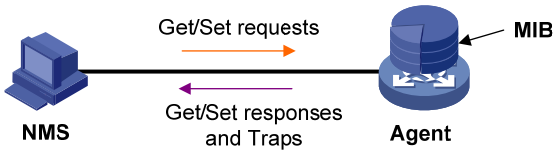
- SNMP v3 提供了基于用户的安全模型（USM，User-Based Security Model）的认证机制。用户可以设置认证和加密功能，认证用于验证报文发送方的合法性，避免非法用户的访问；加密则是对 NMS 和 Agent 之间的传输报文进行加密，以免被窃听。通过有无认证和有无加密等功能组合，可以为 SNMP NMS 和 SNMP Agent 之间的通信提供更高的安全性。

NMS 和 Agent 的 SNMP 版本匹配，是它们之间成功互访的前提条件。Agent 可以同时配置多个版本，与不同的 NMS 交互采用不同的版本。

1.1.3 MIB 简介

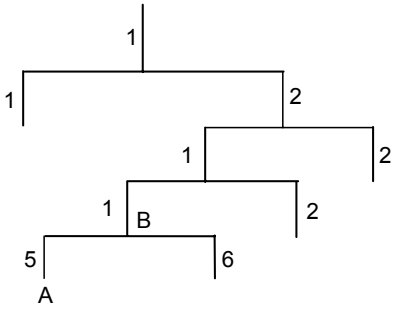
任何一个被管理的资源都表示成一个对象，称为被管理的对象。MIB（Management Information Base，管理信息库）是被管理对象的集合。它定义了被管理对象的一系列的属性：对象的名字、对象的访问权限和对象的数据类型等。每个Agent都有自己的MIB。NMS根据权限可以对MIB中的对象进行读/写操作。NMS、Agent和MIB之间的关系如 图 1-1所示。

图1-1 NMS、Agent 和 MIB 关系图



MIB是以树状结构进行存储的。树的节点表示被管理对象，它可以用从根开始的一条路径唯一地识别（OID）。如 图 1-2所示。被管理对象B可以用一串数字{1.2.1.1}唯一确定，这串数字是被管理对象的OID（Object Identifier，对象标识符）。

图1-2 MIB 树结构



1.2 配置 SNMP 基本参数

由于SNMP v3 版本的配置和SNMP v1 版本、SNMP v2c版本的配置有较大区别，所以下面分两种情况进行SNMP基本功能的配置介绍，详见 表 1-1和 表 1-2。

表1-1 配置 SNMP 基本参数（SNMP v3 版本）

操作	命令	说明
进入系统视图	**system-view**	-
启动 SNMP Agent 服务	**snmp-agent**	可选 缺省情况下，SNMP Agent 服务处于关闭状态 执行此命令或执行 **snmp-agent** 的任何一条配置命令（不含 **display** 命令），都可以启动 SNMP Agent
设置系统信息	**snmp-agent sys-info** { **contact** *sys-contact* | **location** *sys-location* | **version** { **all** | { **v1** | **v2c** | **v3** }* } }	可选 缺省情况下，系统维护联系信息为“Hangzhou H3C Technologies Co.,Ltd.”；物理位置信息为“Hangzhou, China”；版本为 SNMP v3
设置一个 SNMP 组	**snmp-agent group v3** *group-name* [**authentication** | **privacy**] [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*]	必选
计算用户给定明文密码通过加密算法处理后的密文密码	**snmp-agent calculate-password** *plain-password* **mode** { **md5** | **sha** } { **local-engineid** | **specified-engineid** *engineid* }	可选
为一个 SNMP 组添加一个新用户	**snmp-agent usm-user v3** *user-name group-name* [[**cipher**] **authentication-mode** { **md5** | **sha** } *auth-password* [**privacy-mode** { **3des** | **aes128** | **des56** } *priv-password*]] [**acl** *acl-number*]	必选 如果使用 **cipher** 参数，则后面的 *auth-password* 和 *priv-password* 都将被视为密文密码
设置 Agent 能接收/发送的 SNMP 消息包的最大长度	**snmp-agent packet max-size** *byte-count*	可选 缺省情况下，Agent 能接收/发送的 SNMP 消息包长度的最大值为 1500 字节
设置本地 SNMP 实体的引擎 ID	**snmp-agent local-engineid** *engineid*	可选 缺省为公司的企业号+设备信息
创建或更新 MIB 视图内容	**snmp-agent mib-view** { **excluded | included** } *view-name oid-tree* [**mask** *mask-value*]	可选 缺省情况下，视图名为 ViewDefault，OID 为 1

表1-2 配置 SNMP 基本参数（SNMP v1 版本、SNMP v2c 版本）

操作	命令	说明
进入系统视图	**system-view**	-
启动 SNMP Agent 服务	**snmp-agent**	可选 缺省情况下，SNMP Agent 服务处于关闭状态。 执行此命令或执行 **snmp-agent** 的任何一条配置命令，都可以启动 SNMP Agent

操作			命令	说明
设置系统信息			**snmp-agent sys-info** { **contact** *sys-contact* | **location** *sys-location* | **version** { { **v1** | **v2c** | **v3** }* | **all** } }	必选 缺省情况下，系统维护联系信息为“Hangzhou H3C Technologies Co.,Ltd..”；物理位置信息为“Hangzhou, China”;版本为 SNMP v3
设置访问权限	直接设置	创建一个新的 SNMP 团体	**snmp-agent community** { **read** | **write** } *community-name* [**acl** *acl-number* | **mib-view** *view-name*]*	二者必选其一 直接设置是以 SNMP v1 和 v2c 版本的团体名进行设置 间接设置采用与 SNMP v3 版本一致的命令形式，添加的用户到指定的组，即相当于 SNMP v1 和 SNMP v2c 版本的团体名，在 NMS 上配置的团体名需要跟 Agent 上配置的用户名一致
	间接设置	设置一个 SNMP 组	**snmp-agent group** { **v1** | **v2c** } *group-name* [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*]	
		为一个 SNMP 组添加一个新用户	**snmp-agent usm-user** { **v1** | **v2c** } *user-name group-name* [**acl** *acl-number*]	
设置 Agent 能接收/发送的 SNMP 消息包的大小			**snmp-agent packet max-size** *byte-count*	可选 缺省情况下，Agent 能接收/发送的 SNMP 消息包长度的最大值为 1500 字节
设置本地 SNMP 实体的引擎 ID			**snmp-agent local-engineid** *engineid*	可选 缺省为公司的企业号+设备信息
创建或更新 MIB 视图内容			**snmp-agent mib-view** { **excluded** | **included** } *view-name oid-tree* [**mask** *mask-value*]	可选 缺省情况下，视图名为 ViewDefault

注意

USM 用户创建后是否有效，与设备的 SNMP 实体引擎 ID 有关，如果用户创建时的引擎 ID 和当前的引擎 ID 不同，则该用户当前无效。

1.3 配置 SNMP 日志

1.3.1 SNMP 日志介绍

SNMP 日志功能将记录 NMS 对 SNMP Agent 的 GET 和 SET 操作。当进行 GET 操作时，Agent 会记录 NMS 用户的 IP 地址、GET 操作的节点名和节点的 OID。当进行 SET 操作时，Agent 会记录 NMS 用户的 IP 地址、SET 操作的节点名、节点的 OID、设置的值以及 SET 操作返回的错误码和错误索引。这些日志将被发送到设备的信息中心，级别为 informational，即作为设备的一般提示信息。通过设置信息中心的参数，最终决定 SNMP 日志的输出规则（即是否允许输出以及输出方向）。

SNMP 日志记录 GET 请求、SET 请求和 SET 响应信息，不记录 GET 响应信息。

1.3.2 开启 SNMP 日志功能

表1-3 开启 SNMP 日志功能

操作	命令	说明
进入系统视图	**system-view**	-
打开 SNMP 日志开关	**snmp-agent log** { **all** | **get-operation** | **set-operation** }	必选 缺省情况下，SNMP 日志开关处于关闭状态
配置 SNMP 日志输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	可选 缺省情况下，SNMP 日志只会输出到日志主机（loghost）和日志文件（logfile）方向，如果要输出到别的方向（比如控制台、监视终端等）需要使用该命令配置

说明

- 大量的日志记录会占用设备的存储空间，影响设备的性能。正常情况下，建议关闭 SNMP 日志功能。
- SNMP 日志的大小不能超过信息中心日志允许的大小，其中每条日志信息中记录的 node 域和 value 域的长度之和不能超过 1K 字节。否则，超出的部分将不输出。
- 有关系统信息、信息中心以及 **info-center source** 命令的详细介绍请参见“系统分册”中的“信息中心配置”。

1.4 配置 SNMP Trap

1.4.1 开启 Trap 功能

Trap 是 Agent 主动向 NMS 发送的信息，用于报告一些紧急的重要事件（如被管理设备重新启动等）。Trap 报文有两种：通用 Trap 和企业自定义 Trap。设备支持的通用 Trap 包括 **authentication**、**coldstart**、**linkdown**、**linkup** 和 **warmstart** 五种，其它均为企业自定义 Trap。企业自定义 Trap 由模块生成。因为 Trap 信息通常较多，会占用设备内存，从而影响设备性能，所以建议用户根据需要开启指定模块的 Trap 功能，生成相应的 Trap 报文。

开启 Trap 功能后，相应模块生成的 Trap 报文将被发送到设备的信息中心。信息中心有七个输出方向，缺省情况下，允许所有模块的所有 Trap 信息发往控制台（console）、监视终端（monitor）、日志主机（loghost）和日志文件（logfile）；允许所有模块的高于等于 warnings 级别的 Trap 信息发往告警缓冲区（trapbuffer）和 SNMP 模块（snmpagent）；日志缓冲区（logbuffer）方向不能发送 Trap 信息。可以根据各模块生成的 Trap 信息的级别设置信息中心参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。有关信息中心的配置请参见“系统分册”中的“信息中心配置”。

表1-4 开启 Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
在全局下开启 Trap 功能	**snmp-agent trap enable** [**bfd** | **bgp** | **configuration** | **flash** | **fr** | **mpls** | **ospf** [*process-id*] [**ifauthfail** | **ifcfgerror** | **ifrxbadpkt** | **ifstatechange** | **iftxretransmit** | **lsdbapproachoverflow** | **lsdboverflow** | **maxagelsa** | **nbrstatechange** | **originatelsa** | **vifcfgerror** | **virifauthfail** | **virifrxbadpkt** | **virifstatechange** | **viriftxretransmit** | **virnbrstatechange**] * | **standard** [**authentication** | **coldstart** | **linkdown** | **linkup** | **warmstart**] * | **system** | **vrrp** [**authfailure** | **newmaster**]]	可选 缺省情况下，在全局下的 Trap 功能处于开启状态
进入接口视图	**interface** *interface-type interface-number*	-
在端口下开启端口状态变化的 Trap 功能	**enable snmp trap updown**	可选 缺省情况下，端口下端口状态变化的 Trap 功能处于开启状态

注意

如果要求端口在状态发生改变时生成端口 linkUp/linkDown 的 Trap 报文，需要在端口下和全局都开启端口状态变化的 Trap 功能。端口下开启请使用命令 **enable snmp trap updown**，全局下开启请使用命令 **snmp-agent trap enable** [**standard** [**linkdown** \| **linkup**] *]。

1.4.2 配置 Trap 报文发送参数

1. 配置准备

如果要将 Trap 发送给 NMS，则需要进行以下配置准备：

- 完成 SNMP 基本配置（包括版本设置，如果使用 SNMP v1 和 v2c 版本需要设置团体名；如果使用 SNMP v3 版本需要设置用户名和 MIB 视图）。
- 设备与 NMS 建立连接，能够互相操作。

2. 配置过程

当 Trap 报文发送到 SNMP 模块后，SNMP 模块将 Trap 报文存在消息队列里，用户可以设置该队列的长度、Trap 报文在队列里的保存时间，也可以将报文发送给指定的目标主机（通常为 NMS）等。

表1-5 配置 Trap 报文发送参数

操作	命令	说明
进入系统视图	**system-view**	-
设置 Trap 目标主机属性	**snmp-agent target-host trap address udp-domain** { *ip-address* | **ipv6** *ipv6-address* } [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*] **params securityname** *security-string* [**v1** | **v2c** | **v3** [**authentication** | **privacy**]]	可选 **vpn-instance** 参数只支持在 IPv4 网络使用 如果要将 Trap 信息发送给 NMS，该步骤为必选，并将 *ip-address* 指定为 NMS 的 IP 地址

操作	命令	说明
设置发送 Trap 报文中的源地址	**snmp-agent trap source** *interface-type* { *interface-number* | *interface-number.subnumber* }	可选
对 RFC 定义的标准 linkUp/linkDown Trap 报文进行私有扩展	**snmp-agent trap if-mib link extended**	可选 缺省情况下，使用的是 RFC 定义的标准 linkUp/linkDown Trap 报文
设置发送的 Trap 报文消息队列的长度	**snmp-agent trap queue-size** *size*	可选 缺省情况下，Trap 报文的消息队列最多可以存储 100 条 Trap 消息
设置 Trap 报文的保存时间	**snmp-agent trap life** *seconds*	可选 缺省情况下，Trap 报文的保存时间为 120 秒

说明

- 对 linkUp/linkDown Trap 报文进行私有扩展后，设备生成和发送的 linkUp/linkDown Trap 报文将由 RFC 定义的标准 linkUp/linkDown Trap 报文后追加接口描述和接口类型信息构成。如果 NMS 不支持该扩展报文，请禁用私有扩展功能，使设备发送标准的 linkUp/linkDown Trap 报文。
- 当 Trap 报文的发送队列满，又收到新的 Trap 信息时，则系统会自动删除最旧的 Trap 报文。
- Trap 报文的保存时间到达时，系统会自动删除该 Trap 报文。

1.5 SNMP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令，均可以显示配置后 SNMP 的运行情况，通过查看显示信息，来验证配置的效果。

表1-6 SNMP 显示和维护

操作	命令
显示系统维护联络信息、系统位置信息及 SNMP 版本信息	**display snmp-agent sys-info** [**contact** | **location** | **version**]*
显示 SNMP 报文统计信息	**display snmp-agent statistics**
显示设备的 SNMP 实体引擎 ID	**display snmp-agent local-engineid**
显示 SNMP 组信息	**display snmp-agent group** [*group-name*]
显示 Trap 消息队列的基本信息	**display snmp-agent trap queue**
显示系统当前可以发送 Trap 消息的模块及其 Trap 消息的使能状态	**display snmp-agent trap-list**
显示 SNMP v3 用户信息	**display snmp-agent usm-user** [**engineid** *engineid* | **username** *user-name* | **group** *group-name*] *
显示 SNMP v1 或 SNMP v2c 团体信息	**display snmp-agent community** [**read** | **write**]
显示 MIB 视图的信息	**display snmp-agent mib-view** [**exclude** | **include** | **viewname** *view-name*]

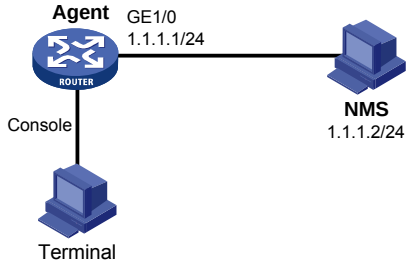
1.6 SNMP 典型配置举例

1. 组网需求

- 网管工作站（NMS）与路由器（Agent）通过以太网相连
- 网管工作站 IP 地址为 1.1.1.2/24
- 路由器以太网口 IP 地址为 1.1.1.1/24
- NMS 通过 SNMP v3 只能对 Agent 的接口状态进行监控管理，Agent 在故障或者出错的时候能够主动向 NMS 报告情况，NMS 接收 Trap 的端口号为 5000

2. 组网图

图1-3 配置 SNMP 组网图



3. 配置步骤

(1) 配置 Agent

设置访问权限。

```
<Sysname> system-view
[Sysname] undo snmp-agent mib-view ViewDefault
[Sysname] snmp-agent mib-view included test interfaces
[Sysname] snmp-agent group v3 managev3group read-view test write-view test
[Sysname] snmp-agent usm-user v3 managev3user managev3group
```

设置路由器的联系人和位置信息，以方便维护。

```
[Sysname] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Sysname] snmp-agent sys-info location telephone-closet,3rd-floor
```

设置以太网端口 GigabitEthernet1/0 的 IP 地址为 1.1.1.1/24。

```
[Sysname] interface GigabitEthernet 1/0
[Sysname-GigabitEthernet1/0] ip address 1.1.1.1 24
```

将以太网端口 GigabitEthernet1/0 上的接口 IP 地址作为 Trap 报文的源地址。

```
[Sysname] snmp-agent trap source GigabitEthernet 1/0
```

允许向网管工作站（NMS）1.1.1.2/24 发送 Trap 报文，使用的团体名为 public。

```
[Sysname] snmp-agent trap enable
[Sysname] snmp-agent target-host trap address udp-domain 1.1.1.2 udp-port 5000 params
securityname public
```

(2) 配置 NMS

SNMP v3 采用认证和加密的安全机制，在 NMS 上需要设置用户名，选择安全级别。根据不同的安全级别，需要分别设置认证方式、认证密码、加密方式、加密密码等。另外，还要设置“超时”时间和“重试次数”。用户可利用网管系统完成对设备的查询和配置操作，具体情况请参考 NMS 的配套手册。

说明

NMS 侧的配置必须和设备侧保持一致，否则无法进行相应操作。

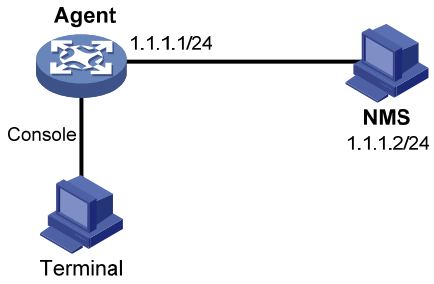
1.7 SNMP 日志典型配置举例

1. 组网需求

- NMS 与 Agent 通过以太网相连
- NMS 的 IP 地址为 1.1.1.2/24
- Agent 的 VLAN 接口 IP 地址为 1.1.1.1/24
- 在 Agent 上进行如下配置：设置团体名和访问权限、SNMP 协议版本

2. 组网图

图1-4 配置 SNMP 日志组网图



3. 配置步骤

说明

下面只列出了与 SNMP 日志相关的配置，NMS 和 Agent 的 SNMP 功能配置略。

打开控制台对日志信息的显示功能（此步骤可省略，缺省为该功能是开启的）。

```
<Sysname> terminal monitor
<Sysname> terminal logging
```

配置信息中心允许输出级别高于等于 informational 的系统信息到 Console 口。

```
<Sysname> system-view
[Sysname] info-center source snmp channel console log level informational
```

打开 Agent 的 SNMP 日志开关，对 NMS 的 GET 和 SET 操作进行记录。

```
[Sysname] snmp-agent log get-operation
[Sysname] snmp-agent log set-operation
```

- NMS 对 Agent 进行 GET 操作时，可以在控制台上看到如下日志信息。

```
%Jan 1 02:49:40:566 2006 Sysname SNMP/6/GET:
seqNO = <10> srcIP = <1.1.1.2> op = <get> node = <sysName(1.3.6.1.2.1.1.5.0)> value=<>
```

- NMS 对 Agent 进行 SET 操作，可以在控制台上看到如下日志信息。

```
%Jan 1 02:59:42:576 2006 Sysname SNMP/6/SET:
seqNO = <11> srcIP = <1.1.1.2> op = <set> errorIndex = <0> errorStatus =<noError> node =
<sysName(1.3.6.1.2.1.1.5.0)> value = <Sysname>
```

表1-7 SNMP 日志输出信息描述表

字段	描述
Jan 1 02:49:40:566 2006	表示 SNMP 日志生成的时间
seqNO	表示该 SNMP 日志的编号（系统会对记录的 SNMP 日志进行自动编号，编号从 0 开始）
srcIP	NMS 的 IP 地址
op	表示 SNMP 操作类型（GET 或者 SET）
node	SNMP 操作节点的名称和实例的 OID
errorIndex	错误索引（其中 0 表示没有错误）
errorstatus	错误状态（其中 noError 表示没有错误）
value	SET 操作时设置的值（GET 操作时此域为空，表示不记录 GET 操作获取的值） 当设置的值为字符串，而且字符串中包含编码范围超出了 ASCII 0~127 或者不可显示的字符时，则以十六进制的方式显示整个字符串，形如：value = <81-43>[hex]



说明

信息中心的系统信息可以输出到控制台或日志缓冲区等方向。本举例是将 SNMP 日志信息输出到控制台，如果要配置别的输出方向，请参见“系统分册”中的“信息中心配置”。

2 MIB 风格配置

H3C 私有 MIB 有两种风格，一种设备 sysOID 在 H3C 企业 ID 25506 下，私有 MIB 在企业 ID 2011 下的 MIB，称为 H3C 兼容风格 MIB；一种设备 sysOID 与私有 MIB 均在 H3C 企业 ID 25506 下的 MIB，称为 H3C 新风格 MIB。这两种风格的 MIB 在管理功能上完全一致，区别仅在于根节点不同。设备在出厂前都会加载好 MIB，设备型号不同，可能 MIB 风格不同。为了支持 NMS 对设备的灵活管理，设备提供了 MIB 风格配置功能，即用户可以任意切换 MIB 的两种风格，但需要保证设备的 MIB 风格和 NMS 的 MIB 风格一致，NMS 才能顺利配套管理设备。

2.1 设置 MIB 风格

表2-1 设置 MIB 风格

操作	命令	说明
进入系统视图	**system-view**	-
设置设备的 MIB 风格	**mib-style [new | compatible]**	可选 缺省情况下，设备的 MIB 风格为 **new**

说明

修改设备的 MIB 风格后需要重启设备才会生效。所以，建议在设置 MIB 风格后，马上重启设备，以保证 MIB 风格修改生效。

2.2 MIB 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令，均可以显示配置后 MIB 的风格，通过查看显示信息，来验证配置的效果。

表2-2 SNMP 显示和维护

操作	命令
查询设备的 MIB 风格	**display mib-style**

目 录

1 文件系统管理配置 1-1
1.1 文件系统管理 1-1
1.1.1 文件系统简介 1-1
1.1.2 文件名参数输入规则 1-1
1.1.3 目录操作 1-2
1.1.4 文件操作 1-3
1.1.5 批处理操作 1-5
1.1.6 存储设备操作 1-5
1.1.7 设置文件系统的提示方式 1-7
1.1.8 文件系统应用举例 1-7
1.2 配置文件管理 1-8
1.2.1 配置文件简介 1-8
1.2.2 保存当前配置 1-9
1.2.3 配置主备板配置文件同步保存功能 1-11
1.2.4 设置下次启动配置文件 1-11
1.2.5 备份下次启动配置文件 1-12
1.2.6 删除设备中的下次启动配置文件 1-12
1.2.7 恢复下次启动配置文件 1-13
1.2.8 设备配置显示和维护 1-13
2 FTP配置 2-1
2.1 FTP简介 2-1
2.1.1 FTP概述 2-1
2.1.2 FTP的实现 2-1
2.2 配置FTP客户端 2-2
2.2.1 建立FTP连接 2-2
2.2.2 FTP客户端基本操作 2-3
2.2.3 FTP客户端配置举例 2-5
2.3 配置FTP服务器 2-7
2.3.1 FTP服务器的基本配置 2-7
2.3.2 配置FTP服务器的验证和授权 2-7
2.3.3 FTP服务器配置举例 2-8
2.4 FTP显示和维护 2-10
3 TFTP配置 3-1
3.1 TFTP简介 3-1
3.1.1 TFTP概述 3-1
3.1.2 TFTP的实现 3-1
3.2 配置TFTP客户端 3-2
3.3 TFTP客户端显示和维护 3-3
3.4 TFTP客户端配置举例 3-3

1 文件系统管理配置

1.1 文件系统管理

1.1.1 文件系统简介

设备运行过程中所需要的文件（如：主机软件、配置文件等）保存在设备的存储设备中，为了方便用户对存储设备进行有效的管理，设备以文件系统的方式对这些文件进行管理。文件系统功能主要包括目录的创建和删除、文件的拷贝和显示等。

缺省情况下，对于有可能丢失数据的命令（比如删除文件、覆盖文件等命令），文件系统将提示用户进行确认。

根据操作对象的不同，可以把文件系统操作分为以下几类：

- 目录操作
- 文件操作
- 批处理操作
- 存储设备操作
- 设置文件系统的提示方式

1.1.2 文件名参数输入规则

在设备上执行文件系统操作时，文件名参数的输入方式需要遵循 表 1-1。

表1-1 设备文件名参数输入规则（集中式设备）

操作	命令	长度	举例
file-name	纯文件名（只有文件名而没有路径），表示当前工作路径下的文件	1～91 个字符	a.cfg 表示当前目录下的 a.cfg 文件
path/*file-name*	文件夹＋纯文件名，表示当前路径指定文件夹下的指定文件。*path* 表示文件夹的名称，*path* 参数可以输入多次，表示多级文件夹下的文件	1～135 个字符	test/a.cfg 表示当前路径下 test 子文件夹下的 a.cfg 文件
drive:/[*path*]/*file-name*	存储介质＋文件夹＋纯文件名，表示设备上某块存储介质上的文件（存储介质的命名请参见 1.1.6 1. 存储设备的命名）。*drive*表示存储介质的名称，通常为flash或者cf 如果设备上只有一个存储介质，可以不用给出存储介质的信息；如果设备上有多个存储介质，需要给出存储介质的信息以确定是哪块存储介质上的文件	1～135 个字符	flash:/test/a.cfg 表示 Flash 根目录下 test 文件夹下的 a.cfg 文件

表1-2 设备文件名参数输入规则（分布式设备）

操作	命令	长度	举例
file-name	纯文件名（只有文件名而没有路径），表示当前工作路径下的文件	1～91 个字符	a.cfg 表示当前目录下的 a.cfg 文件，如果当前工作路径在主用主控板，则 a.cfg 表示主用主控板上的 a.cfg 文件；如果当前工作路径在备用主控板，则 a.cfg 表示备用主控板上的 a.cfg 文件
path/*file-name*	文件夹＋纯文件名，表示当前路径指定文件夹下的指定文件。*path* 表示文件夹的名称，*path* 参数可以输入多次，表示多级文件夹下的文件	1～135 个字符	test/a.cfg 表示当前路径下 test 子文件夹下的 a.cfg 文件
drive:/[*path*]/*file-name*	存储介质＋文件夹＋纯文件名，表示设备上某块存储介质上的文件（存储介质的命名请参见 1.1.6 1. 存储设备的命名）。*drive*表示存储介质的名称，通常为flash或者cf；备用主控板上的存储介质通常表示为slot1#flash或者slot1#cf（其中 1 表示备用主控板所在的槽位号，可以使用**display device**命令查看单板与槽位号的对应关系）	1～135 个字符	flash:/test/a.cfg 表示 Flash 根目录下 test 文件夹下的 a.cfg 文件 如果要读写备用主控板（编号为 1）上 Flash 根目录下的 a.cfg 文件时，命令行文件名参数需要输入 slot1#flash:/a.cfg

1.1.3 目录操作

目录操作包括创建/删除目录、显示当前的工作路径以及显示指定目录或文件信息等。

1. 显示目录信息

表1-3 显示目录信息

操作	命令	说明
显示目录或文件信息	**dir** [**/all**] [*file-url*]	必选 该命令在用户视图下执行

2. 显示当前的工作路径

表1-4 显示当前的工作路径

操作	命令	说明
显示当前的工作路径	**pwd**	必选 该命令在用户视图下执行

3. 改变当前的工作路径

表1-5 改变当前的工作路径

操作	命令	说明
改变当前的工作路径	**cd** { *directory* | **..** | */* }	必选 该命令在用户视图下执行

4. 创建目录

表1-6 创建目录

操作	命令	说明
创建目录	**mkdir** *directory*	必选 该命令在用户视图下执行

5. 删除目录

表1-7 删除目录

操作	命令	说明
删除目录	**rmdir** *directory*	必选 该命令在用户视图下执行



说明

- 被删除的目录必须为空目录（即删除目录前，必须先删除该目录下的所有文件及子目录。文件的删除请参见 **delete** 命令，子目录的删除请参见 **rmdir** 命令）。
- 成功执行 **rmdir** 后，回收站中原来属于该文件夹的文件会自动被彻底删除。

1.1.4 文件操作

文件操作包括显示指定目录或文件信息、显示文件的内容、重命名文件、拷贝文件、移动文件、删除文件、恢复删除的文件、彻底删除文件。



说明

创建文件可以通过拷贝、下载操作或 **save** 命令来辅助完成。

1. 显示文件信息

表1-8 显示文件信息

操作	命令	说明
显示目录或文件信息	**dir** [**/all**] [*file-url*]	必选 该命令在用户视图下执行

2. 显示文件内容

表1-9 显示文件内容

操作	命令	说明
显示文件的内容	**more** *file-url*	必选 目前只支持显示文本文件的内容 该命令在用户视图下执行

3. 重命名文件

表1-10 重命名文件

操作	命令	说明
重命名文件	**rename** *fileurl-source fileurl-dest*	必选 该命令在用户视图下执行

4. 拷贝文件

表1-11 拷贝文件

操作	命令	说明
拷贝文件	**copy** *fileurl-source fileurl-dest*	必选 该命令在用户视图下执行

5. 移动文件

表1-12 移动文件

操作	命令	说明
移动文件	**move** *fileurl-source fileurl-dest*	必选 该命令在用户视图下执行

6. 删除文件

表1-13 删除文件

操作	命令	说明
删除文件	**delete** [**/unreserved**] *file-url*	必选 该命令在用户视图下执行

注意

- 使用 **delete** *file-url* 命令删除文件，被删除的文件被保存在回收站中，仍会占用存储空间。如果用户经常使用该命令删除文件，则可能导致设备的存储空间不足，请用户查看回收站中是否有废弃文件。如果要彻底删除回收站中的某个废弃文件，必须在该文件的原目录下执行 **reset recycle-bin** 命令，才可以回收存储空间。
- 使用 **delete /unreserved** *file-url* 命令删除文件，被删除的文件被彻底删除，不能再恢复。效果等同于执行 **delete** *file-url* 命令后，再在同一目录下执行了 **reset recycle-bin** 命令。

7. 恢复删除文件

表1-14 恢复删除文件

操作	命令	说明
恢复删除文件	**undelete** *file-url*	必选 该命令在用户视图下执行

8. 彻底删除回收站中的文件

表1-15 彻底删除回收站中的文件

操作	命令	说明
进入要删除文件的原路径	**cd** { *directory* | **..** | **/** }	可选 如果要删除的文件的原路径不是当前路径，则该步骤必选 该命令在用户视图下执行
彻底删除回收站中的文件	**reset recycle-bin** [**/force**]	必选 该命令在用户视图下执行

1.1.5 批处理操作

批处理文件是可执行命令的集合，批处理功能用于自动执行批处理文件里的命令，执行过程相当于手工逐条执行这些命令。

批处理操作的推荐步骤：

(1) 在 PC 上编辑批处理文件；

(2) 将批处理文件下载到设备上，如果文件名的后缀不是.bat，必须使用 **rename** 命令修改为.bat；

(3) 执行批处理文件。

表1-16 批处理操作

操作	命令	说明
进入系统视图	**system-view**	-
执行批处理文件	**execute** *filename*	必选

注意

批处理命令不保证每一条命令的执行，比如某命令设置错误或者该命令执行的条件不成熟，系统就会跳过该命令转到下一条。

1.1.6 存储设备操作

1. 存储设备的命名

存储设备的命名遵循以下规则：

- 存储设备的名称由物理设备名称加分区序号组成。

- 存储设备的物理设备名称由存储设备类型加该类型设备的序号组成。设备的序号使用英文字母（如 a、b、c 等）描述。例如对于有多个 CF 卡的设备，则内置 CF 卡的名称为 cfa，外置 CF 卡命名为 cfb，依次类推。SR66 产品内置 CF 卡命名为 cfa，外置 CF 卡命名为 cfb；扁形的 USB 接口（丝印为“0”）命名为 usba。
- 如果设备支持存储设备分区，则分区设备的名称就是存储设备的物理设备名称加分区的序号组成。分区的序号使用数字（如 0、1、2 等）描述。如只有一个 CF 卡的设备上，CF 卡的第 2 个分区的名称为 cfa1；有多个 CF 卡的设备上，第二个 CF 卡的第 3 个分区的名称为 cfb2。如果只有一个分区，则分区的名称为存储设备的物理设备名称后加 0，如 usba0。

2. 管理存储设备的存储空间

由于异常操作等原因，存储设备的某些空间可能不可用，用户可以通过 **fixdisk** 命令来恢复存储设备的空间。如果使用 **format** 命令则会格式化指定的存储设备，该存储设备上的所有数据将被删除。

表1-17 管理存储设备的存储空间

操作	命令	说明
恢复存储设备的空间	**fixdisk** *device*	可选 该命令在用户视图下执行
格式化存储设备	**format** *device* [**FAT16** | **FAT32**]	可选 该命令在用户视图下执行

注意

- 格式化操作将导致存储设备上的所有文件丢失，并且不可恢复；尤其需要注意的是，如果存储设备上有启动配置文件，格式化该存储设备，将丢失启动配置文件。
- 对于分布式设备，只能对主用主控板上的存储设备进行 fixdisk 操作，不能对备用主控板上的存储设备进行 fixdisk 操作。
- 对于支持分区的设备，只能格式化分区设备，不能格式化整个存储设备。

3. 存储设备的挂载/卸载

对于可以支持热插拔的存储设备（如 CF 卡等），可以在用户视图下，使用 **mount** 和 **umount** 命令挂载和卸载该存储设备。当存储设备插入低版本的系统时，系统不能自动识别该存储设备，必须通过挂载操作后，才能对该存储设备执行读写操作。

卸载存储设备是逻辑上让存储设备处于非连接状态，此时，用户可以安全的拔出存储设备；挂载存储设备是让卸载的存储设备重新处于连接状态。

表1-18 存储设备的挂载/卸载

操作	命令	说明
挂载存储设备	**mount** *device*	可选 缺省情况下，存储设备连接到设备后，自动被挂载，处于挂载状态，即存储设备插入时已经处于连接状态，不需挂载就可使用

操作	命令	说明
卸载存储设备	**umount** *device*	可选 缺省情况下，存储设备连接到设备后，自动被挂载，处于挂载状态，在拔出存储设备前，需要先执行卸载操作

注意

- 在执行挂载或卸载操作过程中，禁止对单板或存储设备进行插拔或倒换操作；在进行文件操作过程中也禁止对存储设备进行插拔或倒换操作。否则，可能会引起文件系统的损坏。
- 处于挂载状态的存储设备在拔出系统前，请先执行卸载操作，以免损坏存储设备。
- 对于支持分区的设备，只能挂载/卸载分区设备，不能挂载/卸载整个存储设备。

1.1.7 设置文件系统的提示方式

用户可以通过命令修改当前文件系统的提示方式。文件系统支持两种提示方式：

alert：当用户对文件进行有危险性的操作时，系统会要求用户进行交互确认。

quiet：当用户对文件进行任何操作，系统均不要求用户进行确认。该方式可能会导致一些因误操作而发生的、不可恢复的、对系统造成破坏的情况产生。

表1-19 设置文件系统的提示方式

操作	命令	说明
进入系统视图	**system-view**	-
设置文件系统的提示方式	**file prompt { alert | quiet }**	可选 缺省情况下，文件系统的提示方式为 **alert**

1.1.8 文件系统应用举例

查看当前目录下的文件及子目录。

```
<Sysname> dir
Directory of cfa0:/
   0     -rw-  10944608  Jan 11 2030 23:47:54   sr6602.bin
   1     drw-         -  Apr 26 2000 12:00:02   logfile
   2     -rw-     95805  Jun 30 2007 10:00:28   startup.cfg
   3     -rw-  22745844  Jun 28 2007 20:19:50   sr6608.bin
   4     -rw-      2541  Jul 02 2007 13:53:16   jl.cfg
   5     -rw-  10930192  Apr 26 2000 12:12:38   6602pecl.bin
   6     -rw-      1827  Jan 12 2030 00:03:46   j1.cfg
506336 KB total (461312 KB free)
File system type of cfa0: FAT16
```

进入 logfile 目录，并创建新文件夹 mytest。

```
<Sysname> cd logfile
<Sysname> mkdir mytest
%Created dir cfa0:/logfile/mytest.
```

显示当前的工作路径。

```
<Sysname> pwd
cfa0:/logfile
```

查看 logfile 目录下的文件及子目录。

```
<Sysname> dir
Directory of cfa0:/logfile/
   0     -rw-   1250094  Jul 01 2007 16:09:50   logfile.log
   1     drw-         -  Jan 12 2030 00:35:20   mytest
506336 KB total (461296 KB free)
File system type of cfa0: FAT16
```

返回上一级目录。

```
<Sysname> cd ..
```

显示当前的工作路径。

```
<Sysname> pwd
cfa0:
```

1.2 配置文件管理

配置文件管理是设备提供的用于管理配置文件的一项功能。它具有较好的用户操作界面，方便用户对配置文件进行管理。

1.2.1 配置文件简介

配置文件以文本格式的命令行保存用户对设备进行的配置，记录用户的配置信息。通过配置文件，用户可以非常方便地查阅这些配置信息

1. 配置的类型

设备的配置按其作用的时间域分为两种：

- 起始配置：当设备启动时，根据读取的配置文件，进行初始化工作，该配置称为起始配置（startup-configuration）；如果设备中没有配置文件，则系统在启动过程中使用缺省参数进行初始化。
- 当前配置：与起始配置相对应，系统运行时，用户对设备进行的配置称为当前配置（current-configuration）。当前配置存放在设备的临时存储器中，设备重启之后会失效。

2. 配置文件的内容与格式

配置文件为一个文本文件，其保存规则如下：

- 配置文件的内容为命令，且只保存非缺省配置；
- 配置文件以命令视图为基本框架，同一命令视图的命令组织在一起，形成一节，节与节之间通常用空行或注释行隔开（以#开始的为注释行，空行或注释行可以是一行或多行）；
- 文件中各节的安排顺序通常为：系统配置、接口配置、各种协议配置和用户界面配置；
- 以 return 结束。

3. 多配置文件共存

存储介质上可以同时存放多个配置文件。用户可以将设备在多个使用环境下的配置保存成多个配置文件，当设备在网络中移动时，通过指定下次启动配置文件并重启设备，可以使设备以新的配置迅

速适应切换后的网络环境，节省重新配置的工作量。而且多配置文件可以方便用户的备份、修改、恢复等操作，提高设备的可靠性。

设备将使用多个配置文件中的某个文件来启动设备，但用户可以根据需要以及设备的支持情况配置两个下次启动配置文件，一个为主用（main）下次启动配置文件，一个为备用（backup）下次启动配置文件。设备重启后将使用主用下次启动配置文件来启动设备，当主用下次启动配置文件损坏或者丢失的时候，系统会自动使用备用下次启动配置文件来启动或配置设备。相对于只支持一个下次启动配置文件的设备（即不支持主备用下次启动配置文件的设备），这种双重保护的方式增强了设备的安全性和可靠性。

同一时刻，设备上最多只能有一个主用下次启动配置文件和一个备用下次启动配置文件，这两个文件也可以不指定（不指定时为 NULL），也可以将两个文件指定为同一个配置文件。

主用/备用下次启动配置文件的设置方式有两种：

- 保存当前配置时设置，具体配置请参见 1.2.2 保存当前配置。
- 设置下次启动配置文件时设置，具体配置请参见 1.2.4 设置下次启动配置文件。

4. 配置文件的选择顺序

系统启动时，配置文件的选择遵循以下规则：（不支持主用/备用下次启动配置文件的设备）

(1) 如果用户指定了启动配置文件，且配置文件存在，则以启动配置文件初始化配置。

(2) 如果用户指定的启动配置文件不存在，则以空配置启动。

当设备支持主备用下次启动配置文件，系统启动时，配置文件的选择遵循以下规则：（支持主用/备用下次启动配置文件的设备）

(1) 如果主用下次启动配置文件存在，则以主用下次启动配置文件初始化配置。

(2) 如果主用下次启动配置文件不存在，备用下次启动配置文件存在，则以备用下次启动配置文件初始化配置；如果备用下次启动配置文件也不存在，则以空配置启动（空配置启动指使用设备的出厂配置启动）。

1.2.2 保存当前配置

1. 保存当前配置配置任务简介

用户通过命令行可以修改设备的当前配置，而这些配置是暂时的，如果要使当前配置在系统下次重启时仍然有效，在重启设备前，需要将当前配置保存到配置文件中。

表1-20 保存当前配置配置任务简介

操作	说明	详细配置
使能配置文件同步保存功能	可选	2.
保存当前配置	必选	3.

2. 使能配置文件同步保存功能

- 使能配置文件同步保存功能后，当用户执行“**save [safely] [backup | main]**”或者“**save** *filename* **all**+回车”操作保存当前配置时，主用主控板和备用主控板会同时自动把当前配置保存到指定的配置文件中，并将该文件配置为设备的下次启动文件，以保证主、备板配置文件内容的一致；

- 如果没有使能配置文件同步保存功能，则当用户执行“**save [safely] [backup | main]**”或者“**save** *filename* **all**+回车”操作保存当前配置时，只有主用主控板自动把当前配置保存到指定的配置文件中，并将该文件配置为主用主控板的下次启动文件，备用主控板不会执行保存操作，也不会重新设置备用主控板的下次启动文件。

表1-21 使能配置文件同步保存功能

操作	命令	说明
进入系统视图	**system-view**	-
使能配置文件同步保存功能	**slave auto-update config**	可选 缺省情况下，配置文件同步保存功能处于使能状态


说明

执行“**save** *filename*+回车”操作，系统只按指定路径保存当前配置，备用板备用主控板不执行同步保存操作。

3. 保存当前配置

配置的保存方式有两种：

- 快速保存方式，执行不带 **safely** 参数的 **save** 命令。这种方式保存速度快，但是保存过程中如果出现设备重启、断电等问题，原有配置文件可能会丢失。
- 安全方式，执行带 **safely** 参数的 **save** 命令。这种方式保存速度慢，即使保存过程中出现设备重启、断电等问题，原有配置文件仍然会保存到设备中，不会丢失。

在电源稳定程度较好的环境中，推荐用户使用快速保存方式保存配置文件；在电源环境恶劣或者远程维护等情况，推荐用户使用安全模式保存配置文件。


注意

如果在配置文件保存过程中出现设备重启、断电等问题，可能会造成下次启动配置文件丢失，这时，设备将以空配置启动。设备启动后，用户需要重新设置下次启动文件（请参见“1.2.4 设置下次启动配置文件”）。

表1-22 保存当前配置（集中式设备）

操作	命令	说明
将当前配置保存到指定文件，但不会将该文件设置为下次启动配置文件	**save** *file-url*	二者必选其一 该命令可在任意视图下执行
将当前配置保存到存储介质的根目录下，并将该文件设置为下次启动配置文件	**save [safely] [backup | main]**	

表1-23 保存当前配置（分布式设备）

操作	命令	说明
将当前配置保存到指定文件，但不会将该文件设置为下次启动配置文件	**save** [*file-url* [**all** | **slot** *slot-number*]]	二者必选其一 该命令可在任意视图下执行
将当前配置保存到主用主控板和备用主控板存储介质的根目录，并将该文件设置为下次启动配置文件	**save** [**safely**] [**backup** | **main**]	

说明

- 配置文件名后缀必须为“.cfg”。
- 使用“**save** [**safely**] [**backup** | **main**]”或者“**save** *filename* **all**+回车”操作，是对主用主控板和备用主控板同时生效还是只对主用主控板生效，与配置文件的同步保存功能是否使能有关，详细介绍请参见“1.2.3 配置主备板配置文件同步保存功能”。
- 对于支持 **main** 参数的设备，**save** [**safely**]和 **save** [**safely**] **main** 命令执行效果相同：都会在保存当前配置的同时，将指定文件设置为主用下次启动配置文件。
- 执行**save** [**safely**] [**backup** | **main**]命令时，如果出现设备重启、断电等问题，可能会造成下次启动配置文件丢失，这时，设备将以空配置启动。设备启动后，用户需要重新设置下次启动配置文件（请参见“1.2.4 设置下次启动配置文件”）。

1.2.3 配置主备板配置文件同步保存功能

说明

本特性只有分布式设备支持。

对于分布式设备，用户只能在主用板上执行命令，不能在备用板上执行命令。开启主备板配置文件同步保存功能后，当用户执行“**save**+回车”或者“**save** *filename* **all**+回车”操作保存当前配置时，备用板会自动把当前的配置保存到自己的配置文件中，以保持主备板配置文件内容的一致。

表1-24 配置主备板配置文件同步保存功能

操作	命令	说明
进入系统视图	**system-view**	-
开启主备板配置文件同步保存功能	**slave auto-update config**	可选 缺省情况下，主备自动同步功能处于开启状态

说明

执行“**save** *filename*+回车”操作，系统只按指定路径保存当前配置，备用主控板不执行同步保存操作。

1.2.4 设置下次启动配置文件

下次启动配置文件是指设备下次启动时使用的配置文件。设置下次启动配置文件有两种方式：

- 使用 **save** 命令。将当前配置保存到指定配置文件时，使用交互方式，系统会自动把该文件设置为设备的下次启动配置文件（对支持主备用下次启动配置文件的设备，则设置为主用下次启动配置文件）。
- 使用专用命令，步骤请参见 表 1-25。

表1-25 设置下次启动时的配置文件

操作	命令	说明
配置下次启动时的配置文件名	**startup saved-configuration** *cfgfile* [**backup** | **main**]	必选 该命令在用户视图下执行

注意

配置文件必须以“.cfg”作为扩展名，启动配置文件必须存放在存储设备的根目录下

1.2.5 备份下次启动配置文件

备份下次启动配置文件特性用于将设备下次启动配置文件备份至 TFTP 服务器上。

对于支持主用/备用下次启动配置文件的设备，备份操作的对象是主用下次启动配置文件。

表1-26 备份下次启动配置文件

操作	命令	说明
将设备的下次启动配置文件备份到指定的 TFTP 服务器	**backup startup-configuration to** *dest-addr* [*dest-filename*]	必选 该命令在用户视图下执行

说明

在执行配置文件的备份操作前，请:

- 保证设备与服务器之间的路由可达，服务器端开启了 TFTP 服务，执行备份操作的客户端设备已获得了相应的读写权限。
- 在用户视图下使用 **display startup** 命令查看一下设备是否已经设置了下次启动配置文件，并使用 **dir** 命令查看所设置的下次启动配置文件是否存在。若下次启动配置文件设置为 NULL，或者所设置的配置文件不存在，备份操作将会失败。

1.2.6 删除设备中的下次启动配置文件

用户通过命令可以删除设备中的下次启动配置文件。对于支持主备用下次启动配置文件的设备，用户可以只删除主用下次启动配置文件，或者只删除备用下次启动配置文件，但如果当前设备的主备用下次启动配置文件相同，仅执行一次删除操作，系统只会将相应的下次启动配置文件设置为 NULL，但不会删除该文件。

出现以下情况时，用户可能需要删除设备中的下次启动配置文件：

- 在设备软件升级之后，系统软件和配置文件不匹配。
- 设备中的配置文件被破坏（常见原因是加载了错误的配置文件）。

下次启动配置文件被删除后，设备下次上电时，系统将采用空配置进行初始化。

表1-27 删除设备中的下次启动配置文件擦除设备中配置文件

操作	命令	说明
删除设备中的下次启动配置文件	**reset saved-configuration [backup | main]**	必选 该命令在用户视图下执行

注意

- 本特性会将下次启动配置文件从设备上彻底删除，所以请谨慎使用。（集中式设备）
- 本特性会将配置文件从主用主控板和备用主控板上彻底删除，所以请慎用该命令。（分布式设备）

1.2.7 恢复下次启动配置文件

恢复下次启动配置文件特性用于将 TFTP 服务器上保存的配置文件下载到设备并设置为下次启动配置文件。（集中式设备）

恢复下次启动配置文件特性用于将 TFTP 服务器上保存的指定配置文件下载到主用主控板和备用主控板存储介质的根目录下，并设置为主用主控板和备用主控板的下次启动配置文件。（分布式设备）

表1-28 恢复设备的下次启动配置文件

操作	命令	说明
恢复下次启动配置文件	**restore startup-configuration from** *src-addr src-filename*	必选 该命令在用户视图下执行

说明

- 对于支持主备用下次启动配置文件的设备，恢复操作的对象是主用下次启动配置文件。
- 在执行配置文件的恢复操作前，请保证设备与服务器之间的路由可达，服务器端开启了 TFTP 服务，执行恢复操作的客户端设备已获得了相应的读写权限。
- 该命令执行成功后，用户可以在用户视图下使用 **display startup** 命令查看设备下次启动配置文件名是否与 *filename* 参数保持一致，并使用 **dir** 命令查看恢复的下次启动配置文件是否存在。

1.2.8 设备配置显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示设备的当前配置和起始配置情况。用户可以通过查看显示信息验证配置的效果。

表1-29 设备配置显示和维护

操作	命令
查看设备存储介质中保存的当前正在使用的配置文件的内容	**display saved-configuration [by-linenum]**
显示用于本次及下次启动的配置文件名	**display startup**
查看当前视图下生效的配置	**display this [by-linenum]**

操作	命令
查看当前生效的配置	**display current-configuration** [[**configuration** [*configuration*] | **controller** | **interface** [*interface-type*] [*interface-number*]] [**by-linenum**] [**|** { **begin** | **include** | **exclude** } *text*]]



display this 和 **display current-configuration** 命令的详细介绍请参见“系统分册”中的“系统基本配置命令”。

2 FTP 配置

2.1 FTP 简介

2.1.1 FTP 概述

FTP（File Transfer Protocol，文件传输协议）用于在远端服务器和本地主机之间传输文件，是 IP 网络上传输文件的通用协议。在万维网（WWW，World Wide Web）出现以前，用户使用命令行方式传输文件，最通用的应用程序就是 FTP。虽然目前大多数用户在通常情况下选择使用 Email 和 Web 传输文件，但是 FTP 仍然有着比较广泛的应用。

FTP 协议在 TCP/IP 协议族中属于应用层协议，用于在远端服务器和本地客户端之间传输文件，使用 TCP 端口 20 和 21 进行传输。端口 20 用于传输数据，端口 21 用于传输控制消息。FTP 协议基本操作在 RFC959 中进行了描述。

FTP 有两种文件传输模式：

- 二进制模式，用于传输程序文件（比如后缀名为.app、.bin 和.btm 的文件）；
- ASCII 码模式，用于传输文本格式的文件（比如后缀名为.txt、.bat 和.cfg 的文件）。

2.1.2 FTP 的实现

设备在支持 FTP 协议上有两种方式：

- 设备作为 FTP 客户端：用户在 PC 上通过终端仿真程序或 Telnet 程序连接到设备（设备作为 FTP 客户端），执行 **ftp** 命令，建立设备与远程 FTP 服务器的连接，访问远程 FTP 服务器上的文件。
- 设备作为 FTP 服务器：用户运行 FTP 客户端程序，作为 FTP 客户端，登录到设备上进行访问（用户登录前，网络管理员需要事先配置好 FTP 服务器的 IP 地址）。

图2-1 FTP 配置示意图



设备作为 FTP 客户端时，需要进行如下配置：

表2-1 设备作为 FTP 客户端时的配置

设备	操作	说明
Device（作为 FTP 客户端）	可以直接使用 **ftp** 命令登录远端的 FTP 服务器	如果远端 FTP 服务器支持匿名访问，设备可以直接登录；如果远端 FTP 服务器不支持匿名访问，则必须先获取 FTP 用户名和密码后，才能成功登录远端的 FTP 服务器
PC（作为 FTP 服务器）	启动 FTP 服务器，并作了用户名、密码、用户的权限等相关的配置	-

设备作为 FTP 服务器时，需要进行如下配置：

表2-2 设备作为 FTP 服务器时的配置

设备	操作	说明
Device（作为 FTP 服务器）	启动 FTP 服务器功能	缺省情况下，系统关闭 FTP 服务器功能 可以通过 **display ftp-server** 命令查看设备上 FTP 服务器功能的配置信息
	配置 FTP 服务器的验证和授权	配置 FTP 用户的用户名、密码、授权的工作目录。出于安全考虑，设备不支持匿名登录，用户必须使用合法的用户名和密码。缺省情况下，合法用户可访问的目录为设备的根目录
	配置 FTP 服务器的运行参数	配置 FTP 连接的超时时间等参数
PC（作为 FTP 客户端）	使用 FTP 客户端程序登录设备	用户必须获取 FTP 用户名和密码后，才能成功登录远端的 FTP 服务器

注意

- FTP 功能可以正常使用的条件是 FTP 服务器与 FTP 客户端之间路由可达。
- 当设备作为 FTP 服务器，使用 IE 登录设备时，因为 IE 在登录过程中建立了多个用户连接，而目前设备某一时刻只支持一个用户连接，所以 IE 登录设备时 FTP 的部分功能不支持。

2.2 配置 FTP 客户端

2.2.1 建立 FTP 连接

FTP 客户端要访问 FTP 服务器，必须先与 FTP 服务器建立连接。连接的建立方式有两种，一种是使用 **ftp** 命令直接建立连接，一种是在 FTP 客户端视图下使用 **open** 命令间接建立连接。

源地址绑定的原理是先在设备上状态比较稳定的接口（建议使用 Loopback 接口或 Dailer 接口）上配置好 IP 地址，然后使用该 IP 地址作为 FTP 连接的源 IP 地址。该功能简化了 ACL 规则和安全策略的配置，只要将 ACL 规则的源地址或者目的地址参数指定为该地址，就可以屏蔽接口 IP 地址的差异以及接口状态的影响，实现对设备进出报文的过滤。源地址可以通过配置源接口或源 IP 来实现，源接口下配置的主 IP 地址即为发送报文的源地址。FTP 客户端在与 FTP 服务器通信的时候，发送报文的源地址选取遵循以下规则：

- 如果没有指定 FTP 客户端的源地址，则采用路由决定的源地址进行通信。
- 如果只用 **ftp client source** 或 **ftp** 命令指定了源地址，则采用该地址进行通信。
- 如果执行 **ftp client source** 命令指定了源地址后，又在 **ftp** 命令中指定了源地址，则采用 **ftp** 命令中指定的源地址进行通信。

ftp client source 命令指定的源地址对所有的 **ftp** 连接有效，**ftp** 命令指定的源地址只对当前的 FTP 连接有效。

表2-3 建立 FTP 连接（IPv4 组网环境）

操作	命令	说明
进入系统视图	**system-view**	-
配置 FTP 客户端的源地址	**ftp client source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }	可选 缺省情况下，设备使用路由决定的源地址与 FTP 服务器通信
退回用户视图	**quit**	-
在用户视图下直接登录远程 FTP 服务器	**ftp** [*server-address* [*service-port*] [**source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }]]	二者必选其一 **ftp** 命令直接在用户视图下执行；**open** 命令在 FTP 客户端视图下执行
在 FTP 客户端视图下间接登录远程 FTP 服务器	**ftp**	
	open *server-address* [*service-port*]	

说明

- 如果源接口下没有配置主地址，将导致 FTP 连接建立失败。
- 如果先后使用 **ftp client source** 命令配置了客户端 FTP 报文的源接口和源 IP，则新配置的源 IP 将覆盖现有的源接口配置。反之亦然。

表2-4 建立 FTP 连接（IPv6 组网环境）

操作	命令	说明
在用户视图下直接登录远程 FTP 服务器	**ftp ipv6** [*server-address* [*service-port*] [**source ipv6** *source-ipv6-address*] [**-i** *interface-type interface-number*]]	二者必选其一 **ftp ipv6** 命令直接在用户视图下执行；**open ipv6** 命令在 FTP 客户端视图下执行
在 FTP 客户端视图下间接登录远程 FTP 服务器	**ftp ipv6**	
	open ipv6 *server-address* [*service-port*] [**-i** *interface-type interface-number*]	

2.2.2 FTP 客户端基本操作

当设备作为FTP客户端，与FTP服务器连接建立成功后（连接操作请参见 2.2.1 建立FTP连接），对于已经获得授权的目录，设备可以进行如下操作。

表2-5 FTP 客户端基本操作

操作	命令	说明
显示远程 FTP 服务器支持的 FTP 相关协议命令的帮助信息	**remotehelp** [*protocol-command*]	可选
使能显示 FTP 服务器返回的详细信息	**verbose**	可选 缺省情况下，verbose 开关为开启状态
当设备作为 FTP 客户端时，打开 FTP 调试信息开关	**debugging**	可选 缺省情况下，FTP 客户端调试信息开关处于关闭状态

操作	命令	说明
成功登录 FTP 服务器后，使用其他用户身份重新登录	**user** *username* [*password*]	可选
设置 FTP 文件传输的模式为 ASCII 模式	**ascii**	可选 缺省情况下，文件传输模式为 ASCII 模式
设置 FTP 文件传输的模式为二进制流模式	**binary**	可选 缺省情况下，文件传输模式为 ASCII 模式
切换远程 FTP 服务器上的工作路径	**cd** { *directory* | **..** | **/** }	可选
退出远程 FTP 服务器的当前目录，返回 FTP 服务器的上一级目录	**cdup**	可选
查看远程 FTP 服务器上的目录/文件的详细信息	**dir** [*remotefile* [*localfile*]]	可选
查询远程 FTP 服务器上的目录/文件	**ls** [*remotefile* [*localfile*]]	可选
下载 FTP 服务器上的文件	**get** *remotefile* [*localfile*]	可选
上传本地文件到远程 FTP 服务器	**put** *localfile* [*remotefile*]	可选
显示当前用户正在访问的远程 FTP 服务器上的路径	**pwd**	可选
获取 FTP 客户端本地的工作路径	**lcd**	可选
在远程 FTP 服务器上创建目录	**mkdir** *directory*	可选
设置数据传输的方式为被动方式	**passive**	可选 缺省情况下，数据传输的方式为被动方式
彻底删除远程 FTP 服务器上的指定文件	**delete** *remotefile*	可选
删除 FTP 服务器上指定的目录	**rmdir** *directory*	可选
不退出 FTP 客户端视图的前提下，断开与 FTP 服务器的连接	**disconnect**	可选 等效于 **close** 命令
不退出 FTP 客户端视图的前提下，断开与 FTP 服务器的连接	**close**	可选 等效于 **disconnect** 命令
断开与远程 FTP 服务器的连接，并退回到用户视图	**bye**	可选
终止与远程 FTP 服务器的连接，并退回到用户视图	**quit**	可选 在 FTP 客户端视图下执行，等效于 **bye** 命令

说明

- FTP 传输文件有两种模式：一种是 ASCII 码模式（可用 **ascii** 命令设置），用于传输文本文件；另一种是二进制模式（可用 **binary** 命令设置），用于传输程序文件。
- **ls** 命令只能显示出目录/文件的名称，而 **dir** 命令可以查看与目录/文件相关的信息，如大小，创建日期等。
- 当用户通过设备登录FTP服务器时，只有级别为 3（管理级）的用户登录设备后才能使用 表 2-5 所列的命令，但命令能否执行成功，还将受FTP服务器端的授权限制。

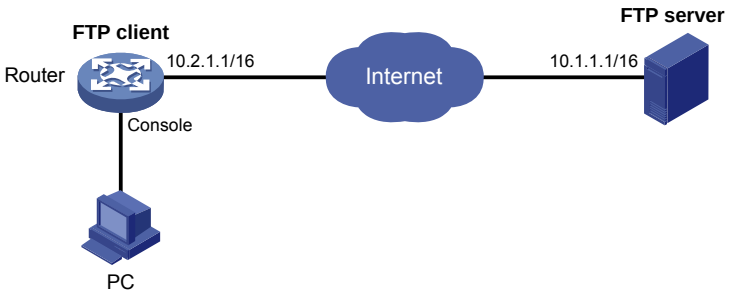
2.2.3 FTP 客户端配置举例

1. 组网需求

- 路由器作为 FTP 客户端，从 FTP 服务器上下载新的启动文件，并将客户端的配置文件上传到服务器进行备份。
- FTP 服务器的 IP 地址为 10.1.1.1/16。
- 服务器上已设置路由器登录 FTP 服务器的用户名为 abc，密码为 pwd。路由器以用户名 abc、密码 pwd 登录 FTP 服务器。
- PC 通过 Console 口对路由器进行操作。

2. 组网图

图2-2 利用 FTP 客户端功能实现平滑升级



3. 配置步骤

注意

如果设备剩余的内存空间不够，请使用 **fixdisk** 命令清理内存或者使用 **delete /unreserved** *file-url* 命令删除部分暂时不用的文件后再执行以下操作。

删除路由器中的多余文件，以保证剩余足够的空间，用于存储新的启动文件。

```
<Sysname> dir
Directory of cfa0:/

   0    drw-           -  Dec 07 2005 10:00:57   filename
   1    drw-           -  Jan 02 2006 14:27:51   logfile
   2    -rw-        1216  Jan 02 2006 14:28:59   config.cfg
```

```
   3   -rw-      1216  Jan 02 2006 16:27:26   backup.cfg

2540 KB total (2511 KB free)
<Sysname> delete /unreserved cfa0:/backup.cfg
```

以 FTP 方式登录服务器，获取系统启动文件。

```
<Sysname> ftp 10.1.1.1
Trying 10.1.1.1 ...
Connected to 10.1.1.1.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(10.1.1.1:(none)):abc
331 Give me your password, please
Password:
230 Logged in successfully
```

将启动文件 aaa.bin 从 PC 下载到设备。

```
[ftp] binary
200 Type set to I.
```

将启动文件 aaa.bin 从 PC 下载到设备，并重命名为 bbb.bin。

```
[ftp] get aaa.bin bbb.bin


227 Entering Passive Mode (10,1,1,1,4,1).
125 BINARY mode data connection already open, transfer starting for /aaa.bin.
.....226 Transfer complete.
FTP: 5805100 byte(s) received in 19.898 second(s) 291.74K byte(s)/sec.
```

将本机的配置文件 config.cfg 上传到服务器进行备份。

```
[ftp] ascii
[ftp] put config.cfg back-config.cfg
227 Entering Passive Mode (10,1,1,1,4,2).
125 ASCII mode data connection already open, transfer starting for /config.cfg.
226 Transfer complete.
FTP: 3494 byte(s) sent in 5.646 second(s), 618.00 byte(s)/sec.
[ftp] bye
```

将 bbb.bin 指定为设备的主用下次启动文件。

```
<Sysname> boot-loader file bbb.bin main
```

重启设备，完成设备软件升级。

```
<Sysname> reboot
```

 注意

下次启动文件必须存放在存储介质的根目录下（对于支持存储设备分区的设备，该文件必须存放在第一个分区内）。可使用文件的拷贝或移动操作来调整文件的路径为根目录。关于 **boot-loader** 命令的详细介绍请参见“系统分册”中的“设备管理命令”。

2.3 配置 FTP 服务器

当设备作为 FTP 服务器时，可以进行如下配置。

2.3.1 FTP 服务器的基本配置

用户登录到 FTP 服务器，使用 **put** 命令上传文件的过程中，FTP 服务器中文件的更新有两种方式，快速更新方式与普通更新方式。

- 快速更新方式，即 FTP 服务器先将上传文件接收到内存中，全部接收完后，再将文件内容写入存储设备。采用这种方式，即使文件传送过程发生断电等异常情况，也不会损坏 FTP 服务器上的现有文件。
- 普通更新方式，即 FTP 服务器一边接收用户的文件，一边将其写入存储设备。采用这种方式，可能会因为断电等异常情况致使 FTP 服务器上的现有文件被损坏。与快速升级方式相比，普通升级方式需要的空闲内存较少。

表2-6 配置 FTP 服务器

操作	命令	说明
进入系统视图	**system-view**	-
启动 FTP 服务器功能	**ftp server enable**	必选 缺省情况下，FTP 服务器功能处于关闭状态
使用 ACL（Access Control List，访问控制列表）限制哪些 FTP 客户端可以访问设备	**ftp server acl** *acl-number*	可选 缺省情况下，没有使用 ACL 限制 FTP 客户端
配置 FTP 服务器的连接空闲时间	**ftp timeout** *minutes*	可选 缺省情况下，连接空闲时间为 30 分钟 如果在连接空闲时间内，FTP 服务器和客户端没有消息交互，则断开它们之间的连接
设置在上传过程中，FTP 服务器更新文件的方式	**ftp update { fast | normal }**	可选 缺省情况下，在给 FTP 服务器上传文件的过程中，FTP 服务器采用 **normal** 方式更新文件
退回到用户视图	**quit**	-
强制释放通过指定用户名建立的 FTP 连接	**free ftp user** *username*	可选 该命令在用户视图下执行

2.3.2 配置 FTP 服务器的验证和授权

FTP 服务器的授权信息包含提供给 FTP 用户的工作目录的路径。只有验证通过和授权成功的用户，才能得到 FTP 服务器的服务。

表2-7 配置 FTP 服务器的验证和授权

操作	命令	说明
进入系统视图	**system-view**	-
进入本地用户视图	**local-user** *user-name*	必选 如果指定的本地用户不存在，则先完成本地用户的创建，再进入该用户的视图 缺省情况下，系统中没有任何本地用户，系统不支持 FTP 匿名用户访问
设置当前本地用户的密码	**password** { **simple** | **cipher** } *password*	必选
设置用户可以使用的服务类型为 FTP	**service-type ftp**	必选 缺省情况下，系统不支持 FTP 匿名用户访问，不对用户授权任何服务；若授权 FTP 服务，缺省授权使用设备的根目录
配置用户的属性	**authentication-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	可选 缺省情况下，FTP/SFTP 用户可以访问设备的根目录，用户的级别为 0。可以使用该命令进行修改

说明

- **local-user**、**password**、**service-type ftp** 和 **authorization-attribute** 命令的详细介绍，请参见“安全分册”中的“AAA 命令”。
- 设备作为 FTP 服务器时，对客户端的访问操作有级别限制：如果要对设备的文件系统执行写操作（比如上传、删除、创建/删除文件夹），则必须将 FTP 登录用户的级别设置为 3；如果执行其它操作（比如普通的查看操作），则 FTP 登录用户的级别不受限制，可以为 0～3 中的任意级别。

2.3.3 FTP 服务器配置举例

1. 组网需求

- 路由器作为 FTP 服务器，PC 作为 FTP 客户端。IP 地址如组网图所示，路由器和 PC 之间路由可达。
- 设备最新版本的启动文件存储在 PC 上，请使用 FTP 功能实现设备的升级以及设备配置文件的备份。
- 客户端登录 FTP 服务器的用户名为 abc，密码为 pwd。

2. 组网图

图2-3 利用 FTP 服务器功能实现升级



3. 配置步骤

(1) 配置 Router（FTP Server）

在路由器上添加一个本地用户 abc，并设置其认证密码为 pwd，用户级别为 3 管理级。

```
<Sysname> system-view
[Sysname] local-user abc
[Sysname-luser-abc] password simple pwd
[Sysname-luser-abc] level 3
```

设置授权访问目录，可以使用的服务类型为 FTP。

```
[Sysname-luser-ftp] authorization-attribute work-directory cfa0:/
[Sysname-luser-abc] service-type ftp
[Sysname-luser-abc] quit
```

启动路由器的 FTP 服务。

```
[Sysname] ftp server enable
[Sysname] quit
```

删除路由器中的多余文件，以保证剩余足够的空间，用于存储新的启动文件。

```
<Sysname> dir
Directory of cfa0:/

   0   drw-         -  Dec 07 2005 10:00:57   filename
   1   drw-         -  Jan 02 2006 14:27:51   logfile
   2   -rw-      1216  Jan 02 2006 14:28:59   config.cfg
   3   -rw-      1216  Jan 02 2006 16:27:26   back.cfg

2540 KB total (2511 KB free)
<Sysname> delete /unreserved cfa0:/back.cfg
```

(2) 配置 PC（FTP Client）

以 FTP 方式登录 FTP 服务器。

```
c:\> ftp 1.1.1.1
Connected to 1.1.1.1.
220 FTP service ready.
User(1.1.1.1:(none)):abc
331 Password required for abc.
Password:
230 User logged in.
```

将路由器的配置文件 config.cfg 下载到 PC 本地进行备份。

```
ftp> get config.cfg back-config.cfg
```

上传启动文件 newest.app 到 Device。

```
ftp> put newest.bin
ftp> bye
```

说明

- 利用 FTP 功能升级配置文件时，操作步骤与上述介绍基本相同，需要注意的是获取后的配置文件同样要放在存储介质的根目录下（对于支持存储设备分区的设备，该文件必须存放在第一个分区内）。
- 利用 FTP 远程升级 Bootrom 程序，文件传送完成后再执行 **bootrom update** 命令来更新升级 Bootrom。

(3) 升级路由器

将 newest.app 指定为设备的主用下次启动文件。

```
<Sysname> boot-loader file newest.bin main
```

重启设备，完成设备软件升级。

```
<Sysname> reboot
```

注意

下次启动文件必须存放在存储介质的根目录下（对于支持存储设备分区的设备，该文件必须存放在第一个分区内）。可使用文件的拷贝或移动操作来调整文件的路径为根目录。关于 **boot-loader** 命令的详细介绍请参见“系统分册”中的“设备管理命令”。

2.4 FTP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 FTP 的运行情况，通过查看显示信息验证配置的效果。

表2-8 FTP 显示和维护

操作	命令
查看 FTP 客户端的当前配置	**display ftp client configuration**
查看 FTP 服务器的配置情况	**display ftp-server**
查看 FTP 登录用户的详细情况	**display ftp-user**

3 TFTP 配置

3.1 TFTP 简介

3.1.1 TFTP 概述

TFTP（Trivial File Transfer Protocol，简单文件传输协议）也是用于在远端服务器和本地主机之间传输文件的，相对于 FTP，TFTP 没有复杂的交互存取接口和认证控制，适用于客户端和服务器之间不需要复杂交互的环境。TFTP 协议的运行基于 UDP 协议，使用 UDP 端口 69 进行数据传输。TFTP 协议基本操作在 RFC1986 中进行了描述。。

TFTP 协议传输是由客户端发起的：

- 当 TFTP 客户端需要从服务器下载文件时，由客户端向 TFTP 服务器发送读请求包，然后从服务器接收数据，并向服务器发送确认；
- 当 TFTP 客户端需要向服务器上传文件时，由客户端向 TFTP 服务器发送写请求包，然后向服务器发送数据，并接收服务器的确认。

TFTP 传输文件有两种模式：

- 二进制模式，用于传输程序文件（比如后缀名为.app、.bin 和.btm 的文件）。
- ASCII 码模式，用于传输文本格式的文件（比如后缀名为.txt、.bat 和.cfg 的文件）。

3.1.2 TFTP 的实现

目前，设备只能作为 TFTP 客户端，不支持作为 TFTP 服务器。

图3-1 TFTP 配置示意图



使用 TFTP 之前，网络管理员需要配置好 TFTP 客户端和服务器的 IP 地址，并且确保客户端和服务器之间的路由可达。

设备作为 TFTP 客户端时，需要进行如下配置：

表3-1 设备作为 TFTP 客户端时的配置

设备	操作	说明
Device（作为 TFTP 客户端）	• 进行 IP 地址和路由功能配置， 确保设备和 TFTP 服务器之间路由可达 • 直接使用 TFTP 命令登录远端的 TFTP 服务器上传或者下载文件	-
PC（作为 TFTP 服务器）	启动 TFTP 服务器，并配置 TFTP 工作目录	-

3.2 配置 TFTP 客户端

当设备作为 TFTP 客户端时，可以把本地设备的文件上传到 TFTP 服务器，还可以从 TFTP 服务器下载文件到本地设备。下载又分为两种：

- 普通下载。在这种方式下，设备将获取的远端文件直接写到存储设备中。这样如果是覆盖性下载系统文件，而且下载失败（如网络断开等原因），则原系统文件已被删除，设备将无法正常启动。
- 安全下载。在这种方式下，设备将获取的远端文件先保存到内存中，等用户文件全部接收完毕，才将它写到存储设备中。这样如果系统文件下载失败（如网络断开等原因），因为原有的系统文件没有被覆盖，设备仍能够启动。这种方法安全系数较高，但需要较大的内存空间。

当操作启动文件或配置文件等重要文件时，建议采用安全下载或者非覆盖性下载（即下载时使用一个当前目录下不存在的文件名作为目标文件名）。

源地址绑定的原理是先在设备上状态比较稳定的接口（建议使用 Loopback 接口或 Dailer 接口）上配置好 IP 地址，然后使用该 IP 地址作为 TFTP 连接的源 IP 地址。该功能简化了 ACL 规则和安全策略的配置，只要将 ACL 规则的源地址或者目的地址参数指定为该地址，就可以屏蔽接口 IP 地址的差异以及接口状态的影响，实现对设备进出报文的过滤。源地址可以通过配置源接口或源 IP 来实现，源接口下配置的主 IP 地址即为发送报文的源地址。TFTP 客户端在与 TFTP 服务器通信的时候，发送报文的源地址选取遵循以下规则：

- 如果没有指定 TFTP 客户端的源地址，则采用路由决定的源地址进行通信。
- 如果只用 **tftp client source** 或 **tftp** 命令指定了源地址，则采用该地址进行通信。
- 如果执行 **tftp client source** 命令指定了源地址后，又在 **tftp** 命令中指定了源地址，则采用 **tftp** 命令中指定的源地址进行通信。

tftp client source 命令指定的源地址对所有的 **tftp** 传输有效，**tftp** 命令指定的源地址只对当前的 **tftp** 传输有效。

表3-2 配置 TFTP 客户端

操作	命令	说明
进入系统视图	**system-view**	-
使用 ACL 限制设备可访问哪些 TFTP 服务器	**tftp-server** [**ipv6**] **acl** *acl-number*	可选 缺省情况下，没有使用 ACL 限制 TFTP 服务器
配置 TFTP 客户端的源地址	**tftp client source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }	可选 缺省情况下，设备使用路由决定的源地址与 TFTP 服务器通信
退回用户视图	**quit**	-
在 IPv4 网络，用 TFTP 下载/上传/安全下载文件	**tftp** *server-address* { **get** | **put** | **sget** } *source-filename* [*destination-filename*] [**source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }]	可选 该命令在用户视图下执行

操作	命令	说明
在 IPv6 网络，用 TFTP 下载/上传文件	**tftp ipv6** *tftp-ipv6-server* [**-i** *interface-type interface-number*] { **get** | **put** } *source-file* [*destination-file*]	可选 该命令在用户视图下执行

说明

- 如果源接口下没有配置主地址，将导致 TFTP 传输失败。
- 如果先后配置了客户端 TFTP 报文的源接口和源 IP，则新配置的源 IP 将覆盖现有的源接口配置。反之亦然。

3.3 TFTP 客户端显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 TFTP 客户端的情况，通过查看显示信息验证配置的效果。

表3-3 TFTP 客户端显示和维护

操作	命令
查看 TFTP 客户端的当前配置	**display tftp client configuration**

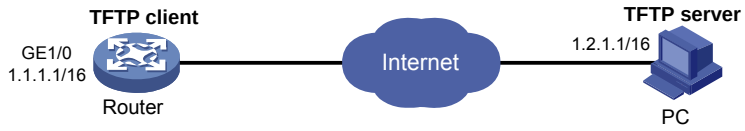
3.4 TFTP 客户端配置举例

1. 组网需求

- 路由器作为 TFTP 客户端，PC 作为 TFTP 服务器。IP 地址如组网图所示，Device 和 PC 之间路由可达。
- 路由器通过 TFTP 协议从 PC 上下载最新启动文件完成设备升级，同时将配置文件 config.cfg 上传到 PC 实现备份。

2. 组网图

图3-2 利用 TFTP 客户端功能实现平滑升级



3. 配置步骤

(1) 配置 PC（TFTP 服务器）（具体配置步骤略）

- 在 PC 上启动 TFTP 服务器功能；
- 配置 TFTP 服务器的工作目录。

(2) 配置路由器（TFTP 客户端）

注意

如果路由器剩余的内存空间不够,请使用 **fixdisk** 命令清理内存或者使用 **delete /unreserved** *file-url* 命令删除部分暂时不用的文件后再执行以下操作。

进入系统视图。

```
<Sysname> system-view
```

将启动文件 aaa.bin 从 PC 下载到路由器，并重命名为 bbb.bin。

```
<Sysname> tftp 1.2.1.1 get aaa.bin bbb.bin
```

将路由器的配置文件 config.cfg 上传到 TFTP 服务器。

```
<Sysname> tftp 1.2.1.1 put config.cfg configback.cfg
```

将 newest.app 指定为设备的主用下次启动文件。

```
<Sysname> boot-loader file bbb.bin main
```

重启设备，完成设备软件升级。

```
<Sysname> reboot
```

注意

下次启动文件必须存放在存储介质的根目录下（对于支持存储设备分区的设备，该文件必须存放在第一个分区内）。可使用文件的拷贝或移动操作来调整文件的路径为根目录。关于 **boot-loader** 命令的详细介绍请参见“系统分册”中的“设备管理命令”。

目 录

1 系统基本配置 1-1
1.1 显示设备的配置 1-1
1.2 系统基本配置 1-1
1.2.1 进入系统视图 1-1
1.2.2 退出当前视图 1-1
1.2.3 退回用户视图 1-2
1.2.4 配置设备名称 1-2
1.2.5 配置系统时间 1-2
1.2.6 使能/禁止显示版权信息 1-4
1.2.7 配置欢迎信息 1-5
1.2.8 配置命令行的快捷键 1-6
1.2.9 配置用户级别和命令级别 1-7
1.2.10 配置并发配置用户数 1-12
1.2.11 系统基本配置显示和维护 1-12
1.3 命令行特性 1-13
1.3.1 命令行接口简介 1-13
1.3.2 命令行在线帮助 1-14
1.3.3 命令行同步信息输出功能 1-15
1.3.4 命令行的undo格式 1-15
1.3.5 命令行编辑功能 1-15
1.3.6 命令行显示 1-16
1.3.7 命令行的历史记录功能 1-18
1.3.8 命令行错误信息 1-19

1 系统基本配置

1.1 显示设备的配置

为了了解设备的当前配置情况，避免重复配置以及配置冲突的情况发生，在进行设备的配置之前可以使用 **display** 命令查看设备的配置。设备的配置大体可以分成如下两种：

- 当前配置：设备上正在运行的配置。
- 保存配置：保存在当前启动配置文件中的配置。

表1-1 显示设备的配置

操作	命令	说明
显示设备当前生效的配置	**display current-configuration** [[**configuration** [*configuration*] | **controller** | **interface** [*interface-type*] [*interface-number*]] [**by-linenum**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]]	该命令在任意视图下执行
显示设备存储介质中保存的配置	**display saved-configuration** [**by-linenum**]	该命令在任意视图下执行

说明

display saved-configuration 命令的详细介绍请参见“系统分册”中的“文件系统管理命令”。

1.2 系统基本配置

1.2.1 进入系统视图

当用户登录到设备后，会自动进入用户视图，此时屏幕显示的提示符是：<*设备名*>。用户视图可执行的操作有限，主要是查看操作、文件操作、FTP 和 Telnet 操作等，需要进入系统视图，才能进一步对设备进行配置。

表1-2 进入或退出系统视图

操作	命令	说明
从用户视图进入系统视图	**system-view**	必选 该命令在用户视图下执行

1.2.2 退出当前视图

系统将命令行接口划分为若干个命令视图，命令视图采用分层结构，比如用户视图下有系统视图，系统视图下又有接口视图、VLAN 视图等。当前视图下的功能配置完成，需要退出当前视图时，可以进行如下操作。

表1-3 退出当前视图

操作	命令	说明
从当前视图返回上一级视图	**quit**	必选 如果当前视图是用户视图，执行该操作会中断用户终端与设备之间的当前连接 该命令可在任意视图下执行

1.2.3 退回用户视图

本特性为用户提供了一种快捷的、从任意的非用户视图返回到用户视图的方式，而不需要多次执行**quit**命令。用户也可以直接按组合键<Ctrl+Z>从当前视图退回到用户视图。

表1-4 退回用户视图

操作	命令	说明
退回用户视图	**return**	必选 该命令可在任意的非用户视图下执行

1.2.4 配置设备名称

设备名称用于在网络中标志某台设备，在系统内部，设备名称对应于命令行接口的提示符，如设备的名称为 Sysname，则用户视图的提示符为<Sysname>。

表1-5 配置设备名称

操作	命令	说明
进入系统视图	**system-view**	-
设置设备名称	**sysname** *sysname*	可选

1.2.5 配置系统时间

1. 系统时间的配置

系统时间是系统信息时间戳显示的时间。该时间由配置的相对时间、时区和夏令时三个参数运算之后联合决定，通过**display clock**命令可以查看。为了保证与其他设备协调工作，用户需要将系统时间配置准确。

表1-6 配置系统时间

操作	命令	说明
配置时间和日期	**clock datetime** *time date*	可选 该命令在用户视图下执行
进入系统视图	**system-view**	-
配置系统所在的时区	**clock timezone** *zone-name* { **add** | **minus** } *zone-offset*	可选

操作	命令	说明
配置夏令时	**clock summer-time** *zone-name* **one-off** *start-time start-date end-time end-date add-time*	二者必选其一
	clock summer-time *zone-name* **repeating** *start-time start-date end-time end-date add-time*	

2. 系统时间的显示

系统时间由**clock datetime**、**clock timezone**和**clock summer-time**三条命令联合决定。如果以上三条命令都不配置，则**display clock**命令显示的为原系统时间。如果把三条以上命令任意组合进行配置，配置后的系统时间请参见 表 1-7。表中配置列各参数的含义为：

- 1：表示执行 **clock datetime** 命令配置了时间 *date-time*；
- 2：表示执行 **clock timezone** 命令配置了时区参数，时间偏移量为 *zone-offset*；
- 3：表示执行 **clock summer-time** 命令配置了夏令时参数，时间偏移量为 *summer-offset*；
- [1]：表示 **clock datetime** 命令是可选配置，可执行也可不执行。

表中举例默认原系统时间为 2005/1/1 1:00:00。

表1-7 系统时间配置和显示关系表

配置	display clock 显示的时间	举例
1	*date-time*	配置：clock datetime 1:00 2007/1/1 显示：01:00:00 UTC Mon 01/01/2007
2	原系统时间±"*zone-offset*"	配置：clock timezone zone-time add 1 显示：02:00:00 zone-time Sat 01/01/2005
1、2	"*date-time*"±"*zone-offset*"	配置：clock datetime 2:00 2007/2/2 和 clock timezone zone-time add 1 显示：03:00:00 zone-time Fri 02/02/2007
[1]、2、1	*date-time*	配置：clock timezone zone-time add 1 和 clock datetime 3:00 2007/3/3 显示：03:00:00 zone-time Sat 03/03/2007
3	原系统时间不在夏令时段内，显示：原系统时间	配置：clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 显示：01:00:00 UTC Sat 01/01/2005
	原系统时间在夏令时段内，显示：原系统时间+"*summer-offset*"	配置：clock summer-time ss one-off 00:30 2005/1/1 1:00 2005/8/8 2 显示：03:00:00 ss Sat 01/01/2005
1、3	date-time 不在夏令时段内，显示：date-time	配置：clock datetime 1:00 2007/1/1 和 clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 显示：01:00:00 UTC Mon 01/01/2007
	date-time 在夏令时段内，显示："date-time"+"summer-offset"	配置：clock datetime 8:00 2007/1/1 和 clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 显示：10:00:00 ss Mon 01/01/2007

配置	display clock 显示的时间	举例
[1]、3、1	date-time 不在夏令时段内，显示：date-time	配置：clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 和 clock datetime 1:00 2008/1/1 显示：01:00:00 UTC Tue 01/01/2008
	date-time 在夏令时段内，再根据“date-time”-”summer-offset”的值是否在夏令时段内来判断。如果该值不在夏令时段内，则显示“date-time”-”summer-offset”；如果在夏令时段内，显示 date-time	配置：clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 和 clock datetime 1:30 2007/1/1 显示：23:30:00 UTC Sun 12/31/2006
		配置：clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 和 clock datetime 3:00 2007/1/1 显示：03:00:00 ss Mon 01/01/2007
2、3 或者 3、2	如果原系统时间±"*zone-offset*"的值不在夏令时段内，则显示：原系统时间±"zone-offset"	配置：clock timezone zone-time add 1 和 clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 显示：02:00:00 zone-time Sat 01/01/2005
	如果原系统时间±"*zone-offset*"的值在夏令时段内，显示原系统时间±"zone-offset"+”summer-offset”	配置：clock timezone zone-time add 1 和 clock summer-time ss one-off 1:00 2005/1/1 1:00 2005/8/8 2 显示：04:00:00 ss Sat 01/01/2005
1、2、3 或者 1、3、2	如果"*date-time*"±"*zone-offset*"的值不在夏令时段内，则显示："date-time"±"zone-offset"	配置：clock datetime 1:00 2007/1/1、clock timezone zone-time add 1 和 clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 显示：02:00:00 zone-time Mon 01/01/2007
	如果"*date-time*"±"*zone-offset*"的值在夏令时段内，显示"date-time"±"zone-offset"+”summer-offset”	配置：clock datetime 1:00 2007/1/1、clock timezone zone-time add 1 和 clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 显示：04:00:00 ss Mon 01/01/2007
[1]、2、3、1 或者 [1]、3、2、1	date-time 不在夏令时段内，显示：date-time	配置：clock timezone zone-time add 1、clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 和 clock datetime 1:00 2007/1/1 显示：01:00:00 zone-time Mon 01/01/2007
	date-time 在夏令时段内，再根据“date-time”-”summer-offset”的值是否在夏令时段内来判断。如果不在夏令时段内，则显示“date-time”-”summer-offset”；如果在夏令时段内，显示 date-time	配置：clock timezone zone-time add 1、clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 和 clock datetime 1:30 2008/1/1 显示：23:30:00 zone-time Mon 12/31/2007
		配置：clock timezone zone-time add 1、clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 和 clock datetime 3:00 2008/1/1 显示：03:00:00 ss Tue 01/01/2008

1.2.6 使能/禁止显示版权信息

- 使能显示版权信息后，使用 Telnet 或 SSH 方式登录设备时会显示版权信息，使用 Console 口、AUX 口或异步串口方式登录设备再退出用户视图时会显示版权信息，其它情况不显示版权信息。显示的版权信息形如：

```
**************************************************************************
* Copyright (c) 2004-2008 Hangzhou H3C Tech. Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                             *
* no decompiling or reverse-engineering shall be allowed.                *
**************************************************************************
```

- 禁止显示版权信息后，在任何情况下都不会显示版权信息。

表1-8 使能/禁止显示版权信息

操作	命令	说明
进入系统视图	**system-view**	-
使能显示版权信息	**copyright-info enable**	可选 缺省情况下，显示版权信息处于使能状态
禁止显示版权信息	**undo copyright-info enable**	可选 缺省情况下，显示版权信息处于使能状态

1.2.7 配置欢迎信息

1. 欢迎信息简介

欢迎信息是用户在连接到设备、进行登录验证以及开始交互配置时系统显示的一段提示信息。管理员可以根据需要，设置相应的提示信息。

目前，系统支持的欢迎信息有五种：

- shell 欢迎信息，也称 session 条幅。进入控制台会话时显示；
- 用户接口欢迎信息，也称 incoming 条幅。主要用于 Modem 激活用户接口时显示；
- 登录欢迎信息，也称 login 条幅。主要用于配置密码验证和 scheme 认证时显示；
- MOTD（Message Of The Day）欢迎信息，在启动认证前显示；
- 授权欢迎信息，也称 legal 条幅。系统在用户登录前会给出一些版权或者授权信息，然后显示 legal 条幅，并等待用户确认是否继续进行认证或者登录。如果用户输入“Y”或者直接按<Enter>键，则进入认证或登录过程；如果输入“N”，则退出认证或登录过程。“Y”和“N”不区分大小写。

2. 配置欢迎信息

配置欢迎信息时，系统支持两种输入方式：一种方式为欢迎信息的所有内容紧接在命令关键字后面输入，输入内容的开始字符和结束字符必须相同，但这两字符不作为欢迎信息的内容，此时包括命令关键字、开始和结束字符在内总共可以输入 510 个字符；另一种方式为欢迎信息的所有内容通过按<Enter>键分多行输入，此时可以输入长达 2000 个字符。

多行输入又分为三种方式：

- 命令关键字后直接回车（按<Enter>键）：以“%”结束设置，“回车”及“%”符号不属于欢迎信息的内容。
- 命令关键字后输入一个字符后回车：以首行输入的字符结束设置，首行的字符及结束符不属于欢迎信息的内容。
- 命令关键字后输入多个字符（首尾不相同）后回车：以首行的第一个字符结束设置。首行的第一个字符及结束符不属于欢迎信息的内容。

表1-9 配置欢迎信息

操作	命令	说明
进入系统视图	**system-view**	-
配置登录进入用户视图时的欢迎信息（Modem 登录方式适用）	**header incoming** *text*	可选
配置登录验证时的欢迎信息	**header login** *text*	可选
配置登录终端界面前的授权信息	**header legal** *text*	可选
配置登录进入用户视图时的欢迎信息（非 Modem 登录方式适用）	**header shell** *text*	可选
配置登录终端界面前的欢迎信息	**header motd** *text*	可选

1.2.8 配置命令行的快捷键

为便于用户对常用命令进行快捷操作，系统提供了五个快捷键供用户自定义。只要用户按下相应的快捷键，系统即可执行对应的命令。

表1-10 配置命令行的快捷键

操作	命令	说明
进入系统视图	**system-view**	-
配置命令行的快捷键	**hotkey** { **CTRL_G** | **CTRL_L** | **CTRL_O** | **CTRL_T** | **CTRL_U** } *command*	可选 缺省情况请参见表下方的说明
显示系统中快捷键的分配信息	**display hotkey**	可在任意视图下执行，系统保留的快捷键请参见 表 1-11

说明

缺省情况下，系统为<Ctrl+G>、<Ctrl+L>、<Ctrl+O>三个快捷键指定了对应的命令，其它两个快捷键<Ctrl+T>、<Ctrl+U>缺省值为 NULL（空）。

- <Ctrl+G>对应命令 **display current-configuration**（显示当前配置）；
- <Ctrl+L>对应命令 **display ip routing-table**（显示 IPv4 路由表信息）；
- <Ctrl+O>对应命令 **undo debugging all**（关闭所有模块的调试信息开关）。

表1-11 系统保留的快捷键

快捷键	功能
<Ctrl+A>	将光标移动到当前行的开头
<Ctrl+B>	将光标向左移动一个字符
<Ctrl+C>	停止当前正在执行的功能
<Ctrl+D>	删除当前光标所在位置的字符
<Ctrl+E>	将光标移动到当前行的末尾
<Ctrl+F>	将光标向右移动一个字符
<Ctrl+H>	删除光标左侧的一个字符

快捷键	功能
<Ctrl+K>	终止呼出的连接
<Ctrl+N>	显示历史命令缓冲区中的后一条命令
<Ctrl+P>	显示历史命令缓冲区中的前一条命令
<Ctrl+R>	重新显示当前行信息
<Ctrl+V>	粘贴剪贴板的内容
<Ctrl+W>	删除光标左侧连续字符串内的所有字符
<Ctrl+X>	删除光标左侧所有的字符
<Ctrl+Y>	删除光标右侧所有的字符
<Ctrl+Z>	退回到用户视图
<Ctrl+]>	终止呼入的连接或重定向连接
<Esc+B>	将光标移动到左侧连续字符串的首字符处
<Esc+D>	删除光标所在位置及其右侧连续字符串内的所有字符
<Esc+F>	将光标向右移到下一个连续字符串之前
<Esc+N>	将光标向下移动一行（键入回车前有效）
<Esc+P>	将光标向上移动一行（键入回车前有效）
<Esc+<>	将光标所在位置指定为剪贴板的开始位置
<Esc+>>	将光标所在位置指定为剪贴板的结束位置

说明

以上快捷键为设备所定义，当用户使用终端软件与设备进行交互时，这些快捷键可能在终端软件中被定义为其他指令。此时，快捷键的操作优先遵从终端软件的定义，而不会对设备生效。

1.2.9 配置用户级别和命令级别

1. 级别简介

为了限制不同用户对设备的访问权限，系统对用户进行了分级管理。用户的级别与命令级别对应，不同级别的用户登录后，只能使用等于或低于自己级别的命令。

命令的级别由低到高分为访问级、监控级、系统级和管理级四种，分别对应级别值 0、1、2、3。详细介绍请见 表 1-12：

表1-12 命令级别简介

级别值	级别名称	描述
0	访问级	用于网络诊断等功能的命令、从本设备出发访问外部设备的命令。该级别命令配置后不允许保存，设备重启后，该级别命令会恢复到缺省状态 缺省情况下，访问级的命令包括：**ping**、**tracert**、**telnet**、**ssh2** 等
1	监控级	用于系统维护、业务故障诊断等功能的命令。该级别命令配置后不允许保存，设备重启后，该级别命令会恢复到缺省状态 缺省情况下，监控级的命令包括：**debugging**、**terminal**、**refresh**、**reset**、**send** 等

级别值	级别名称	描述
2	系统级	业务配置命令，包括路由、各个网络层次的命令，这些命令用于向用户提供直接网络服务 缺省情况下，系统级的命令包括：所有配置命令（管理级的命令除外）
3	管理级	关系到系统的基本运行、系统支撑模块功能的命令，这些命令对业务提供支撑作用 缺省情况下，管理级的命令包括：文件系统命令、FTP 命令、TFTP 命令、XModem 命令下载、用户管理命令、级别设置命令、系统内部参数设置命令（非协议规定、非 RFC 规定）等

2. 配置用户级别

用户级别可以通过 AAA 认证参数或者用户界面来配置：

(1) 通过 AAA 认证参数配置用户级别

如果用户登录时使用的用户界面的认证方式为 **sheme**，并且用户登录时需要输入用户名和密码，则用户级别以及用户可使用的命令，在配置 AAA 认证时指定。

表1-13 通过 AAA 认证参数配置用户级别

操作		命令	说明
进入系统视图		**system-view**	-
进入用户界面视图		**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **tty** | **vty** } *first-num2* [*last-num2*] }	-
设置登录用户界面的认证方式为 **scheme**		**authentication-mode scheme** [**command-authorization**]	必选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console、TTY 用户界面不需要认证
退回系统视图		**quit**	-
配置 SSH 用户的认证方式为 password 方式		相关内容请参见“安全分册”中的“SSH2.0 配置”	如果用户使用 SSH 方式登录，并且认证时要求输入用户名和密码，则该步骤必选
通过认证参数设置用户的级别	使用本地认证时	• 使用 **local-user** 命令创建本地用户并进入本地用户视图 • 使用 **authorization-attribute** 命令的 **level** 参数配置用户本身的级别	二者必选其一 使用本地认证时，如果没有配置用户级别，则用户级别为 0，即只能使用 0 级别的命令 使用远程认证时，如果没有配置用户级别，则用户级别由认证服务器的缺省配置决定
	使用远程认证（RADIUS 认证、HWTACACS 认证及 LDAP 认证）时	在认证服务器上进行配置	

说明

- 关于用户界面的介绍请参见“系统分册”中的“用户界面配置”。有关 **user-interface**、**authentication-mode**、**user privilege level** 命令的介绍请参见“系统分册”中的“用户界面命令”。
- 有关 AAA 认证的介绍请参见“安全分册”中的“AAA 配置”。**local-user** 和 **authorization-attribute** 命令的介绍请参见“安全分册”中的“AAA 命令”。
- 有关 SSH 的介绍请参见“安全分册”中的“SSH2.0 配置”。

(2) 通过 AAA 认证参数配置用户级别举例

设置使用 VTY 1 的 Telnet 用户登录设备时，需要本地验证用户名和口令，用户级别为 3。

```
<Sysname> system-view
[Sysname] user-interface vty 1
[Sysname-ui-vty1] authentication-mode scheme
[Sysname-ui-vty1] quit
[Sysname] local-user test
[Sysname-luser-test] password cipher 123
[Sysname-luser-test] service-type telnet
```

通过以上配置，用户使用 VTY 1 Telnet 登录设备时，需要输入用户名 test，密码 123，认证通过后只能使用级别为 0 的命令，想要使用级别为 0、1、2、3 的命令还需要配置：

```
[Sysname-luser-test] authorization-attribute level 3
```

(3) 通过用户界面配置用户级别

如果用户登录时使用的用户界面的认证方式为 **scheme**，而且是 SSH 的 publickey 认证方式时（该方式只需要输入用户名，不需要输入密码），则用户级别等于用户界面的级别；如果用户登录时使用的用户界面的认证方式为 **none** 或者 **password**（即不需要输入用户名），用户级别也等于用户界面的级别。

表1-14 通过用户界面配置用户级别（SSH 的 publickey 认证方式）

操作	命令	说明
配置 SSH 用户的认证方式为 publickey 方式	相关内容请参见“安全分册”中的“SSH2.0 配置”	如果用户使用 SSH 方式登录，并且认证时只要输入用户名，不用输入密码，则该步骤必选 配置该步骤后，相应的用户界面的认证方式必须设置为 **scheme**
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | **vty** *first-num2* [*last-num2*] }	-
设置用户使用当前用户界面登录设备时的认证方式	**authentication-mode scheme** [**command-authorization**]	可选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console、TTY 用户界面不需要认证
配置从当前用户界面登录系统的用户的级别	**user privilege level** *level*	可选 缺省情况下，通过 Console 口登录系统的用户级别是 3，通过其它用户界面登录系统的用户级别是 0

表1-15 通过用户界面配置用户级别（none 或者 password 认证方式）

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **tty** | **vty** } *first-num2* [*last-num2*] }	-
设置用户使用当前用户界面登录设备时的认证方式	**authentication-mode** { **none** | **password** }	可选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console、TTY 用户界面不需要认证
配置从当前用户界面登录系统的用户的级别	**user privilege level** *level*	可选 缺省情况下，通过 Console 口登录系统的用户级别是 3，通过其它用户界面登录系统的用户级别是 0

(4) 通过用户界面配置用户级别举例

- 设置所有的 Telnet 用户登录设备时，不需要身份认证，用户级别为 1。（该配置可用于实验室环境，不建议对网络上的设备进行如下配置，因为存在安全隐患）

```
<Sysname> system-view
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode none
[Sysname-ui-vty0-4] user privilege level 1
```

缺省情况下，用户使用 Telnet 方式登录设备，通过口令验证后，只能使用以下命令：

```
<Sysname> ?
User view commands:
  display  Display current system information
  ping     Ping function
  quit     Exit from current command view
  rsh      Establish one RSH connection
  ssh2     Establish a secure shell client connection
  super    Set the current user priority level
  telnet   Establish one TELNET connection
  tftp     Open TFTP connection
  tracert  Trace route function
```

通过用户界面设置级别后，用户不需要认证就可以 Telnet 登录设备并使用以下命令：

```
<Sysname> ?
User view commands:
  debugging      Enable system debugging functions
  dialer         Dialer disconnect
  display        Display current system information
  ping           Ping function
  quit           Exit from current command view
  refresh        Do soft reset
  reset          Reset operation
  rsh            Establish one RSH connection
  screen-length  Specify the lines displayed on one screen
```

```
send            Send information to other user terminal interface
ssh2            Establish a secure shell client connection
super           Set the current user priority level
telnet          Establish one TELNET connection
terminal        Set the terminal line characteristics
tftp            Open TFTP connection
tracert         Trace route function
undo            Cancel current setting
```

- 设置所有的 Telnet 用户登录设备时，需要验证口令，用户级别为 2。

```
<Sysname> system-view
[Sysname] user-interface vty 0 4
[Sysname-ui-vty1] authentication-mode password
[Sysname-ui-vty0-4] set authentication password cipher 123
[Sysname-ui-vty0-4] user privilege level 2
```

缺省情况下，用户使用 Telnet 方式登录设备，通过口令验证后，只能使用级别为 0 的命令。通过用户界面设置级别后，用户使用 Telnet 方式登录设备，输入密码 123 后，就可以使用级别为 0、1、2 的命令。

3. 切换用户级别

切换用户级别是指在不退出当前登录、不断开当前连接的前提下暂时的修改用户级别，级别修改后不需要重新登录、不需要重新认证，可以继续配置设备，只是可以执行的配置会不一样，比如用户的级别为 3，可以对系统参数进行设置，如果将用户的级别切换到 0，则只能执行简单的 **ping**、**tracert** 和很少一部分 **display** 命令等。切换后的级别是临时的，只对当前登录生效，用户重新登录后，又会恢复到原有级别。

为了防止对设备的误操作，通常情况下建议管理员使用较低级别的用户登录设备、查看设备运行参数，当需要对设备进行维护时，再临时切换到较高的级别；当管理员需要暂时离开设备或者将设备暂时交给其他人代为管理时，为了安全起见，可以临时切换到较低的级别，来限制其他人员的操作。

从高级别切换到低级别时，不要输入级别切换密码；从低级别切换到高级别时，为了保证操作的安全性，只有使用 Console 用户界面登录的用户不需要输入切换密码，使用 AUX、TTY 或 VTY 用户界面登录的用户均需要输入级别切换密码。该密码不同于用户登录认证时输入的密码，它只用于临时级别的切换。如果输入的密码错误或者没有配置级别切换密码，切换操作将会失败。因此，在进行切换操作前，请先配置级别切换密码。

表1-16 修改用户的级别

操作	命令	说明
进入系统视图	**system-view**	-
配置切换用户级别的密码	**super password** [**level** *user-level*] { **simple** | **cipher** } *password*	必选 缺省情况下，没有设置切换用户级别的密码
退回到用户视图	**quit**	-
切换用户级别	**super** [*level*]	必选 用户的登录设备时，已经具有了一定的级别，该级别由用户界面或者认证用户级别决定

注意

- 在使用 **super password** 命令时，若不指定用户级别，则配置的是切换到 3 级用户的密码。
- 用户级别的切换密码的显示方式有两种，建议用户使用密文（**cipher**）形式，明文（**simple**）密码容易被盗取。

4. 修改命令级别

缺省情况，各个视图下的每条命令都有指定的级别（如 表 1-17所示）。管理员也可以根据用户需要改变命令的级别：实现低级别用户可以使用部分高级别命令的需求；或者将命令的级别提高，增加设备的安全性。

表1-17 修改命令的级别

操作	命令	说明
进入系统视图	**system-view**	-
设置指定视图下的命令的级别	**command-privilege level** *level* **view** *view command*	可选 缺省情况请参见 表 1-12

注意

通常情况下，建议用户不要修改缺省的命令级别或者在专业人员的指导下进行修改，以免造成操作和维护上的不便甚至给设备带来安全隐患。

1.2.10 配置并发配置用户数

用户可以使用 **configure-user count** 命令来设置可以同时进入系统视图的用户数。

表1-18 配置并发配置用户数

操作	命令	说明
进入系统视图	**system-view**	-
配置并发配置用户数	**configure-user count** *number*	可选 缺省情况下，允许两个用户在系统视图下进行配置

说明

- 当多个用户同时进入系统视图并对某个特性进行配置的时候，系统以最新配置为准。
- 如果当前并发配置用户数目已达到所设置的限制数目，新登录的用户将无法再进入系统视图。

1.2.11 系统基本配置显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以查看系统相应的配置信息和运行信息。

表1-19 系统基本配置显示和维护

操作	命令
显示系统版本信息	**display version**
显示系统当前的时间和日期	**display clock**
显示终端用户的信息	**display users [all]**
显示已经登录设备且当前命令视图不是用户视图的用户的信息	**display configure-user**
显示当前视图下生效的配置	**display this [by-linenum]**
显示剪贴板的内容	**display clipboard**
显示或保存系统当前多个功能模块运行的统计信息	**display diagnostic-information**

在日常维护或系统出现故障时，为了便于问题定位，用户需要查看各个功能模块的运行信息。因为各个功能模块都有其对应的运行信息，所以一般情况下，用户需要逐条运行相应的 **display** 命令。为便于一次性收集更多信息，用户可以在任意视图下执行 **display diagnostic-information** 命令，显示或保存系统当前多个功能模块运行的统计信息。**display diagnostic-information** 命令一次性收集的信息等效于依次执行 **display clock**、**display version**、**display device**、**display current-configuration** 等命令后终端显示的信息。

说明

- **display users** 命令的详细介绍请参见“系统分册”中的“用户界面命令”。
- 以上只介绍了与全局有关的 **display** 命令，有关各协议和各种接口的 **display** 命令请参见对应章节。

1.3 命令行特性

1.3.1 命令行接口简介

命令行接口是设备与用户之间的交互界面。通过命令行接口，用户可以输入命令对设备进行配置，并可以通过查看输出的信息确认配置结果，方便用户配置和管理设备。

为提高设备的可管理性和可操作性，命令行接口还提供了如下的特性：

- 支持对命令进行分级保护，未授权用户不能使用高于其等级的命令对设备进行配置，详情请参见“1.2.9 配置用户级别和命令级别”；
- 用户随时可以键入“?”以获得在线帮助；
- 提供种类丰富、内容详尽的调试信息，帮助用户诊断、定位网络故障；
- 提供命令历史记录功能，用户可以方便地查看曾经执行过的命令，并再次执行；
- 支持不完整关键字输入。在执行命令时，用户可以输入关键字的部分字符而无须输入关键字的全部字符，但为了让用户进一步确认即将执行的操作，用户输入的字符必须能精确匹配一条命令才会执行（如果第一个字符匹配不了，就需要输入第二个，以此类推）。比如以以 s 开头的命令有 save、startup saved-configuration、system-view 等，如果要保存当前配置，至少要输入 sa；如果要设置下次启动文件，至少要输入 st s；如果要进入系统视图，至少要

输入 sy，而不能只输入 s。关键字的全部字符可以按<Tab>键由系统自动补全，也可以手工全部输入。

1.3.2 命令行在线帮助

命令行接口提供如下几种在线帮助：

- 完全帮助
- 部分帮助

通过上述各种在线帮助能够获取到帮助信息，分别描述如下：

(1) 在任意视图下，键入<?>获取该视图下所有的命令及其简单描述。

```
<Sysname> ?
User view commands:
  backup              Backup next startup-configuration file to TFTP server
  boot-loader         Set boot loader
  bootrom             Update/read/backup/restore bootrom
  cd                  Change current directory
  clock               Specify the system clock
  cluster             Run cluster command
  copy                Copy from one file to another
  debugging           Enable system debugging functions
  delete              Delete a file
  dir                 List files on a file system
  display             Show running system information
```

……略……

(2) 键入一条命令，后接以空格分隔的<?>，如果该位置为关键字，则列出全部关键字及其简单描述。

```
<Sysname> terminal ?
  debugging  Send debug information to terminal
  logging    Send log information to terminal
  monitor    Send information output to current terminal
  trapping   Send trap information to terminal
```

(3) 键入一命令，后接以空格分隔的<?>，如果该位置为参数，则列出有关的参数描述。

```
<Sysname> system-view
[Sysname] interface vlan-interface ?
  <1-4094>  VLAN interface number
[Sysname] interface vlan-interface 1 ?
  <cr>
[Sysname] interface vlan-interface 1
```

<cr>表示该位置无参数，在紧接着的下一个命令行该命令被复述，直接键入回车即可执行。

(4) 键入字符串，其后紧接<?>，列出以该字符串开头的所有命令。

```
<Sysname> c?
   cd
   clock
   copy
```

(5) 键入命令，后接字符串，字符串紧接<?>，列出命令以该字符串开头的所有关键字。

```
<Sysname> display cl?
  clipboard
  clock
```

(6) 键入命令的某个关键字的前几个字母，按下<Tab>键，如果以输入字母开头的关键字唯一，则可以显示出完整的关键字；如果不唯一，则显示最先匹配到的完整的关键字（匹配规则为：将下一个字母按字母序排列，排在最前面的最先被匹配），反复按<Tab>键，则可以循环显示所有以输入字母开头的关键字，用户可以从中选择所需要的关键字。

1.3.3 命令行同步信息输出功能

同步信息输出是指当用户的输入被系统信息输出打断时，系统将回显用户的上一步输入或操作，用户可以接着输入或者执行上一步的操作。

使用 **info-center synchronous** 命令可以开启同步信息输出功能，关于该功能的详细介绍请参见“系统分册”中的“信息中心配置”。

1.3.4 命令行的 undo 格式

在命令前加 **undo** 关键字，即为命令的 undo 形式。几乎每条配置命令都有对应的 undo 形式，**undo** 命令一般用来恢复缺省情况、禁用某个功能或者删除某项设置。例如，**info-center enable** 命令用来开启信息中心功能；**undo info-center enable** 命令用来关闭信息中心功能。（缺省情况下，信息中心处于开启状态。）

1.3.5 命令行编辑功能

命令行接口提供了基本的命令编辑功能，支持多行编辑（每行达到指定字符时，系统会自动换行。不能按<Enter>键换行，输入<Enter>键后，系统会自动执行该命令），每条命令的最大长度为 510 个字符，如 表 1-20所示。

表1-20 编辑功能表

按键	功能
普通按键	若编辑缓冲区未满，则插入到当前光标位置，并向右移动光标
退格键<Backspace>	删除光标位置的前一个字符，光标前移
左光标键←或<Ctrl+B>	光标向左移动一个字符位置
右光标键→或<Ctrl+F>	光标向右移动一个字符位置
上光标键↑或<Ctrl+P>	显示历史命令
下光标键↓或<Ctrl+N>	
<Tab>键	输入不完整的关键字后按下<Tab>键，系统自动执行部分帮助： ● 如果与之匹配的关键字唯一，则系统用此完整的关键字替代原输入并换行显示； ● 如果与之匹配的关键字不唯一，则反复按<Tab>键，可以循环显示所有以输入字母开头的关键字； ● 如果没有与之匹配的关键字，系统会不作任何修改，重新换行显示原输入

说明

编辑命令行时，除了 表 1-20中的快捷键，用户还可以使用其它快捷键（详细介绍请参见 表 1-11）或者自定义快捷键（详细介绍请参见 1.2.8 配置命令行的快捷键”）。

1.3.6 命令行显示

使用过滤显示功能可以帮助用户迅速查找到自己关心的信息。如果显示的信息较多，一屏不能全部显示时，系统会自动进行分屏显示。分屏显示的时候也可以使用过滤显示功能提取出用户关心的信息。

1. 过滤显示

设备提供了过滤显示功能，用户通过指定正则表达式（即指定显示规则），可以帮助迅速查找所需的信息。

过滤显示的使用方法有两种：

- 在命令行中通过输入 **begin**、**exclude** 或 **include** 关键字加正则表达式的方式来过滤显示；
- 在分屏显示时，使用“/”、“-”或“＋”符号加正则表达式的方式，它将对剩余的信息使用正则表达式进行过滤显示。其中，“/”等同关键字 **begin**；“-”等同关键字 **exclude**；“＋”等同关键字 **include**。

begin、**exclude** 或 **include** 关键字的含义如下：

- **begin**：显示特定行及其以后的所有行，该特定行必须包含指定正则表达式。
- **exclude**：显示不包含指定正则表达式的所有行。
- **include**：只显示包含指定正则表达式的所有行。

正则表达式为 1～256 个字符的字符串，区分大小写，它还支持多种特殊字符，特殊字符的匹配规则如 表 1-21所示。

表1-21 正则表达式中的特殊字符

特殊字符	含义	使用说明
^string	行首匹配符，string 只能出现在每行的开始	如：^user 只能匹配以 user 开始的行，不能匹配以 Auser 开始的行
string $	行尾匹配符，string 只能出现在每行的末尾	如： user$只能匹配以 user 结尾的行，不能匹配以 userA 结尾的行
.	句点，通配符，匹配任何一个字符，包括单个字符、特殊字符和空格等	如：.l 可以配置 vlan 和 mpls 等
*	星号，匹配星号前面的字符或字符组零次或多次	如：zo*可以匹配 z 以及 zoo；(zo)*可以匹配 zo 以及 zozo
+	加号，匹配加号前面的字符或字符组一次或多次	如：zo+可以匹配 zo 以及 zoo，但不能匹配 z
|	竖线，匹配|左边的整个字符串或者右边的整个字符串	如：def|int 只能匹配包含 def 或者 int 的字符串
_	下划线，该字符出现在表达式的开头或结尾时，等效于行首匹配符或行尾匹配符（即特殊字符^或$），其他情况下等效于逗号、空格或者作为普通字符时的左括号、右括号、左大括号、右大括号	如：a_b 可以匹配 a b 和 a(b 等；_ab 只能匹配以 ab 开头的行；ab_只能匹配以 ab 结束的行
-	连接符，用于连接两个数值或字母（小的在前，大的在后），与“[]”符号连用表示一个范围	如：从 1 到 9 表示为 1-9（包括 1 和 9）；从 a 到 h 表示为 a-h（包括 a 和 h）

特殊字符	含义	使用说明
[ ]	表示字符选择范围，将以选择范围内的单个字符为条件进行匹配，只要字符串里包含该范围的某个字符就能匹配到	如：[16A]表示可以匹配到的字符串只需要包含 1、6 或 A 中任意一个；[1-36A] 表示可以匹配到的字符串只需要包含 1、2、3、6 或 A 中任意一个（-为连接符） 表示如果]需要作为普通字符出现在[ ]内时，必须把]写在[ ]的最前面，形如[]string]，才能匹配到]。[没有这样的限制
( )	表示字符组，一般与“+”或“*”等符号一起使用	如：(123A)表示字符组 123A；408(12)+可以匹配 40812 或 408121212 等字符串，但不能匹配 408
\index	表示重复一次指定字符组，字符组是指\前用()括起来的字符串，index 对应\前字符组的顺序号（按从左至右的顺序从 1 开始编号）：如果/前面只有一个字符组，则 index 只能为 1；如果/前面有 n 个字符组，则 index 可以为 1 到 n 中的任意整数	如：(string)\1 表示把 string 重复一次，匹配的字符串必须包含 stringstring；(string1)(string2)\2 表示把 string2 重复一次，匹配的字符串必须包含 string1string2string2；(string1)(string2)\1\2 表示先把 string1 重复一次，再重复一次 string2，匹配的字符串必须包含 string1string2string1string2
[^]	表示选择范围外的字符，将以单个字符为条件进行匹配，只要字符串里包含该范围外的某个字符就能匹配到	如：[^16A]表示可匹配的字符串只需要包含 1、6 和 A 之外的任意字符，该字符串也可以包含字符 1、6 或 A，但不能只包含这三个字符。比如[^16A]可以匹配 abc、m16，不能匹配 1、16、16A
\<string	匹配以 string 开头的字符串	如：\<do 可以匹配单词 domain，还可以匹配字符串 doa
string\>	匹配以 string 结尾的字符串	如：do\>可以匹配单词 undo，还可以匹配字符串 abcdo
\bcharacter2	匹配 characte1character2，characte1 可以是除了数字、字母和下划线外的任意字符，\b 等效于[^A-Za-z0-9_]	如：\ba 可以匹配-a，-为 characte1，a 为 character2，但是不能匹配 2a 和 ba 等
\Bcharacter	匹配到的字符串中必须包含字符 character，且 character 前不能是空格	如：\Bt 可以匹配 install 里的 t 而不能匹配 big top 中的 t
character1\w	匹配 characte1character2，character2 必须是数字、字母或下划线。\w 相当于[A-Za-z0-9_]	如：v\w 能匹配到 vlan，v 为 characte1，l 为 character2，v\w 还能匹配 service，i 为 character2
\W	等效于\b	如：\Wa 可以匹配-a，-为 characte1，a 为 character2，但是不能匹配 2a 和 ba 等
\	转义操作符，\后紧跟本表列的单个特殊字符时，将去除特殊字符的特定含义	如：\\可以匹配包含\的字符串，\^可以匹配包含^的字符串，\\b 可以匹配包含\b 的字符串

2. 分屏显示

当显示信息较多时，系统会自动将信息分屏显示。通常情况下，一屏将显示 24 行信息，用户也可以使用 **screen-length** 命令设置用户界面下一屏显示的行数（**screen-length** 命令的详细介绍请参见“系统分册”中的“用户界面命令”）。当然，用户也可以通过以下配置禁用当前登录用户的分屏显示功能。

表1-22 禁止分屏显示

操作	命令	说明
禁用当前用户的分屏显示功能	**screen-length disable**	必选 缺省情况下，用户登录后将遵循用户界面下的 **screen-length** 设置。**screen-length** 设置的缺省情况为：允许分屏显示，下一屏显示 24 行数据 该操作在用户视图下执行，仅对当前用户有效，用户重登录后将恢复到缺省情况

3. 显示功能键

命令行接口提供了这样的显示特性：在一次显示信息超过一屏时，提供了暂停功能，这时用户可以有三种选择，如 表 1-23所示。

表1-23 显示功能表

按键或命令	功能
暂停显示时键入空格键	继续显示下一屏信息
暂停显示时键入回车键	继续显示下一行信息
暂停显示时键入<Ctrl+C>	停止显示和命令执行
<Ctrl+E>	将光标移动到当前行的末尾
<PageUp>	显示上一页信息
<PageDown>	显示下一页信息

1.3.7 命令行的历史记录功能

命令行接口将用户最近使用的历史命令自动保存到历史命令缓存区，用户可以随时了解最近执行成功了哪些操作以及调用保存的历史命令，并重复执行。

缺省情况下，命令行接口为每个用户保存 10 条历史命令，但用户可以通过 **history-command max-size** 命令来设置当前用户界面历史命令缓冲区的容量（关于 **history-command max-size** 命令的详细介绍请参见“系统分册”中的“用户界面命令”）。另外，

- 设备保存的历史命令与用户输入的命令格式相同，如果用户使用了命令的不完整形式，保存的历史命令也是不完整形式。
- 如果用户连续多次执行同一条命令，设备的历史命令中只保留最早的一次。但如果执行时输入的形式不同，将作为不同的命令对待。例如：多次执行 display cu 命令，历史命令中只保存一条。如果执行 **display cu** 和 **display current-configuration**，将保存为两条历史命令。

表1-24 访问历史命令

操作	命令或按键	结果
显示历史命令	**display history-command**	显示用户输入的有效历史命令
访问上一条历史命令	上光标键↑或<Ctrl+P>	如果还有更早的历史命令，则取出上一条历史命令
访问下一条历史命令	下光标键↓或<Ctrl+N>	如果还有更晚的历史命令，则取出下一条历史命令

说明

用光标键对历史命令进行访问，在 Windows 200X 及 XP 的 Terminal 和 Telnet 下都是有效的，但对于 Windows 9X 超级终端，↑、↓光标键会无效，这是由于 Windows 9X 的超级终端对这两个键作了不同解释所致，这时可以用组合键<Ctrl+P>和<Ctrl+N>来代替↑、↓光标键达到同样目的。

1.3.8 命令行错误信息

所有用户键入的命令，如果通过语法检查，则正确执行，否则向用户报告错误信息，常见错误信息参见 表 1-25。

表1-25 命令行常见错误信息表

英文错误信息	错误原因
% Unrecognized command found at '^' position.	没有查找到命令
	没有查找到关键字
	参数类型错误
	参数值越界
% Incomplete command found at '^' position.	输入命令不完整
% Ambiguous command found at '^' position.	输入参数不明确，存在二义性
Too many parameters	输入参数太多
% Wrong parameter found at '^' position.	输入参数错误

目 录

1 系统维护与调试 1-1
1.1 Ping功能 1-1
1.1.1 Ping功能简介 1-1
1.1.2 Ping配置 1-2
1.2 Tracert功能 1-2
1.2.1 Tracert功能简介 1-2
1.2.2 Tracert配置 1-3
1.3 系统调试功能 1-3
1.3.1 系统调试功能简介 1-3
1.3.2 系统调试配置 1-4
1.4 Ping和Tracert配置举例 1-5

1 系统维护与调试

在日常的系统维护中，用户可以使用 **ping** 命令和 **tracert** 命令来检查当前网络的连接情况。在日常的系统调试中，用户可以使用 **debug** 命令来打开调试信息开关，通过调试信息来诊断系统故障。

1.1 Ping 功能

1.1.1 Ping 功能简介

通过使用 **ping** 命令，用户可以检查指定地址的设备是否可达，测试网络连接是否出现故障。

Ping 功能是基于 ICMP 协议来实现的：源端向目的端发送 ICMP 回显请求（ECHO-REQUEST）报文；源端根据是否收到目的端的 ICMP 回显应答（ECHO-REPLY）报文来判断目的端是否可达，根据发送报文个数、接收到响应报文个数来判断链路的质量，根据 ping 报文的往返时间来判断源端与目的端之间的“距离”。

H3C设备的Ping功能提供了-r选项，使用该选项可以了解源端到目的端的路由细节。ping –r的原理如 图 1-1所示。

(1) 源端（Device A）发送 RR 选项为空的 ICMP 回显请求。

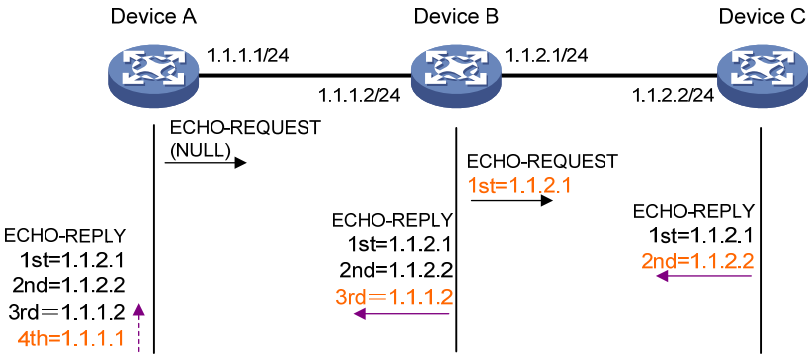
(2) 中间设备（Device B）将自己出接口的 IP 地址添加到 ICMP 回显请求报文的 RR 选项中，并转发该报文。

(3) 目的端收到请求后，发送 ICMP 回显响应报文，响应报文会拷贝请求报文的 RR 选项，并将自己出接口的 IP 地址添加到 RR 选项中。

(4) 中间设备（Device B）将自己出接口的 IP 地址添加到 RR 选项中，并转发该报文。

(5) 源端收到 ICMP 回显响应报文，将自己入接口的 IP 地址添加到 RR 选项中，上送给 CPU 处理。CPU 通过终端将路由细节呈现给用户。

图1-1 ping –r 原理示意图



1.1.2 Ping 配置

表1-1 Ping 配置

操作	命令	说明
检查 IP 网络中的指定地址是否可达	**ping** [**ip**] [**-a** *source-ip* | **-c** *count* | **-f** | **-h** *ttl* | **-i** *interface-type interface-number* | **-m** *interval* | **-n** | **-p** *pad* | **-q** | **-r** | **-s** *packet-size* | **-t** *timeout* | **-tos** *tos* | **-v** | **-vpn-instance** *vpn-instance-name*] * *host*	二者必选其一 **ping** 命令用于 IPv4 网络环境，**ping ipv6** 命令用于 IPv6 网络环境 两条命令均可在任意视图下执行
	ping ipv6 [**-a** *source-ipv6* | **-c** *count* | **-m** *interval* | **-s** *packet-size* | **-t** *timeout*] * *host* [**-i** *interface-type interface-number*]	

说明

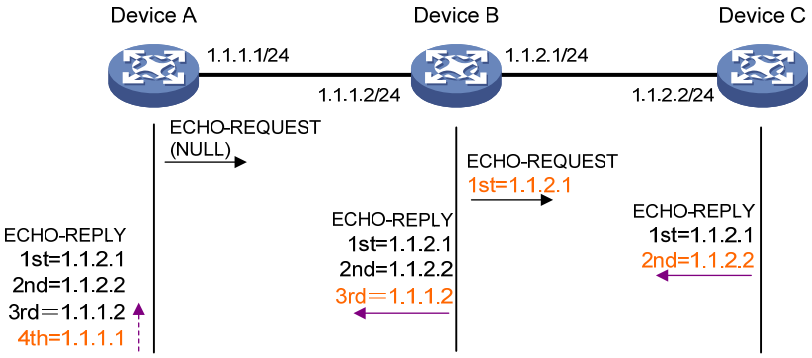
- 如果网络传输速度较慢，用户在配置 **ping** 命令的超时时间参数**-t** 时，可以适当增大超时时间。
- 在配置 **ping** 命令，使用**-i** 参数指定出接口的情况下，只能 ping 直连网段地址。
- 使用 **debugging ip icmp** 命令打开 ICMP 模块的调试开关，可以查看 Ping 过程中 ICMP 报文的交互情况。
- **ping lsp** 命令的详细介绍请参见“MPLS 分册/MPLS 基本配置命令”。

1.2 Tracert 功能

1.2.1 Tracert 功能简介

通过使用 **tracert** 命令，用户可以查看 IP 报文从源端传送到目的端所经过的三层设备，从而检查网络连接是否可用。当网络出现故障时，用户可以使用该命令分析出现故障的网络节点。

图1-2 Tracert 原理示意图



Tracert功能也是基于ICMP协议来实现的，如 图 1-2所示，Tracert功能的原理为：

(1) 源端（Device A）发送一个 TTL 值为 1 的 IP 数据报文给目的端（Device D）（其中报文的 UDP 端口号是目的端的任何一个应用程序都不可能使用的端口号）；

(2) 第一跳（Device B）（即该报文所到达的第一个三层设备）回应一个 TTL 超时的 ICMP 错误信息（该报文中含有第一跳的 IP 地址 1.1.1.2），这样源端就得到了第一个三层设备的地址（1.1.1.2）；

(3) 源端重新发送一个 TTL 值为 2 的报文给目的端；

(4) 第二跳（Device C）回应一个 TTL 超时的 ICMP 错误信息，这样源端就得到了第二个三层设备的地址（1.1.2.2）；

(5) 以上过程不断进行，直到该报文到达目的端，因目的端没有应用程序使用该 UDP 端口，目的端返回一个端口不可达的 ICMP 错误消息（携带了目的端的 IP 地址 1.1.3.2）。

(6) 当源端收到这个端口不可达的 ICMPv6 错误消息后，就知道报文已经到达了目的端，从而得到数据报文从源端到目的端所经历的路径（1.1.1.2；1.1.2.2；1.1.3.2）。

1.2.2 Tracert 配置

表1-2 Tracert 配置

操作	命令	说明
进入系统视图	**system-view**	-
开启设备的 ICMP 超时报文的发送功能	**ip ttl-expires enable**	必选 缺省情况下，ICMP 超时报文发送功能处于关闭状态
开启 ICMP 目的不可达报文发送功能	**ip unreachables enable**	必选 缺省情况下，ICMP 目的不可达报文发送功能处于关闭状态
查看源端到目的端的路由	**tracert** [**-a** *source-ip* | **-f** *first-ttl* | **-m** *max-ttl* | **-p** *port* | **-q** *packet-number* | **-vpn-instance** *vpn-instance-name* | **-w** *timeout*] * *host* **tracert ipv6** [**-f** *first-ttl* | **-m** *max-ttl* | **-p** *port* | **-q** *packet-number* | **-w** *timeout*] * *host*	二者必选其一 **tracert** 命令用于 IPv4 网络环境，**tracert ipv6** 命令用于 IPv6 网络环境 两条命令均可在任意视图下执行

说明

- 使用 **debugging ip icmp** 命令打开 ICMP 模块的调试开关，可以查看 Tracert 过程中 ICMP 报文的交互情况。
- **tracert lsp** 命令的详细介绍请参见“MPLS 分册/MPLS 基本配置命令”。

1.3 系统调试功能

1.3.1 系统调试功能简介

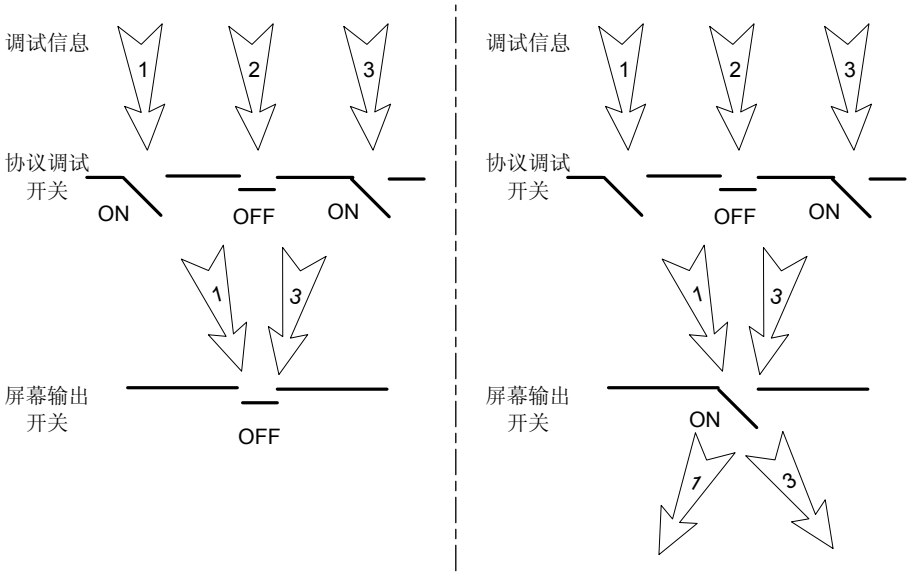
设备提供了种类丰富的调试功能，对于设备所支持的绝大部分协议和功能，系统都提供了相应的调试信息，帮助用户对错误进行诊断和定位。

调试信息的输出可以由两个开关控制：

- 协议调试开关，控制是否生成某协议的调试信息。
- 屏幕输出开关，控制是否在某个用户屏幕上显示调试信息。

如 图 1-3所示：假设设备可以为 1、2、3 三个模块提供调试信息，用户只有将两个开关都打开，调试信息才会在终端显示出来。

图1-3 系统调试开关关系图



1.3.2 系统调试配置

因为调试信息的生成、输出会影响系统的运行效率，所以，本特性一般在维护人员进行网络故障诊断时使用。调试结束后，请及时关闭相应的调试开关，或者使用 **undo debugging all** 命令，关闭所有模块的调试开关。

调试信息的输出与各协议/功能模块的 **debugging** 命令以及信息中心的配置有关。在访问终端（包括 Console 控制台和 VTY 访问终端）上显示是最常用的调试信息输出方式，用户还可以将调试信息发送到别的输出方向，具体配置请参见“系统分册”中的“信息中心配置”。缺省情况下，使用以下配置步骤，可以将调试信息输出到访问终端。

表1-3 系统调试操作

操作	命令	说明
开启终端对系统信息的监视功能	**terminal monitor**	可选 缺省情况下，控制台的监视功能处于开启状态，监视终端的监视功能处于关闭状态 该操作在用户视图下执行
开启终端对调试信息的显示功能	**terminal debugging**	必选 缺省情况下，控制台对调试信息的显示功能处于关闭状态 该操作在用户视图下执行
打开系统中指定模块的调试开关	**debugging** { **all** [**timeout** *time*] | *module-name* [*option*] }	必选 缺省情况下，所有功能项的调试开关均处于关闭状态 该操作在用户视图下执行
显示系统中已经打开的调试开关	**display debugging** [**interface** *interface-type interface-number*] [*module-name*]	可选 可在任意视图下执行

说明

只有同时配置了 **debugging**、**terminal debugging** 和 **terminal monitor** 命令，才能在终端显示调试信息的具体内容。**terminal debugging** 和 **terminal monitor** 命令的详细介绍请参见“系统分册”中的“信息中心命令”。

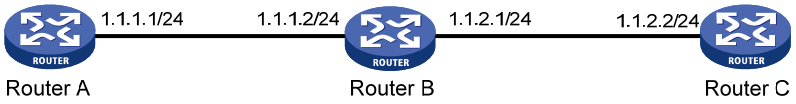
1.4 Ping 和 Tracert 配置举例

1. 组网需求

Router A 使用 Telnet 登录 Router C 失败，现需要确认 Router A 与 Router C 之间是否路由可达，如果路由不可达，需要确定故障的网络节点。

2. 组网图

图1-4 Ping 和 Tracert 应用组网图



3. 配置步骤

使用 ping 命令查看 Router A 和 Router C 之间路由是否可达。

```
<RouterA> ping 1.1.2.2
  PING 1.1.2.2: 56  data bytes, press CTRL_C to break
    Request time out
    Request time out
    Request time out
    Request time out
    Request time out

  --- 1.1.2.2 ping statistics ---
    5 packet(s) transmitted
    0 packet(s) received
    100.00% packet loss
```

路由不可达，使用 tracert 命令确定故障的网络节点。

```
<RouterA> system-view
[RouterA] ip ttl-expires enable
[RouterA] ip unreachables enable
[RouterA] tracert 1.1.2.2
 traceroute to 1.1.2.2(1.1.2.2) 30 hops max,40 bytes packet, press CTRL_C to bre
ak
 1  1.1.1.2 14 ms 10 ms 20 ms
 2  * * *
 3  * * *
 4  * * *
 5
```

```
<RouterA>
```

从上面结果可以看出，Router A 和 Router C 之间路由不可达，Router A 和 Router B 之间路由可达，Router B 和 Router C 之间的连接出了问题。此时可以在 Router A 和 Router C 上使用 **debugging ip icmp** 命令打开 ICMP 报文的调试开关，查看设备有没有收发指定的 ICMP 报文。或者使用 **display ip routing-table** 查看有没有到对端的路由。

目 录

1 信息中心配置 1-1
1.1 信息中心简介 1-1
1.1.1 信息中心概述 1-1
1.1.2 系统信息的格式 1-4
1.2 配置信息中心 1-6
1.2.1 信息中心配置任务简介 1-6
1.2.2 配置信息发送到控制台 1-6
1.2.3 配置信息发送到监视终端 1-7
1.2.4 配置信息发送到日志主机 1-8
1.2.5 配置信息发送到告警缓冲区 1-9
1.2.6 配置信息发送到日志缓冲区 1-9
1.2.7 配置信息发送到SNMP模块 1-10
1.2.8 配置信息保存到日志文件 1-11
1.2.9 配置同步信息输出功能 1-11
1.2.10 禁止端口生成Link up/Link down日志信息 1-12
1.3 信息中心显示和维护 1-12
1.4 信息中心典型配置举例 1-13
1.4.1 日志发送到Unix日志主机的配置举例 1-13
1.4.2 日志发送到Linux日志主机的配置举例 1-15
1.4.3 日志发送到控制台的配置举例 1-16

1 信息中心配置

1.1 信息中心简介

1.1.1 信息中心概述

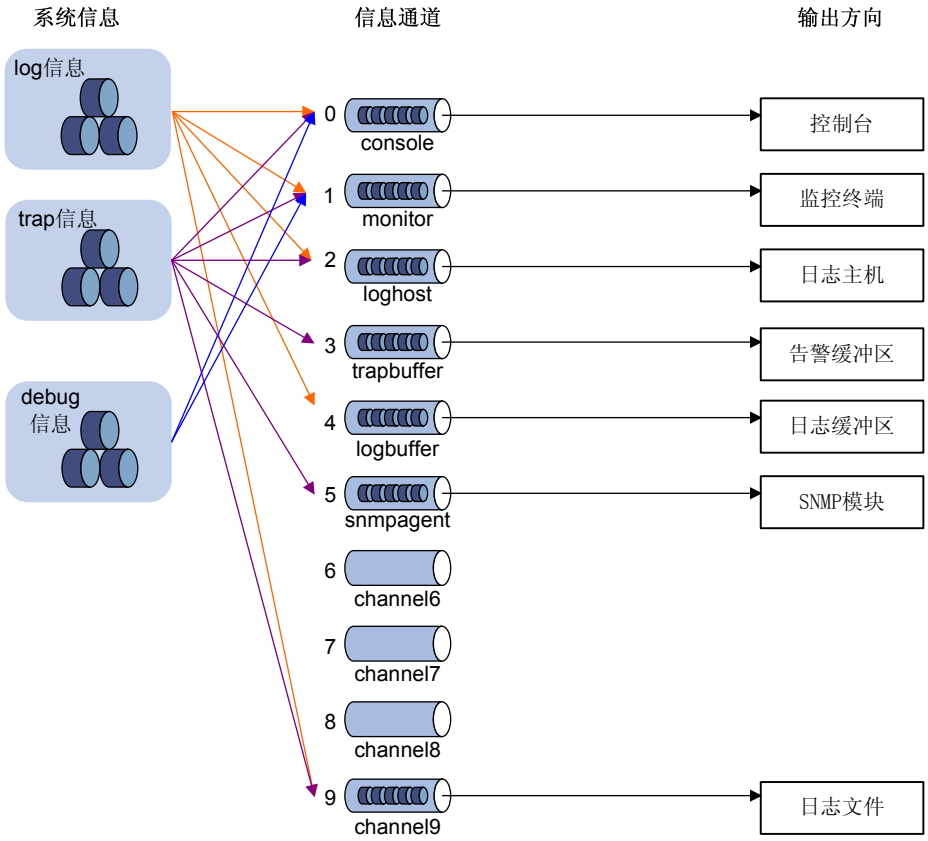
信息中心是系统的信息枢纽，它能够对所有的系统信息进行分类、管理，为网络管理员和开发人员监控网络运行情况和诊断网络故障提供了强有力的支持。

信息中心的工作过程如下：

- 接收各模块生成的日志信息（log）、告警信息（trap）和调试信息（debug）。
- 根据用户设置的输出规则，将信息输出到信息通道。
- 根据信息通道和输出方向的关联，将信息输出到不同方向。

概括来说，信息中心的主要工作就是将三种信息，按照八个严重等级，分配到十个信息通道中，再输出到多个方向，图 1-1就描述了缺省情况下，三种信息分别可以分配到哪些信息通道，每个信息通道对应哪个输出方向。下面将进行详细介绍。

图1-1 信息中心功能示意图



说明

信息中心缺省情况下处于开启状态。在信息中心开启时，特别是在处理信息较多时，由于信息分类、输出的原因，对系统性能有一定的影响。

1. 系统信息的分类

信息中心的系统信息共分为三类信息：

- log 类，日志类信息
- trap 类，告警类信息
- debug 类，调试类信息

2. 系统信息的八个等级

信息按严重性划分为八个等级，严重性由高到底的顺序依次为：emergencies、alerts、critical、errors、warnings、notifications、informational 和 debugging。根据信息级别输出信息时，将会输出信息级别高于或等于所设置级别的信息。比如，输出规则中指定允许级别为 informational 的信息输出，则级别为 emergencies～informational 的信息均会输出。

表1-1 信息级别列表

信息级别	数值	描述
emergencies	0	系统不可用信息
alerts	1	需要立刻做出反应的信息
critical	2	严重信息
errors	3	错误信息
warnings	4	警告信息
notifications	5	正常出现但是重要的信息
informational	6	需要记录的通知信息
debugging	7	调试过程产生的信息

3. 系统信息的七个输出方向和十个通道

系统可以向以下七个方向发送系统信息：控制台（console）、监视终端（monitor）、日志缓冲区（logbuffer）、日志主机（loghost）、告警缓冲区（trapbuffer）、SNMP 模块和日志文件（logfile）七个方向。

系统支持十个通道，缺省情况下，其中 0～5、9 七个通道已经定义了通道名、输出规则和输出方向。可以通过命令改变通道名、输出规则和输出方向，用户还可以在不改变七个通道缺省配置的情况下，配置 6、7、8 这三个富余通道，将他们与输出方向关联，以便对输出的系统信息实施配置的过滤规则。

表1-2 信息通道和输出方向

通道编号	通道的缺省名称	通道的缺省输出方向	说明
0	console	控制台	可以接收 log、trap、debug 信息
1	monitor	监视终端	可以接收 log、trap、debug 信息，方便远程维护
2	loghost	日志主机	可以接收 log、trap、debug 信息，通常系统信息在日志主机以文件的方式存储，以便查询
3	trapbuffer	告警缓冲区	可以接收 trap 信息，是设备内部的一块缓存，用于记录信息

通道编号	通道的缺省名称	通道的缺省输出方向	说明
4	logbuffer	日志缓冲区	可以接收 log 和 debug 信息，是设备内部的一块缓存，用于记录信息
5	snmpagent	SNMP 模块	可以接收 trap 信息
6	channel6	未指定	可以接收 log、trap、debug 信息
7	channel7	未指定	可以接收 log、trap、debug 信息
8	channel8	未指定	可以接收 log、trap、debug 信息
9	channel9	日志文件	可以接收 log、trap、debug 信息



说明

七个方向的设置相互独立，但首先需要开启信息中心，其余的设置才会生效。

4. 按来源模块输出系统信息

系统由众多的协议模块、单板驱动程序、配置模块构成，系统信息可按来源模块进行划分，并过滤输出。系统支持的来源模块可以通过 **info-center source ?** 命令的帮助信息获取。

5. 系统信息的缺省输出规则

缺省输出规则规定了各个输出方向可以输出的信息模块、输出的信息类型以及输出的信息级别，如表 1-3所示。该表表明，缺省情况下：

- 允许所有模块的所有 log 信息发往日志文件；允许所有模块的高于或等于 informational 级别的 log 信息发往日志主机；允许所有模块的高于或等于 warnings 级别的 log 信息发往控制台、监视终端和日志缓冲区；所有模块的所有 log 信息不允许发往告警缓冲区和 SNMP 模块方向。
- 允许所有模块的所有 Trap 信息发往控制台、监视终端、日志主机和日志文件；允许所有模块的高于或等于 warnings 级别的 Trap 信息发往告警缓冲区和 SNMP 模块；所有模块的所有 Trap 信息不允许发往日志缓冲区方向。
- 允许所有模块的所有 debug 信息发往控制台和监视终端；所有模块的所有 debug 信息不允许发往日志主机、告警缓冲区、日志缓冲区、SNMP 模块和日志文件。

表1-3 输出方向的缺省输出规则

输出方向	允许输出的模块	LOG		TRAP		DEBUG	
		开关	级别	开关	级别	开关	级别
控制台	default（所有模块）	开	warnings	开	debugging	开	debugging
监视终端	default（所有模块）	开	warnings	开	debugging	开	debugging
日志主机	default（所有模块）	开	informational	开	debugging	关	debugging
告警缓冲区	default（所有模块）	关	informational	开	warnings	关	debugging

输出方向	允许输出的模块	LOG		TRAP		DEBUG	
		开关	级别	开关	级别	开关	级别
日志缓冲区	default（所有模块）	开	warnings	关	debugging	关	debugging
SNMP 模块	default（所有模块）	关	debugging	开	warnings	关	debugging
日志文件	default（所有模块）	开	debugging	开	debugging	关	debugging

1.1.2 系统信息的格式

按照系统信息的输出方向不同，系统信息可能有不同格式。

- 当输出方向为非日志主机（控制台、监视终端、日志缓冲区、告警缓冲区、SNMP 模块和日志文件）时，系统信息格式大致为：

```
timestamp sysname module/level/digest:content
```

对应的中文格式为：

时间戳 主机名 模块名/信息级别/信息摘要:信息文本

比如，监视终端与设备相连，当有终端登录设备时，在监视终端可以看到格式如下的日志信息：

```
%Jun 26 17:08:35:809 2008 Sysname SHELL/4/LOGIN: VTY login from 1.1.1.1
```

- 当输出方向为日志主机时，按照 RFC 3164（The BSD syslog Protocol），系统信息格式为：

```
<Int_16>timestamp sysname %%nnmodule/level/digest: source content
```



说明

- 以上格式中的尖括号（< >）、空格、斜杠（/）、冒号（:）是必须的。
- 以上格式是设备给各个输出方向发送的原始信息的格式，可能与用户看到的信息格式有差异，最终显示格式与用户使用的查看工具有关，请以实际情况为准。

下面对每一个字段做详细说明。

1. Int_16（优先级）

优先级的计算按如下公式：facility*8+severity。facility表示工具名称，在设置日志主机相关参数的时候可以设置，参数取值为local0～local7，对应的十进制数值为 16～23，缺省取值为local7。主要用于在日志主机端标志不同的日志来源，查找、过滤对应日志源的日志。severity（信息级别）的取值范围为 0～7，信息级别具体含义请参见 表 1-1。

本字段只有在信息发往日志主机时才有效。

2. timestamp（时间戳）

时间戳记录了系统信息产生的时间，方便用户查看和定位系统事件用户可使用 **info-center timestamp** 命令设置是否带时间戳，以及所带时间戳的格式。发往日志主机的系统信息的时间戳不带毫秒信息，发送到其它方向的系统信息会精确到毫秒。

3. sysname（主机名）

主机名是本机的系统名。用户可使用 **sysname** 命令修改主机名（具体配置请参见“系统分册”中的“系统基本配置命令”）。

主机名与模块名之间以一个空格隔开。

4. %%

该字段表示标志厂商的前导字符。

本字段只有在信息发往日志主机时才有效。

5. nn

该字段表示 syslog 的版本标识。

本字段只有在信息发往日志主机时才有效。

6. module（模块名）

该字段表示产生信息的功能模块的名称。模块列表可以通过在系统视图下输入命令 **info-center source ?** 查看到。模块名与信息级别之间以一个斜杠（/）隔开。

7. level（信息级别）

系统信息的级别共分为 8 级，从 0～7，它们的定义和说明请参见 表 1-1。各模块生成的系统信息的级别在开发阶段已经确定，用户不能更改，但可以使用**info-center source**命令选择指定级别的信息输出。

8. digest（信息摘要）

信息摘要是一个不超过 32 个字符的字符串，表示该信息的内容大意。

- 对于输出方向为日志主机的系统信息，如果该字符串以“(l)”结尾则表示该信息为 Log 信息，如果该字符串以“(t)”结尾则表示该信息为 Trap 信息，如果该字符串以“(d)”结尾则表示该信息为 Debug 信息。
- 对于输出方向为非日志主机的系统信息，时间戳前面会有百分号（%）或井字符号（#）或米字符号（*），分别代表该条信息是日志信息或告警信息或调试信息。

9. source（定位信息）

该字段表示该信息的产生者，可能是单板槽号、IRF 的成员设备编号或者日志发送者的源 IP，该字段可选。

本字段只有在信息发往日志主机时才有效。

10. content（信息文本）

该字段表示了该条系统信息的具体内容。

1.2 配置信息中心

1.2.1 信息中心配置任务简介

表1-4 信息中心配置任务简介

配置任务	说明	详细配置
配置信息发送到控制台	可选	1.2.2
配置信息发送到监视终端	可选	1.2.3
配置信息发送到日志主机	可选	1.2.4
配置信息发送到告警缓冲区	可选	1.2.5
配置信息发送到日志缓冲区	可选	1.2.6
配置信息发送到 SNMP 模块	可选	1.2.7
配置信息保存到日志文件	可选	1.2.8
配置同步信息输出功能	可选	1.2.9

1.2.2 配置信息发送到控制台

1. 配置信息发送到控制台

表1-5 配置信息发送到控制台

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2
设置系统向控制台输出信息的通道	**info-center console channel** { *channel-number* | *channel-name* }	可选 缺省情况下，系统使用 0 号（console）通道向控制台输出信息
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 5.
设置时间戳输出格式	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	可选 缺省情况下，log、trap 和 debug 信息的时间戳输出格式均为 **date** 格式

2. 开启控制台对系统信息的显示功能

在配置将系统信息发送到控制台后，为了能在控制台上观察到输出的信息，还要开启控制台对相应信息的显示功能。请在用户视图下进行以下操作。

表1-6 开启控制台对系统信息的显示功能

操作	命令	说明
开启控制台对系统信息的监视功能	**terminal monitor**	可选 缺省情况下，控制台的监视功能处于开启状态，监视终端的监视功能处于关闭状态
开启控制台对调试信息的显示功能	**terminal debugging**	必选 缺省情况下，控制台对调试信息的显示功能处于关闭状态
开启控制台对日志信息的显示功能	**terminal logging**	可选 缺省情况下，控制台对日志信息的显示功能处于开启状态
开启控制台对告警信息的显示功能	**terminal trapping**	可选 缺省情况下，控制台对告警信息的显示功能处于开启状态

1.2.3 配置信息发送到监视终端

系统信息可以发往控制台，也可以发往监视终端。监视终端是指以 AUX、VTY、TTY 类型用户界面登录的用户终端。

1. 配置信息发送到监视终端

表1-7 配置信息发送到监视终端

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2
设置系统向监视终端输出系统信息的通道	**info-center monitor channel** { *channel-number* | *channel-name* }	可选 缺省情况下，系统使用 1 号（monitor）通道向监视终端输出系统信息
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 1.
设置时间戳输出格式	**info-center timestamp** { **debugging | log | trap** } { **boot** | **date** | **none** }	可选 缺省情况下，log、trap 和 debug 信息的时间戳输出格式均为 **date** 格式

2. 开启监视终端对系统信息的显示功能

在配置将系统信息发送到监视终端后，为了能在监视终端上观察到输出的信息，还要开启监视终端对相应信息的显示功能。

表1-8 开启监视终端对系统信息的显示功能

操作	命令	说明
开启监视终端对系统信息的监视功能	**terminal monitor**	必选 缺省情况下，控制台的监视功能处于开启状态，监视终端的监视功能处于关闭状态
开启监视终端对调试信息的显示功能	**terminal debugging**	必选 缺省情况下，监视终端对调试信息的显示功能处于关闭状态
开启监视终端对日志信息的显示功能	**terminal logging**	可选 缺省情况下，监视终端对日志信息的显示功能处于开启状态
开启监视终端对告警信息的显示功能	**terminal trapping**	可选 缺省情况下，监视终端对告警信息的显示功能处于开启状态

1.2.4 配置信息发送到日志主机

表1-9 配置信息发送到日志主机

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2
指定日志主机，设置系统向日志主机输出信息时的参数	**info-center loghost** *host-ip* [**channel** { *channel-number* | *channel-name* } | **facility** *local-number*]*	必选 缺省情况下，系统不向日志主机输出信息。如果使能系统向日志主机输出信息，则系统默认使用 2 号（loghost）通道
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* }* | **log** { **level** *severity* | **state** *state* }* | **trap** { **level** *severity* | **state** *state* }*]*	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 1.
配置发送的日志信息的源 IP 地址	**info-center loghost source** *interface-type interface-number*	可选 缺省情况下，将根据路由来确定发送日志信息的出接口，使用该接口的主 IP 地址作为发送的日志信息的源 IP 地址
设置发往日志主机的系统信息的时间戳输出格式	**info-center timestamp loghost** { **date** | **no-year-date** | **none** }	可选 缺省情况下，发往日志主机的系统信息的时间戳输出格式为 **date** 格式

1.2.5 配置信息发送到告警缓冲区

说明

用户可以设置将 log、trap 和 buffer 三种类型的信息发送给告警缓冲区，但告警缓冲区只接收 trap 信息，其它类型的信息将被丢弃。

表1-10 配置信息发送到告警缓冲区

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2
设置系统向告警缓冲区输出信息以及相关参数	**info-center trapbuffer** [**channel** { *channel-number* | *channel-name* } | **size** *buffersize*] *	可选 缺省情况下，系统使用 3 号（trapbuffer）通道向告警缓冲区输出信息，缓冲区可存储 256 条信息
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 5.
设置时间戳输出格式	**info-center timestamp** { **debugging | log | trap** } { **boot** | **date** | **none** }	可选 缺省情况下，log、trap 和 debug 信息的时间戳输出格式均为 **date** 格式

1.2.6 配置信息发送到日志缓冲区

说明

用户可以设置将 log、trap 和 debug 三种类型的信息发送给日志缓冲区，但日志缓冲区只能接收 log 和 debug 信息，trap 信息将被丢弃。

表1-11 配置信息发送到日志缓冲区

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2

操作	命令	说明
设置系统向日志缓冲区输出信息以及相关参数	**info-center logbuffer** [**channel** { *channel-number* | *channel-name* } | **size** *buffersize*] *	可选 缺省情况下，系统使用 4 号（logbuffer）通道向日志缓冲区输出信息，缓冲区可存储 512 条信息
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 5.
设置时间戳输出格式	**info-center timestamp** { **debugging | log | trap** } { **boot** | **date** | **none** }	可选 缺省情况下，log、trap 和 debug 信息的时间戳输出格式均为 **date** 格式

1.2.7 配置信息发送到 SNMP 模块

说明

用户可以设置将 log、trap 和 debug 三种类型的信息发送给 SNMP 模块，但 SNMP 模块只接收 trap 信息，其它类型的信息将被丢弃。

为了监控设备的运行状况，通常会将 Trap 信息发送到 SNMP 网络管理工作站（NMS）。此时，首先需要将信息发送给 SNMP 模块，再设置 SNMP 模块 Trap 发送参数，对 Trap 报文进行进一步的处理（详细介绍请参见“系统分册”中的“SNMP 配置”）。

表1-12 配置信息发送到 SNMP 模块

操作	命令	说明
进入系统视图	**system-view**	-
开启信息中心	**info-center enable**	可选 缺省情况下，信息中心处于开启状态
为指定编号的信息通道命名	**info-center channel** *channel-number* **name** *channel-name*	可选 通道的缺省名称请参见 表 1-2
设置系统向 SNMP 模块输出信息的通道	**info-center snmp channel** { *channel-number* | *channel-name* }	可选 缺省情况下，系统使用 5 号（snmpagent）通道向 SNMP 模块送系统信息
配置系统信息的输出规则	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* }* | **log** { **level** *severity* | **state** *state* }* | **trap** { **level** *severity* | **state** *state* }*]*	可选 缺省情况下，系统信息的输出规则请参见 1.1.1 5.
设置时间戳输出格式	**info-center timestamp** { **debugging | log | trap** } { **boot** | **date** | **none** }	可选 缺省情况下，log、trap 和 debug 信息的时间戳输出格式均为 **date** 格式

1.2.8 配置信息保存到日志文件

通过日志文件（logfile）特性，用户可以将系统产生的日志信息，按照指定的频率保存到该设备的指定路径下，便于用户在本地随时查看历史操作，保障设备的正常运行。

日志在保存到日志文件前，先保存在日志文件缓冲区（logfile buffer）。系统会按照指定的频率将缓冲区的内容写入日志文件，频率一般配置为 24 小时，在设备比较空闲的时候（比如清晨）进行保存，用户也可以手工触发保存。成功保存后，保存之前的日志文件缓冲区里的内容会被清空。

日志文件有容量限制，当日志文件的大小达到最大值时，系统会自动创建新的日志文件来保存新信息，日志文件的名称为 logfile1.log、logfile2.log、……。当日志文件的个数或者存储介质的空间达到设备支持的最大值时，系统会删除最旧的日志文件再创建新的日志文件，日志文件的路径为 /logfile/。

表1-13 配置信息保存到日志文件

操作	命令	说明
进入系统视图	**system-view**	-
开启日志文件特性	**info-center logfile enable**	可选 缺省情况下，日志文件特性处于开启状态
设置自动保存日志文件的频率	**info-center logfile frequency** *freq-sec*	可选 缺省情况下，设备自动保存日志文件的频率是 86400 秒。
设置单个日志文件最大能占用的存储空间的大小	**info-center logfile size-quota** *size*	可选 缺省情况下，日志文件最大能占用的存储空间的大小为 8MB
设置存储日志文件的目录	**info-center logfile switch-directory** *dir-name*	可选 缺省情况下，存放日志文件的目录为存储设备根目录下的 logfile 目录（存储设备根目录由设备指定）
手动将日志文件缓冲区中的内容保存到日志文件	**logfile save**	可选 任意视图均可执行 缺省情况下，系统将按照 **info-center logfile frequency** 命令所设置的频率自动保存日志文件



说明

- 为了保证设备的正常运行，**info-center logfile size-quota** 设置的日志文件的大小最小不能低于 1MB，最大不能超过 10MB。
- **info-center logfile switch-directory** 命令通常用于日志文件的备份或者迁移前的准备工作，其配置会在设备重启或主备倒换后失效。

1.2.9 配置同步信息输出功能

同步信息输出是指当用户在输入时有日志、告警、调试等系统信息输出，则在输出后会回显命令行提示符（在命令编辑状态回显提示符，交互状态回显“[Y/N]”字符串）和用户已有的输入。

此功能用于用户在进行操作（操作还没有完成）却被大量的系统信息打断时，回显用户的上一步操作，用户可以接着执行上一步的操作。

表1-14 配置同步信息输出功能

操作	命令	说明
进入系统视图	**system-view**	-
打开同步信息输出功能	**info-center synchronous**	必选 缺省情况下，同步信息输出功能处于关闭状态

说明

- 在当前命令行提示符下，如果用户没有任何输入，此时若有日志等系统信息输出，输出后将不会回显命令行提示符；
- 当处在交互状态，需要用户输入一些交互信息时（非 Y/N 确认信息），因为情况各异，所以若有系统信息输出，输出后不再回显提示信息，而只是将用户已有的输入换行打印出来。

1.2.10 禁止端口生成 Link up/Link down 日志信息

缺省情况下，设备的所有端口在端口状态改变时都会生成端口 Link up 和 Link down 的日志信息。如果用户只关心某个或某些端口的状态，这时可以使用该功能，禁止其它端口生成 Link up/Link down 日志信息；或者某个接口的状态不稳定，总在频繁地改变，这时，会生成大量的 Link up/Link down 日志信息，为了屏蔽这些冗余信息，可以使用该功能，禁止该端口生成 Link up/Link down 日志信息。

表1-15 禁止端口生成 Link up/Link down 日志信息

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
禁止端口生成 Link up/Link down 日志信息	**undo enable log updown**	必选 缺省情况下，允许所有端口在状态发生改变时生成端口 Link up 和 Link down 的日志信息

说明

使用本特性后，如果端口状态改变，将不再生成端口 Link up 和 Link down 的日志信息。这样，可能会影响用户监控端口状态，所以，一般情况下，建议采用缺省配置。

1.3 信息中心显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后信息中心的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset logbuffer**、**reset trapbuffer** 命令可以分别将日志缓冲区和告警缓冲区的统计信息清除。

表1-16 信息中心显示和维护

操作	命令
显示信息通道的信息	**display channel** [*channel-number* | *channel-name*]
显示各个输出方向的信息	**display info-center**
显示系统日志缓冲区的状态和缓冲区记录的日志信息（集中式设备）	**display logbuffer** [**reverse**] [**level** *severity* | **size** *buffersize*]* [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示系统日志缓冲区的状态和缓冲区记录的日志信息（分布式设备）	**display logbuffer** [**reverse**] [**level** *severity* | **size** *buffersize* | **slot** *slot-number*]* [**|** { **begin** | **exclude** | **include** } *regular-expression*]
显示系统日志缓冲区的概要信息（集中式设备）	**display logbuffer summary** [**level** *severity*]
显示系统日志缓冲区的概要信息（分布式设备）	**display logbuffer summary** [**level** *severity* | **slot** *slot-number*] *
显示日志文件缓冲区内容	**display logfile buffer**
显示日志文件配置	**display logfile summary**
显示系统告警缓冲区的状态和缓冲区记录的告警信息	**display trapbuffer** [**reverse**] [**size** *buffersize*]
清除日志缓冲区内的信息	**reset logbuffer**
清除告警缓冲区内的信息	**reset trapbuffer**

1.4 信息中心典型配置举例

1.4.1 日志发送到 Unix 日志主机的配置举例

1. 组网需求

- 将系统的日志信息发送到 Unix 日志主机；
- 日志主机的 IP 地址为 1.2.0.1/16；
- 信息级别高于等于 informational 的日志信息将会发送到日志主机上；
- 允许输出日志信息的模块为 ARP 和 IP。

2. 组网图

图1-2 配置信息中心组网图（Unix 日志主机）



3. 配置步骤

配置前请确保 Device 和 PC 之间路由可达，具体配置步骤略。

(1) 设备上的配置

开启信息中心。

```
<Sysname> system-view
```

```
[Sysname] info-center enable
```

将 IP 地址为 1.2.0.1/16 的主机用作日志主机，使用 loghost 通道发送信息（可选，系统缺省为 loghost 通道），使用 local4 作为日志主机记录工具。

```
[Sysname] info-center loghost 1.2.0.1 channel loghost facility local4
```

关闭 loghost 通道所有模块 log、trap、debug 信息的输出开关。

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```

 注意

由于系统对各通道允许输出的系统信息的缺省情况不一样，所以配置前必须将所有模块的需求通道（本例为 loghost）上 log、trap、debug 信息的输出开关关闭，再根据当前的需求配置输出规则，以免输出太多不需要的信息。

配置输出规则：允许 ARP 和 IP 模块的、级别高于等于 informational 的日志信息输出到日志主机（注意：允许输出信息的模块由产品决定）。

```
[Sysname] info-center source arp channel loghost log level informationalstate on
[Sysname] info-center source ip channel loghost log level informationalstate on
```

(2) 日志主机上的配置

下面的配置示例是在 SunOS 4.0 上完成的，在其它厂商的 Unix 操作系统上的配置操作基本类似。

第一步：以超级用户（root）的身份登录日志主机。

第二步：在/var/log/路径下为 Device 创建同名日志文件夹 Device，在该文件夹创建文件 info.log，用来存储来自 Device 的日志。

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

第三步：编辑文件/etc/syslog.conf，添加以下内容。

```
# Device configuration messages
local4.info     /var/log/Device/info.log
```

以上配置中，local4 表示日志主机接收日志的工具名称，info 表示信息级别。Unix 系统会把级别高于等于 informational 的日志记录到/var/log/Device/info.log 文件中。

 说明

在编辑/etc/syslog.conf 时应注意以下问题:

- 注释必须独立成行，并以字符#开头。
- 在文件名之后不得有多余的空格。
- /etc/syslog.conf 中指定的工具名称及信息级别与设备上 **info-center loghost** 和 **info-center source** 命令的相应参数的指定值要保持一致，否则日志信息可能无法正确输出到日志主机上。

第四步：当日志文件 info.log 建立且/etc/syslog.conf 文件被修改之后，应通过执行以下命令查看系统守护进程 syslogd 的进程号，中止 syslogd 进程，并重新用-r 选项在后台启动 syslogd，使修改后配置生效。

```
# ps -ae | grep syslogd
```

```
147
# kill -HUP 147
# syslogd -r &
```

进行以上操作之后，系统就可以在相应的文件中记录日志信息了。

1.4.2 日志发送到 Linux 日志主机的配置举例

1. 组网需求

- 系统的日志信息发送到 Linux 日志主机上，日志主机的 IP 地址为 1.2.0.1/16；
- 信息级别高于等于 informational 的日志信息将会发送到日志主机上；
- 所有模块均允许输出日志信息。

2. 组网图

图1-3 配置信息中心组网图（Linux 日志主机）



3. 配置步骤

配置前请确保 Device 和 PC 之间路由可达，具体配置步骤略。

(1) 设备上的配置

开启信息中心。

```
<Sysname> system-view
[Sysname] info-center enable
```

将 IP 地址为 1.2.0.1/16 的主机用作日志主机，使用 loghost 通道发送信息（可选，系统缺省为 loghost 通道），使用 local5 作为日志主机记录工具。

```
[Sysname] info-center loghost 1.2.0.1 channel loghost
```

关闭 loghost 通道所有模块 log、trap、debug 信息的输出开关。

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```

注意

由于系统对各通道允许输出的系统信息的缺省情况不一样，所以配置前必须将所有模块的需求通道（本例为 loghost）上 log、trap、debug 信息的输出开关关闭，再根据当前的需求配置输出规则，以免输出太多不需要的信息。

配置输出规则：允许所有模块、级别高于等于 informational 的日志信息输出到日志主机。

```
[Sysname] info-center source default channel loghost log level informational state on
```

(2) 日志主机上的配置

第一步：以超级用户（root）的身份登录日志主机。

第二步：在/var/log/路径下为 Device 创建同名日志文件夹 Device，在该文件夹创建文件 info.log，用来存储来自 Device 的日志。

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

第三步：编辑文件/etc/syslog.conf，添加以下内容。

```
# Device configuration messages
local5.info     /var/log/Device/info.log
```

以上配置中，local5 表示日志主机接收日志的工具名称，info 表示信息级别。Linux 系统会把级别高于等于 informational 的日志记录到/var/log/Device/info.log 文件中。

说明

在编辑/etc/syslog.conf 时应注意以下问题:

- 注释必须独立成行，并以字符#开头。
- 在文件名之后不得有多余的空格。
- /etc/syslog.conf 中指定的工具名称及信息级别与设备上 **info-center loghost** 和 **info-center source** 命令的相应参数的指定值要保持一致，否则日志信息可能无法正确输出到日志主机上。

第四步：当日志文件 info.log 建立且/etc/syslog.conf 文件被修改之后，应通过执行以下命令查看系统守护进程 syslogd 的进程号，中止 syslogd 进程，并重新用-r 选项在后台启动 syslogd，使修改后配置生效。

```
# ps -ae | grep syslogd
147
# kill -9 147
# syslogd -r &
```

说明

对 Linux 日志主机，必须保证 syslogd 进程是以-r 选项启动。

进行以上操作之后，系统就可以在相应的文件中记录日志信息了。

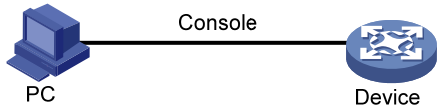
1.4.3 日志发送到控制台的配置举例

1. 组网需求

- 将信息级别高于等于 informational 的日志信息发送到控制台上；
- 允许输出日志信息的模块为 ARP 和 IP。

2. 组网图

图1-4 配置信息中心组网图（控制台）



3. 配置步骤

开启信息中心。

```
<Sysname> system-view
[Sysname] info-center enable
```

使用 console 通道向控制台输出日志信息（可选，缺省情况下系统向控制台输出信息的通道就为 console 通道）。

```
[Sysname] info-center console channel console
```

关闭控制台通道所有模块 log、trap、debug 信息的输出开关。

```
[Sysname] info-center source default channel console debug state off log state off trap state off
```

 注意

由于系统对各通道允许输出的系统信息的缺省情况不一样，所以配置前必须将所有模块的需求通道（本例为 console）上 log、trap、debug 信息的输出开关关闭，再根据当前的需求配置输出规则，以免输出太多不需要的信息。

配置输出规则：允许 ARP 和 IP 模块的、级别高于等于 informational 的日志信息输出（注意：允许输出的信息模块由产品决定）。

```
[Sysname] info-center source arp channel console log level informational state on
[Sysname] info-center source ip channel console log level informational state on
[Sysname] quit
```

打开终端显示功能（可选，缺省已经打开了该功能）。

```
<Sysname> terminal monitor
Info: Current terminal monitor is on.
<Sysname> terminal logging
Info: Current terminal logging is on.
```

以上命令配置成功后，如果指定的模块产生了日志信息，信息中心会自动把这些日志发送到控制台，在控制台的屏幕上显示。

目 录

1 用户界面配置 1-1
1.1 用户界面简介 1-1
1.1.1 用户界面概述 1-1
1.1.2 用户与用户界面的关系 1-1
1.1.3 用户界面的编号 1-1
1.2 用户界面配置任务简介 1-2
1.3 配置AUX口属性 1-2
1.4 配置终端属性 1-3
1.5 配置Modem属性 1-4
1.6 配置自动执行命令 1-5
1.7 通过用户界面配置用户级别 1-5
1.8 配置VTY用户界面访问限制 1-6
1.9 配置VTY用户界面支持的协议 1-7
1.10 配置AUX口的重定向功能 1-7
1.11 配置登录用户界面的认证方式 1-8
1.12 配置启动终端会话及终止当前运行任务的快捷键 1-10
1.13 向指定的用户界面发送消息 1-10
1.14 释放指定用户界面上建立的连接 1-10
1.15 用户界面显示和维护 1-11

1 用户界面配置

1.1 用户界面简介

1.1.1 用户界面概述

当用户使用 Console 口、AUX 口、Telnet 或者 SSH 方式登录设备的时候，系统会分配一个用户界面（也称为 Line）用来管理、监控设备和用户间的当前会话。每个用户界面有对应的用户界面视图（User-interface view），在用户界面视图下网络管理员可以配置一系列参数，比如用户登录时是否需要认证、是否重定向到别的设备以及用户登录后的级别等，当用户使用该用户界面登录的时候，将受到这些参数的约束，从而达到统一管理各种用户会话连接的的目的。

目前系统支持的命令行配置方式有：

- Console 口本地配置
- AUX 口本地或远程配置
- Telnet 或 SSH 本地或远程配置

与这些配置方式对应的是三种类型的用户界面：

- Console 用户界面：用来管理和监控通过 Console 口登录的用户。Console 口是一种线设备端口。设备提供 Console 口，端口类型为 EIA/TIA-232 DCE。
- AUX 用户界面：用来管理和监控通过 AUX 口登录的用户。AUX 口（Auxiliary port，辅助端口）也是一种线设备端口。设备提供 AUX 口，端口类型为 EIA/TIA-232 DTE，通常用于通过 Modem 进行拨号访问。
- VTY（Virtual Type Terminal，虚拟类型终端）用户界面：用来管理和监控通过 VTY 方式登录的用户。VTY 口属于逻辑终端线，用于对设备进行 Telnet 或 SSH 访问。目前每台设备最多支持 5 个 VTY 用户同时访问。

1.1.2 用户与用户界面的关系

用户界面的管理和监控对象是使用某种方式登录的用户，虽然单个用户界面某一时刻只能被一个用户使用，但它并不针对某个用户。比如用户 A 使用 Console 口登录设备时，将受到 Console 用户界面视图下配置的约束，当使用 VTY 1 登录设备时，将受到 VTY 1 用户界面视图下配置的约束。

根据设备的硬件配备情况，一台设备上可能有多个 Console 口、AUX 口和（或）以太网接口，所以设备可能支持多个 Console、AUX、VTY 用户界面，这些用户界面与用户并没有固定的对应关系。用户登录时，系统会根据用户的登录方式，自动给用户分配一个当前空闲的、编号最小的某类型的用户界面，整个登录过程将受该用户界面视图下配置的约束。同一用户登录的方式不同，分配的用户界面不同；同一用户登录的时机不同，分配的用户界面可能不同。

1.1.3 用户界面的编号

用户界面的编号有两种方式：绝对编号方式和相对编号方式。

1. 绝对编号方式

使用绝对编号方式，可以唯一的指定一个用户界面或一组用户界面。绝对编号从 0 开始自动编号，每次增长 1，先给所有 Console 用户界面编号，然后是所有 AUX 用户界面，最后是所有 VTY 用户界面。使用 **display user-interface**（不带参数）可查看到设备当前支持的用户界面以及它们的绝对编号。

2. 相对编号方式

相对编号是每种类型用户界面的内部编号。该方式只能指定某种类型的用户界面中的一个或一组，而不能跨类型操作。

相对编号方式的形式是：“用户界面类型 编号”，遵守如下规则：

- 控制台的编号：CON 0，第二个为 CON 1，依次类推。
- 辅助接口的编号：AUX 0，第二个为 AUX 1，依次类推。
- VTY 的编号：第一个为 VTY 0，第二个为 VTY 1，依次类推。

1.2 用户界面配置任务简介

表1-1 用户界面配置

配置任务	说明	详细配置
配置 AUX 口属性	可选	1.3
配置终端属性	可选	1.4
配置 Modem 属性	可选	1.5
配置自动执行命令	可选	1.6
配置用户界面所能访问的命令级别	可选	1.7
配置 VTY 用户界面访问限制	可选	1.8
配置 VTY 用户界面支持的协议	可选	1.9
配置 AUX 口的重定向功能	可选	1.10
配置登录用户界面的认证方式	可选	1.11
配置启动终端会话及终止当前运行任务的快捷键	可选	1.12
向指定的用户界面发送消息	可选	1.13
释放指定用户界面上建立的连接	可选	1.14

1.3 配置 AUX 口属性

AUX 口通信最重要的参数包括：

- 传输速率：表示设备给访问终端每秒钟传送的比特的个数，它是一个衡量通信速度的参数。传输速率和距离成反比，高传输速率常常用于放置的很近的仪器间的通信。
- 数据位：是指通信双方使用多少比特位来表示一个字符，如何设置取决于需要传送的信息。比如，如果传送的是标准的 ASCII 码，则可以将数据位设置为 7，如果传输的是扩展的 ASCII 码，则需要将数据位设置为 8。

- 奇偶校验位：在串口通信中一种简单的检错方式，用来检查数据在传输过程种是否出错。
- 停止位：用来表示单个包的结束。停止位的位数越多，传输效率越低。

对于两个通信的端口，这些参数必须一致才能正常通信。用户可以通过以下步骤，来配置 AUX 口属性，使得设备与访问终端的特性能够匹配。

表1-2 配置 AUX 口属性

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface aux 0**	-
设置传输速率	**speed** *speed-value*	可选 缺省情况下，传输速率为 9600bps
配置流量控制方式	**flow-control** { { **hardware** [*flow-control-type1*] | **software** [*flow-control-type2*] } * | **none** }	可选
设置校验位的解析和生成方式	**parity** { **even** | **mark** | **none** | **odd** | **space** }	可选 缺省情况下，设备校验位的校验方式为 **none**，即不进行校验
设置停止位的个数	**stopbits** { **1** | **1.5** | **2** }	可选 缺省情况下，停止位为 1
设置数据位	**databits** { **5** | **6** | **7** | **8** }	可选 缺省情况下，用户界面的数据位为 8 位

说明

访问终端和设备相应用户界面下传输速率、数据位、校验位、停止位、流量控制的设置必须一致，双方才能正常通信。

1.4 配置终端属性

表1-3 配置终端属性

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-
启动终端服务	**shell**	可选 缺省情况下，系统在所有的用户界面上启动终端服务
设置用户连接的超时时间	**idle-timeout** *minutes* [*seconds*]	可选 缺省情况下，用户连接的超时时间为 10 分钟
设置下一屏显示的行数	**screen-length** *screen-length*	可选 缺省情况下，下一屏显示 24 行数据

操作	命令	说明
设置当前用户界面下的终端显示类型	**terminal type** { **ansi** | **vt100** }	可选 缺省情况下，终端显示类型为 ANSI
设置历史命令缓冲区可存放的历史命令的条数	**history-command max-size** *size-value*	可选 缺省情况下，历史命令缓冲区可存放 10 条历史命令
退回用户视图	**return**	-
锁定用户界面，防止未授权的用户操作该界面	**lock**	可选 缺省情况下，系统不会自动锁住当前用户界面



说明

设备支持两种终端显示类型：ANSI 和 VT100。当设备的终端类型与客户端（如超级终端或者 Telnet 客户端等）的终端类型不一致，或者均设置为 ANSI，并且当前编辑的命令行的总字符数超过 80 个字符时，客户端会出现光标错位、终端屏幕不能正常显示的现象。建议两端都设置为 VT100 类型。

1.5 配置 Modem 属性



说明

本特性只对工作在流方式下的 AUX 口有效，对 Console 口无效。

用户在 AUX 口下使用 Modem 接入设备，通过用户界面视图可管理和配置 Modem 的有关参数。

表1-4 配置 Modem 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **vty** } *first-num2* [*last-num2*] }	-
设置呼入连接建立时，用户从摘机到拨号的有效间隔时间	**modem timer answer** *time*	可选 缺省情况下，拨号超时时间为 60 秒
设置 Modem 为自动应答方式	**modem auto-answer**	可选 缺省情况下，Modem 的应答方式为手动应答
使能 Modem 的呼入/呼出功能	**modem** { **both** | **call-in** | **call-out** }	可选 缺省情况下，Modem 的呼入和呼出功能处于禁止状态

1.6 配置自动执行命令

配置自动执行命令后，用户在登录时，系统会自动执行已经配置好的命令，执行完命令后，自动断开用户连接。如果这条命令引发起了一个任务，系统会等这个任务执行完毕后再断开连接。该命令通常用来配置 Telnet 命令，使用户登录时自动连接到指定的主机。

表1-5 配置自动执行命令

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **vty** } *first-num2* [*last-num2*] }	-
配置自动执行命令	**auto-execute command** *command*	必选 缺省情况下，未设定自动执行命令

auto-execute command 命令使用时有如下的限制：

- Console 口不支持 **auto-execute command**。
- 如果设备上只有一个 AUX 口，没有 Console 口（Console 口和 AUX 口共用），则此 AUX 口也不支持 **auto-execute command**。
- 对其他类型的接口不作限制。

注意

- 使用 **auto-execute command** 命令后，可能导致用户不能通过该终端线对本系统进行常规配置，需谨慎使用。
- 在配置 **auto-execute command** 命令并保存配置（执行 **save** 操作）之前，要确保可以通过其他 VTY、CON、AUX 用户登录进来更改配置，以便出现问题后，能删除该配置。

1.7 通过用户界面配置用户级别

用户级别用来限制不同用户对设备的访问权限。如果用户登录时使用的用户界面的认证方式为 **sheme**（即需要输入用户名和密码），而且是 SSH 的 publickey 认证方式时，则用户级别等于用户界面的级别，该级别在用户界面视图下配置，缺省情况下为 0；如果用户登录时使用的用户界面的认证方式为 **none** 或者 **password**（即不需要输入用户名），则用户级别等于用户界面的级别。

表1-6 配置用户界面所能访问的命令级别

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-

操作	命令	说明
配置从当前用户界面登录系统的用户所能访问的命令级别	**user privilege level** *level*	可选 缺省情况下，通过 Console 口登录系统的用户所能访问的命令级别是 3，通过其它用户界面登录系统所能访问的命令级别是 0



说明

- 关于用户级别的详细介绍请参见“系统分册”中的“系统基本配置操作”。
- 用户级别可以通过用户界面也可以通过 AAA 认证参数来配置。哪种配置对用户生效由用户登录时使用的认证方式决定。详细介绍请参见“系统分册”中的“系统基本配置操作”。

1.8 配置 VTY 用户界面访问限制

该配置通过引用 ACL，对 VTY 用户界面的访问权限进行限制。ACL 的相关内容请参见“安全分册”中的“ACL 配置”。

- 如果 VTY 用户界面下没有配置 ACL，则使用该用户界面建立 Telnet 或者 SSH 连接时不进行限制；
- 如果 VTY 用户界面下配置了 ACL，分两种情况：如果能够匹配上 ACL，则按规则中定义的是 permit 还是 deny 来决定是否允许建立连接；如果没有匹配上 ACL，则直接拒绝建立连接。

表1-7 配置 VTY 用户界面访问限制

操作		命令	说明
进入系统视图		**system-view**	-
进入 VTY 用户界面视图		**user-interface** { *first-num1* [*last-num1*] | **vty** *first-num2* [*last-num2*] }	-
配置系统对 VTY 用户界面的访问权限进行限制	引用基本/高级 ACL 对访问权限进行限制	**acl** [**ipv6**] *acl-number* { **inbound** | **outbound** }	二者必选其一 缺省情况下，系统不对访问权限进行限制
	引用二层 ACL 对访问权限进行限制	**acl** *acl-number* **inbound**	



说明

系统将带 **inbound** 参数的基本/高级 ACL、带 **outbound** 参数的基本/高级 ACL、二层 ACL 看成是三种不同类型的 ACL，在同一个 VTY 用户界面下，不同类型的 ACL 可以共存，如果同时配置了不同类型的 ACL，则匹配的顺序由先到后为基本/高级 ACL、二层 ACL；相同类型的 ACL 只能配置一条，以最新的配置为准。

1.9 配置 VTY 用户界面支持的协议

表1-8 配置 VTY 用户界面支持的协议

操作	命令	说明
进入系统视图	**system-view**	-
进入 VTY 用户界面视图	**user-interface** { *first-num1* [*last-num1*] | **vty** *first-num2* [*last-num2*] }	-
设置所在用户界面支持的协议	**protocol inbound** { **all** | **ssh** | **telnet** }	可选 缺省情况下，系统支持 Telnet 和 SSH 协议

注意

- 如果配置用户界面支持 SSH 协议，为确保登录成功，请您务必先配置相应的认证方式为 **scheme**；若配置认证方式为 **password** 或 **none**，则 **protocol inbound ssh** 配置结果将失败。相关配置请参见“系统分册”中的“用户界面命令”的 **authentication-mode** 命令。
- 使用 **protocol inbound** 命令配置的协议将在用户下次使用该用户界面登录时生效。

1.10 配置 AUX 口的重定向功能

利用重定向终端服务功能，当用户 A 通过 Telnet 客户端程序以特定的端口号登录路由器 B 时，可以实现重定向连接，登录到与路由器 AUX 口相连的设备 C 上，用户看到的配置界面就是设备 C 的，用户所作的操作对设备 C 生效。

表1-9 配置 AUX 口的重定向功能

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface aux 0**	-
使能 AUX 口的重定向功能	**redirect enable**	必选 缺省情况下，AUX 口重定向功能被禁止
设置 Telnet 重定向的空闲超时时间	**redirect timeout** *time*	可选 缺省情况下，Telnet 重定向的空闲超时时间为 360 秒
设置在建立 Telnet 重定向连接时不进行 Telnet 选项协商	**redirect refuse-negotiation**	可选 缺省情况下，进行 Telnet 选项协商
设置在建立 Telnet 重定向连接时不对 ACSII 值 0xff 进行转义	**redirect refuse-teltransfer**	可选 缺省情况下，在建立 Telnet 重定向连接时会将 ACSII 值 0xff 转义为 0xff 0xff
设置 Telnet 重定向的监听端口	**redirect listen-port** *port-number*	可选 缺省情况下，Telnet 重定向的监听端口号为用户界面的绝对编号加 2000

操作	命令	说明
设置 Telnet 重定向设备对从 Telnet 客户端接收到的回车符进行处理	**redirect return-deal from-telnet**	可选 缺省情况下，设备不对从 Telnet 客户端接收到的回车符进行处理
设置 Telnet 重定向设备对从终端接收到的回车符进行处理	**redirect return-deal from-terminal**	可选 缺省情况下，设备不对从终端接收到的回车符进行处理
强制断开已经建立的 Telnet 重定向连接	**redirect disconnect**	可选

说明

- 只有先执行 **redirect enable** 命令使能了用户界面的重定向功能后，才能执行其它 **redirect** 命令。
- 重定向设备与被登录设备相连端口对应的用户界面的停止位设置（stopbits）必须相同，否则重定向将失败。

1.11 配置登录用户界面的认证方式

通过设置用户界面的认证方式，可以实现用户在使用指定用户界面登录时是否需要认证，以提高设备的安全性。设备支持的认证方式有 **none**、**password** 和 **scheme** 三种。

- 如果指定认证方式为 **none**，则下次使用该用户界面登录时不需要进行用户名和密码认证，这种情况可能会带来安全隐患。
- 如果指定认证方式为 **password**，则下次使用该用户界面登录时需要进行密码认证，输入空的或者错误密码，均会导致登录失败。如果没有设置认证密码：对于 AUX、VTY 及 MODEM 拨号用户，重新登录时，系统将提示“Login password has not been set !”，登录失败；对于其他用户界面，重新登录时（比如 Console），为了保证设备的可操作性，可以直接登录，不需要输入密码。关闭连接前，请务必设置该用户界面的登录密码。
- 如果指定认证方式为**scheme**，则下次使用该用户界面登录时需要进行用户名和密码认证，用户名或密码错误，均会导致登录失败。关闭连接前，请务设置认证用户名和密码。用户认证又分为本地认证和远程认证，如果采用本地认证，则需要配置本地用户及相应参数（如 表 1-12 所示）；如果采用远程认证，则需要在远程认证服务器上配置用户名和密码。有关用户认证方式及参数的详细介绍请参见“安全分册/AAA RADIUS HWTACACS配置”。缺省情况下，设备对访问用户采用本地认证方式。当用户使用SSH方式登录设备时，该段描述只适用于password认证方式，不适用于publickey认证方式，有关SSH的详细介绍请参见“安全分册”的“SSH2.0 配置”。

表1-10 配置登录用户界面的认证方式为 none

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-

操作	命令	说明
设置登录用户界面的认证方式为 **none**	**authentication-mode none**	必选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console 用户界面不需要认证

表1-11 配置登录用户界面的认证方式为 **password**

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-
设置登录用户界面的认证方式为 **password**	**authentication-mode password**	必选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console 用户界面不需要认证
设置本地认证的密码	**set authentication password** { **cipher** | **simple** } *password*	必选 缺省情况下，没有设置本地认证的密码

表1-12 配置登录用户界面的认证方式为 **scheme**（本地认证）

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-
设置登录用户界面的认证方式为 **scheme**	**authentication-mode scheme** [**command-authorization**]	必选 缺省情况下，VTY、AUX 用户界面认证方式为 **password**，Console 用户界面不需要认证
配置用户界面所能访问的命令级别	请参见“1.7 通过用户界面配置用户级别”	可选 缺省情况下，通过 Console 口登录系统的用户所能访问的命令级别是 3，通过其它用户界面登录系统所能访问的命令级别是 0
设置认证的用户名，并进入本地用户视图	**local-user** *user-name*	必选 缺省情况下，设备中没有设置本地用户
设置认证的密码	**password** { **cipher** | **simple** } *password*	必选
设置用户可以使用的服务类型	**service-type** { **ssh** | **telnet** | **terminal** } * [**level** *level*] }	必选 当使用 VTY 用户界面用户登录时服务类型需要配置为 **ssh** 或 **telnet**，当使用 Console 或者 Aux 用户界面用户登录时服务类型需要配置 **terminal**

说明

local-user、**password** 和 **service-type** 命令的详细介绍请参见“安全分册/AAA 命令”。

1.12 配置启动终端会话及终止当前运行任务的快捷键

表1-13 配置启动终端会话及终止当前运行任务的快捷键

操作	命令	说明
进入系统视图	**system-view**	-
进入用户界面视图	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	-
配置启动终端会话的快捷键	**activation-key** *character*	可选 缺省情况下，按<Enter>键启动终端会话
配置终止当前运行任务的快捷键	**escape-key** { **default** | *character* }	可选 缺省情况下，按<Ctrl+C>组合键终止当前运行的任务

说明

VTY 用户界面不支持 **activation-key** 命令。

1.13 向指定的用户界面发送消息

表1-14 向指定的用户界面发送消息

操作	命令	说明
向指定的用户界面发送消息	**send** { **all** | *num1* | { **aux** | **console** | **vty** } *num2* }	必选 该命令在用户视图下执行

1.14 释放指定用户界面上建立的连接

系统支持多个用户同时对设备进行配置，当管理员在维护设备时，其他在线用户的配置影响到管理员的操作，或者管理员正在进行一些重要配置不想被其他用户干扰的话，管理员可以使用以下命令强制断开该用户的连接。

表1-15 释放指定用户界面上建立的连接

操作	命令	说明
释放指定用户界面上建立的连接	**free user-interface** { *num1* | { **aux** | **console** | **vty** } *num2* }	必选 该命令在用户视图下执行

说明

不能使用该命令释放用户当前自己使用的连接。

1.15 用户界面显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后用户界面的运行情况，通过查看显示信息验证配置的效果。

表1-16 用户界面显示和维护

操作	命令
显示当前正在使用的用户界面以及用户的相关信息	**display users**
显示设备支持的所有用户界面以及用户的相关信息	**display users all**
显示用户界面的相关信息	**display user-interface** [*num1* | { **aux** | **console** | **vty** } *num2*] [**summary**]
显示当前用户配置过的历史命令	**display history-command**

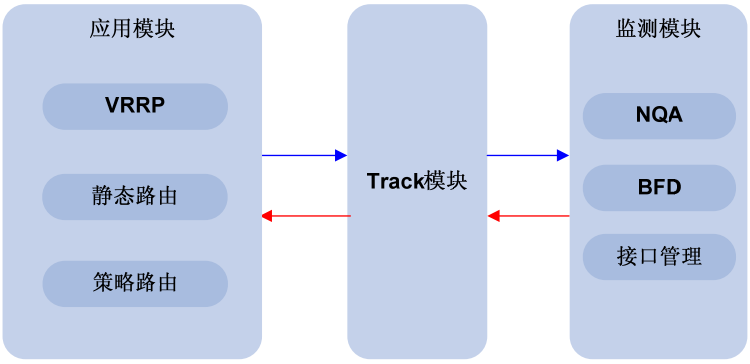
目 录

1 Track配置 1-1
1.1 Track简介 1-1
1.1.1 Track模块与监测模块联动 1-1
1.1.2 Track模块与应用模块联动 1-1
1.2 Track配置任务简介 1-2
1.3 配置Track与监测模块联动 1-2
1.3.1 配置Track与NQA联动 1-2
1.3.2 配置Track与BFD联动 1-2
1.3.3 配置Track与接口管理联动 1-3
1.4 配置Track与应用模块联动 1-3
1.4.1 配置Track与VRRP联动 1-3
1.4.2 配置Track与静态路由联动 1-4
1.4.3 配置Track与策略路由联动 1-5
1.5 Track显示和维护 1-6
1.6 Track典型配置举例 1-6
1.6.1 VRRP、Track与NQA联动配置举例 1-6
1.6.2 静态路由、Track与NQA联动配置举例 1-9
1.6.3 静态路由、Track与BFD联动配置举例 1-12

1 Track 配置

1.1 Track 简介

图1-1 联动功能实现示意图



Track 的用途是实现联动功能。

联动功能由应用模块、Track 模块和监测模块三部分组成。联动功能是指通过建立联动项，实现不同模块之间的联动，即由监测模块通过 Track 模块触发应用模块执行某种操作。监测模块负责对链路状态、网络性能等进行探测，并通过 Track 模块将探测结果通知给应用模块。应用模块感知到网络状态的变化后，及时进行相应的处理，从而避免通信的中断或服务质量的降低。

Track 模块位于应用模块和监测模块之间，主要功能是屏蔽不同监测模块的差异，为应用模块提供统一的接口。

1.1.1 Track 模块与监测模块联动

用户通过配置，建立 Track 模块和监测模块之间的联动关系。监测模块负责对接口状态、链路状态等进行探测，并将探测结果通知给 Track 模块，以便及时改变 Track 项的状态：

- 如果探测成功，则对应 Track 项的状态为 Positive；
- 如果探测失败，则对应 Track 项的状态为 Negative。

目前，监测模块包括 NQA（Network Quality Analyzer，网络质量分析）、BFD（Bidirectional Forwarding Detection，双向转发检测）和接口管理：

- NQA 的详细介绍，请参见“系统分册”中的“NQA 配置”；
- BFD 的详细介绍，请参见“IP 路由分册”中的“BFD 配置”；
- 接口管理用来监视接口的物理状态或网络层协议状态。接口的物理状态或网络层协议状态为 up 时，Track 项的状态为 Positive；接口的物理状态或网络层协议状态为 down 时，Track 项的状态为 Negative。

1.1.2 Track 模块与应用模块联动

用户通过配置，建立 Track 模块和应用模块之间的联动关系。Track 项的状态发生变化后，Track 模块将通知应用模块进行相应的处理。

目前，可以与 Track 模块实现联动功能的应用模块包括：

- VRRP
- 静态路由
- 策略路由

1.2 Track 配置任务简介

为了实现联动功能，需要在 Track 与监测模块、Track 与应用模块之间建立联动关系。

表1-1 Track 配置任务简介

配置任务		说明	详细配置
配置 Track 与监测模块联动	配置 Track 与 NQA 联动	三者必选其一	1.3.1
	配置 Track 与 BFD 联动		1.3.2
	配置 Track 与接口管理联动		1.3.3
配置 Track 与应用模块联动	配置 Track 与 VRRP 联动	三者必选其一	1.4.1
	配置 Track 与静态路由联动		1.4.2
	配置 Track 与策略路由联动		1.4.3

1.3 配置 Track 与监测模块联动

1.3.1 配置 Track 与 NQA 联动

通过该配置可以在 Track 模块和 NQA 之间建立联动关系。由 NQA 对链路状态等进行探测，并将探测结果通知给 Track 模块。

表1-2 配置 Track 与 NQA 联动

操作	命令	说明
进入系统视图	**system-view**	-
创建与 NQA 测试组中指定联动项关联的 Track 项	**track** *track-entry-number* **nqa entry** *admin-name operation-tag* **reaction** *item-num*	必选 缺省情况下，没有创建与 NQA 测试组中指定联动项关联的 Track 项



说明

配置 Track 项时，引用的 NQA 测试组和联动项可以不存在，此时该 Track 项的状态为 Invalid。

1.3.2 配置 Track 与 BFD 联动

通过该配置可以在 Track 模块和 BFD 之间建立联动关系。由 BFD 进行探测，并将探测结果通知给 Track 模块。

表1-3 配置 Track 与 BFD 联动

操作	命令	说明
进入系统视图	**system-view**	-
创建和 BFD 会话关联的 Track 项	**track** *track-entry-number* **bfd echo interface** *interface-type interface-number* **remote ip** *remote-ip* **local ip** *local-ip*	必选 缺省情况下，没有创建和 BFD 会话关联的 Track 项

1.3.3 配置 Track 与接口管理联动

通过该配置可以在 Track 模块和接口管理之间建立联动关系，接口管理负责将接口物理状态或网络层协议状态的变化通知给 Track 模块。

表1-4 配置 Track 与接口管理联动

操作	命令	说明
进入系统视图	**system-view**	-
创建和接口管理关联的 Track 项，监视接口的物理状态	**track** *track-entry-number* **interface** *interface-type interface-number*	二者必选其一 缺省情况下，没有创建和接口管理关联的 Track 项
创建和接口管理关联的 Track 项，监视三层接口的网络层协议状态	**track** *track-entry-number* **interface** *interface-type interface-number* **protocol** { **ipv4** | **ipv6** }	

1.4 配置 Track 与应用模块联动

1.4.1 配置 Track 与 VRRP 联动

通过在 Track 模块和 VRRP 之间建立联动，可以实现：

- 根据上行链路的状态，改变路由器的优先级。当上行链路出现故障，局域网内的主机无法通过路由器访问外部网络时，被监视 Track 项的状态会变为 negative，并将路由器的优先级降低指定的数额。从而，使得备份组内其它路由器的优先级高于这个路由器的优先级，成为 Master 路由器，保证局域网内主机与外部网络的通信不会中断。
- 在 Backup 路由器上监视 Master 路由器的状态。当 Master 路由器出现故障时，工作在切换模式的 Backup 路由器能够迅速成为 Master 路由器，以保证通信不会中断。

1. 配置准备

在配置 VRRP 监视 Track 项之前，需要先在接口上创建备份组并配置虚拟 IP 地址。

2. 配置过程

表1-5 配置 Track 与 VRRP 联动

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
创建备份组，并配置备份组的虚拟 IP 地址	**vrrp** [**ipv6**] **vrid** *virtual-router-id* **virtual-ip** *virtual-address*	必选 缺省情况下，没有创建备份组
配置 VRRP 监视指定的 Track 项	**vrrp** [**ipv6**] **vrid** *virtual-router-id* **track** *track-entry-number* [**reduced** *priority-reduced* | **switchover**]	必选 缺省情况下，没有指定被监视的 Track 项

说明

- 不允许对 IP 地址拥有者进行监视指定 Track 项的配置。
- 被监视 Track 项的状态由 Negative 变为 Positive 后，对应的路由器优先级会自动恢复。
- 被监视的 Track 项可以是未创建的 Track 项。可以通过 **vrrp vrid track** 命令指定监视的 Track 项后，再通过 **track** 命令创建该 Track 项。
- VRRP 配置的详细介绍，请参见“系统分册”中的“VRRP 配置”。

1.4.2 配置 Track 与静态路由联动

通过在 Track 模块和静态路由之间建立联动，可以实现静态路由有效性的实时判断。

如果在配置静态路由时只指定了下一跳而没有指定出接口，可以通过将静态路由与 Track 项关联，根据 Track 项的状态来判断静态路由的有效性：

- 当 Track 项状态为 Positive 或 Invalid 时，配置的静态路由将生效；
- 当 Track 项状态为 Negative 时，配置的静态路由无效。

表1-6 配置 Track 与静态路由联动

操作	命令	说明
进入系统视图	**system-view**	-
配置通过 Track 与静态路由联动，检测静态路由下一跳是否可达	**ip route-static** *dest-address* { *mask* | *mask-length* } { *next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } **track** *track-entry-number* [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	二者必选其一 缺省情况下，没有配置 Track 与静态路由联动
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* | *mask-length* } { *next-hop-address* **track** *track-entry-number* [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address* **track** *track-entry-number* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

说明

- 配置静态路由与 Track 联动时，该条静态路由可以是已经创建的，也可以是未创建的。对于已经创建的静态路由，只是将静态路由与 Track 项关联；对于未创建的静态路由，首先要生成该静态路由，然后将其与 Track 项关联。
- 静态路由关联的 Track 项可以是未创建的 Track 项。通过 **track** 命令创建 Track 项后，联动功能开始生效。
- 如果 Track 模块通过 NQA 探测私网静态路由中下一跳的可达性，静态路由下一跳的 VPN 实例名与 NQA 测试组配置的实例名必须相同，才能进行正常的探测。
- 需要注意在静态路由进行迭代时，Track 项监测的应该是静态路由真正的下一跳，而不是配置的下一跳。否则，可能导致错误地将有效路由判断为无效路由。
- 静态路由配置的详细介绍，请参见“IP 路由分册”中的“静态路由配置”。

1.4.3 配置 Track 与策略路由联动

IP 单播策略路由通过关联 Track 机制，增强了应用的灵活性，以及策略路由对网络环境的动态感知能力。

策略路由可以在配置报文的发送接口、缺省发送接口、下一跳、缺省下一跳时与 Track 项关联，通过 Track 项的状态来动态地决定策略的可用性。当应用动作增加 Track 关联后：

- 如果事件发生时，Track 项状态为 Positive 或 Invalid，则该配置项有效，可以指导转发；
- 如果事件发生时，Track 项状态为 Negative，则该配置项无效，转发时忽略该配置项。

1. 配置准备

配置 Track 与策略路由联动前，需要先创建策略或一个策略节点，并配置匹配规则。

2. 配置过程

表1-7 配置 Track 与策略路由联动

操作	命令	说明
进入系统视图	**system-view**	-
创建策略或一个策略节点，并进入该策略视图	**policy-based-route** *policy-name* { **deny** | **permit** } **node** *node-number*	必选
设置 IP 报文长度匹配条件	**if-match packet-length** *min-len max-len*	二者必选其一
设置 ACL 匹配条件	**if-match acl** *acl-number*	
设置报文的发送接口，并与 Track 项关联	**apply output-interface** *interface-type interface-number* [**track** *track-entry-number*] [*interface-type interface-number* [**track** *track-entry-number*]]	四者中至少选择其一
设置报文的下一跳，并与 Track 项关联	**apply ip-address next-hop** *ip-address* [**track** *track-entry-number*] [*ip-address* [**track** *track-entry-number*]]	
设置报文缺省发送接口，并与 Track 项关联	**apply default output-interface** *interface-type interface-number* [**track** *track-entry-number*] [*interface-type interface-number* [**track** *track-entry-number*]]	
设置报文缺省下一跳，并与 Track 项关联	**apply ip-address default next-hop** *ip-address* [**track** *track-entry-number*] [*ip-address* [**track** *track-entry-number*]]	

说明

- 策略路由关联的 Track 项可以是未创建的 Track 项。通过 **track** 命令创建 Track 项后，联动功能开始生效。
- 策略路由配置的详细介绍，请参见“IP 业务分册”中的“IP 单播策略路由配置”。

1.5 Track 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 Track 的运行情况，通过查看显示信息验证配置的效果。

表1-8 Track 显示和维护

操作	命令
显示 Track 项的信息	**display track** { *track-entry-number* | **all** }

1.6 Track 典型配置举例

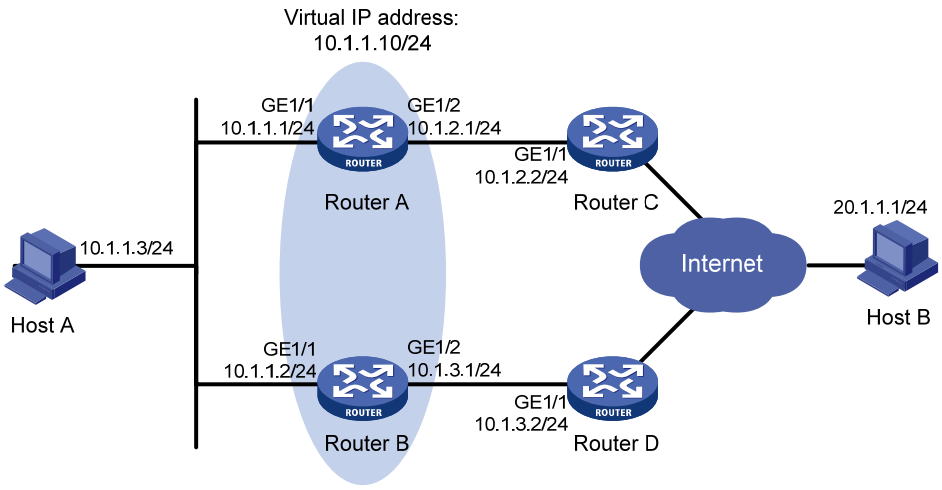
1.6.1 VRRP、Track 与 NQA 联动配置举例

1. 组网需求

- Host A 需要访问 Internet 上的 Host B，Host A 的缺省网关为 10.1.1.10/24；
- Router A 和 Router B 属于虚拟 IP 地址为 10.1.1.10 的备份组 1；
- 当 Router A 正常工作时，Host A 发送给 Host B 的报文通过 Router A 转发；当通过 NQA 监测到 Router A 上行链路不通时，Host A 发送给 Host B 的报文通过 Router B 转发。

2. 组网图

图1-2 VRRP、Track 与 NQA 联动配置组网图



3. 配置步骤

(1) 按照 图 1-2配置各接口的IP地址，具体配置过程略。

(2) 在 Router A 上配置 NQA 测试组

```
<RouterA> system-view
```

创建管理员名为 admin、操作标签为 test 的 NQA 测试组。

```
[RouterA] nqa entry admin test
```

配置测试类型为 ICMP-echo。

```
[RouterA-nqa-admin-test] type icmp-echo
```

配置目的地址为 10.1.2.2。

```
[RouterA-nqa-admin-test-icmp-echo] destination ip 10.1.2.2
```

测试频率为 100ms。

```
[RouterA-nqa-admin-test-icmp-echo] frequency 100
```

配置联动项 1（连续失败 5 次触发联动）。

```
[RouterA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail threshold-type
consecutive 5 action-type trigger-only
[RouterA-nqa-admin-test-icmp-echo] quit
```

启动探测。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

(3) 在 Router A 上配置 Track 项

配置 Track 项 1，关联 NQA 测试组（管理员为 admin，操作标签为 test）的联动项 1。

```
[RouterA] track 1 nqa entry admin test reaction 1
```

(4) 在 Router A 上配置 VRRP

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 10.1.1.10。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] vrrp vrid 1 virtual-ip 10.1.1.10
```

设置 Router A 在备份组 1 中的优先级为 110。

```
[RouterA-GigabitEthernet1/1] vrrp vrid 1 priority 110
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterA-GigabitEthernet1/1] vrrp vrid 1 authentication-mode simple hello
```

设置 Master 路由器发送 VRRP 报文的间隔时间为 5 秒。

```
[RouterA-GigabitEthernet1/1] vrrp vrid 1 timer advertise 5
```

设置 Router A 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterA-GigabitEthernet1/1] vrrp vrid 1 preempt-mode timer delay 5
```

设置监视 Track 项。

```
[RouterA-GigabitEthernet1/1] vrrp vrid 1 track 1 reduced 30
```

(5) 在 Router B 上配置 VRRP

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/1
```

创建备份组 1，并配置备份组 1 的虚拟 IP 地址为 10.1.1.10。

```
[RouterB-GigabitEthernet1/1] vrrp vrid 1 virtual-ip 10.1.1.10
```

设置备份组的认证方式为 SIMPLE，认证字为 hello。

```
[RouterB-GigabitEthernet1/1] vrrp vrid 1 authentication-mode simple hello
```

设置 Master 路由器发送 VRRP 报文的间隔时间为 5 秒。

```
[RouterB-GigabitEthernet1/1] vrrp vrid 1 timer advertise 5
```

设置 Router B 工作在抢占方式，抢占延迟时间为 5 秒。

```
[RouterB-GigabitEthernet1/1] vrrp vrid 1 preempt-mode timer delay 5
```

(6) 验证配置结果

配置完成后，在 Host A 上可以 ping 通 Host B。通过 **display vrrp** 命令查看配置后的结果。

显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/1] display vrrp verbose
IPv4 Standby Information:
     Run Mode       : Standard
     Run Method     : Virtual MAC
 Total number of virtual routers : 1
   Interface GigabitEthernet1/1
     VRID           : 1                  Adver Timer  : 5
     Admin Status   : Up                 State        : Master
     Config Pri     : 110                Running Pri  : 110
     Preempt Mode   : Yes                Delay Time   : 5
     Auth Type      : Simple             Key          : hello
     Virtual IP     : 10.1.1.10
     Virtual MAC    : 0000-5e00-0101
     Master IP      : 10.1.1.1
   VRRP Track Information:
     Track Object   : 1                  State : Positive   Pri Reduced : 30
```

显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/1] display vrrp verbose
IPv4 Standby Information:
     Run Mode       : Standard
     Run Method     : Virtual MAC
 Total number of virtual routers : 1
   Interface GigabitEthernet1/1
     VRID           : 1                  Adver Timer  : 5
     Admin Status   : Up                 State        : Backup
     Config Pri     : 100                Running Pri  : 100
     Preempt Mode   : Yes                Delay Time   : 5
     Auth Type      : Simple             Key          : hello
     Virtual IP     : 10.1.1.10
     Master IP      : 10.1.1.1
```

以上显示信息表示在备份组 1 中 Router A 为 Master 路由器，Router B 为 Backup 路由器，Host A 发送给 Host B 的报文通过 Router A 转发。

Router A 与 Router C 不通时，在 Host A 上仍然可以 ping 通 Host B。通过 **display vrrp** 命令查看备份组的信息。

Router A 与 Router C 不通时，显示 Router A 上备份组 1 的详细信息。

```
[RouterA-GigabitEthernet1/1] display vrrp verbose
IPv4 Standby Information:
     Run Mode       : Standard
     Run Method     : Virtual MAC
 Total number of virtual routers : 1
   Interface GigabitEthernet1/1
     VRID           : 1                  Adver Timer   : 5
     Admin Status   : Up                 State         : Backup
     Config Pri     : 110                Running Pri   : 80
     Preempt Mode   : Yes                Delay Time    : 5
     Auth Type      : Simple             Key           : hello
     Virtual IP     : 10.1.1.10
     Master IP      : 10.1.1.2
   VRRP Track Information:
     Track Object   : 1                  State : Negative   Pri Reduced : 30
```

Router A 与 Router C 不通时，显示 Router B 上备份组 1 的详细信息。

```
[RouterB-GigabitEthernet1/1] display vrrp verbose
IPv4 Standby Information:
     Run Mode       : Standard
     Run Method     : Virtual MAC
 Total number of virtual routers : 1
   Interface GigabitEthernet1/1
     VRID           : 1                  Adver Timer   : 5
     Admin Status   : Up                 State         : Master
     Config Pri     : 100                Running Pri   : 100
     Preempt Mode   : Yes                Delay Time    : 5
     Auth Type      : Simple             Key           : hello
     Virtual IP     : 10.1.1.10
     Virtual MAC    : 0000-5e00-0101
     Master IP      : 10.1.1.2
```

以上显示信息表示 Router A 与 Router C 不通时，Router A 的优先级降低为 80，成为 Backup 路由器，Router B 成为 Master 路由器，Host A 发送给 Host B 的报文通过 Router B 转发。

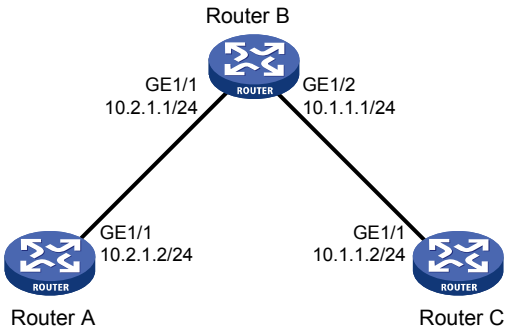
1.6.2 静态路由、Track 与 NQA 联动配置举例

1. 组网需求

- Router A 到达 Router C 的静态路由下一跳为 Router B。
- 在 Router A 上通过静态路由、Track 与 NQA 联动，对到达 Router C 的静态路由有效性进行实时判断。

2. 组网图

图1-3 静态路由、Track 与 NQA 联动配置组网图



3. 配置步骤

(1) 按照 图 1-3配置各接口的IP地址，具体配置过程略。

(2) 在 Router A 上配置静态路由，并与 Track 项关联

配置到达 Router C 的静态路由下一跳地址为 10.2.1.1，并配置静态路由与 Track 项 1 关联。

```
<RouterA> system-view
[RouterA] ip route-static 10.1.1.2 24 10.2.1.1 track 1
```

(3) 在 Router A 上配置 NQA 测试组

创建管理员名为 admin、操作标签为 test 的 NQA 测试组。

```
[RouterA] nqa entry admin test
```

配置测试类型为 ICMP-echo。

```
[RouterA-nqa-admin-test] type icmp-echo
```

配置目的地址为 10.2.1.1。

```
[RouterA-nqa-admin-test-icmp-echo] destination ip 10.2.1.1
```

测试频率为 100ms。

```
[RouterA-nqa-admin-test-icmp-echo] frequency 100
```

配置联动项 1（连续失败 5 次触发联动）。

```
[RouterA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail threshold-type
consecutive 5 action-type trigger-only
[RouterA-nqa-admin-test-icmp-echo] quit
```

启动探测。

```
[RouterA] nqa schedule admin test start-time now lifetime forever
```

(4) 在 Router A 上配置 Track 项

配置 Track 项 1，关联 NQA 测试组（管理员为 admin，操作标签为 test）的联动项 1。

```
[RouterA] track 1 nqa entry admin test reaction 1
```

(5) 验证配置结果

显示 Router A 上 Track 项的信息。

```
[RouterA] display track all
Track ID: 1
```

```
  Status: Positive
  Reference object:
    NQA entry: admin test
    Reaction: 1
```

显示 Router A 的路由表。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         Static 60   0            10.2.1.1        GE1/1
10.2.1.0/24         Direct 0    0            10.2.1.2        GE1/1
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

以上显示信息表示，NQA 测试的结果为下一跳地址 10.2.1.1 可达（Track 项状态为 Positive），配置的静态路由生效。

在 Router B 上删除接口 GigabitEthernet1/1 的 IP 地址。

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] undo ip address
```

显示 Router A 上 Track 项的信息。

```
[RouterA] display track all
Track ID: 1
  Status: Negative
  Reference object:
    NQA entry: admin test
    Reaction: 1
```

显示 Router A 的路由表。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 4        Routes : 4

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.2.1.0/24         Direct 0    0            10.2.1.2        GE1/1
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

以上显示信息表示，NQA 测试的结果为下一跳地址 10.2.1.1 不可达（Track 项状态为 Negative），配置的静态路由无效。

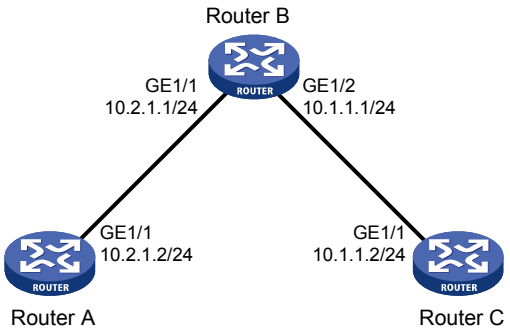
1.6.3 静态路由、Track 与 BFD 联动配置举例

1. 组网需求

- Router A 到达 Router C 的静态路由下一跳为 Router B。
- 在 Router A 上通过静态路由、Track 与 BFD 联动，对到达 Router C 的静态路由有效性进行实时判断。

2. 组网图

图1-4 静态路由、Track 与 BFD 联动配置组网图



3. 配置步骤

(1) 按照 图 1-4配置各接口的IP地址，具体配置过程略。

(2) 在 Router A 上配置静态路由，并与 Track 项关联。

配置到达 Router C 的静态路由下一跳地址为 10.2.1.1，并配置静态路由与 Track 项 1 关联。

```
<RouterA> system-view
[RouterA] ip route-static 10.1.1.2 24 10.2.1.1 track 1
```

(3) 在 Router A 上配置 BFD 基本功能

配置 BFD echo 报文的源地址为 10.10.10.10。

```
[RouterA] bfd echo-source-ip 10.10.10.10
```

(4) 在 Router A 上创建和 BFD 会话关联的 Track 项

创建和 BFD 会话关联的 Track 项 1，检测 Router A 是否可以与静态路由的下一跳 Router B 互通。

```
[RouterA] track 1 bfd echo interface gigabitethernet 1/1 remote ip 10.2.1.1 local ip 10.2.1.2
```

(5) 验证配置结果

显示 Router A 上 Track 项的信息。

```
[RouterA] display track all
Track ID: 1
  Status: Positive
  Reference object:
    BFD Session:
    Packet type: Echo
    Interface  : GigabitEthernet1/1
    Remote IP  : 10.2.1.1
    Local IP   : 10.2.1.2
```

显示 Router A 的路由表。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         Static 60   0            10.2.1.1        GE1/1
10.2.1.0/24         Direct 0    0            10.2.1.2        GE1/1
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

以上显示信息表示，BFD 检测的结果为下一跳地址 10.2.1.1 可达（Track 项状态为 Positive），配置的静态路由生效。

在 Router B 上删除接口 GigabitEthernet1/1 的 IP 地址。

```
<RouterB> system-view
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] undo ip address
```

显示 Router A 上 Track 项的信息。

```
[RouterA] display track all
Track ID: 1
  Status: Negative
  Reference object:
    BFD Session:
    Packet type: Echo
    Interface  : GigabitEthernet1/1
    Remote IP  : 10.2.1.1
    Local IP   : 10.2.1.2
```

显示 Router A 的路由表。

```
[RouterA] display ip routing-table
Routing Tables: Public
         Destinations : 4        Routes : 4

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.2.1.0/24         Direct 0    0            10.2.1.2        GE1/1
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

以上显示信息表示，BFD 检测的结果为下一跳地址 10.2.1.1 不可达（Track 项状态为 Negative），配置的静态路由无效。

目 录

1 IPv6 NetStream配置 1-1
1.1 IPv6 NetStream简介 1-1
1.1.1 IPv6 NetStream概述 1-1
1.1.2 IPv6 Netstream聚合功能简介 1-1
1.1.3 IPv6 NetStream实现 1-2
1.2 IPv6 NetStream配置任务简介 1-2
1.3 配置IPv6 NetStream统计功能 1-2
1.4 配置IPv6 NetStream聚合统计功能 1-3
1.4.1 配置准备 1-3
1.4.2 配置IPv6 NetStream聚合统计功能 1-3
1.5 配置IPv6 NetStream统计输出报文的属性 1-3
1.5.1 配置IPv6 Netstream统计输出报文的属性 1-3
1.5.2 配置IPv6 NetStream统计输出报文版本 9 模板的刷新率 1-4
1.6 配置IPv6 NetStream的流老化 1-4
1.6.1 流老化简介 1-4
1.6.2 配置IPv6 NetStream的流老化 1-5
1.7 IPv6 NetStream显示和维护 1-5
1.8 IPv6 NetStream典型配置举例 1-6
1.8.1 配置IPv6 NetStream分类统计功能 1-6
1.8.2 配置IPv6 NetStream统计输出功能 1-7

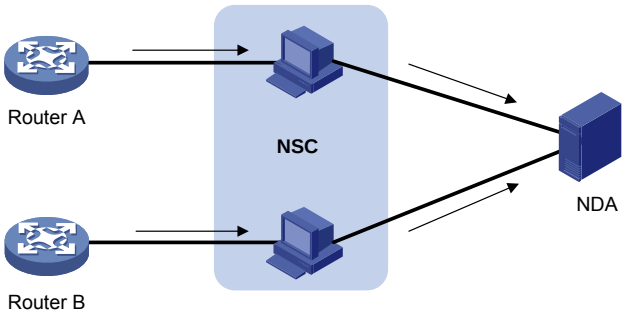
1 IPv6 NetStream 配置

1.1 IPv6 NetStream 简介

1.1.1 IPv6 NetStream 概述

IPv6 NetStream 提供报文统计功能，它根据 IPv6 报文的目的 IP 地址、源 IP 地址、目的端口号、源端口号、协议号、流标签以及输入或者输出接口来区分流，并针对不同的流进行独立的数据统计。

图1-1 NetStream 数据采集和分析过程示意图



如上图所示，IPv6 NetStream 进行数据采集和分析的过程如下：

(1) 配置了 IPv6 NetStream 功能的设备把采集到的关于流的详细信息定期发送给 NSC（NetStream Collector，网络流数据收集器）；

(2) 信息由 NSC 初步处理后发送给 NDA（NetStream Data Analyzer，网络流数据分析器）；

(3) NDA 对数据进行分析，以用于计费、网络规划等应用。

1.1.2 IPv6 Netstream 聚合功能简介

IPv6 NetStream 支持聚合功能。老化的流在输出前先按照一定的规则进行分类，生成聚合的信息后再发送出去。

在目前的实现中，支持 6 种聚合方式，如下表所示。

表1-1 IPv6 NetStream 的 6 种聚合方式

聚合方式	分类依据
自治系统聚合	源 AS 号、目的 AS 号、输入接口索引、输出接口索引
协议-端口聚合	协议号、源端口、目的端口
源前缀聚合	源 AS 号、源掩码长度、源前缀、输入接口索引
目的前缀聚合	目的 AS 号、目的掩码长度、目的前缀、输出接口索引
源和目的前缀聚合	源 AS 号、目的 AS 号、源掩码长度、目的掩码长度、源前缀、目的前缀、输入接口索引、输出接口索引
BGP 下一跳聚合	BGP 下一跳地址、输出接口索引

说明

- 系统根据所选择聚合方式的分类依据，将符合对应聚合方式的多条流合并为一条聚合流，对应一条聚合记录。6 种聚合方式相互独立，可以同时配置。
- 在统计 AS 域号时如果流量没有按照 BGP 的路由表进行转发则系统无法统计出 AS 域号。
- 在统计 BGP 下一跳地址时如果流量没有按照 BGP 的路由表进行转发则系统无法统计出 BGP 下一跳地址。

1.1.3 IPv6 NetStream 实现

当设备启用 IPv6 NetStream 功能后，IPv6 流信息首先被存储在设备的 IPv6 NetStream 缓冲区中，当设定的老化时间超时或者执行了强制老化命令后，设备就将流信息通过 IPv6 NetStream 统计输出报文（即，流信息经过 UDP 封装以后产生的报文）发送给 NSC，并从设备的 IPv6 NetStream 缓冲区中将该流信息删除。

IPv6 NetStream 统计只输出版本 9。如果没有配置 IPv6 NetStream 聚合方式，设备将直接通过版本 9 报文将老化后的 IPv6 流信息发送给 NSC；如果配置了 IPv6 NetStream 聚合方式，设备将按照一定的规则对 IPv6 流信息进行分类、合并后生成聚合信息，再通过版本 9 报文发送给 NSC。MPLS 流信息通过版本 9 报文发送。

1.2 IPv6 NetStream 配置任务简介

表1-2 IPv6 NetStream 配置任务简介

配置任务	说明	详细配置
配置 IPv6 NetStream 统计功能	必选	1.3
配置 IPv6 NetStream 聚合统计功能	可选	1.4
配置 IPv6 NetStream 统计输出报文的属性	必选	1.5
配置 IPv6 NetStream 的流老化	可选	1.6

1.3 配置 IPv6 NetStream 统计功能

表1-3 配置 IPv6 NetStream 统计功能

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 NetStream 流缓存区大小，即所能容纳 IPv6 NetStream 流的最大数量，以及置在达到最大流表个数时，是进行强制老化，还是禁止创建流	**ipv6 netstream max-entry** { *max-entries* | **aging** | **disable-caching** }	可选 缺省情况下，NetStream 流缓存区大小为 20000，在达到最大流表个数时进行强制老化 该命令仅 SR6602 支持
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置当前接口下的 IPv6 NetStream 统计功能	**ipv6 netstream { inbound | outbound }**	必选 缺省情况下，当前接口下未使能 IPv6 NetStream 统计功能

1.4 配置 IPv6 NetStream 聚合统计功能

1.4.1 配置准备

在配置此统计功能前，必须配置 IPv6 NetStream 统计功能。

1.4.2 配置 IPv6 NetStream 聚合统计功能

表1-4 配置 IPv6 NetStream 聚合统计功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 NetStream 聚合视图	**ipv6 netstream aggregation { as | destination-prefix | prefix | protocol-port | source-prefix | bgp-nexthop }**	必选
使能该类的 IPv6 NetStream 聚合功能	**enable**	必选 缺省情况下，不使能任何 IPv6 NetStream 聚合功能

1.5 配置 IPv6 NetStream 统计输出报文的属性

1.5.1 配置 IPv6 Netstream 统计输出报文的属性

用户可以对 IPv6 NetStream 统计输出报文的源接口和目的 IP 地址及目的端口号进行设置，其中源接口 IP 地址将作为此报文的源 IP 地址。

IPv6 NetStream 统计输出报文支持 BGP 的自治系统选项，用户可以选择以起始自治系统号或者以邻接自治系统号作为给定 IP 地址所在的自治系统号。

IPv6 NetStream 统计输出报文还支持 BGP 下一跳选项，用户可以选择是否在流信息中记录 BGP 下一跳地址。

表1-5 配置 IPv6 NetStream 统计输出报文的属性

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 NetStream 统计输出报文的目的 IP 地址、端口号、私网路由	**ipv6 netstream export host** *ip-address udp-port* **[vpn-instance** *vpn-instance-name* **]**	必选 缺省情况下，系统视图下没有配置目的地址和目的端口号，聚合视图下目的地址和目的端口号为系统视图下所配置的值
配置 IPv6 NetStream 统计输出报文的源接口	**ipv6 netstream export source interface** *interface-type interface-number*	可选 缺省情况下，采用统计报文输出的接口作为源接口

操作	命令	说明
配置 IPv6 NetStream 统计输出报文以及其自治系统选项、BGP 下一跳选项	**ipv6 netstream export version 9 [origin-as | peer-as] [bgp-nexthop]**	可选 缺省情况下，IPv6 普通流信息、IPv6 聚合统计流信息和带 IPv6 选项信息的 MPLS 流信息都通过版本 9 的 NetStream 统计输出报文发送。自治系统选项使用邻接自治系统号（**peer-as**），流信息中不记录 BGP 下一跳地址
配置输出速率限制，并设定 IPv6 NetStream 统计输出报文的最大输出速率	**ipv6 netstream export rate** *rate*	可选 缺省情况下，IPv6 NetStream 统计输出报文的输出速率不受限制

注意

- 在聚合视图下，用户配置的 NetStream 统计输出报文的属性，仅影响聚合报文；而系统视图下的配置不仅影响普通报文，并且当聚合视图下没有配置以上属性时，会对聚合报文生效。
- 自治系统相关知识，请参考“IP 路由分册”中的“BGP 配置”。

1.5.2 配置 IPv6 NetStream 统计输出报文版本 9 模板的刷新率

MPLS 流信息通过版本 9 的 NetStream 统计输出报文发送。

表1-6 配置 IPv6 NetStream 统计输出报文版本 9 模板的刷新率

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 NetStream 统计输出报文版本 9 模板的包刷新率	**ipv6 netstream export v9-template refresh-rate packet** *packets*	可选 缺省情况下，每隔 20 个包发送一次模板
配置 IPv6 NetStream 统计输出报文版本 9 模板的时间刷新率	**ipv6 netstream export v9-template refresh-rate time** *minutes*	可选 缺省情况下，每隔 30 分钟发送一次模板

注意

- 网管软件不可能把接收到的所有的版本 9 数据模板都永久保存下来，由于容量问题原有的模板会被老化，需要适时更新模板，用户可以配置版本 9 模板的包刷新率，及时更新模板。
- 包刷新率和时间刷新率其中只要满足其中任意条件，设备会发送模板给网管软件。

1.6 配置 IPv6 NetStream 的流老化

1.6.1 流老化简介

在实际网络环境中，可能在很短的时间内产生大量流，这就需要根据一定的算法把当前一部分流从缓冲区中删除，称为老化（aging）。

NetStream 的老化机制有三种：按时老化、强制老化、TCP 的 FIN 和 RST 报文触发老化（该形式的老化在 TCP 连接拆除时自动进行）。

1. 按时老化

有两种按时老化方式：

- 按不活跃时间老化：对从最后一个报文流过到当前的时间超过 **inactive timeout** 的流进行老化。
- 按活跃时间老化：对从第一个报文流过到当前的时间超过 **active timeout** 的流进行老化。

2. 强制老化

执行强制老化命令，用户可以将 IPv6 NetStream 缓冲区中所有流老化，并清除 IPv6 NetStream 缓冲区的状态信息、IPv6 NetStream 输出报文信息、IPv6 NetStream v9 模板的状态信息，输出报文信息及 v9 模板的状态信息从强制老化时刻开始重新统计。

1.6.2 配置 IPv6 NetStream 的流老化

表1-7 配置 IPv6 NetStream 的流老化

操作		命令	说明
进入系统视图		**system-view**	-
配置 IPv6 NetStream 的老化方式	配置 IPv6 NetStream 活跃老化时间	**ipv6 netstream timeout active** *minutes*	可选 缺省情况下，IPv6 普通流信息的活跃老化时间为 30 分钟
	配置 IPv6 NetStream 不活跃老化时间	**ipv6 netstream timeout inactive** *seconds*	可选 缺省情况下，IPv6 普通流信息的不活跃老化时间为 30 秒

注意

IPv6 NetStream 活跃老化时间和不活跃老化时间可以同时配置，满足任一个老化时间就会对流进行老化，时间精度为 10 秒钟。

1.7 IPv6 NetStream 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 NetStream 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv6 NetStream 的统计信息。

表1-8 IPv6 NetStream 显示和维护

操作		命令
查看 IPv6 NetStream 流缓存区的配置和状态信息	集中式设备	**display ipv6 netstream cache** [**verbose**]
	分布式设备	**display ipv6 netstream cache** [**slot** *slot-id*] [**verbose**]
查看 IPv6 NetStream 统计输出报文信息		**display ipv6 netstream export**
查看 IPv6 NetStream V9 模板的配置和状态信息	集中式设备	**display ipv6 netstream template**
	分布式设备	**display ipv6 netstream template** [**slot** *slot-id*]

操作	命令
清除将流缓存区中所有流强制老化，并清除 IPv6 NetStream 缓冲区的状态信息和输出报文信息	**reset ipv6 netstream statistics**

1.8 IPv6 NetStream 典型配置举例

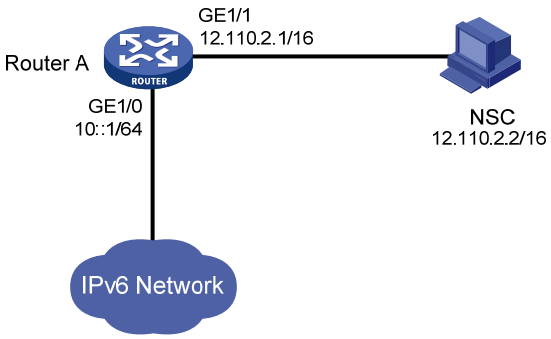
1.8.1 配置 IPv6 NetStream 分类统计功能

1. 组网需求

如 图 1-2所示，在Router A上配置IPv6 NetStream入统计和出统计，并将统计结果发送到NSC 12.110.2.2/16 的 5000 端口。

2. 组网图

图1-2 配置 IPv6 NetStream 分类统计功能组网图



3. 配置步骤

配置 GigabitEthernet1/0，在此接口上启动 IPv6 NetStream 入统计。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ipv6 address 10::1/64
[RouterA-GigabitEthernet1/0] ipv6 netstream inbound
[RouterA-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1，在此接口上启动 IPv6 NetStream 出统计。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 12.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/1] ipv6 address 20::1/64
[RouterA-GigabitEthernet1/1] ipv6 netstream outbound
[RouterA-GigabitEthernet1/1] quit
```

配置 IPv6 NetStream 统计输出报文的目的 IP 地址和目的端口（即 NSC 的地址以及端口），源接口采用缺省配置。

```
[RouterA] ipv6 netstream export host 12.110.2.2 5000
```

1.8.2 配置 IPv6 NetStream 统计输出功能

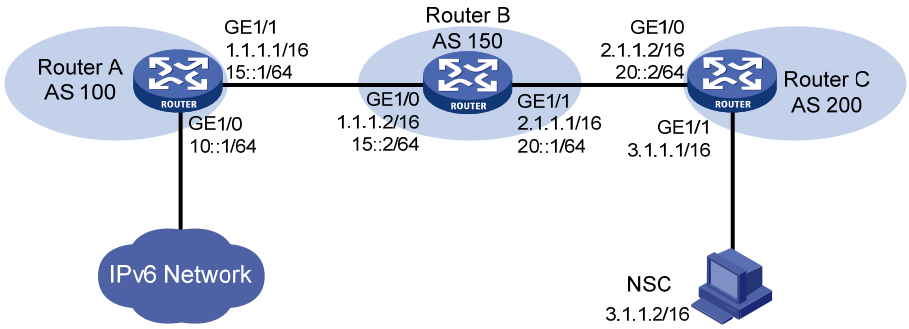
1. 组网需求

在 Router A 上配置 IPv6 NetStream，普通流统计输出报文使用版本 9 格式输出到 NSC 3.1.1.2/16 的 5000 端口，5 种 IPv6 NetStream 聚合（**as**、**protocol-port**、**source-prefix**、**destination-prefix** 和 **prefix**）的聚合流统计输出报文使用版本 9 格式分别输出到该 NSC 的 2000、3000、4000、6000 和 7000 端口。

所有路由器均运行 IPv4 BGP 和 IPv6 BGP 路由协议。Router A、Router B 和 Router C 之间建立了 EBGP 邻居关系。

2. 组网图

图1-3 配置 IPv6 NetStream 统计输出组网图



3. 配置步骤

(1) 配置 Router A

配置 GigabitEthernet1/0，在出、入方向启动 IPv6 NetStream 统计。

```
<RouterA> system-view
[RouterA] ipv6
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ipv6 address 10::1/64
[RouterA-GigabitEthernet1/0] ipv6 netstream inbound
[RouterA-GigabitEthernet1/0] ipv6 netstream outbound
[RouterA-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 1.1.1.1 255.255.0.0
[RouterA-GigabitEthernet1/1] ipv6 address 15::1/64
[RouterA-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterA] bgp 100
[RouterA-bgp] router-id 10.1.1.1
[RouterA-bgp] peer 1.1.1.2 as-number 150
[RouterA-bgp] network 11.110.0.0 255.255.0.0
[RouterA-bgp] ipv6-family
[RouterA-bgp-af-ipv6] peer 15:2 as-number 150
[RouterA-bgp-af-ipv6] quit
```

```
[RouterA-bgp] quit
```

配置 IPv6 NetStream 自治系统选项和 BGP 下一跳选项

```
[RouterA] ipv6 netstream export version 9 origin-as bgp-nexthop
```

配置 IPv6 NetStream 最大流数。

```
[RouterA] ipv6 netstream max-entry 5000
```

配置普通流的输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream export host 3.1.1.2 5000
[RouterA] ipv6 netstream export source interface ethernet 1/0
```

配置自治系统聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream aggregation as
[RouterA-ns6-aggregation-as] enable
[RouterA-ns6-aggregation-as] ipv6 netstream export host 3.1.1.2 2000
[RouterA-ns6-aggregation-as] ipv6 netstream export source interface gigabitethernet 1/1
[RouterA-ns6-aggregation-as] quit
```

配置协议－端口聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream aggregation protocol-port
[RouterA-ns6-aggregation-protport] enable
[RouterA-ns6-aggregation-protport] ipv6 netstream export host 3.1.1.2 3000
[RouterA-ns6-aggregation-protport] ipv6 netstream export source interface gigabitethernet
1/1
[RouterA-ns6-aggregation-protport] quit
```

配置源前缀聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream aggregation source-prefix
[RouterA-ns6-aggregation-srcpre] enable
[RouterA-ns6-aggregation-srcpre] ipv6 netstream export host 3.1.1.2 4000
[RouterA-ns6-aggregation-srcpre] ipv6 netstream export source interface gigabitethernet 1/1
[RouterA-ns6-aggregation-srcpre] quit
```

配置目的前缀聚合模式，以及该模式下输出的报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream aggregation destination-prefix
[RouterA-ns6-aggregation-dstpre] enable
[RouterA-ns6-aggregation-dstpre] ipv6 netstream export host 3.1.1.2 6000
[RouterA-ns6-aggregation-dstpre] ipv6 netstream export source interface gigabitethernet 1/1
[Sysname-ns6-aggregation-dstpre] quit
```

配置前缀聚合模式，以及该模式下输出报文的目的 IP 地址、目的端口和源接口。

```
[RouterA] ipv6 netstream aggregation prefix
[RouterA-ns6-aggregation-prefix] enable
[RouterA-ns6-aggregation-prefix] ipv6 netstream export host 3.1.1.2 7000
[RouterA-ns6-aggregation-prefix] ipv6 netstream export source interface gigabitethernet 1/1
[RouterA-ns6-aggregation-prefix] quit
```

(2) 配置 Router B

配置 GigabitEthernet1/0。

```
<RouterB> system-view
[RouterB] ipv6
```

```
[RouterB] interface gigabitethernet 1/0
[RouterB-GigabitEthernet1/0] ip address 1.1.1.2 255.255.0.0
[RouterB-GigabitEthernet1/0] ipv6 address 15::2/64
[RouterB-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] ip address 2.1.1.1 255.255.0.0
[RouterB-GigabitEthernet1/1] ipv6 address 20::1/64
[RouterB-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterB] bgp 150
[RouterB-bgp] peer 1.1.1.1 as-number 100
[RouterB-bgp] peer 2.1.1.2 as-number 200
[RouterB-bgp] ipv6-family
[RouterB-bgp-af-ipv6] peer 15:1 as-number 100
[RouterB-bgp-af-ipv6] peer 20:2 as-number 200
[RouterB-bgp-af-ipv6] quit
[RouterB-bgp] quit
```

(3) 配置 Router C

配置 GigabitEthernet1/0。

```
<RouterC> system-view
[RouterC] ipv6
[RouterC] interface gigabitethernet 1/0
[RouterC-GigabitEthernet1/0] ip address 2.1.1.2 255.255.0.0
[RouterC-GigabitEthernet1/0] ipv6 address 20::2/64
[RouterC-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1。

```
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] ip address 3.1.1.1 255.255.0.0
[RouterC-GigabitEthernet1/1] quit
```

配置 BGP。

```
[RouterC] bgp 200
[RouterC-bgp] peer 2.1.1.1 as-number 150
[RouterC-bgp] ipv6-family
[RouterC-bgp-af-ipv6] peer 20:1 as-number 150
[RouterC-bgp-af-ipv6] quit
[RouterC-bgp] quit
```

目 录

1 Sampler配置 1-1
1.1 Sampler简介 1-1
1.2 创建采样器功能 1-1
1.3 Sampler显示和维护 1-1
1.4 Sampler典型配置举例 1-2
1.4.1 Sampler与NetStream配合使用 1-2

1 Sampler 配置

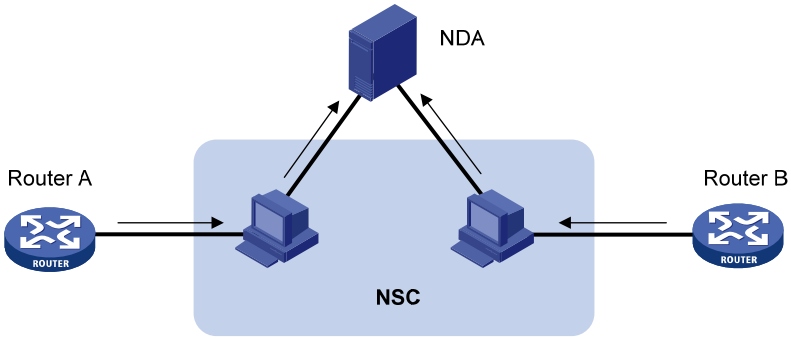
1.1 Sampler 简介

Sampler 提供报文采样功能，从一组固定数量的报文中选出一个报文，送交后续业务模块处理。Sampler 支持随机采样和固定采样两种模式：

- 固定采样：每组报文的第一个被选中。
- 随机采样：每组报文中任意一个报文都有可能被选中，由随机采样数决定。

Sampler可与NetStream配合使用，对网络流量进行采样收集后将上送给NetStream来进行统计，以减轻网络流统计给设备带来的负荷。图 1-1是NetStream特性的一个示意图，采样可以在NSC设备上进行配置，对设备入方向或出方向的流量进行采样。

图1-1 NetStream 数据采集和分析过程示意图



说明

目前 Sampler 只与 NetStream 配合使用。

1.2 创建采样器功能

表1-1 创建采样器功能

操作	命令	说明
进入系统视图	**system-view**	-
创建采样器	**sampler** *sampler-name* **mode** { **fixed** | **random** } **packet-interval** *rate*	必选
NetStream 引用采样器	**ip netstream sampler** *sampler-name* { **inbound** | **outbound** }	必选

1.3 Sampler 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 Sampler 的配置信息和运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 Sampler 的统计信息。

表1-2 Sampler 显示和维护

操作	命令
查看 Sampler 配置和状态信息（集中式设备）	**display sampler** [*sampler-name*]
查看 Sampler 配置和状态信息（分布式设备）	**display sampler** [*sampler-name*] [**slot** *slot-id*]
清除 Sampler 的状态信息	**reset sampler statistics** [*sampler-name*]

1.4 Sampler 典型配置举例

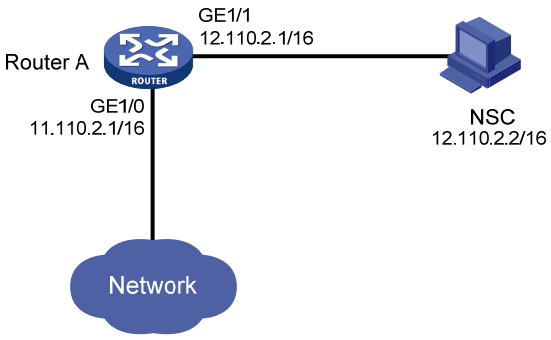
1.4.1 Sampler 与 NetStream 配合使用

1. 组网需求

如 图 1-2所示，在Router A上配置IPv4 NetStream入统计和出统计，并将统计结果发送到NSC 12.110.2.2/16 的 5000 端口。要求入方向使能固定采样，每 256 个报文选中一个进行NetStream统计；要求出方向使能随机采样，每 1024 个报文选中一个进行NetStream统计。

2. 组网图

图1-2 Sampler 与 NetStream 功能组网图



3. 配置步骤

创建一个名为 256 的采样器，采用固定采样方式，设置统计报文的间隔为 8，即 256（2 的 8 次方）个报文中采样 1 个报文。

```
<RouterA> system-view
[RouterA] sampler 256 mode fixed packet-interval 8
```

创建一个名为 1024 的采样器，采用随机采样方式，设置统计报文的间隔为 10，即 1024（2 的 10 次方）个报文中采样 1 个报文。

```
[RouterA] sampler 1024 mode random packet-interval 10
```

配置 GigabitEthernet1/0，在此接口上启动 IPv4 NetStream 入统计并使用采样器 256。

```
[RouterA] interface gigabitethernet 1/0
[RouterA-GigabitEthernet1/0] ip address 11.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/0] ip netstream inbound
[RouterA-GigabitEthernet1/0] ip netstream sampler 256 inbound
[RouterA-GigabitEthernet1/0] quit
```

配置 GigabitEthernet1/1，在此接口上启动 IPv4 NetStream 出统计并使用采样器 1024。

```
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] ip address 12.110.2.1 255.255.0.0
[RouterA-GigabitEthernet1/1] ip netstream outbound
[RouterA-GigabitEthernet1/1] ip netstream sampler 1024 outbound
[RouterA-GigabitEthernet1/1] quit
```

配置 IPv4 NetStream 统计输出报文的目的 IP 地址和目的端口（即 NSC 的地址以及端口），源接口采用缺省配置。

```
[RouterA] ip netstream export host 12.110.2.2 5000
```

目 录

1 热补丁配置 1-1
1.1 热补丁简介 1-1
1.1.1 热补丁基本概念 1-1
1.1.2 补丁状态 1-1
1.2 热补丁配置任务简介 1-4
1.3 配置准备 1-4
1.4 一步式安装补丁 1-5
1.5 分步安装补丁 1-6
1.5.1 配置任务简介 1-6
1.5.2 设置补丁文件的加载路径 1-6
1.5.3 加载补丁文件 1-6
1.5.4 激活补丁 1-7
1.5.5 确认运行补丁 1-7
1.6 一步式卸载补丁 1-8
1.7 分步卸载补丁 1-8
1.7.1 配置任务简介 1-8
1.7.2 停止运行补丁 1-9
1.7.3 删除补丁 1-9
1.8 热补丁显示和维护 1-9
1.9 热补丁典型配置举例（集中式设备） 1-10
1.10 热补丁典型配置举例（分布式设备） 1-10

1 热补丁配置

1.1 热补丁简介

热补丁是一种快速、低成本修复产品软件版本缺陷的方式。和升级软件版本相比，热补丁的主要优势是不会使设备当前正在运行的业务中断，即在不重启设备的情况下，可以对设备当前软件版本的缺陷进行修复。

1.1.1 热补丁基本概念

1. 补丁和补丁文件

补丁（又被称为补丁单元）是用来修复某个缺陷的程序包，通常以补丁文件的形式发布，一个补丁文件可能包含一个或多个补丁，不同的补丁具有不同的功能。当补丁文件被用户从存储介质加载到内存补丁区中时，补丁文件中的补丁将被分配一个在此内存补丁区中唯一的单元序号，用于标志、管理、操作各补丁，补丁的单元序号从 1 开始顺序编号，如某补丁文件中有 3 个补丁单元，那合法的补丁单元号为 1、2 和 3。

2. 增量补丁

补丁文件中的补丁均为增量补丁。所谓增量补丁，是指各补丁单元对于在其前面的补丁单元有依赖性。也就是说，如果补丁文件中有 3 个补丁单元，则 3 号补丁必须在 1 号和 2 号补丁生效之后才能运行，而不能直接单独运行 3 号补丁。

3. 正式补丁和临时补丁

补丁分为正式补丁和临时补丁两种：

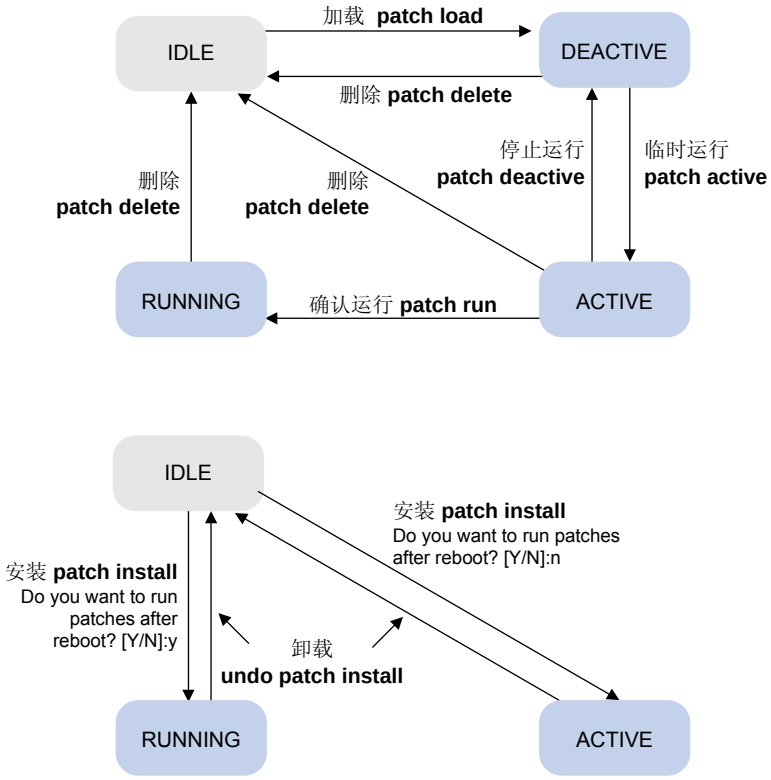
- 正式补丁（Common patches）是通过版本发布流程发布的补丁。
- 临时补丁（Temporary patches）是未通过版本发布流程发布，用于临时解决紧急问题和需求的补丁。

正式补丁总会包含前面临时补丁的功能，从而替代前面的临时补丁。补丁的类型只对补丁加载（Load）过程产生影响——系统在加载正式补丁之前会先将系统中所有临时补丁删除。

1.1.2 补丁状态

每个补丁都有自身的状态，只有在用户命令行的干预下才能发生切换。补丁状态切换与命令操作关系如 图 1-1所示，其中IDLE、DEACTIVE、ACTIVE和RUNNING表示补丁的不同状态，加载、临时运行、确认运行、停止运行、删除、安装、卸载表示补丁操作，分别对应命令**patch load**、**patch active**、**patch run**、**patch deactive**、**patch delete**和**patch install**、**undo patch install**，箭头方向表示状态的转变方向，比如对DEACTIVE状态的补丁执行**patch active**操作，补丁的状态就会变为ACTIVE。

图1-1 补丁状态切换与命令操作关系图



说明

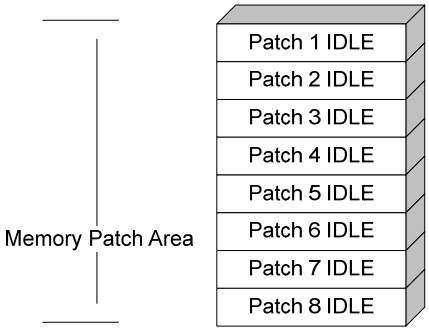
补丁文件的状态信息内容保存在存储介质中的 patchstate 文件里，建议用户不要对该文件进行操作。

1. 初始状态（IDLE）

表示尚未加载补丁，无法进行安装、运行等补丁操作，如 图 1-2所示（假设系统补丁区中最多可以加载 8 个补丁）。

系统重启后，所有处于 IDLE 状态的补丁仍为 IDLE 状态。

图1-2 补丁未加载



说明

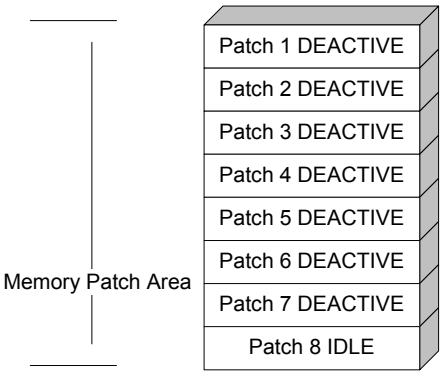
当前系统补丁区中支持的最大补丁数为 200 个。

2. 未激活状态（DEACTIVE）

表示补丁已经被加载到内存补丁区，但尚未运行。假设用户将加载的补丁文件包含 7 个补丁，则这 7 个补丁将在经过版本校验及CRC校验之后被加载到内存补丁区，加载成功的补丁处于DEACTIVE状态，此时系统中补丁状态如 图 1-3所示。

系统重启后，所有处于 DEACTIVE 状态的补丁仍为 DEACTIVE 状态。

图1-3 补丁文件被加载

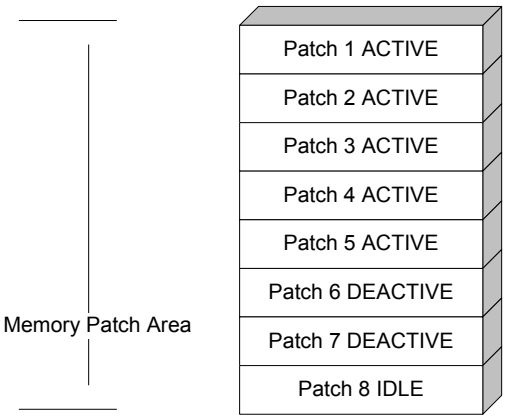


3. 激活状态（ACTIVE）

表示补丁已经被临时运行，即该补丁在设备重启之后不再生效。对于 图 1-3中 7 个处于DEACTIVE状态的补丁，用户如果激活前 5 个补丁，则前 5 个补丁的状态将由DEACTIVE状态变成ACTIVE状态，此时系统中补丁状态如 图 1-4所示。

系统重启后，所有处于 ACTIVE 状态的补丁将变成 DEACTIVE 状态。

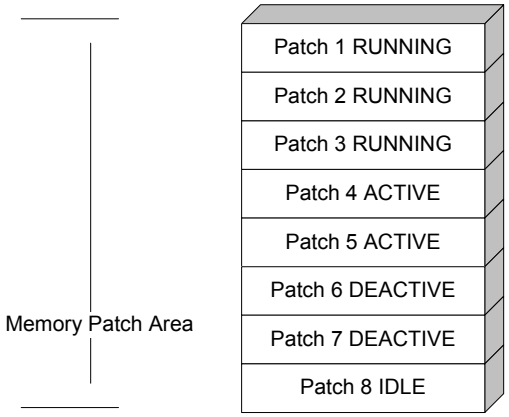
图1-4 补丁被激活



4. 确认运行状态（RUNNING）

表示补丁已经被永久运行，即该补丁在设备重启之后会继续生效。对于 图 1-4中 5 个处于ACTIVE状态的补丁，用户确认运行前 3 个补丁后，前 3 个补丁的状态将由ACTIVE状态变成RUNNING状态，此时系统中补丁状态如 图 1-5所示。

图1-5 补丁被确认运行



系统重启后，所有处于 RUNNING 状态的补丁将保持为 RUNNING 状态。

1.2 热补丁配置任务简介

表1-1 热补丁配置任务简介

配置任务		说明	详细配置
安装补丁	一步式安装补丁	二者任选其一 一步式配置和分步配置效果等同，分步配置时用户可以控制安装过程中的各个状态。用户可以根据实际需要选择任意一种安装方式	1.4
	分步安装补丁		1.5
卸载补丁	一步式卸载补丁	二者任选其一 一步式配置和分步配置效果等同，分步配置时用户可以控制卸载过程中的各个状态。用户可以根据实际需要选择任意一种卸载方式	1.6
	分步卸载补丁		1.7

1.3 配置准备

补丁文件是根据安装补丁的设备或单板类型按需发布。在进行补丁操作前，需要通过 FTP/TFTP 等方式将补丁文件保存到设备的存储介质上，保存时请注意：

- 补丁版本必须与设备型号、设备的软件版本相匹配。如果不匹配，则会造成补丁操作失败。
- 补丁文件的名称必须符合要求，否则系统无法找到补丁文件，造成补丁操作失败。设备是根据补丁标志（PATCH-FLAG）生成补丁文件名，再根据文件名将补丁从存储介质（默认为CF卡）中读取到补丁区进行加载或安装。设备在出厂时已经定义了PATCH-FLAG，不同类型设备或单板的PATCH-FLAG不同。**display patch information**命令显示信息中Version字段的前3个字符就是PATCH-FLAG的后缀，补丁文件名称的格式为“patch_*后缀*.bin”。比如设备的PATCH-FLAG为PATCH-RPU，则它对应的补丁文件必须命名为patch_rpu.bin。单板补丁文件缺省名称如 表 1-2所示。

表1-2 部分产品缺省补丁文件命名参照表

产品	单板类型	PATCH-FLAG	补丁缺省文件名
SR8800	srpxn	PATCH-MXN	patch_mxn.bin
	srp2e	PATCH-M2E	patch_m2e.bin
	lmcue	PATCH-LPE	patch_lpe.bin
	lmcua	PATCH-LPA	patch_lpa.bin
	oam(8248)	PATCH-OAM	patch_oam.bin
S7500E/S7600	mpu	PATCH-MPU	patch_mpu.bin
	lpu	PATCH-LPU	patch_lpu.bin
8062	6602	PATCH-RPU	patch-66020100.bin
	mpua	PATCH-RPE	patch-66080100.bin
	fipl	PATCH-100	patch-66080200.bin
	fiph	PATCH-200	patch-66080300.bin
8042	所有单板	PATCH-HFW	patchmain.bin
S5500-EI	-	PATCH-XXX	patch_xxx.bin

说明

对于分布式设备，加载和安装操作的对象是设备上所有在位单板和辅助 CPU，所以，执行加载和安装操作前，请将主用主控板和接口板的补丁文件保存到主用主控板存储介质的根目录下，备用主控板的补丁文件保存到备用主控板存储介质的根目录下，主用主控板和备用主控板上补丁文件必须一致（包括补丁文件的名称和内容）。

1.4 一步式安装补丁

用户可以使用多条命令完成热补丁的安装，也可以通过 **patch install** 命令来一步式安装补丁。在执行 **patch install** 命令时，系统会提示“Do you want to continue running patches after reboot? [Y/N]:”，

- 如果输入<Y>或者<y>，则会安装指定路径下的所有补丁，设备或单板重启后这些补丁继续生效，补丁状态从 IDLE 转换到 RUNNING，相当于依次执行了 **patch location**、**patch load**、**patch active** 和 **patch run** 命令。
- 如果输入<N>或者<n>，也会安装指定路径下的所有补丁，但设备或单板重启后这些补丁不再继续生效，补丁状态从 IDLE 转换到 ACTIVE，相当于依次执行了 **patch location**、**patch load** 和 **patch active** 命令。

表1-3 一步式安装补丁

操作	命令	说明
进入系统视图	**system-view**	-
一步式完成补丁的安装	**patch install** *patch-location*	必选

说明

- 指定路径下的补丁必须与单板的型号和版本匹配，否则系统不能正确识别补丁文件。
- 执行补丁安装操作后，系统会将补丁文件的加载路径（**patch location**）修改成“**patch install** *patch-location*”中 *patch-location* 参数所指向的位置。

1.5 分步安装补丁

1.5.1 配置任务简介

表1-4 分步安装补丁配置任务简介

配置任务	说明	详细配置
设置补丁文件的加载路径	可选	1.5.2
加载补丁文件	必选	1.5.3
激活补丁	必选	1.5.4
确认运行补丁	可选	1.5.5

1.5.2 设置补丁文件的加载路径

当设备支持多个存储介质时，用户可以将补丁文件保存到其它存储介质。此时，需要使用该命令告知系统补丁文件的位置，执行补丁加载操作时，系统会将该路径下指定后缀的补丁加载到对应的单板。因此，在执行补丁加载之前，请确保补丁文件存放位置与设置好的路径一致。如果设备上只有一个存储介质，则不用设置该参数。

表1-5 设置补丁文件的加载路径

操作	命令	说明
进入系统视图	**system-view**	-
设置补丁文件的加载路径	**patch location** *patch-location*	可选 缺省情况下，补丁文件的加载路径为 cfa0:

说明

- 对于分布式设备，*patch-location* 参数指定的路径必须在主用主控板和备用主控板上都存在，如果备用主控板上不存在该路径，则该命令在备用主控板上不生效。
- 执行 **patch install** 命令会修改补丁文件的加载路径。比如，先配置了 patch location xxx，再执行 patch install yyy，此时系统会自动将补丁文件的加载路径从 xxx 改为 yyy。

1.5.3 加载补丁文件

只有正确加载补丁文件后才能进行补丁状态的管理。系统默认从 CF 卡中读取补丁文件，如果 CF 卡中找不到补丁文件，系统将尝试从其他存储介质中读取补丁文件。

注意

当使用 FTP 或 TFTP 方式将补丁文件上传/下载到设备的 CF 卡中前，请将文件传输的模式设置为二进制模式，以免不能正确解析补丁文件。

表1-6 加载补丁文件（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
将补丁文件从存储介质加载到内存补丁区中	**patch load**	必选

表1-7 加载补丁文件（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
将补丁文件从存储介质加载到指定的内存补丁区中	**patch load slot** *slot-number*	必选

1.5.4 激活补丁

激活补丁后，补丁会立即生效，处于试运行阶段。设备复位或重启后，该补丁不再生效，需要重新激活。

补丁的激活状态主要是提供一个缓冲带，如果补丁本身有问题，可以重启设备，以消除该补丁的作用，从而防止因为补丁错误而导致系统连续运行故障。

表1-8 激活补丁（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
激活指定补丁	**patch active** *patch-number*	必选

表1-9 激活补丁（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
激活指定补丁	**patch active** *patch-number* **slot** *slot-number*	必选

1.5.5 确认运行补丁

确认运行补丁后，补丁的状态将变为 RUNNING，处于正式运行阶段。设备复位或重启后，该补丁会继续生效。

表1-10 确认运行补丁（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
确认运行指定补丁	**patch run** *patch-number*	必选

表1-11 确认运行补丁（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
确认运行指定补丁	**patch run** *patch-number* [**slot** *slot-number*]	必选


说明

该操作只对处于激活状态的补丁有效，对于处于其它状态的补丁，该操作无效。

1.6 一步式卸载补丁

- 集中式设备

通过一条命令完成补丁卸载，命令结束后，补丁状态都进入 IDLE 状态，相当于依次执行了 **patch deactive** 和 **patch delete** 命令。

- 分布式设备

通过一条命令完成所有在位单板和辅助 CPU 的补丁卸载，命令结束后，所有在位单板的补丁状态都进入 IDLE 状态，相当于在每个单板（包括辅助 CPU）上依次执行了 **patch deactive** 和 **patch delete** 命令。

表1-12 一步式卸载补丁

操作	命令	说明
进入系统视图	**system-view**	-
卸载补丁	**undo patch install**	必选

1.7 分步卸载补丁

1.7.1 配置任务简介

表1-13 分步卸载补丁配置任务简介

配置任务	说明	详细配置
停止运行补丁	必选	1.7.2
删除补丁	必选	1.7.3

1.7.2 停止运行补丁

停止运行补丁后，补丁将进入未激活状态（DEACTIVE）。系统按打补丁前的特性运行。

表1-14 停止运行补丁（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
停止运行指定补丁	**patch deactive** *patch-number*	必选

表1-15 停止运行补丁（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
停止运行指定补丁	**patch deactive** *patch-number* **slot** *slot-number*	必选

1.7.3 删除补丁

执行该操作，只是将补丁从内存补丁区中删除，并不会将补丁从存储介质中删除，补丁回到初始状态（IDLE）。补丁删除后，系统按打补丁前的特性运行。

表1-16 删除补丁（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
将补丁从内存补丁区中删除	**patch delete** *patch-number*	必选

表1-17 删除补丁（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
将指定补丁从内存补丁区中删除	**patch delete** *patch-number* **slot** *slot-number*	必选

1.8 热补丁显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示热补丁的当前状态，通过查看显示信息验证配置的效果。

表1-18 热补丁显示和维护

操作	命令
显示热补丁的信息	**display patch information**

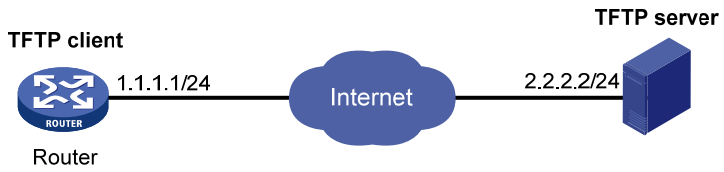
1.9 热补丁典型配置举例（集中式设备）

1. 组网需求

- Router 当前的软件版本存在一些缺陷，需要使用补丁文件进行修复。
- 补丁文件 patch_rpu.bin 保存在 TFTP server 上。
- Router 的 IP 地址为 1.1.1.1/24，TFTP server 的 IP 地址为 2.2.2.2/24，Router 与 TFTP server 之间路由可达。

2. 组网图

图1-6 热补丁升级组网图



3. 配置步骤

(1) TFTP server 的配置（不同应用程序的配置方式不同，具体步骤略）

- 启动 TFTP server 功能。
- 将 patch_rpu.bin 文件放到 TFTP server 的工作路径。

(2) Router 的配置

注意

配置前需确认设备的 CF 卡有足够的剩余空间来存储补丁文件。

开始升级前，执行 **save** 命令保存当前配置（配置步骤略）。

将 TFTP server 上的文件 patch_rpu.bin 下载到设备存储介质的根目录下。

```
<Router> tftp 2.2.2.2 get patch_rpu.bin
```

安装补丁。

```
<Router> system-view
[Router] patch install cfa:/
Patches will be installed. Continue? [Y/N]:y
Do you want to continue running patches after reboot? [Y/N]:y
Installing patches........
Installation completed, and patches will continue to run after reboot.
```

1.10 热补丁典型配置举例（分布式设备）

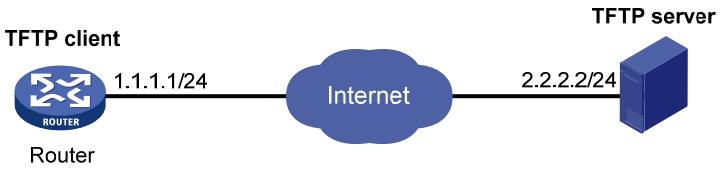
1. 组网需求

- Router 当前的软件版本存在一些缺陷，请使用补丁文件进行修复。

- 最新补丁文件 patch_rpe.bin（主控板的补丁文件）和 patch_200.bin（接口板的补丁文件）都保存在 TFTP server 上。
- Router 的 IP 地址为 1.1.1.1/24，TFTP server 的 IP 地址为 2.2.2.2/24，Router 与 TFTP server 之间路由可达。

2. 组网图

图1-7 热补丁升级组网图



3. 配置步骤

(1) TFTP server 的配置（不同应用程序的配置方式不同，具体步骤略）

- 启动 TFTP server 功能。
- 将 patch_rpe.bin 和 patch_200.bin 文件放到 TFTP server 的工作路径。

(2) Router 的配置

注意

配置前需确认设备的 CF 卡有足够的剩余空间来存储补丁文件。

开始升级前，执行 **save** 命令保存当前配置（配置步骤略）。

将 TFTP server 上的文件 patch_rpe.bin 下载到主用主控板上。

```
<Router> tftp 2.2.2.2 get patch_rpe.bin
<Router> tftp 2.2.2.2 get patch_200.bin
```

将补丁文件拷贝到备用板的根目录下，备用主控板所在的槽位号为 1。

```
<Router> copy patch_rpe.bin slot1#cfa0:/
```

安装补丁。

```
<Router> system-view
[Router] patch install cfa0:/
Patches will be installed. Continue? [Y/N]:y
Do you want to continue running patches after reboot? [Y/N]:y
Installing patches........
Installation completed, and patches will continue to run after reboot.
```

目 录

1 Flow日志配置 1-1
1.1 Flow日志简介 1-1
1.1.1 Flow日志概述 1-1
1.1.2 Flow日志的版本 1-1
1.2 Flow日志配置任务简介 1-2
1.3 配置Flow日志的版本 1-2
1.4 配置Flow日志报文的源地址 1-3
1.5 输出Flow日志 1-3
1.5.1 配置Flow日志发送到日志服务器 1-3
1.5.2 配置Flow日志输出到信息中心 1-4
1.6 Flow日志显示和维护 1-4
1.7 Flow日志典型配置举例 1-5
1.7.1 Flow日志配置举例（集中式设备） 1-5
1.7.2 Flow日志配置举例（分布式设备） 1-6
1.8 常见配置错误举例 1-7

1 Flow 日志配置

1.1 Flow 日志简介

1.1.1 Flow 日志概述

Flow 日志是指用户访问外部网络流信息的相关记录。设备根据报文的 5 元组（源 IP 地址、目的 IP 地址、源端口、目的端口、协议号）对用户访问外部网络的流进行分类统计，并生成用户流（Flow）日志。Flow 日志会记录报文的 5 元组和发送、接收的字节数等信息。网络管理员利用这些信息可以实时跟踪、记录用户访问网络的情况，增强网络的可用性和安全性。

1.1.2 Flow 日志的版本

Flow日志有Flow1.0 和Flow3.0 两个版本。两种Flow日志的格式稍有不同，具体差别请参见 表 1-1 和 表 1-2。

表1-1 Flow1.0 日志信息

字段	描述
SIP	源 IP 地址
DIP	目的 IP 地址
SPORT	TCP/UDP 源端口号
DPORT	TCP/UDP 目的端口号
STIME	流起始时间，以秒为单位，从 1970/1/1 0:0 开始计算
ETIME	流结束时间，以秒为单位，从 1970/1/1 0:0 开始计算
PROT	IP 承载的协议类型
OPERATOR	操作字，主要指流结束原因
RESERVED	保留

表1-2 Flow3.0 日志信息

字段	描述
Prot	IP 承载的协议类型
Operator	操作字，主要指流结束原因
IpVersion	IP 报文版本
TosIPv4	IPv4 报文的 Tos 字段
SourceIP	源 IP 地址
SrcNatIP	NAT 转换后的源 IP 地址
DestIP	目的 IP 地址
DestNatIP	NAT 转换后的目的 IP 地址
SrcPort	TCP/UDP 源端口号
SrcNatPort	NAT 转换后的 TCP/UDP 源端口号

字段	描述
DestPort	TCP/UDP 目的端口号
DestNatPort	NAT 转换后的 TCP/UDP 目的端口号
StartTime	流起始时间，以秒为单位，从 1970/01/01 00:00 开始计算
EndTime	流结束时间，以秒为单位，从 1970/01/01 00:00 开始计算
InTotalPkg	接收的报文包数
InTotalByte	接收的报文字节数
OutTotalPkg	发出的报文包数
OutTotalByte	发出的报文字节数
Reserved1	对于 0x02 版本（FirewallV200R001）保留； 对于 0x03 版本（FirewallV200R005）第一个字节为源 VPN ID，第二个字节为目的 VPN ID，第三、四个字节保留
Reserved2	保留
Reserved3	保留

1.2 Flow 日志配置任务简介

表1-3 Flow 日志配置任务简介

配置任务		说明	详细配置
配置 Flow 日志的版本		可选	1.3
配置 Flow 日志报文的源地址		可选	1.4
输出 Flow 日志	配置日志发送到日志服务器	二者只能选其一	1.5.1
	配置 Flow 日志输出到信息中心		1.5.2

1.3 配置 Flow 日志的版本

请根据日志接收设备的实际能力配置 Flow 日志的版本，如果接收设备不支持某个版本的 Flow 日志，则收到日志后，它不能正确解析。

表1-4 配置 Flow 日志的版本

操作	命令	说明
进入系统视图	**system-view**	-
配置 Flow 日志报文的版本号	**userlog flow export version** *version-number*	可选 缺省情况下，Flow 日志报文的版本号为 1.0

说明

设备支持 Flow1.0 和 Flow3.0 两个版本，但同一时刻只能使用一个版本。所以，如果多次使用该命令配置版本，则最新的配置生效。

1.4 配置 Flow 日志报文的源地址

为了安全性，用户可以通过使用指定的 IP 地址作为 Flow 日志报文的源 IP 地址。当设备向日志接收设备发送 Flow 日志时，就使用这个 IP 地址作为报文的源 IP 地址，而不是使用报文出接口的真正地址，从而可以屏蔽真正的源 IP 地址。而且该功能还简化了 ACL 规则和安全策略的配置，只要将 ACL 规则中定义的源地址或者目的地址参数指定为该地址，就可以屏蔽接口 IP 地址的差异以及接口状态的影响，实现对 Flow 日志报文的过滤。

表1-5 配置 Flow 日志报文的源地址

操作	命令	说明
进入系统视图	**system-view**	-
配置 Flow 日志报文的源地址	**userlog flow export source-ip** *ip-address*	可选 缺省情况下，Flow 日志报文的源地址为发送该报文的接口的 IP 地址

1.5 输出 Flow 日志

Flow 日志有两种输出方式：

- 设备可以将Flow日志封装成UDP报文直接发送给网络中的日志服务器，如 图 1-2所示。日志服务器可以对Flow日志进行解析和分类显示，以达到远程监控的目的。
- 设备可以将 Flow 日志以系统信息的格式输出到信息中心，再通过设置信息中心的输出方向，最终决定 Flow 日志的输出方向。关于信息中心的详细介绍请参见“系统分册/信息中心配置”。

说明

Flow 日志的两种输出方式互斥，同一时刻只能选择一种输出方式。如果同时配置了两种输出方式，则系统会自动选择输出到信息中心，而不会发送到日志服务器。

1.5.1 配置 Flow 日志发送到日志服务器

表1-6 配置 Flow 日志发送到日志服务器（集中式设备）

操作	命令	说明
进入系统视图	**system-view**	-
设置 Flow 日志服务器的 IP 地址和 UDP 端口号	**userlog flow export** [**vpn-instance** *vpn-instance-name*] **host** *ip-address udp-port*	必选 缺省情况下，没有配置 Flow 日志服务器的 IP 地址和 UDP 端口号

表1-7 配置 Flow 日志发送到日志服务器（分布式设备）

操作	命令	说明
进入系统视图	**system-view**	-
设置 Flow 日志服务器的 IP 地址和 UDP 端口号	**userlog flow export slot** *slot-number* [**vpn-instance** *vpn-instance-name*] **host** *ip-address udp-port*	必选 缺省情况下，没有配置 Flow 日志服务器的 IP 地址和 UDP 端口号

说明

- 如果即将执行的命令的 IP 地址与当前生效的日志服务器的 IP 地址相同，其它信息不同，则即将执行的命令会覆盖原有配置。
- 在集中式设备上，最多可指定两个 Flow 日志服务器来接收 Flow 日志，达到最大个数后必须先删除一条配置才能配置新的 Flow 日志服务器。
- 在分布式设备上，每个接口板可以单独配置 Flow 日志服务器，起到分担 Flow 日志服务器负载的作用。每个接口板可以指定两个 Flow 日志服务器来接收 Flow 日志，达到最大个数后必须先删除一条配置才能配置新的 Flow 日志服务器。

1.5.2 配置 Flow 日志输出到信息中心

表1-8 配置 Flow 日志输出到信息中心

操作	命令	说明
进入系统视图	**system-view**	-
配置Flow日志输出到信息中心	**userlog flow syslog**	必选 缺省情况下，Flow 日志输出到 Flow 日志服务器

说明

- 日志输出到信息中心会占用设备的存储空间，所以，建议在日志量较小的情况下，使用该输出方向。
- 日志输出至信息中心时，日志信息的优先级为 informational，即作为设备的一般提示信息。

1.6 Flow 日志显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后的 Flow 日志的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 Flow 日志信息。

表1-9 Flow 日志显示和维护（集中式设备）

操作	命令
查看日志的配置和统计信息	**display userlog export**
清除 Flow 日志的统计信息	**reset userlog flow export**
清除设备缓存区中的 Flow 日志	**reset userlog flow logbuffer**

表1-10 Flow 日志显示和维护（分布式设备）

操作	命令
查看日志的配置和统计信息	**display userlog export slot** *slot-number*
清除 Flow 日志的统计信息	**reset userlog flow export slot** *slot-number*
清除设备缓存区中的 Flow 日志	**reset userlog flow logbuffer slot** *slot-number*

注意

清除日志缓存中的记录会造成日志信息的丢失，正常情况下，建议不要进行清除操作。当确定不需要缓存中的日志信息时，可清除所有的 Flow 日志。

1.7 Flow 日志典型配置举例

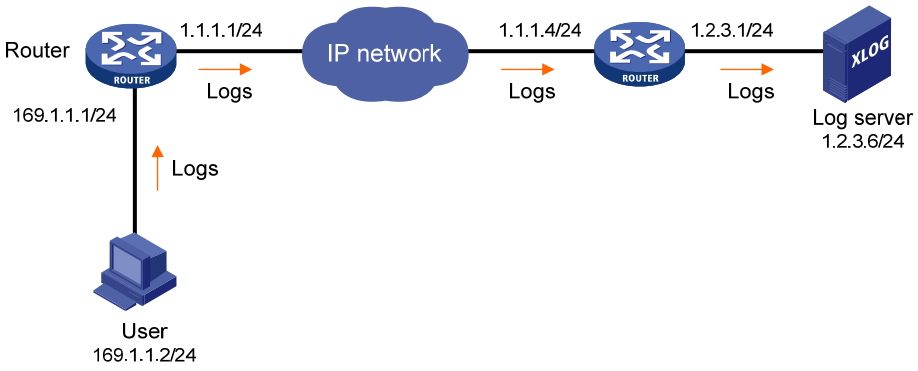
1.7.1 Flow 日志配置举例（集中式设备）

1. 组网需求

如 图 1-2所示，在日志服务器上（Log server）对User的上网活动进行监控。Router为集中式设备。

2. 组网图

图1-1 配置 Flow 日志组网图



3. 配置步骤

请在 Router 上进行如下配置。

将 Flow 日志报文版本号设为 3.0。

```
<Router> system-view
```

```
[Router] userlog flow export version 3
```

将设备产生的 Flow 日志信息发送给 Flow 日志服务器（地址为 1.2.3.6:2000）。

```
[Router] userlog flow export host 1.2.3.6 2000
```

将 169.1.1.2 配置为承载 Flow 日志的 UDP 报文的源 IP 地址，以标志 169.1.1.2 是该日志记录的实际用户。

```
[Router] userlog flow export source-ip 169.1.1.2
```

4. 验证结果

查看设备日志的配置和统计信息。

```
<Router> display userlog export
flow:
   Export Version 3 logs to log server : enabled
   Source address of exported logs   : 169.1.1.2
   Address of log server             : 1.2.3.6 (port: 2000)
   total Logs/UDP packets exported   : 112/87
   Logs in buffer                    : 6
```

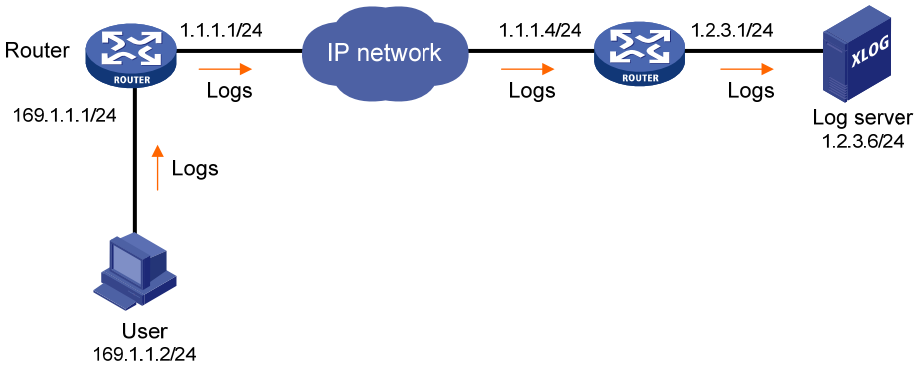
1.7.2 Flow 日志配置举例（分布式设备）

1. 组网需求

如 图 1-2所示，在日志服务器上（Log server）对User的上网活动进行监控。Router为分布式设备。

2. 组网图

图1-2 配置 Flow 日志组网图



3. 配置步骤

请在 Router 上进行如下配置。

将 Flow 日志报文版本号设为 3.0。

```
<Router> system-view
[Router] userlog flow export version 3
```

在分布式设备上，将 2 号接口板上的 Flow 日志信息发送给 Flow 日志服务器（地址为 1.2.3.6:2000）。

```
[Router] userlog flow export slot 2 host 1.2.3.6 2000
```

将 169.1.1.2 配置为承载 Flow 日志的 UDP 报文的源 IP 地址。

```
[Router] userlog flow export source-ip 169.1.1.2
```

4. 验证结果

查看设备 2 号槽位接口板的日志的配置和统计信息。

```
<Router> display userlog export slot 2
flow:
   Export Version 3 logs to log server : enabled
   Source address of exported logs   : 169.1.1.2
   Address of log server             : 1.2.3.6 (port: 2000)
   total Logs/UDP packets exported   : 128/91
   Logs in buffer                    : 10
```

1.8 常见配置错误举例

1. 故障之一：日志无法输出

- 原因：可能是既没有配置信息中心输出方向，也没指定日志服务器的地址。
- 解决方法：配置日志输出到信息中心，或者配置日志输出到日志服务器。

2. 故障之二：日志无法输出到日志服务器

- 原因：可能是既配置了向日志服务器输出，又配置了向信息中心方向输出。
- 解决方法：可以先恢复到缺省配置，再配置日志服务器的 IP 地址和 UDP 端口号。

目 录

1 User Profile配置 1-1
1.1 User Profile简介 1-1
1.2 User Profile配置任务简介 1-1
1.3 创建User Profile 1-1
1.3.1 配置准备 1-1
1.3.2 创建User Profile 1-2
1.4 配置User Profile 1-2
1.5 激活User Profile 1-2
1.6 User Profile显示和维护 1-2

1 User Profile 配置

1.1 User Profile 简介

User Profile（用户配置文件）提供一个配置模板，能够保存预设配置（一系列配置的集合）。用户可以根据不同的应用场景为 User Profile 配置不同的内容，比如 CAR（Committed Access Rate，承诺访问速率）策略和 QoS（Quality of Service，服务质量）策略等。

用户访问设备时，需要先进行身份认证。在认证过程中，认证服务器会将 User Profile 名称下发给设备，设备会立即启用 User Profile 里配置的具体内容。当用户通过认证访问设备时，设备将通过这些具体内容限制用户的访问行为。当用户下线时，系统会自动禁用 User Profile 下的配置项，从而取消 User Profile 对用户的限定。因此，User Profile 适用于限制上线用户的访问行为，没有用户上线（可能是没有用户接入、或者用户没有通过认证、或者用户下线）时，User Profile 是预设配置，并不生效。

使用 User Profile 之后，可以：

- 更精确地利用系统资源。比如，使用 User Profile 之前，只能基于接口/VLAN/全局等应用 QoS 策略，这个 QoS 策略限制的是一群用户；使用 User Profile 之后，可以基于用户应用 QoS 策略，这个 QoS 策略针对的是单个用户。
- 更灵活地限制用户访问系统资源。比如，使用 User Profile 之前，只能基于 CAR 列表/ACL 或者对当前接口的所有流进行流量监管，当用户的物理位置移动时（比如从另一个接口接入），则需要在另外的接口配置流量监管功能；使用 User Profile 之后，可以基于用户进行流量监管，只要用户上线，认证服务器会自动下发相应的 User Profile（配置了 CAR 策略），当用户下线，系统会自动取消相应的配置，不需要再进行手工调整。

1.2 User Profile 配置任务简介

表1-1 User Profile 配置任务简介

配置任务	说明	详细配置
创建 User Profile	必选	1.3
配置 User Profile	必选	1.4
激活 User Profile	必选	1.5

1.3 创建 User Profile

1.3.1 配置准备

在创建 User Profile 之前，需配置认证参数。目前，User Profile 特性只支持 PPPoE 接入认证方式，用户在访问网络时可以根据当前的网络规划选择一种方式进行认证。但需要在客户端、本设备和认证服务器上进行相应的配置（比如用户名、密码、认证方案、域以及 User Profile 和用户的绑定关系等）。

1.3.2 创建 User Profile

表1-2 创建 User Profile

操作	命令	说明
进入系统视图	**system-view**	-
创建 User Profile 并进入 user-profile PPP 视图	**user-profile** *profile-name* **ppp**	必选，如果指定的 User Profile 已经存在，则直接进入 user-profile 视图，不需要再创建 User Profile 进入 user-profile 视图后，下面进行的配置只在 User Profile 处于激活状态，且用户成功上线后才生效



说明

- user-profile 视图下应用的策略不能为空策略，因为空策略的 User Profile 不能被激活。
- 关于 PPP 认证类型的介绍请参见“接入分册”中的“PPP 配置”。

1.4 配置 User Profile

User Profile 创建之后，需要在 User Profile 视图下配置具体的内容才能对上线用户进行限制。目前支持 QoS 策略，相关内容请参见“QoS 分册”中的“QoS 配置”。

1.5 激活 User Profile

User Profile 创建后，必须被激活，才能生效。User Profile 配置生效后，认证服务器才会通知用户认证成功。所以，如果不激活 User Profile，会导致被绑定的用户不能上线。

表1-3 激活 User Profile

操作	命令	说明
进入系统视图	**system-view**	-
激活已创建的 User Profile	**user-profile** *profile-name* **enable**	必选 缺省情况下，创建的 User Profile 处于未激活状态



注意

- 激活的 User Profile 才能被用户使用，但其中的配置内容不允许修改，禁用后才可以被修改或删除。
- 禁用 User Profile 将导致使用该 User Profile 用户强制下线。

1.6 User Profile 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 User Profile 的运行情况，通过查看显示信息验证配置的效果。

表1-4 User Profile 显示和维护

操作	命令
显示当前已创建的全部 User Profile 的信息	**display user-profile**

目 录

1 组播概述 1-1
1.1 组播简介 1-1
1.1.1 三种信息传输方式的比较 1-1
1.1.2 组播传输的特点 1-3
1.1.3 组播中常用的表示法 1-4
1.1.4 组播的优点和应用 1-4
1.2 组播模型分类 1-4
1.3 组播框架结构 1-5
1.3.1 组播地址 1-5
1.3.2 组播协议 1-7
1.4 组播报文的转发机制 1-8
1.5 多实例组播 1-9
1.5.1 多实例简介 1-9
1.5.2 多实例在组播中的应用 1-10

1 组播概述

说明

本手册着重介绍 IP 组播技术及设备操作。如果不加特别说明，本手册中出现的组播均指 IP 组播。

1.1 组播简介

作为一种与单播（Unicast）和广播（Broadcast）并列的通信方式，组播（Multicast）技术能够有效地解决单点发送、多点接收的问题，从而实现了网络中点到多点的高效数据传送，能够节约大量网络带宽、降低网络负载。

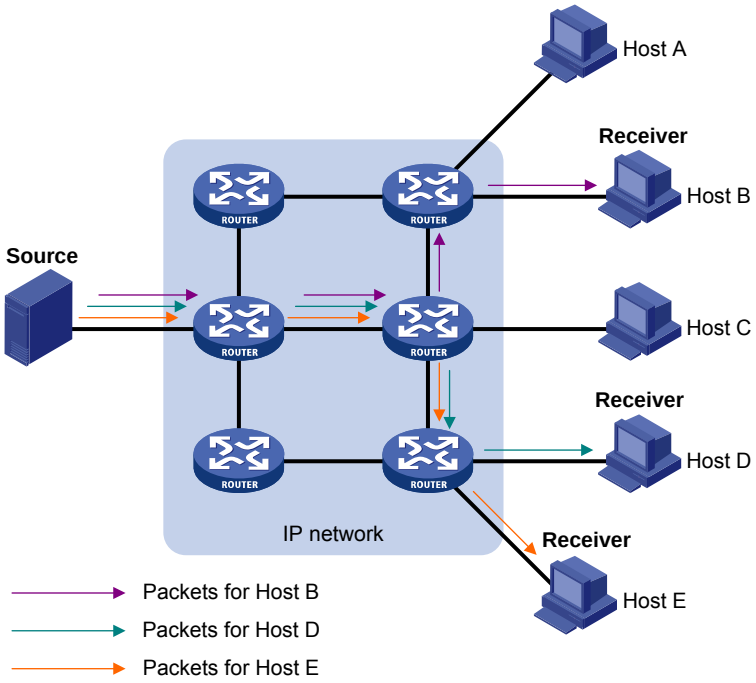
利用组播技术可以方便地提供一些新的增值业务，包括在线直播、网络电视、远程教育、远程医疗、网络电台、实时视频会议等对带宽和数据交互的实时性要求较高的信息服务。

1.1.1 三种信息传输方式的比较

1. 单播方式的信息传输

如 图 1-1所示，在IP网络中若采用单播的方式，信息源（即Source）要为每个需要信息的主机（即Receiver）都发送一份独立的信息拷贝。

图1-1 单播方式的信息传输



假设 Host B、Host D 和 Host E 需要信息，则 Source 要与 Host B、Host D 和 Host E 分别建立一条独立的信息传输通道。

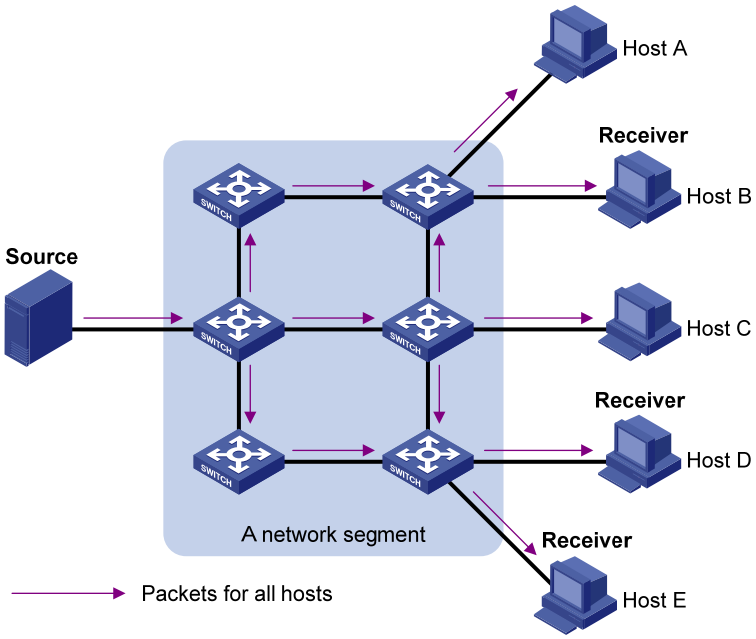
采用单播方式时，网络中传输的信息量与需要该信息的用户量成正比，因此当需要该信息的用户数量较大时，信息源需要将多份内容相同的信息发送给不同的用户，这对信息源以及网络带宽都将造成巨大的压力。

从单播方式的信息传播过程可以看出，该传输方式不利于信息的批量发送。

2. 广播方式的信息传输

如 图 1-2所示，在一个网段中若采用广播的方式，信息源（即Source）将把信息传送给该网段中的所有主机，而不管其是否需要该信息。

图1-2 广播方式的信息传输



假设只有 Host B、Host D 和 Host E 需要信息，若将该信息在网段中进行广播，则原本不需要信息的 Host A 和 Host C 也将收到该信息，这样不仅信息的安全性得不到保障，而且会造成同一网段中信息的泛滥。

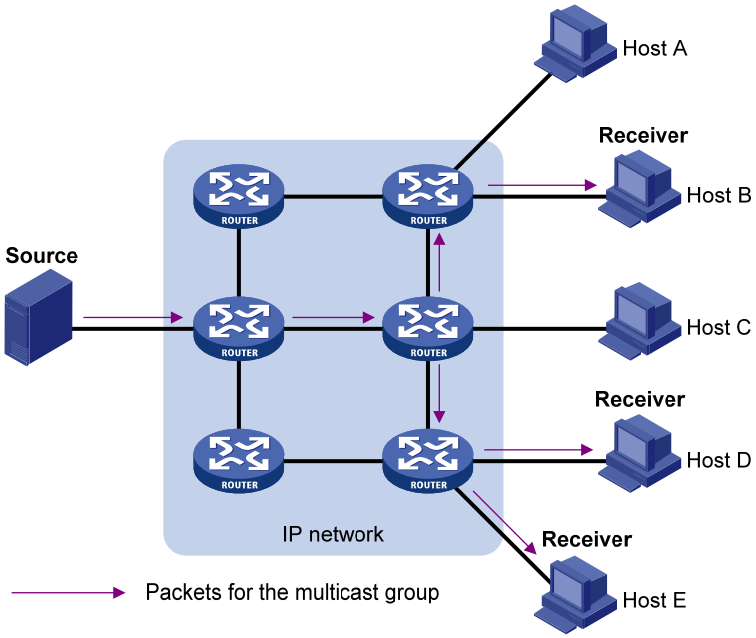
因此，广播方式不利于与特定对象进行数据交互，并且还浪费了大量的带宽。

3. 组播方式的信息传输

综上所述，传统的单播和广播的通信方式均不能以最小的网络开销实现单点发送、多点接收的问题，IP 组播技术的出现及时解决了这个问题。

如 图 1-3所示，当IP网络中的某些主机（即Receiver）需要信息时，若采用组播的方式，组播源（即Source）仅需发送一份信息，借助组播路由协议建立组播分发树，被传递的信息在距离组播源尽可能远的网络节点才开始复制和分发。

图1-3 组播方式的信息传输



假设只有 Host B、Host D 和 Host E 需要信息，采用组播方式时，可以让这些主机加入同一个组播组（Multicast group），组播源向该组播组只需发送一份信息，并由网络中各路由器根据该组播组中各成员的分布情况对该信息进行复制和转发，最后该信息会准确地发送给 Host B、Host D 和 Host E。

综上所述，组播的优势归纳如下：

- 相比单播来说，组播的优势在于：由于被传递的信息在距信息源尽可能远的网络节点才开始被复制和分发，所以用户的增加不会导致信息源负载的加重以及网络资源消耗的显著增加。
- 相比广播来说，组播的优势在于：由于被传递的信息只会发送给需要该信息的接收者，所以不会造成网络资源的浪费，并能提高信息传输的安全性；另外，广播只能在同一网段中进行，而组播可以实现跨网段的传输。

1.1.2 组播传输的特点

组播传输的特点归纳如下：

- “组播组”是一个用 IP 组播地址进行标识的接收者集合，主机通过加入某组播组成为该组播组的成员，从而可以接收发往该组播组的组播数据。组播源通常不需要加入组播组。
- 信息的发送者称为“组播源”，如 图 1-3中的Source。一个组播源可以同时向多个组播组发送信息，多个组播源也可以同时向一个组播组发送信息。
- 所有加入某组播组的主机便成为该组播组的成员，如 图 1-3中的Receiver。组播组中的成员是动态的，主机可以在任何时刻加入或离开组播组。组播组成员可以广泛地分布在网络中的任何地方。
- 支持三层组播功能的路由器或三层交换机统称为“组播路由器”或“三层组播设备”。组播路由器不仅能够提供组播路由功能，也能够在与用户连接的末梢网段上提供组播组成员的管理功能。组播路由器本身也可能是组播组的成员。

为了更好地理解，可以将组播方式的信息传输过程类比于电视节目的传送过程，如 表 1-1所示。

表1-1 组播信息传输与电视节目传输的类比

步骤	电视节目的传送过程	组播方式的信息传输过程
1	电视台 S 通过频道 G 传送电视节目	组播源 S 向组播组 G 发送组播数据
2	用户 U 将电视机的频道调至频道 G	接收者 U 加入组播组 G
3	用户 U 能够收看到由电视台 S 通过频道 G 传送的电视节目了	接收者 U 能够收到由组播源 S 发往组播组 G 的组播数据了
4	用户 U 关闭电视机或切换到其它频道	接收者 U 离开组播组 G

1.1.3 组播中常用的表示法

在组播中，经常出现以下两种表示方式：

- （*，G）：通常用来表示共享树，或者由任意组播源发往组播组 G 的组播报文。其中的“*”代表任意组播源，“G”代表特定组播组 G。
- （S，G）：也称为“组播源组”，通常用来表示最短路径树，或者由组播源 S 发往组播组 G 的组播报文。其中的“S”代表特定组播源 S，“G”代表特定组播组 G。


说明

有关共享树和最短路径树的详细介绍，请参见“IP 组播分册”中的“PIM 配置”。

1.1.4 组播的优点和应用

1. 组播的优点

组播技术的优点主要在于：

- 提高效率：减轻信息源服务器和网络设备 CPU 的负荷；
- 优化性能：减少冗余流量；
- 分布式应用：使用最少的网络资源实现点到多点应用。

2. 组播的应用

组播技术主要应用于以下几个方面：

- 多媒体、流媒体的应用，如：网络电视、网络电台、实时视/音频会议；
- 培训、联合作业场合的通信，如：远程教育、远程医疗；
- 数据仓库、金融应用（股票）；
- 其它任何“点到多点”的数据发布应用。

1.2 组播模型分类

根据接收者对组播源处理方式的不同，组播模型分为以下三类：

1. ASM 模型

简单地说，ASM（Any-Source Multicast，任意信源组播）模型就是任意源组播模型。

在 ASM 模型中，任意一个发送者都可以作为组播源向某组播组地址发送信息。众多接收者通过加入由该组播组地址标识的组播组以获得发往该组播组的组播信息。

在 ASM 模型中，接收者无法预先知道组播源的位置，但可以在任意时间加入或离开该组播组。

2. SFM 模型

SFM（Source-Filtered Multicast，信源过滤组播）模型继承了 ASM 模型，从发送者角度来看，两者的组播组成员关系完全相同。

SFM 模型在功能上对 ASM 模型进行了扩展。在 SFM 模型中，上层软件对收到的组播报文的源地址进行检查，允许或禁止来自某些组播源的报文通过。因此，接收者只能收到来自部分组播源的组播数据。从接收者的角度来看，只有部分组播源是有效的，组播源被经过了筛选。

3. SSM 模型

在现实生活中，用户可能只对某些组播源发送的组播信息感兴趣，而不愿接收其它源发送的信息。SSM（Source-Specific Multicast，指定信源组播）模型为用户提供了一种能够在客户端指定组播源的传输服务。

SSM 模型与 ASM 模型的根本区别在于：SSM 模型中的接收者已经通过其它手段预先知道了组播源的具体位置。SSM 模型使用与 ASM/SFM 模型不同的组播地址范围，直接在接收者与其指定的组播源之间建立专用的组播转发路径。

1.3 组播框架结构

对于 IP 组播，需要关注下列问题：

- 组播源将组播信息传输到哪里？即组播寻址机制；
- 网络中有哪些接收者？即主机注册；
- 这些接收者需要从哪个组播源接收信息？即组播源发现；
- 组播信息如何传输？即组播路由。

IP 组播属于端到端的服务，组播机制包括以下四个部分：

(1) 寻址机制：借助组播地址，实现信息从组播源发送到一组接收者；

(2) 主机注册：允许接收者主机动态加入和离开某组播组，实现对组播成员的管理；

(3) 组播路由：构建组播报文分发树（即组播数据在网络中的树型转发路径），并通过该分发树将报文从组播源传输到接收者；

(4) 组播应用：组播源与接收者必须安装支持视频会议等组播应用的软件，TCP/IP 协议栈必须支持组播信息的发送和接收。

1.3.1 组播地址

为了让组播源和组播组成员进行通信，需要提供网络层组播地址，即 IP 组播地址。同时必须存在一种技术将 IP 组播地址映射为链路层的组播 MAC 地址。

1. IPv4 组播地址

IANA（Internet Assigned Numbers Authority，互联网编号分配委员会）将D类地址空间分配给IPv4组播使用，范围从 224.0.0.0 到 239.255.255.255，具体分类及其含义如 表 1-2所示。

表1-2 IPv4 组播地址的范围及含义

地址范围	含义
224.0.0.0～224.0.0.255	永久组地址。除 224.0.0.0 保留不做分配外，其它地址供路由协议、拓扑查找和协议维护等使用，常用的永久组地址及其含义如 表 1-3所示。对于以该范围内组播地址为目的地址的数据包来说，不论其TTL（Time to Live，生存时间）值为多少，都不会被转发出本地网段
224.0.1.0～238.255.255.255	用户组地址，全网范围内有效。包含两种特定的组地址： • 232.0.0.0/8：SSM 组地址 • 233.0.0.0/8：GLOP 组地址
239.0.0.0～239.255.255.255	本地管理组地址，仅在本地管理域内有效。使用本地管理组地址可以灵活定义组播域的范围，以实现不同组播域之间的地址隔离，从而有助于在不同组播域内重复使用相同组播地址而不会引起冲突。详情请参见 RFC 2365

说明

- 组播组中的成员是动态的，主机可以在任何时刻加入或离开组播组。
- GLOP 是一种 AS（Autonomous System，自治系统）之间的组播地址分配机制，将 AS 号填入该范围内组播地址的中间两个字节中，每个 AS 都可以得到 255 个组播地址。有关 GLOP 的详细介绍请参见 RFC 2770。

表1-3 常用永久组地址及其含义

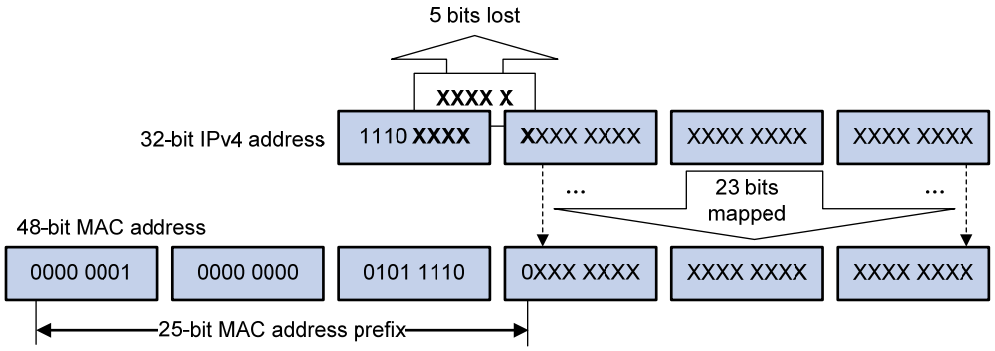
永久组地址	含义
224.0.0.1	所有系统，包括主机与路由器
224.0.0.2	所有组播路由器
224.0.0.3	未分配
224.0.0.4	DVMRP（Distance Vector Multicast Routing Protocol，距离矢量组播路由协议）路由器
224.0.0.5	OSPF（Open Shortest Path First，开放最短路径优先）路由器
224.0.0.6	OSPF 指定路由器/备用指定路由器
224.0.0.7	ST（Shared Tree，共享树）路由器
224.0.0.8	ST 主机
224.0.0.9	RIP-2（Routing Information Protocol version 2，路由信息协议版本 2）路由器
224.0.0.11	移动代理
224.0.0.12	DHCP（Dynamic Host Configuration Protocol，动态主机配置协议）服务器/中继代理
224.0.0.13	所有 PIM（Protocol Independent Multicast，协议无关组播）路由器
224.0.0.14	RSVP（Resource Reservation Protocol，资源预留协议）封装
224.0.0.15	所有 CBT（Core-Based Tree，有核树）路由器
224.0.0.16	指定 SBM（Subnetwork Bandwidth Management，子网带宽管理）
224.0.0.17	所有 SBM
224.0.0.18	VRRP（Virtual Router Redundancy Protocol，虚拟路由器冗余协议）

2. IPv4 组播 MAC 地址

以太网传输单播 IP 报文的时候，目的 MAC 地址使用的是接收者的 MAC 地址。但是在传输组播数据包时，其目的地不再是一个具体的接收者，而是一个成员不确定的组，所以要使用组播 MAC 地址。

IANA规定，IPv4 组播MAC地址的高 24 位为 0x01005E，第 25 位为 0，低 23 位为IPv4 组播地址的低 23 位。IPv4 组播地址与MAC地址的映射关系如 图 1-4所示。

图1-4 IPv4 组播地址与 MAC 地址的映射关系



由于 IPv4 组播地址的高 4 位是 1110，代表组播标识，而低 28 位中只有 23 位被映射到 IPv4 组播 MAC 地址，这样 IPv4 组播地址中就有 5 位信息丢失。于是，就有 32 个 IPv4 组播地址映射到了同一个 IPv4 组播 MAC 地址上，因此在二层处理过程中，设备可能要接收一些本 IPv4 组播组以外的组播数据，而这些多余的组播数据就需要设备的上层进行过滤了。

1.3.2 组播协议

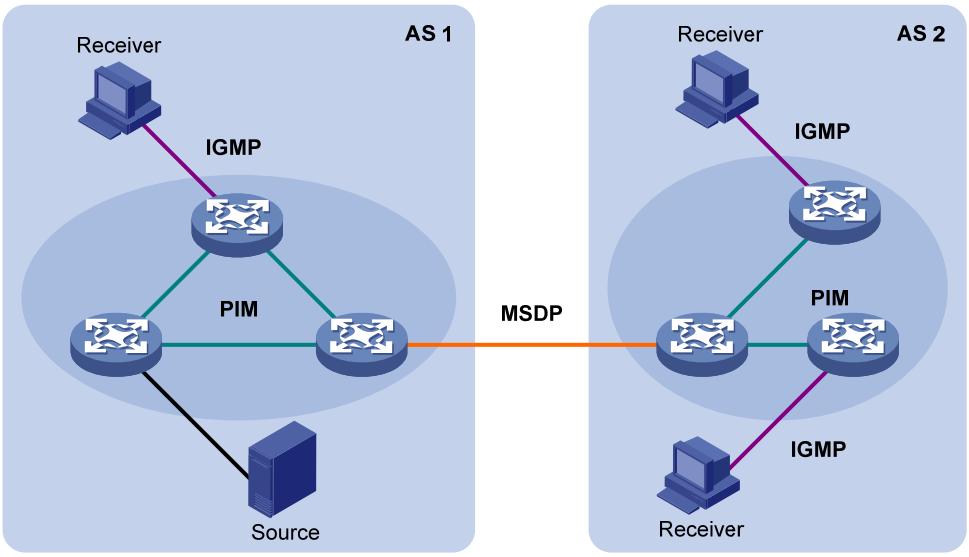
说明

- 通常，我们把工作在网络层的 IP 组播称为“三层组播”，相应的组播协议称为“三层组播协议”，包括 IGMP、PIM、MSDP 等；把工作在数据链路层的 IP 组播称为“二层组播”，相应的组播协议称为“二层组播协议”，包括组播 VLAN 等。
- 组播 VLAN、IGMP、PIM、MSDP 应用于 IPv4;
- 本节主要针对二、三层组播协议在网络中的应用位置和功能进行总体介绍，有关各协议的详细介绍请分别参见“IP 组播分册”中相应的配置手册。

1. 三层组播协议

三层组播协议包括组播组管理协议和组播路由协议两种类型，它们在网络中的应用位置如 图 1-5所示。

图1-5 三层组播协议的应用位置



(1) 组播组管理协议

在主机和与其直接相连的三层组播设备之间通常采用组播组的管理协议 IGMP（Internet Group Management Protocol，互联网组管理协议）或 MLD（Multicast Listener Discovery Protocol，组播侦听者发现协议），该协议规定了主机与三层组播设备之间建立和维护组播组成员关系的机制。

(2) 组播路由协议

组播路由协议运行在三层组播设备之间，用于建立和维护组播路由，并正确、高效地转发组播数据包。组播路由建立了从一个数据源端到多个接收端的无环（loop-free）数据传输路径，即组播分发树。

对于 ASM 模型，可以将组播路由分为域内和域间两大类：

- 域内组播路由用来在 AS 内部发现组播源并构建组播分发树，从而将组播信息传递到接收者。在众多域内组播路由协议中，PIM（Protocol Independent Multicast，协议无关组播）是目前较为典型的一个。按照转发机制的不同，PIM 可以分为 DM（Dense Mode，密集模式）和 SM（Sparse Mode，稀疏模式）两种模式。
- 域间组播路由用来实现组播信息在 AS 之间的传递，目前比较成型的解决方案有：MSDP（Multicast Source Discovery Protocol，组播源发现协议）能够跨越 AS 传播组播源的信息；而 MP-BGP（MultiProtocol Border Gateway Protocol，多协议边界网关协议）的组播扩展 MBGP（Multicast BGP）则能够跨越 AS 传播组播路由。

对于 SSM 模型，没有域内和域间的划分。由于接收者预先知道组播源的具体位置，因此只需要借助 PIM-SM 构建的通道即可实现组播信息的传输。

1.4 组播报文的转发机制

在组播模型中，IP 报文的目的地址字段为组播组地址，组播源向以此目的地址所标识的主机群组传送信息。因此，转发路径上的组播路由器为了将组播报文传送到各个方位的接收站点，往往需要将从一个入接口收到的组播报文转发到多个出接口。与单播模型相比，组播模型的复杂性就在于此：

- 为了保证组播报文在网络中的传输，必须依靠单播路由表或者单独提供给组播使用的路由表（如 MBGP 路由表）来指导转发；
- 为了处理同一设备在不同接口上收到来自不同对端的相同组播信息，需要对组播报文的入接口进行 RPF（Reverse Path Forwarding，逆向路径转发）检查，以决定转发还是丢弃该报文。RPF 检查机制是大部分组播路由协议进行组播转发的基础。

说明

有关 RPF 机制的详细介绍，请参见“IP 组播分册”中的“组播路由与转发配置”或“IPv6 组播路由与转发配置”。

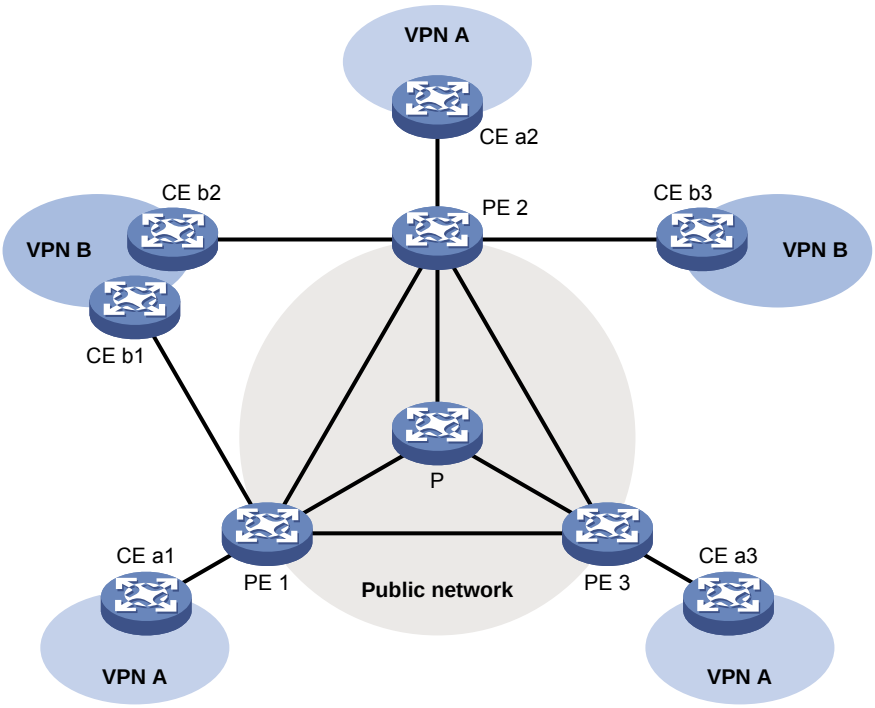
1.5 多实例组播

多实例组播是指在 VPN（Virtual Private Network，虚拟专用网）中应用的组播。

1.5.1 多实例简介

各VPN网络之间、VPN网络与公共网络之间要求信息隔离。如 图 1-6所示，VPN A和VPN B通过PE设备接入公共网络。

图1-6 VPN 典型组网



- P 设备专属于公网，各 CE 设备则专属于某一 VPN。每台设备只为其专属的网络服务，仅维护一套转发机制。
- PE 设备同时接入公网和 VPN 网络，同时为多个网络服务。在设备上必须严格区分各个网络的信息并为各个网络独立维护一套转发机制。这时，PE 设备上为同一网络服务的一套软硬件设施统称为一个实例（Instance）。PE 设备上同时存在多个实例，同一实例分布在多台 PE 设备上。

1.5.2 多实例在组播中的应用

在 PE 上应用多实例后，其具备以下功能：

- 为每个实例独立维护一套组播转发机制：支持各种组播协议，拥有各自独立的 PIM 邻居列表、组播路由表等信息。每个实例转发组播数据时只查找本实例的转发表或路由表；
- 保证各实例之间相互隔离；
- 实现公网实例和 VPN 实例之间的信息交流和数据转换。

多实例组播是实现跨越VPN传播组播数据的基础。以 图 1-6中的VPN实例A为例，组播VPN是指：当VPN A中的组播源向某组播组发送组播数据时，在网络中所有可能的接收者中，只有属于VPN A的组播组成员才能收到该组播源发来的组播数据。组播数据在VPN A中以组播方式进行传输，在公网中也以组播方式进行传输。



说明

- 非 PE 设备上只运行一套统一的组播业务，统称为“公网实例”。
- 在 VPN 实例视图下所作的配置只对该 VPN 实例接口有效，而不属于任何 VPN 实例的接口则统称为“公网实例接口”。

有关多实例组播的详细介绍，请参见“IP 组播分册”中的“组播 VPN 配置”。

目 录

1 组播路由与转发配置 1-1
1.1 组播路由与转发简介 1-1
1.1.1 RPF检查机制 1-1
1.1.2 组播静态路由 1-3
1.1.3 GRE隧道在组播转发中的应用 1-4
1.1.4 组播路径跟踪 1-5
1.2 组播路由与转发配置任务简介 1-6
1.3 使能IP组播路由 1-6
1.4 配置组播路由与转发 1-7
1.4.1 配置准备 1-7
1.4.2 配置组播静态路由 1-7
1.4.3 配置组播路由策略 1-8
1.4.4 配置组播转发范围 1-8
1.4.5 配置组播转发表容量 1-9
1.4.6 跟踪组播数据的传输路径 1-9
1.5 组播路由与转发显示和维护 1-10
1.6 组播路由与转发典型配置举例 1-11
1.6.1 改变RPF路由配置举例 1-11
1.6.2 衔接RPF路由配置举例 1-13
1.7 常见配置错误举例 1-15
1.7.1 组播静态路由失败 1-15

1 组播路由与转发配置

说明

本文所涉及的路由器和路由器图标，代表了一般意义下的路由器或运行了 IP 组播路由协议的三层交换机。

1.1 组播路由与转发简介

在组播实现中，组播路由和转发分为三种表：

- 每个组播路由协议都有一个协议自身的路由表，如 PIM 路由表（PIM Routing-Table）；
- 各组播路由协议的组播路由信息经过综合形成一个总的组播路由表（Multicast Routing-Table）；
- 组播转发表（Multicast Forwarding-Table）直接用于控制组播数据包的转发。

组播路由表由一组（S，G）表项组成，其中（S，G）表示由源 S 向组播组 G 发送组播数据的路由信息。如果路由器支持多种组播路由协议，则其组播路由表中将包括由多种协议生成的组播路由。路由器根据组播路由和转发策略，从组播路由表中选出最优的组播路由，并下发到组播转发表中。

1.1.1 RPF 检查机制

组播路由协议依赖于现有的单播路由信息、MBGP 路由或组播静态路由来创建组播路由表项。组播路由协议在创建组播路由表项时，运用了 RPF（Reverse Path Forwarding，逆向路径转发）检查机制，以确保组播数据能够沿正确的路径传输，同时还能避免由于各种原因而造成的环路。

1. RPF 检查过程

执行 RPF 检查的依据是单播路由、MBGP 路由或组播静态路由：

- 单播路由表中汇集了到达各个目的网段的最短路径；
- MBGP 路由表直接提供组播路由信息；
- 组播静态路由表中列出了用户通过手工静态配置指定的 RPF 路由信息。

在执行 RPF 检查时，路由器同时查找单播路由表、MBGP 路由表和组播静态路由表，具体过程如下：

(1) 首先，分别从单播路由表、MBGP 路由表和组播静态路由表中各选出一条最优路由：

- 以“报文源”的 IP 地址为目的地址查找单播路由表，自动选取一条最优单播路由。对应表项中的出接口为 RPF 接口，下一跳为 RPF 邻居。路由器认为来自 RPF 邻居且由该 RPF 接口收到的组播报文所经历的路径是从源 S 到本地的最短路径。
- 以“报文源”的 IP 地址为目的地址查找 MBGP 路由表，自动选取一条最优 MBGP 路由。对应表项中的出接口为 RPF 接口，下一跳为 RPF 邻居。
- 以“报文源”的 IP 地址为指定源地址查找组播静态路由表，自动选取一条最优组播静态路由。对应表项明确指定了 RPF 接口和 RPF 邻居。

(2) 然后，从这三条最优路由中选择一条作为 RPF 路由：

- 如果配置了按照最长匹配选择路由，则从这三条路由中选出最长匹配的那条路由；如果这三条路由的掩码一样，则选择其中优先级最高的那条路由；如果它们的优先级也相同，则按照组播静态路由、MBGP 路由、单播路由的顺序进行选择。
- 如果没有配置按照最长匹配选择路由，则从这三条路由中选出优先级最高的那条路由；如果它们的优先级相同，则按照组播静态路由、MBGP 路由、单播路由的顺序进行选择。



说明

根据组播报文传输的具体情况不同，“报文源”所代表的具体含义也不同：

- 如果当前报文沿从组播源到接收者或 RP（Rendezvous Point，汇集点）的 SPT（Shortest Path Tree，最短路径树）进行传输，则以组播源为“报文源”进行 RPF 检查；
- 如果当前报文沿从 RP 到接收者的 RPT（Rendezvous Point Tree，共享树）进行传输，则以 RP 为“报文源”进行 RPF 检查；
- 如果当前报文为 BSR（BootStrap Router，自举路由器）报文，沿从 BSR 到各路由器的路径进行传输，则以 BSR 为“报文源”进行 RPF 检查。

有关 SPT、RPT、RP 和 BSR 的详细介绍，请参见“IP 组播分册”中的“PIM 配置”。

---

2. RPF 检查在组播转发中的应用

对每一个收到的组播数据报文都进行 RPF 检查会给路由器带来较大负担，而利用组播转发表可以解决这个问题。在建立组播路由和转发表时，会把组播数据报文（S，G）的 RPF 接口记录为（S，G）表项的入接口。当路由器收到组播数据报文（S，G）后，查找组播转发表：

(1) 如果组播转发表中不存在（S，G）表项，则对该报文执行 RPF 检查，将其 RPF 接口作为入接口，结合相关路由信息创建相应的表项，并下发到组播转发表中：

- 若该报文实际到达的接口正是其 RPF 接口，则 RPF 检查通过，向所有的出接口转发该报文；
- 若该报文实际到达的接口不是其 RPF 接口，则 RPF 检查失败，丢弃该报文。

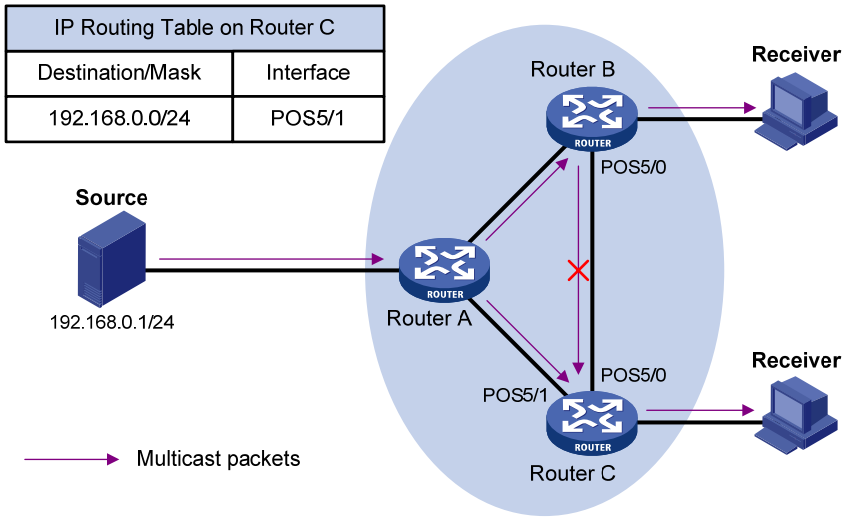
(2) 如果组播转发表中已存在（S，G）表项，且该报文实际到达的接口与入接口相匹配，则向所有的出接口转发该报文。

(3) 如果组播转发表中已存在（S，G）表项，但该报文实际到达的接口与入接口不匹配，则对此报文执行 RPF 检查：

- 若其 RPF 接口与入接口一致，则说明（S，G）表项正确，丢弃这个来自错误路径的报文；
- 若其 RPF 接口与入接口不符，则说明（S，G）表项已过时，于是把入接口更新为 RPF 接口。如果该报文实际到达的接口正是其 RPF 接口，则向所有的出接口转发该报文，否则将其丢弃。

如 图 1-1所示，假设网络中单播路由畅通，未配置MBGP，Router C上也未配置组播静态路由。组播报文（S，G）沿从组播源（Source）到接收者（Receiver）的SPT进行传输。假定Router C上的组播转发表中已存在（S，G）表项，其记录的入接口为POS5/1。

图1-1 RPF 检查过程



- 如果该组播报文从接口 POS5/1 到达 Router C，与（S，G）表项的入接口相匹配，则向所有的出接口转发该报文。
- 如果该组播报文从接口 POS5/0 到达 Router C，与（S，G）表项的入接口不匹配，则对其执行 RPF 检查：通过查找单播路由表发现到达 Source 的出接口（即 RPF 接口）是 POS5/1，与（S，G）表项的入接口一致。这说明（S，G）表项是正确的，该报文来自错误的路径，RPF 检查失败，于是丢弃该报文。

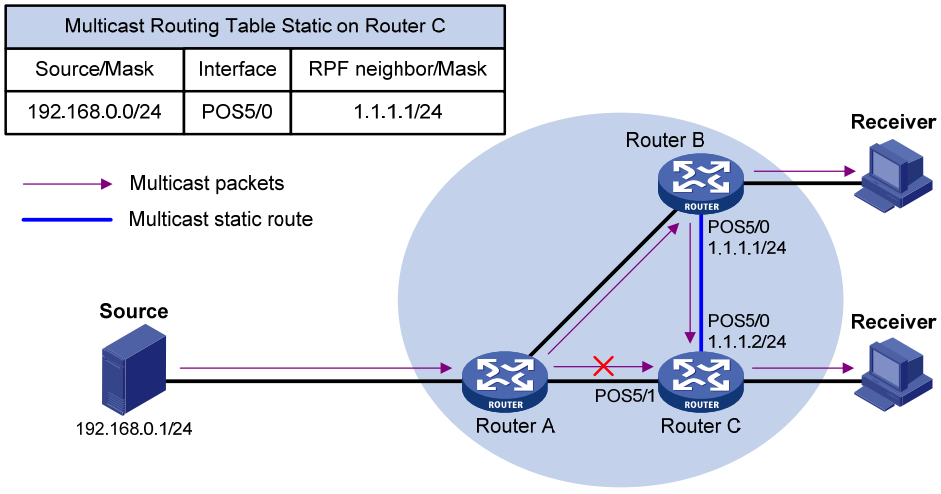
1.1.2 组播静态路由

组播静态路由是 RPF 检查的重要依据之一。根据具体应用环境的不同，组播静态路由有以下两种主要用途：

1. 改变 RPF 路由

通常，组播的网络拓扑结构与单播相同，组播数据的传输路径也与单播相同。可以通过配置组播静态路由以改变 RPF 路由，从而为组播数据创建一条与单播不同的传输路径。

图1-2 改变 RPF 路由示意图



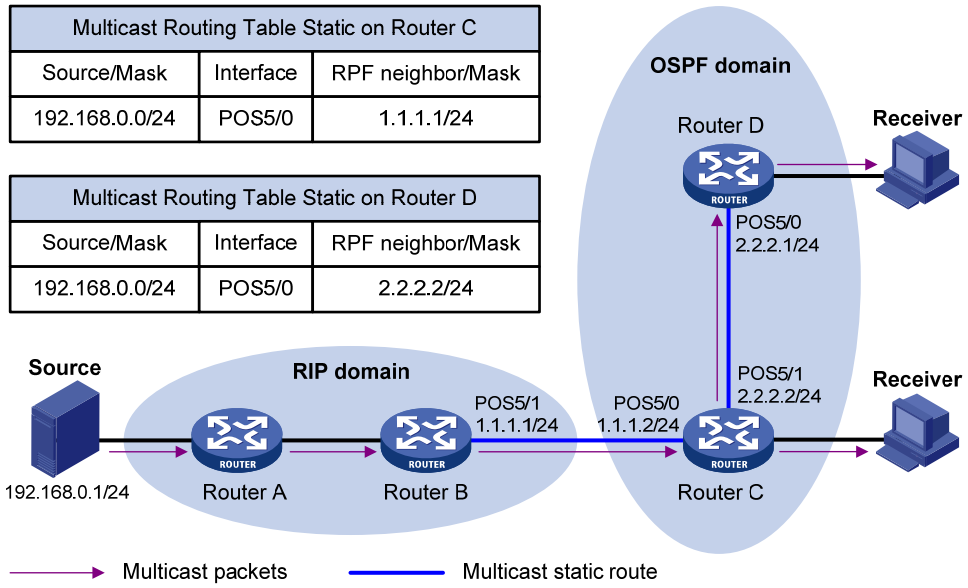
如 图 1-2所示，当网络中没有配置组播静态路由时，Router C到组播源（Source）的RPF邻居为Router A，从Source发出的组播信息沿Router A—Router C的路径传输，与单播路径一致；当在Router C

上配置了组播静态路由，指定从Router C到Source的RPF邻居为Router B之后，从Source发出的组播信息将改变传输路径，沿Router A—Router B—Router C的新路径传输。

2. 衔接 RPF 路由

当网络中的单播路由被阻断时，由于没有 RPF 路由而无法进行包括组播数据在内的数据转发。可以通过配置组播静态路由以生成 RPF 路由，从而创建组播路由表项以指导组播数据的转发。

图1-3 衔接 RPF 路由示意图



如 图 1-3所示，RIP域与OSPF域之间实行单播路由隔离。当网络中没有配置组播静态路由时，OSPF域内的接收者（Receiver）不能收到RIP域内的组播源（Source）所发出的组播信息；当在Router C和Router D上均配置了组播静态路由，分别指定从Router C到Source的RPF邻居为Router B、从Router D到Source的RPF邻居为Router C之后，Receiver便能收到Source发出的组播信息了。

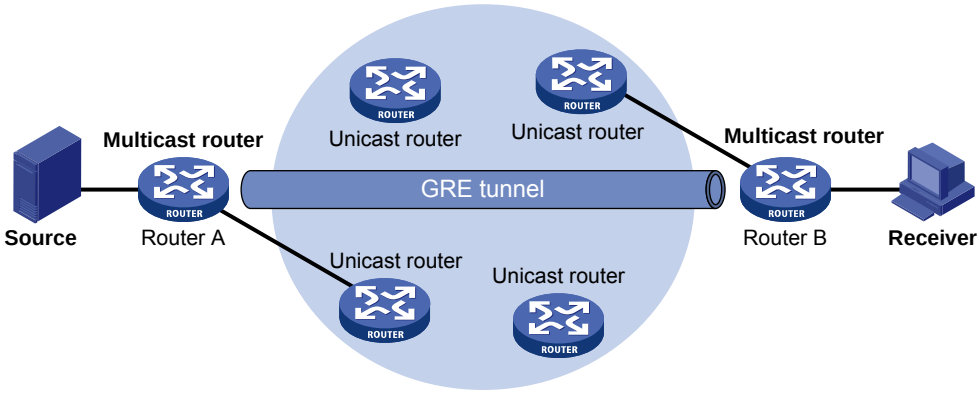
说明

- 组播静态路由的作用只在于影响 RPF 检查，而不能用于指导组播数据转发，故又称为 RPF 静态路由；
- 组播静态路由仅在所配置的组播路由器上生效，不会以任何方式被广播或者引入给其它路由器。

1.1.3 GRE 隧道在组播转发中的应用

网络中可能存在不支持组播协议的路由器。从组播源（Source）发出的组播数据沿组播路由器逐跳转发，当下一跳路由器不支持组播协议时，组播转发路径将被阻断。此时，通过在处于单播网段两端的组播路由器之间建立 GRE（Generic Routing Encapsulation，通用路由封装）隧道，可以实现跨越单播网段的组播数据交换。有关 GRE 隧道的详细介绍，请参见“VPN 分册”中的“GRE 配置”。

图1-4 使用隧道方式传送组播数据



如 图 1-4所示，在Router A和Router B之间建立起GRE隧道。Router A将组播数据包封装在单播IP报文中，经由单播路由器转发，传送到隧道另一端的Router B。然后，Router B将单播IP报文头剥掉，继续进行组播传输。

如果在隧道的两端配置了单播静态路由，则任意单播数据包都可以通过该隧道传输。为了将该隧道专用于组播数据包的传输，可以在隧道两端只配置组播静态路由，从而使单播数据包的传输不能再使用此隧道。

1.1.4 组播路径跟踪

Multicast traceroute（组播路径跟踪）用来跟踪组播数据从第一跳到最后一跳路由器所经过的路径。

1. Multicast traceroute 基本概念

(1) 最后一跳路由器（Last-hop Router）：如果某路由器有一个接口的 IP 地址与指定地址在同一个网段内，具备组播功能，且能够向该网段转发特定组播源发来的组播流，则称该路由器为最后一跳路由器。

(2) 第一跳路由器（First-hop Router）：与组播源直连的路由器。

(3) 查询器（Querier）：触发组播路径跟踪的路由器。

2. Multicast traceroute 报文简介

Multicast traceroute 报文是一种特殊的 IGMP 报文，与普通 IGMP 报文的区别在于其类型字段为 0x1F/0x1E，且其目的 IP 地址为单播地址。Multicast traceroute 报文分为以下三种类型：

- Query 报文：类型字段为 0x1F
- Request 报文：类型字段为 0x1F
- Response 报文：类型字段为 0x1E

3. Multicast traceroute 执行步骤

(1) 查询器向最后一跳路由器发送 Query 报文；

(2) 最后一跳路由器在收到的 Query 报文后加上本地响应数据块转换成 Request 报文，查找到上游邻居后向其单播发送该 Request 报文；

(3) 最后一跳路由器到组播源之间的每一跳都在 Request 报文之后附加一个响应数据块，并向其上游邻居单播转发；

(4) 第一跳路由器在收到 Request 报文后，将其报文类型改为 Response 报文，向查询器单播发送完整的报文。

1.2 组播路由与转发配置任务简介

表1-1 组播路由与转发配置任务简介

配置任务		说明	详细配置
使能 IP 组播路由		必选	1.3
配置组播路由与转发	配置组播静态路由	可选	1.4.2
	配置组播路由策略	可选	1.4.3
	配置组播转发范围	可选	1.4.4
	配置组播转发表容量	可选	1.4.5
	跟踪组播数据的传输路径	可选	1.4.6

注意

IP 组播不支持从 IP 地址网段应用。也就是说，当为接口配置了从 IP 地址后，组播数据并不能通过从 IP 地址所属的网段进行路由和转发，而只能通过主 IP 地址进行路由和转发。有关主、从 IP 地址的详细介绍，请参见“IP 业务分册”中的“IP 地址配置”。

1.3 使能 IP 组播路由

在配置各项三层组播功能之前，必须首先使能 IP 组播路由。

1. 使能公网实例中的 IP 组播路由

表1-2 使能公网实例中的 IP 组播路由

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态

2. 使能 VPN 实例中的 IP 组播路由

表1-3 使能 VPN 实例中的 IP 组播路由

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD

操作	命令	说明
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态



说明

有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。

1.4 配置组播路由与转发

1.4.1 配置准备

在配置组播路由与转发之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 PIM-DM（或 PIM-SM）

在配置组播路由与转发之前，需准备以下数据：

- 组播转发的最小 TTL 值
- 单条组播转发表项的最大下行节点数目
- 组播转发表的最大表项数

1.4.2 配置组播静态路由

通过配置组播静态路由，可以为来自特定组播源的组播报文指定 RPF 接口或 RPF 邻居。

表1-4 配置组播静态路由

操作	命令	说明
进入系统视图	**system-view**	-
配置组播静态路由	**ip rpf-route-static** [**vpn-instance** *vpn-instance-name*] *source-address* { *mask* | *mask-length* } [*protocol* [*process-id*]] [**route-policy** *policy-name*] { *rpf-nbr-address* | *interface-type interface-number* } [**preference** *preference*] [**order** *order-number*]	必选 缺省情况下，没有配置组播静态路由



注意

在配置组播静态路由时，若 RPF 邻居的接口类型是以太网接口、三层聚合接口、Loopback 接口时，不能使用指定接口（*interface-type interface-number*）的方式指定 RPF 邻居，而只能使用指定地址（*rpf-nbr-address*）的方式指定 RPF 邻居。

1.4.3 配置组播路由策略

可以配置组播路由器按照最长匹配原则来选择RPF路由，有关RPF路由选择的详细介绍，请参见“1.1.1 1. RPF检查过程”一节。此外，通过配置根据组播源或组播源组进行组播流量的负载分担，还可以优化存在多条组播数据流时的网络流量。

1. 配置公网实例中的组播路由策略

表1-5 配置公网实例中的组播路由策略

操作	命令	说明
进入系统视图	**system-view**	-
配置按照最长匹配选择 RPF 路由	**multicast longest-match**	可选 缺省情况下，选择优先级最高的路由作为 RPF 路由
配置对组播流量进行负载分担	**multicast load-splitting { source | source-group }**	可选 缺省情况下，不对组播流量进行负载分担

2. 配置 VPN 实例中的组播路由策略

表1-6 配置 VPN 实例中的组播路由策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置按照最长匹配选择 RPF 路由	**multicast longest-match**	可选 缺省情况下，选择优先级最高的路由作为 RPF 路由
配置对组播流量进行负载分担	**multicast load-splitting { source | source-group }**	可选 缺省情况下，不对组播流量进行负载分担

1.4.4 配置组播转发范围

组播信息在网络中的转发并不是漫无边际的，每个组播组对应的组播信息都必须在确定的范围内传递。确定充当组播转发边界的接口，以形成一个封闭的组播转发区域；

可以在所有支持组播转发的接口上配置针对某个组播组的转发边界。组播转发边界为指定范围的组播组划定了边界条件，如果组播报文的目的地址与边界条件匹配，就停止转发。当在一个接口上配置了组播转发边界后，将不能从该接口转发组播报文（包括本机发出的组播报文），也不能从该接口接收组播报文。

表1-7 配置组播转发范围

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置组播转发边界	**multicast boundary** *group-address* { *mask* | *mask-length* }	必选 缺省情况下，没有配置组播转发边界

1.4.5 配置组播转发表容量

路由器为每个收到的组播数据报文都维护相应的转发表项。但是，过量的组播转发表项可能会耗尽路由器内存，从而导致路由器性能下降。用户可以根据实际组网情况和业务性能要求对组播转发表中的表项数量进行限制。如果组播转发表最大表项数的配置值小于当前值，则超出数目的表项并不会立刻被删除，而必须由组播路由协议来删除，同时也无法添加新的组播转发表项。

路由器为每个下行节点复制一份组播数据报文并发送出去，每个下行节点就形成组播分发树的一条分支。用户可以根据实际组网情况和业务性能要求对组播转发表中单条表项的下行节点数目（即出接口数目）进行限制，以缓解路由器的复制压力。如果单条组播转发表项的最大下行节点数目的配置值小于当前值，则超出数目的下行节点并不会被立刻删除，而必须由组播路由协议来删除，同时新增的下行节点将无法添加到该表项中。

1. 配置公网实例中的组播转发表容量

表1-8 配置公网实例中的组播转发表容量

操作	命令	说明
进入系统视图	**system-view**	-
配置组播转发表的最大表项数	**multicast forwarding-table route-limit** *limit*	可选 缺省情况下，组播转发表的最大表项数为系统所允许的最大值，取值范围为 0～4096
配置单条组播转发表项的最大下行节点数目	**multicast forwarding-table downstream-limit** *limit*	可选 缺省情况下，单条组播转发表项的最大下行节点数目为系统所允许的最大值，取值范围为 0～128

2. 配置 VPN 实例中的组播转发表容量

表1-9 配置 VPN 实例中的组播转发表容量

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置组播转发表的最大表项数	**multicast forwarding-table route-limit** *limit*	可选 缺省情况下，组播转发表的最大表项数为系统所允许的最大值，取值范围为 0～4096
配置单条组播转发表项的最大下行节点数目	**multicast forwarding-table downstream-limit** *limit*	可选 缺省情况下，单条组播转发表项的最大下行节点数目为系统所允许的最大值，取值范围为 0～128

1.4.6 跟踪组播数据的传输路径

可在任意视图下执行 **mtracert** 命令来跟踪组播数据从第一跳到最后一跳路由器所经过的路径。

表1-10 跟踪组播数据的传输路径

操作	命令	说明
跟踪组播数据的传输路径	**mtracert** *source-address* [[*last-hop-router-address*] *group-address*]	必选 可在任意视图下执行

1.5 组播路由与转发显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后组播路由与转发的信息，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除组播路由与转发的统计信息。

表1-11 组播路由与转发显示和维护

操作		命令
查看组播边界信息		**display multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **boundary** [*group-address* [*mask* | *mask-length*]] [**interface** *interface-type interface-number*]
查看组播转发表信息	集中式设备	**display multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **forwarding-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }]] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics**] * [**port-info**]
	分布式设备	**display multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **forwarding-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }]] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics** | **slot** *slot-number*] * [**port-info**]
查看组播路由表信息		**display multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **routing-table** [*source-address* [**mask** { *mask* | *mask-length* }]] | *group-address* [**mask** { *mask* | *mask-length* }]] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } }] *
查看组播静态路由信息		**display multicast routing-table** [**vpn-instance** *vpn-instance-name* | **all-instance**] **static** [**config**] [*source-address* { *mask-length* | *mask* }]
查看组播源的 RPF 信息		**display multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **rpf-info** *source-address* [*group-address*]
清除组播转发表中的转发项		**reset multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **forwarding-table** { { *source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }]] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }
清除组播路由表中的路由项		**reset multicast** [**vpn-instance** *vpn-instance-name* | **all-instance**] **routing-table** { { *source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }]] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }

注意

- 执行 **reset** 命令将清除组播路由表或组播转发表中的信息，可能导致组播信息无法正常传输；
- 清除组播路由表中的路由项后，组播转发表中的相应表项也将随之删除；
- 清除组播转发表中的转发项后，组播路由表中的相应表项也将随之删除。

1.6 组播路由与转发典型配置举例

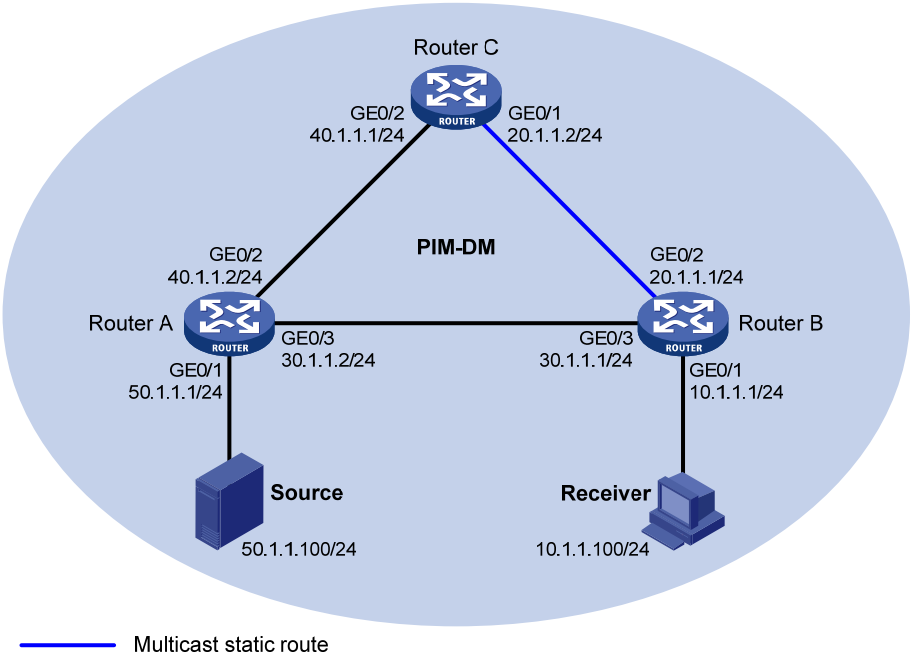
1.6.1 改变 RPF 路由配置举例

1. 组网需求

- 网络中运行 PIM-DM，所有路由器都支持组播功能；
- Router A、Router B 和 Router C 之间运行 OSPF 协议；
- 通常情况下，Receiver 能通过 Router A—Router B 这条与单播路径相同的路径接收来自 Source 的组播信息；
- 要求通过配置，使 Receiver 能通过 Router A—Router C—Router B 这条与单播路径不同的路径接收来自 Source 的组播信息。

2. 组网图

图1-5 改变 RPF 路由配置举例



3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-5配置各接口的IP地址和掩码，具体配置过程略。

配置 PIM-DM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-DM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-DM 和 IGMP

在 Router B 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在主机侧接口 GigabitEthernet 0/1 上使能 IGMP。

```
<RouterB> system-view
[RouterB] multicast routing-enable
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] igmp enable
[RouterB-GigabitEthernet0/1] pim dm
[RouterB-GigabitEthernet0/1] quit
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] pim dm
[RouterB-GigabitEthernet0/2] quit
[RouterB] interface GigabitEthernet 0/3
[RouterB-GigabitEthernet0/3] pim dm
[RouterB-GigabitEthernet0/3] quit
```

在 Router A 上使能 IP 组播路由，并在各接口上使能 PIM-DM。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim dm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] pim dm
[RouterA-GigabitEthernet0/2] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] pim dm
[RouterA-GigabitEthernet0/3] quit
```

Router C 上的配置与 Router A 相似，配置过程略。

在 Router B 上使用 **display multicast rpf-info** 命令查看到 Source 的 RPF 信息。

```
[RouterB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Ethernet1/3, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: igp
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

Router B 上当前的 RPF 路由来源于单播路由，RPF 邻居是 Router A。

(3) 配置组播静态路由

在 Router B 上配置组播静态路由，到 Source 的 RPF 邻居为 Router C。

```
[RouterB] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

(4) 检验配置效果

在 Router B 上使用 **display multicast rpf-info** 命令查看到 Source 的 RPF 信息。

```
[RouterB] display multicast rpf-info 50.1.1.100
```

```
RPF information about source 50.1.1.100:
    RPF interface: Ethernet1/2, RPF neighbor: 20.1.1.2
    Referenced route/mask: 50.1.1.0/24
    Referenced route type: multicast static
    Route selection rule: preference-preferred
    Load splitting rule: disable
```

与配置组播静态路由前相比，Router B 上的 RPF 路由已经产生了变化，其来源变为组播静态路由，RPF 邻居变为 Router C。

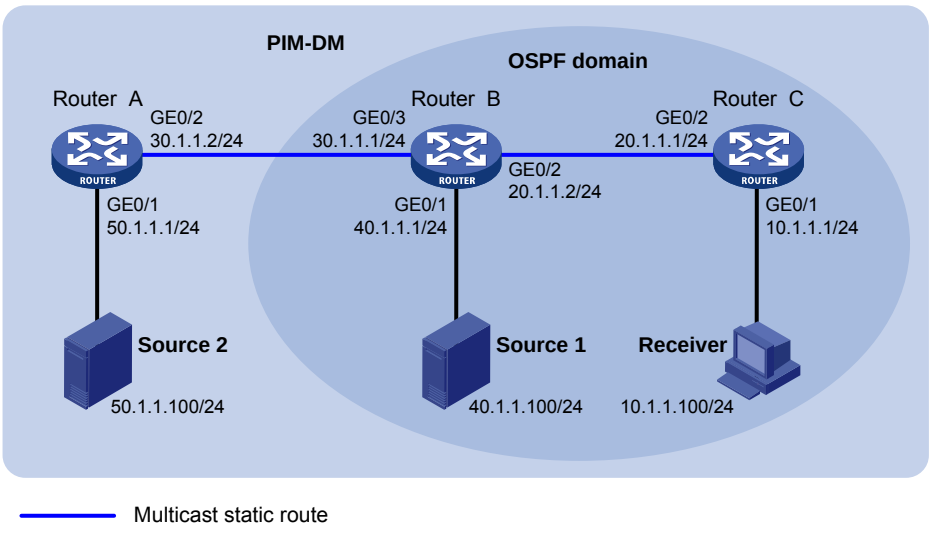
1.6.2 衔接 RPF 路由配置举例

1. 组网需求

- 网络中运行 PIM-DM，所有路由器都支持组播功能；
- Router B 和 Router C 之间运行 OSPF 协议，并与 Router A 单播路由隔离；
- 通常情况下，Receiver 能接收来自 OSPF 域内 Source 1 的组播信息；
- 要求通过配置，使 Receiver 也可以接收来自 OSPF 域外 Source 2 的组播信息。

2. 组网图

图1-6 衔接 RPF 路由配置组网图



3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-6配置各接口的IP地址和掩码，具体配置过程略。

配置 Router B 和 Router C 之间采用 OSPF 协议进行互连，确保 Router B 和 Router C 之间在网络层互通，并且能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-DM 和 IGMP

在 Router C 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在主机侧接口 Ethernet1/1 上使能 IGMP。

```
<RouterC> system-view
[RouterC] multicast routing-enable
```

```
[RouterC] interface GigabitEthernet 0/1
[RouterC-GigabitEthernet0/1] igmp enable
[RouterC-GigabitEthernet0/1] pim dm
[RouterC-GigabitEthernet0/1] quit
[RouterC] interface GigabitEthernet 0/2
[RouterC-GigabitEthernet0/2] pim dm
[RouterC-GigabitEthernet0/2] quit
```

在 Router A 上使能 IP 组播路由，并在各接口上使能 PIM-DM。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim dm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] pim dm
[RouterA-GigabitEthernet0/2] quit
```

Router B 上的配置与 Router A 相似，配置过程略。

在 Router B 和 Router C 上分别使用 **display multicast rpf-info** 命令查看到 Source 2 的 RPF 信息。

```
[RouterB] display multicast rpf-info 50.1.1.100
[RouterC] display multicast rpf-info 50.1.1.100
```

没有显示信息输出，说明在 Router B 和 Router C 上都没有到 Source 2 的 RPF 路由。

(3) 配置组播静态路由

在 Router B 上配置组播静态路由，到 Source 2 的 RPF 邻居为 Router A。

```
[RouterB] ip rpf-route-static 50.1.1.100 24 30.1.1.2
```

在 Router C 上配置组播静态路由，到 Source 2 的 RPF 邻居为 Router B。

```
[RouterC] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

(4) 检验配置效果

在 Router B 和 Router C 上分别使用 **display multicast rpf-info** 命令查看到 Source 2 的 RPF 信息。

```
[RouterB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Ethernet1/3, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
[RouterC] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Ethernet1/2, RPF neighbor: 20.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

与配置组播静态路由前相比，Router B 和 Router C 上都有了到 Source 2 的 RPF 路由，且其均来源于组播静态路由。

1.7 常见配置错误举例

1.7.1 组播静态路由失败

1. 故障现象

路由器没有配置动态路由协议，接口的物理状态与链路层协议状态都显示为 up；但是组播静态路由失败。

2. 分析

- 如果没有正确配置或更新与当前网络情况相匹配的组播静态路由，则组播路由表中不存在此路由项以及组播静态路由的配置信息；
- 如果查询到了最佳路由，也可能导致组播静态路由失败。

3. 处理过程

(1) 使用 **display multicast routing-table static config** 命令在组播路由表中查看组播静态路由的配置信息，以确定是否正确配置了对应的路由并存在于组播路由表中；

(2) 使用 **display multicast routing-table static** 命令在组播路由表中查看组播静态路由的信息，以确定是否正确配置了对应的路由并存在于组播路由表中；

(3) 检查组播静态路由下一跳接口的接口类型。若为非点到点接口，则配置组播静态路由时，出接口必须使用下一跳地址的形式配置；

(4) 检查是否匹配指定的路由协议。如果指定了协议，使用 **display ip routing-table** 命令检查是否该协议添加了相同的路由；

(5) 检查是否匹配指定的路由策略。如果指定了路由策略，使用 **display route-policy** 命令检查配置的路由策略。

目 录

1 IGMP配置 1-1
1.1 IGMP简介 1-1
1.1.1 IGMP的版本 1-1
1.1.2 IGMPv1 工作机制 1-1
1.1.3 IGMPv2 的改进 1-3
1.1.4 IGMPv3 的改进 1-3
1.1.5 IGMP SSM Mapping 1-5
1.1.6 IGMP Proxying 1-6
1.1.7 多实例的IGMP 1-7
1.1.8 协议规范 1-7
1.2 IGMP配置任务简介 1-7
1.3 配置IGMP基本功能 1-8
1.3.1 配置准备 1-8
1.3.2 使能IGMP 1-8
1.3.3 配置IGMP版本 1-9
1.3.4 配置静态加入 1-9
1.3.5 配置组播组过滤器 1-10
1.3.6 配置组播组数量限制 1-10
1.4 调整IGMP性能 1-10
1.4.1 配置准备 1-11
1.4.2 配置IGMP报文选项 1-11
1.4.3 配置IGMP查询和响应 1-12
1.4.4 配置IGMP快速离开 1-14
1.5 配置IGMP SSM Mapping 1-15
1.5.1 配置准备 1-15
1.5.2 使能IGMP SSM Mapping 1-15
1.5.3 配置IGMP SSM Mapping规则 1-16
1.6 配置IGMP Proxying 1-16
1.6.1 配置准备 1-16
1.6.2 使能IGMP Proxying 1-16
1.6.3 配置下行接口的组播转发能力 1-17
1.7 IGMP显示和维护 1-18
1.8 IGMP典型配置举例 1-19
1.8.1 IGMP基本功能配置举例 1-19
1.8.2 IGMP SSM Mapping功能配置举例 1-21
1.8.3 IGMP Proxying功能配置举例 1-24
1.9 常见配置错误举例 1-25
1.9.1 接收者侧路由器上无组成员信息 1-25
1.9.2 同一网段各路由器上组成员关系不一致 1-26

1 IGMP 配置

说明

本文所涉及的路由器和路由器图标，代表了一般意义下的路由器或运行了 IGMP 协议的三层交换机。

1.1 IGMP 简介

IGMP 是 Internet Group Management Protocol（互联网组管理协议）的简称。它是 TCP/IP 协议族中负责 IP 组播成员管理的协议，用来在 IP 主机和与其直接相邻的组播路由器之间建立、维护组播组成员关系。

1.1.1 IGMP 的版本

到目前为止，IGMP 有三个版本：

- IGMPv1（由 RFC 1112 定义）
- IGMPv2（由 RFC 2236 定义）
- IGMPv3（由 RFC 3376 定义）

所有版本的 IGMP 都支持 ASM（Any-Source Multicast，任意信源组播）模型；IGMPv3 可以直接应用于 SSM（Source-Specific Multicast，指定信源组播）模型，而 IGMPv1 和 IGMPv2 则需要在 IGMP SSM Mapping 技术的支持下才能应用于 SSM 模型。

说明

有关 ASM 和 SSM 模型的介绍，请参见“IP 组播分册”中的“组播概述”。

1.1.2 IGMPv1 工作机制

IGMPv1 主要基于查询和响应机制来完成对组播组成员的管理。

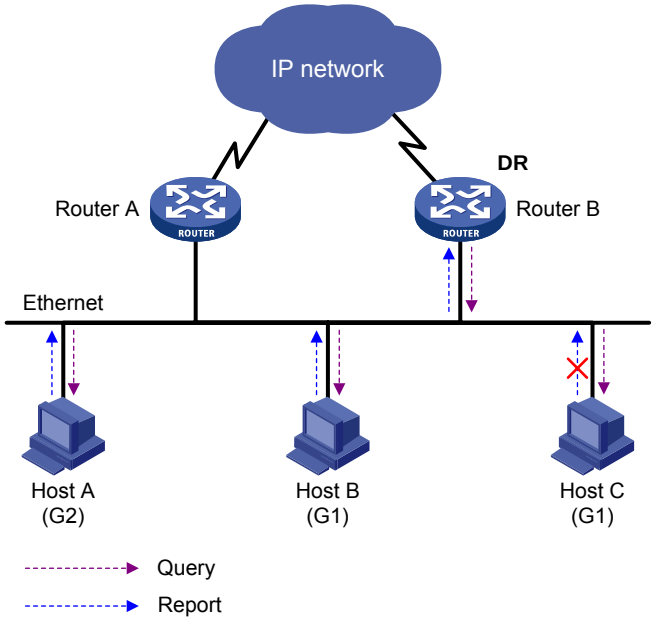
当一个网段内有多台组播路由器时，由于它们都能从主机那里收到 IGMP 成员关系报告报文（Membership Report Message），因此只需要其中一台路由器发送 IGMP 查询报文（Query Message）就足够了。这就需要有一个查询器（Querier）的选举机制来确定由哪台路由器作为 IGMP 查询器。

对于 IGMPv1 来说，由组播路由协议（如 PIM）选举出唯一的组播信息转发者 DR（Designated Router，指定路由器）作为 IGMP 查询器。

说明

有关 DR 的介绍，请参见“IP 组播分册”中的“PIM 配置”。

图1-1 IGMP 查询响应示意图



如 图 1-1所示，假设Host B与Host C想要收到发往组播组G1 的组播数据，而Host A想要收到发往组播组G2 的组播数据，那么主机加入组播组以及IGMP查询器（Router B）维护组播组成员关系的基本过程如下：

(1) 主机会主动向其要加入的组播组发送 IGMP 成员关系报告报文以声明加入，而不必等待 IGMP 查询器发来的 IGMP 查询报文；

(2) IGMP 查询器周期性地以组播方式向本地网段内的所有主机与路由器发送 IGMP 查询报文（目的地址为 224.0.0.1）；

(3) 在收到该查询报文后，关注 G1 的 Host B 与 Host C 其中之一（这取决于谁的延迟定时器先超时）——譬如 Host B 会首先以组播方式向 G1 发送 IGMP 成员关系报告报文，以宣告其属于 G1。由于本地网段中的所有主机和路由器都能收到 Host B 发往 G1 的报告报文，因此当 Host C 收到该报告报文后，将不再发送同样针对 G1 的报告报文，因为 IGMP 路由器（Router A 和 Router B）已知道本地网段中有对 G1 感兴趣的主机了。这个机制称为主机上的 IGMP 成员关系报告抑制机制，该机制有助于减少本地网段的信息流量；

(4) 与此同时，由于 Host A 关注的是 G2，所以它仍将以组播方式向 G2 发送报告报文，以宣告其属于 G2；

(5) 经过以上的查询和响应过程，IGMP 路由器了解到本地网段中有 G1 和 G2 的成员，于是由组播路由协议（如 PIM）生成（*，G1）和（*，G2）组播转发项作为组播数据的转发依据，其中的“*”代表任意组播源；

(6) 当由组播源发往 G1 或 G2 的组播数据经过组播路由到达 IGMP 路由器时，由于 IGMP 路由器上存在（*，G1）和（*，G2）组播转发项，于是将该组播数据转发到本地网段，接收者主机便能收到该组播数据了。

IGMPv1 没有专门定义离开组播组的报文。当运行 IGMPv1 的主机离开某组播组时，将不会向其要离开的组播组发送报告报文。当网段中不再存在该组播组的成员后，IGMP 路由器将收不到任何发往该组播组的报告报文，于是 IGMP 路由器在一段时间之后便删除该组播组所对应的组播转发项。

1.1.3 IGMPv2 的改进

与 IGMPv1 相比，IGMPv2 增加了查询器选举机制和离开组机制。

1. 查询器选举机制

在 IGMPv1 中，当某共享网段上存在多个组播路由器时，由组播路由协议（如 PIM）选举的指定路由器充当查询器。

在 IGMPv2 中，增加了独立的查询器选举机制，其选举过程如下：

(1) 所有 IGMPv2 路由器在初始时都认为自己是查询器，并向本地网段内的所有主机和路由器发送 IGMP 普遍组查询（General Query）报文（目的地址为 224.0.0.1）；

(2) 本地网段中的其它 IGMPv2 路由器在收到该报文后，将报文的源 IP 地址与自己的接口地址作比较。通过比较，IP 地址最小的路由器将成为查询器，其它路由器成为非查询器（Non-Querier）；

(3) 所有非查询器上都会启动一个定时器（即其它查询器存在时间定时器 Other Querier Present Timer）。在该定时器超时前，如果收到了来自查询器的 IGMP 查询报文，则重置该定时器；否则，就认为原查询器失效，并发起新的查询器选举过程。

2. 离开组机制

在 IGMPv1 中，主机离开组播组时不会向组播路由器发出任何通知，导致组播路由器只能依靠组播组成员查询的响应超时来获知组播组成员的离开。

而在 IGMPv2 中，当一个主机离开某组播组时：

(1) 该主机向本地网段内的所有组播路由器（目的地址为 224.0.0.2）发送离开组（Leave Group）报文；

(2) 当查询器收到该报文后，向该主机所声明要离开的那个组播组发送特定组查询（Group-Specific Query）报文（目的地址字段和组地址字段均填充为所要查询的组播组地址）；

(3) 如果该网段内还有该组播组的其它成员，则这些成员在收到特定组查询报文后，会在该报文中所设定的最大响应时间（Max Response Time）内发送成员关系报告报文；

(4) 如果在最大响应时间内收到了该组播组其它成员发送的成员关系报告报文，查询器就会继续维护该组播组的成员关系；否则，查询器将认为该网段内已无该组播组的成员，于是不再维护这个组播组的成员关系。

1.1.4 IGMPv3 的改进

IGMPv3 在兼容和继承 IGMPv1 和 IGMPv2 的基础上，进一步增强了主机的控制能力，并增强了查询和报告报文的功能。

1. 主机控制能力的增强

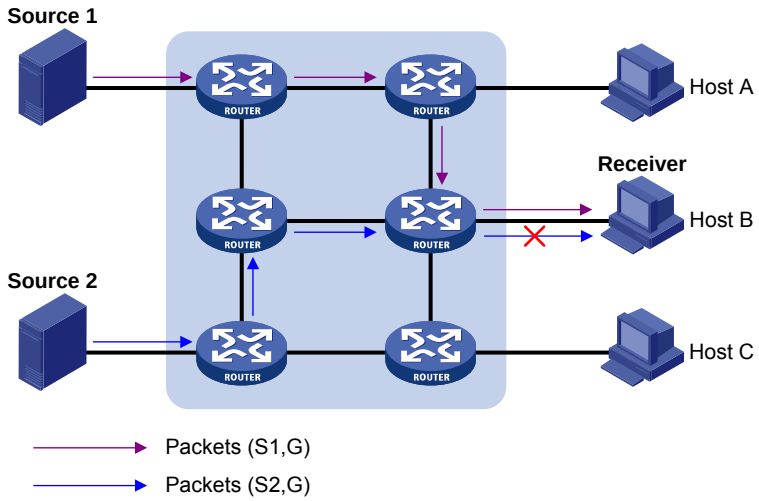
IGMPv3 增加了针对组播源的过滤模式（INCLUDE/EXCLUDE），使主机在加入某组播组 G 的同时，能够明确要求接收或拒绝来自某特定组播源 S 的组播信息。当主机加入组播组时：

- 若要求只接收来自指定组播源如 S1、S2、……的组播信息，则其报告报文中可以标记为 INCLUDE Sources（S1，S2，……）；

- 若拒绝接收来自指定组播源如 S1、S2、……的组播信息，则其报告报文中可以标记为 EXCLUDE Sources（S1，S2，……）。

如 图 1-2所示，网络中存在Source 1（S1）和Source 2（S2）两个组播源，均向组播组G发送组播报文。Host B仅对从Source 1 发往G的信息感兴趣，而对来自Source 2 的信息没有兴趣。

图1-2 指定源组的组播流路经



如果主机与路由器之间运行的是 IGMPv1 或 IGMPv2，Host B 加入组播组 G 时无法对组播源进行选择，因此无论 Host B 是否需要，来自 Source 1 和 Source 2 的组播信息都将传递给 Host B。

当主机与路由器之间运行了 IGMPv3 之后，Host B 就可以要求只接收来自 Source 1、发往 G 的组播信息（S1，G），或要求拒绝来自 Source 2、发往 G 的组播信息（S2，G），这样就只有来自 Source 1 的组播信息才能传递给 Host B 了。

2. 查询和报告报文功能的增强

(1) 携带源地址的查询报文

IGMPv3 不仅支持 IGMPv1 的普遍组查询和 IGMPv2 的特定组查询，而且还增加了对特定源组查询的支持：

- 普遍组查询报文中，既不携带组地址，也不携带源地址；
- 特定组查询报文中，携带组地址，但不携带源地址；
- 特定源组查询报文中，既携带组地址，还携带一个或多个源地址。

(2) 包含多组记录的报告报文

IGMPv3 报告报文的目的地址为 224.0.0.22，可以携带一个或多个组记录。在每个组记录中，包含有组播组地址和组播源地址列表。组记录可以分为多种类型，如下：

- IS_IN：表示组播组与组播源列表之间的过滤模式为 INCLUDE，即只接收从指定组播源列表发往该组播组的组播数据。
- IS_EX：表示组播组与组播源列表之间的过滤模式为 EXCLUDE，即只接收从指定组播源列表之外的组播源发往该组播组的组播数据。
- TO_IN：表示组播组与组播源列表之间的过滤模式由 EXCLUDE 转变为 INCLUDE。
- TO_EX：表示组播组与组播源列表之间的过滤模式由 INCLUDE 转变为 EXCLUDE。

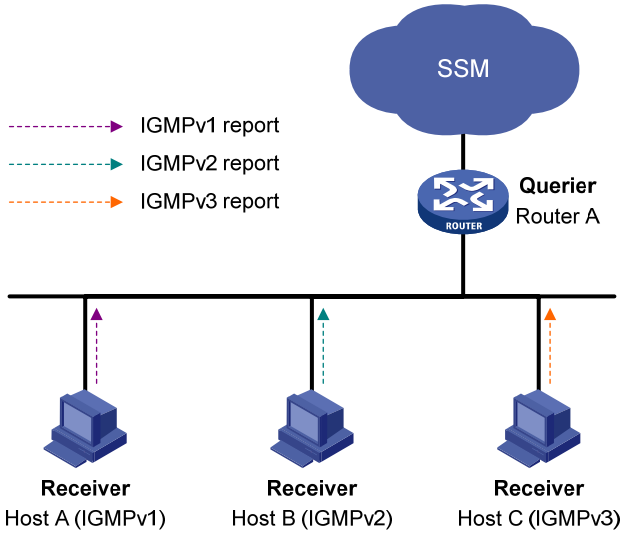
- ALLOW：表示在现有状态的基础上，还希望从某些组播源接收组播数据。如果当前的对应关系为 INCLUDE，则向现有组播源列表中添加这些组播源；如果当前的对应关系为 EXCLUDE，则从现有组播源列表中删除这些组播源。
- BLOCK：表示在现有状态的基础上，不再希望从某些组播源接收组播数据。如果当前的对应关系为 INCLUDE，则从现有组播源列表中删除这些组播源；如果当前的对应关系为 EXCLUDE，则向现有组播源列表中添加这些组播源。

1.1.5 IGMP SSM Mapping

IGMP SSM Mapping 通过在路由器上配置 SSM 静态映射规则，从而为运行 IGMPv1 或 IGMPv2 的接收者主机提供对 SSM 模型的支持。

SSM 模型要求在接收者主机所在的网段，路由器能够了解主机加入组播组时所指定的组播源。如果接收者主机上运行的是 IGMPv3，则可以在 IGMPv3 的报告报文中直接指定组播源的地址；如果某些接收者主机只能运行 IGMPv1 或 IGMPv2，则在 IGMPv1 或 IGMPv2 的报告报文中无法指定组播源的地址。这种情况下需要通过在路由器上配置 IGMP SSM Mapping 功能，将 IGMPv1 或 IGMPv2 报告报文中所包含的（*，G）信息映射为（G，INCLUDE，（S1，S2...））信息。

图1-3 IGMP SSM Mapping 组网图



在如 图 1-3所示的SSM网络中，Host A、Host B和Host C上分别运行IGMPv1、IGMPv2 和IGMPv3。在不允许将Host A和Host B升级为IGMPv3 的情况下，若要为Host A和Host B也提供SSM组播服务，则需在Router A上使能IGMP SSM Mapping并配置相应的映射规则。

配置完成后，当 Router A 收到来自主机的 IGMPv1 或 IGMPv2 报告报文时，首先检查该报文中所携带的组播组地址 G，然后根据检查结果的不同分别进行处理：

(1) 如果 G 不在 SSM 组地址范围内，则提供 ASM 组播服务。

(2) 如果 G 在 SSM 组地址范围内：

- 若 Router A 上没有 G 对应的 IGMP SSM Mapping 规则，则无法提供 SSM 组播服务，丢弃该报文；
- 若 Router A 上有 G 对应的 IGMP SSM Mapping 规则，则依据规则将报告报文中所包含的（*，G）信息映射为（G，INCLUDE，（S1，S2...））信息，可以提供 SSM 组播服务。

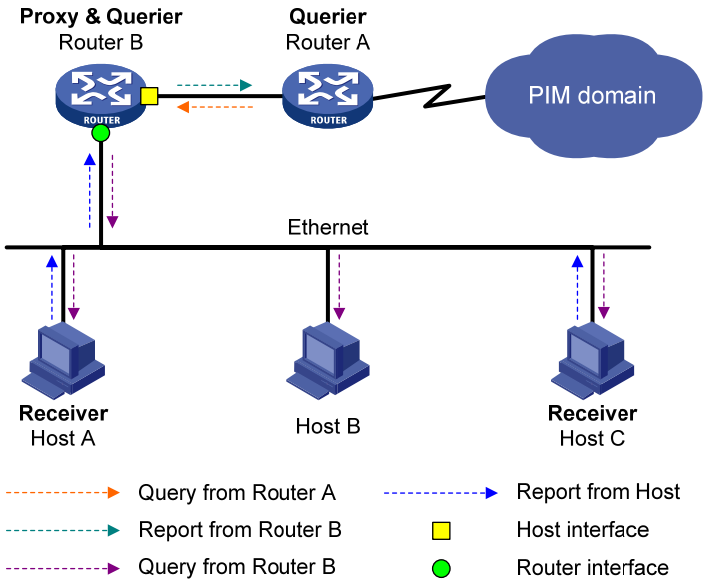
说明

- IGMP SSM Mapping 不对 IGMPv3 的报告报文进行处理。
- 有关 SSM 组地址范围的介绍，请参见“IP 组播分册”中的“PIM 配置”。

1.1.6 IGMP Proxying

在一些简单的树型网络拓扑中，边缘设备上并不需要运行复杂的组播路由协议（如 PIM），可以通过在这些设备上配置 IGMP Proxying（IGMP 代理）功能，使其代理下游主机来发送 IGMP 报文及维护组成员关系，并基于该关系进行组播转发。在上游设备看来，配置了 IGMP Proxying 功能的设备（称为 IGMP 代理设备）不再是一个 PIM 邻居，而只是一台主机。

图1-4 IGMP Proxying 组网图



如 图 1-4所示，IGMP Proxying中定义了以下两种接口类型：

- 上行接口：又称代理接口，指 IGMP 代理设备上运行 IGMP Proxying 功能的接口，即朝向组播分发树树根方向的接口。由于该接口执行 IGMP 协议的主机行为，因此也称为主机接口（Host Interface）。
- 下行接口：指 IGMP 代理设备上除上行接口外其它运行 IGMP 协议的接口，即背向组播分发树树根方向的接口。由于该接口执行 IGMP 协议的路由器行为，因此也称为路由器接口（Router Interface）。

IGMP 代理设备上维护着一个组成员关系数据库（Membership Database），将所有下行接口维护的组成员关系记录都存到这个数据库中。组成员关系记录的结构如下：（Multicast-address，Filter-mode，Source-list），每条记录都是各下行接口上具有相同组地址的成员关系记录的合集。

上行接口正是依据这个数据库来执行主机行为——当收到查询报文时根据当前数据库状态响应报告报文，或者当数据库变化时主动发送报告或离开报文；而下行接口则执行路由器行为——参与查询器的选举、发送查询报文并根据报告报文维护组成员关系等。

1.1.7 多实例的 IGMP

IGMP 依据接口来维护组成员关系，VPN 实例的 IGMP 根据接口所属的实例来处理协议报文的收发。当路由器收到 IGMP 报文时，需要区分该报文所属的实例，并在该实例范围内对其进行处理。当某 VPN 实例的 IGMP 需要和其它组播协议交互信息时，只会通知本 VPN 实例内的其它组播协议。

1.1.8 协议规范

与 IGMP 相关的协议规范有：

- RFC 1112：Host Extensions for IP Multicasting
- RFC 2236：Internet Group Management Protocol, Version 2
- RFC 3376：Internet Group Management Protocol, Version 3
- RFC 4605：Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")

1.2 IGMP 配置任务简介

表1-1 IGMP 配置任务简介

配置任务		说明	详细配置
配置 IGMP 基本功能	使能 IGMP	必选	1.3.2
	配置 IGMP 版本	可选	1.3.3
	配置静态加入	可选	1.3.4
	配置组播组过滤器	可选	1.3.5
	配置组播组数量限制	可选	1.3.6
调整 IGMP 性能	配置 IGMP 报文选项	可选	1.4.2
	配置 IGMP 查询和响应	可选	1.4.3
	配置 IGMP 快速离开	可选	1.4.4
配置 IGMP SSM Mapping	使能 IGMP SSM Mapping	可选	1.5.2
	配置 IGMP SSM Mapping 规则	可选	1.5.3
配置 IGMP Proxying	使能 IGMP Proxying	可选	1.6.2
	配置下行接口的组播转发能力	可选	1.6.3



说明

- IGMP 视图下的配置具有全局性，接口视图下的配置仅对所在接口有效。
- 若没有在接口视图下进行配置，则该接口继承 IGMP 视图下的全局配置；若两个视图都进行了配置，则优先采用该接口视图下所作的配置。

1.3 配置 IGMP 基本功能

1.3.1 配置准备

在配置 IGMP 基本功能之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 PIM-DM（或 PIM-SM）

在配置 IGMP 基本功能之前，需准备以下数据：

- IGMP 的版本
- 以静态方式加入的组播组和组播源的地址
- 组播组过滤的 ACL 规则
- 组播组数量限制值

1.3.2 使能 IGMP

在需要建立和维护组播组成员关系的接口上使能 IGMP。

1. 使能公网实例中的 IGMP

表1-2 使能公网实例中的 IGMP

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 IGMP	**igmp enable**	必选 缺省情况下，IGMP 处于关闭状态

2. 使能 VPN 实例中的 IGMP

表1-3 使能 VPN 实例中的 IGMP

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 IGMP	**igmp enable**	必选 缺省情况下，IGMP 处于关闭状态

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.3.3 配置 IGMP 版本

由于不同版本 IGMP 协议的报文结构与种类不同，因此需要为同一网段上的所有路由器配置相同版本的 IGMP，否则 IGMP 将不能正常运行。

1. 全局配置 IGMP 版本

表1-4 全局配置 IGMP 版本

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 IGMP 视图	**igmp** [**vpn-instance** *vpn-instance-name*]	-
配置 IGMP 的版本	**version** *version-number*	可选 缺省情况下，IGMP 的版本为 IGMPv2

2. 在接口上配置 IGMP 版本

表1-5 在接口上配置 IGMP 版本

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 IGMP 的版本	**igmp version** *version-number*	可选 缺省情况下，IGMP 的版本为 IGMPv2

1.3.4 配置静态加入

在配置了静态加入组播组或组播源组后，接口将作为该组播组的虚拟组成员存在，从而可以接收发往该组的组播数据，以测试组播数据的转发。

表1-6 配置静态加入

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置静态加入组播组或组播源组	**igmp static-group** *group-address* [**source** *source-address*]	必选 缺省情况下，接口没有以静态方式加入任何组播组或组播源组

说明

- 在运行 PIM-SM 的设备上配置静态加入时，如果待配接口上使能了 PIM-SM，则该接口必须为 PIM-SM 的 DR，否则该接口将不能加入组播组或组播源组；如果待配接口上未使能 PIM-SM 但使能了 IGMP，则该接口必须为 IGMP 查询器，否则该接口也不能加入组播组或组播源组。有关 PIM-SM 和 DR 的介绍，请参见“IP 组播分册”中的“PIM 配置”。
- 在配置了静态加入后，接口并不会对 IGMP 查询器发出的查询报文进行响应；在配置或取消静态加入时，接口也不会主动发送 IGMP 成员关系报告报文或 IGMP 离开组报文。也就是说，该接口并没有真正成为该组播组的成员。

1.3.5 配置组播组过滤器

如果不希望接口所在网段上的主机加入某些组播组，可以在该接口上设置一个 ACL 规则作为过滤器，对收到的报告报文进行过滤，以限制接口所服务的组播组范围。

表1-7 配置组播组过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置组播组过滤器	**igmp group-policy** *acl-number* [*version-number*]	必选 缺省情况下，没有配置组播组过滤器

1.3.6 配置组播组数量限制

通过配置允许接口加入的组播组最大数量，可以方便用户灵活控制接口所能加入的组播组个数。

表1-8 配置组播组数量限制

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置允许接口加入的组播组最大数量	**igmp group-limit** *limit*	必选 缺省情况下，允许接口加入的组播组最大数量为 1024 个

1.4 调整 IGMP 性能

说明

对于本节中的配置任务来说：

- 在 IGMP 视图下所作的配置对所有接口生效，在接口视图下所作的配置仅对当前接口生效；
- 如果在这两个视图下进行了相同功能或参数的配置，则不论配置先后，接口视图下的配置将被优先采用。

1.4.1 配置准备

在调整 IGMP 性能之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 IGMP 基本功能

在调整 IGMP 性能之前，需准备以下数据：

- IGMP 查询器的启动查询间隔
- IGMP 查询器的启动查询次数
- 发送 IGMP 普遍组查询报文的时间间隔
- IGMP 查询器的健壮系数
- IGMP 普遍组查询的最大响应时间
- 发送 IGMP 特定组查询报文的时间间隔
- IGMP 其它查询器的存在时间

1.4.2 配置 IGMP 报文选项

由于 IGMP 特定组查询报文和特定源组查询报文的存在，而组播组是千变万化的，设备不可能加入所有的组，因此 IGMP 需要借助 Router-Alert（路由器报警）选项来将送达本地、但没有加入的组播报文送往上层协议进行处理。有关 Router-Alert 选项的详细介绍，请参考 RFC 2113。

对于 IP 头中是否携带 Router-Alert 选项的 IGMP 报文，设备会做出不同的处理：

- 出于兼容性考虑，缺省情况下设备不对 Router-Alert 选项进行检查，即处理所有收到的 IGMP 报文。此时，IGMP 报文中无论是否携带有 Router-Alert 选项，设备都会将其送给上层协议进行处理。
- 为了提高设备性能、减少不必要的开支，同时出于协议安全性的考虑，可以配置设备丢弃未携带 Router-Alert 选项的 IGMP 报文，此时，当设备收到 IGMP 报文时，会检查该报文的 Router-Alert 选项，如果没有携带该选项，就丢弃该报文。

1. 全局配置 IGMP 报文选项

表1-9 全局配置 IGMP 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 IGMP 视图	**igmp** [**vpn-instance** *vpn-instance-name*]	-
配置丢弃未携带 Router-Alert 选项的 IGMP 报文	**require-router-alert**	可选 缺省情况下，设备不对 Router-Alert 选项进行检查
配置在发送的 IGMP 报文中携带 Router-Alert 选项	**send-router-alert**	可选 缺省情况下，在发送的 IGMP 报文中携带 Router-Alert 选项

2. 在接口上配置 IGMP 报文选项

表1-10 在接口上配置 IGMP 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置丢弃未携带 Router-Alert 选项的 IGMP 报文	**igmp require-router-alert**	可选 缺省情况下，设备不对 Router-Alert 选项进行检查
配置在发送的 IGMP 报文中携带 Router-Alert 选项	**igmp send-router-alert**	可选 缺省情况下，在发送的 IGMP 报文中携带 Router-Alert 选项

1.4.3 配置 IGMP 查询和响应

当 IGMP 查询器启动时，会以“IGMP 查询器启动查询间隔”为时间间隔发送“IGMP 查询器启动查询次数”次 IGMP 普遍组查询报文。

当 IGMP 查询器启动后，会周期性地发送 IGMP 普遍组查询报文，以判断网络上是否有组播组成员，发送间隔即为“发送 IGMP 普遍组查询报文的时间间隔”。可以根据网络的实际情况来修改周期性发送 IGMP 普遍组查询报文的时间间隔。

当 IGMP 查询器收到 IGMP 离开组报文后，也会发送“健壮系数”次 IGMP 特定组查询报文，发送间隔即为“发送 IGMP 特定组查询报文的时间间隔”。由此可见，IGMP 查询器的健壮系数越大，IGMP 查询器就越“健壮”，但是组播组超时所需的时间也就越长。

在收到 IGMP 查询报文（包括普遍组查询和特定组查询）后，主机会为其所加入的每个组播组都启动一个延迟定时器，其值在 0 到最大响应时间（该时间值从 IGMP 查询报文的最大响应时间字段获得）中随机选定，当定时器的值减为 0 时，主机就会向该定时器对应的组播组发送 IGMP 成员关系报告报文。

合理配置 IGMP 查询的最大响应时间，既可以使主机对 IGMP 查询报文做出快速响应，又可以减少由于定时器同时超时，造成大量主机同时发送报告报文而引起的网络拥塞：

- 对于 IGMP 普遍组查询报文来说，通过配置 IGMP 普遍组查询的最大响应时间来填充其最大响应时间字段；
- 对于 IGMP 特定组查询报文来说，所配置的发送 IGMP 特定组查询报文的时间间隔将被填充到其最大响应时间字段。也就是说，IGMP 特定组查询的最大响应时间从数值上与发送 IGMP 特定组查询报文的时间间隔相同。

当同一网段上有多台组播路由器时，由 IGMP 查询器负责发送 IGMP 查询报文。如果非查询器在“其它查询器存在时间”超时前没有收到来自查询器的 IGMP 查询报文，就会认为原有查询器失效，从而触发新的查询器选举过程；否则，非查询器将重置“其它查询器存在时间定时器”。

1. 全局配置 IGMP 查询和响应

表1-11 全局配置 IGMP 查询和响应

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 IGMP 视图	**igmp** [**vpn-instance** *vpn-instance-name*]	-
配置 IGMP 查询器的启动查询间隔	**startup-query-interval** *interval*	可选 关于IGMP查询器启动查询间隔的缺省取值请参见 表 1-12后的说明
配置 IGMP 查询器的启动查询次数	**startup-query-count** *value*	可选 关于IGMP查询器启动查询次数的缺省取值请参见 表 1-12后的说明
配置发送 IGMP 普遍组查询报文的时间间隔	**timer query** *interval*	可选 缺省情况下，发送 IGMP 普遍组查询报文的时间间隔为 60 秒
配置 IGMP 查询器的健壮系数	**robust-count** *robust-value*	可选 缺省情况下，IGMP 查询器的健壮系数为 2
配置 IGMP 普遍组查询的最大响应时间	**max-response-time** *interval*	可选 缺省情况下，IGMP 普遍组查询的最大响应时间为 10 秒
配置发送 IGMP 特定组查询报文的时间间隔	**last-member-query-interval** *interval*	可选 缺省情况下，发送 IGMP 特定组查询报文的时间间隔为 1 秒
配置 IGMP 其它查询器的存在时间	**timer other-querier-present** *interval*	可选 关于IGMP其它查询器存在时间的缺省取值请参见 表 1-12后的说明

2. 在接口上配置 IGMP 查询和响应

表1-12 在接口上配置 IGMP 查询和响应

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 IGMP 查询器的启动查询间隔	**igmp startup-query-interval** *interval*	可选 关于IGMP查询器启动查询间隔的缺省取值请参见 表 1-12后的说明
配置 IGMP 查询器的启动查询次数	**igmp startup-query-count** *value*	可选 关于IGMP查询器启动查询次数的缺省取值请参见 表 1-12后的说明
配置发送 IGMP 普遍组查询报文的时间间隔	**igmp timer query** *interval*	可选 缺省情况下，发送 IGMP 普遍组查询报文的时间间隔为 60 秒
配置 IGMP 查询器的健壮系数	**igmp robust-count** *robust-value*	可选 缺省情况下，IGMP 查询器的健壮系数为 2

操作	命令	说明
配置 IGMP 普遍组查询的最大响应时间	**igmp max-response-time** *interval*	可选 缺省情况下，IGMP 普遍组查询的最大响应时间为 10 秒
配置发送 IGMP 特定组查询报文的时间间隔	**igmp last-member-query-interval** *interval*	可选 缺省情况下，发送 IGMP 特定组查询报文的时间间隔为 1 秒
配置 IGMP 其它查询器的存在时间	**igmp timer other-querier-present** *interval*	可选 关于IGMP其它查询器存在时间的缺省取值请参见 表 1-12后的说明

说明

对于 IGMP 查询器的启动查询间隔和次数，以及 IGMP 其它查询器的存在时间这三个值，若进行了手工配置则取其各自的配置值，否则：

- 对于 IGMP 查询器的启动查询间隔，其取值为发送 IGMP 普遍组查询报文时间间隔的 1/4。缺省情况下，发送 IGMP 普遍组查询报文的时间间隔为 60 秒，则此时 IGMP 查询器的启动查询间隔 = 60 ÷ 4 = 15（秒）。
- 对于 IGMP 查询器的启动查询次数，其取值等于 IGMP 查询器的健壮系数。缺省情况下，IGMP 查询器的健壮系数为 2，则此时 IGMP 查询器的启动查询次数也为 2 次。
- 对于 IGMP 其它查询器的存在时间，其取值 = 发送 IGMP 普遍组查询报文的时间间隔 × IGMP 查询器的健壮系数 + IGMP 普遍组查询的最大响应时间 ÷ 2。缺省情况下，以上三个参数的取值分别为 60 秒、2 次和 10 秒，则此时 IGMP 其它查询器的存在时间 = 60 × 2 + 10 ÷ 2 = 125（秒）。

注意

- 应确保 IGMP 其它查询器的存在时间大于发送 IGMP 普遍组查询报文的时间间隔，否则有可能导致网络内的 IGMP 查询器反复变化。
- 应确保发送 IGMP 普遍组查询报文的时间间隔大于 IGMP 普遍组查询的最大响应时间，否则有可能造成对组播组成员的误删。
- 对 IGMP 普遍组查询的最大响应时间、发送 IGMP 特定组查询报文的时间间隔以及 IGMP 其它查询器的存在时间所做的配置，只有当设备运行在 IGMPv2 或 IGMPv3 时才有效。

1.4.4 配置 IGMP 快速离开

在某些应用（如 ADSL 拨号上网）中，IGMP 查询器的一个端口唯一对应着一台接收者主机，当主机在多个组播组间频繁切换（如进行电视选台）时，为了快速响应主机的离开组报文，可以在 IGMP 查询器上开启 IGMP 快速离开功能。

在开启了 IGMP 快速离开功能之后，当 IGMP 查询器收到来自主机的离开组报文时，不再发送 IGMP 特定组查询报文，而是直接向上游发送离开通告，这样一方面减小了响应延迟，另一方面也节省了网络带宽。

1. 全局配置 IGMP 快速离开

表1-13 全局配置 IGMP 快速离开

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 IGMP 视图	**igmp** [**vpn-instance** *vpn-instance-name*]	-
配置组播组成员快速离开	**fast-leave** [**group-policy** *acl-number*]	必选 缺省情况下，组播组成员快速离开功能是关闭的

2. 在接口上配置 IGMP 快速离开

表1-14 在接口上配置 IGMP 快速离开

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置组播组成员快速离开	**igmp fast-leave** [**group-policy** *acl-number*]	必选 缺省情况下，组播组成员快速离开功能是关闭的

注意

只有当设备运行在 IGMPv2 或 IGMPv3 时，配置 IGMP 快速离开功能才有效。

1.5 配置 IGMP SSM Mapping

在 SSM 网络中，由于各种可能的限制，某些接收者主机只能运行 IGMPv1 或 IGMPv2。为了向这些仅支持 IGMPv1 或 IGMPv2 的接收者主机提供 SSM 服务，可以在路由器上配置 IGMP SSM Mapping 功能。

1.5.1 配置准备

在配置 IGMP SSM Mapping 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 IGMP 基本功能

1.5.2 使能 IGMP SSM Mapping

表1-15 使能 IGMP SSM Mapping

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能 IGMP SSM Mapping 功能	**igmp ssm-mapping enable**	必选 缺省情况下，接口上的 IGMP SSM Mapping 功能处于关闭状态

说明

为保证本网段内运行任意版本 IGMP 的接收者主机都能得到 SSM 服务，建议在处于该网段的接口上运行 IGMPv3。

1.5.3 配置 IGMP SSM Mapping 规则

通过多次配置可以实现同一组播组到多个组播源的映射。

表1-16 配置 IGMP SSM Mapping 规则

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例的 IGMP 视图	**igmp** [**vpn-instance** *vpn-instance-name*]	-
配置 IGMP SSM Mapping 规则	**ssm-mapping** *group-address* { *mask* | *mask-length* } *source-address*	必选 缺省情况下，未配置 IGMP SSM Mapping 规则

注意

对于同时支持 IGMP Snooping 和 IGMP 的交换机来说，若 VLAN 接口上运行的是 IGMPv3，则在使用 **igmp-snooping host-join** 命令配置模拟主机加入时即便不指定组播源，模拟主机也仍将发送 IGMPv3 的报告报文，这种情况下不会依据 IGMP SSM Mapping 规则创建相应的组播组。有关 **igmp-snooping host-join** 命令的详细介绍，请参见“IP 组播分册”中的“IGMP Snooping 命令”。

1.6 配置 IGMP Proxying

1.6.1 配置准备

在配置 IGMP Proxying 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 使能 IP 组播路由

1.6.2 使能 IGMP Proxying

在设备朝向组播分发树树根方向的接口上使能了 IGMP 代理功能之后，该设备就成为了 IGMP 代理设备。

表1-17 使能 IGMP Proxying

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 IGMP 代理功能	**igmp proxying enable**	必选 缺省情况下，接口上的 IGMP 代理功能处于关闭状态

说明

- 一台设备上只允许有一个接口使能 IGMP 代理功能（多实例情况下，每个实例中只允许有一个接口使能 IGMP 代理功能）。
- 在进行IGMP的相关配置时，在已使能IGMP代理功能的接口上不允许再使能IGMP，且只有 **igmp require-router-alert**、**igmp send-router-alert** 和 **igmp version** 命令可在该接口上生效。
- 在已使能 IGMP 代理功能的设备上不允许再使能其它组播路由协议（如不能在接口上使能 PIM-DM 或 PIM-SM，但在 PIM 视图下配置的 **source-lifetime**、**source-policy** 和 **ssm-policy** 命令仍会生效），反之亦然；同时，由于 IGMPv1 的查询器要由 PIM 协议选举出的 DR 来充当，因此若 IGMP 代理设备的下行接口运行的是 IGMPv1，那么该接口将无法成为 DR，从而也就无法充当 IGMP 查询器。
- 若某 VLAN 内已使能了 IGMP Snooping，则该 VLAN 对应的 VLAN 接口上不能再使能 IGMP 代理功能，反之亦然。

1.6.3 配置下行接口的组播转发能力

通常，组播数据只能被查询器转发，非查询器不具备组播转发能力，这样可以避免组播数据被重复转发。对于 IGMP 代理设备也同样如此，只有当其下行接口是查询器时，组播数据才能通过该接口转发给下游主机。

但在某些情况下，IGMP 代理设备的下行接口不能在查询器竞选中获胜，此时应使能该下行接口在非查询器状态下的组播转发能力。

表1-18 配置下行接口的组播转发能力

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能下行接口在非查询器状态下的组播转发能力	**igmp proxying forwarding**	必选 缺省情况下，当 IGMP 代理设备的下行接口处于非查询器状态时，不转发组播数据

注意

在共享网段内存在多台 IGMP 代理设备的情况下，当其中一台 IGMP 代理设备的下行接口已竞选为查询器时，不应再在其它 IGMP 代理设备的下行接口上进行此配置，否则该网段内将收到多份重复的组播数据流。

1.7 IGMP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IGMP 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IGMP 的统计信息。

表1-19 IGMP 显示和维护

操作	命令
查看 IGMP 组播组的信息	**display igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **group** [*group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]
查看接口上 IGMP 配置和运行信息	**display igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **interface** [*interface-type interface-number*] [**verbose**]
查看 IGMP 代理组的信息	**display igmp** [**all-instance** | **vpn-instance** *vpn-instance-name*] **proxying group** [*group-address*] [**verbose**]
查看 IGMP 路由表的信息	**display igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **routing-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }] | **flags** { **act** | **suc** }] *
查看 IGMP SSM Mapping 规则	**display igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **ssm-mapping** *group-address*
查看依据 IGMP SSM Mapping 规则创建的组播组信息	**display igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **ssm-mapping group** [*group-address* | **interface** *interface-type interface-number*] [**verbose**]
清除 IGMP 组播组的信息	**reset igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **group** { **all** | **interface** *interface-type interface-number* { **all** | *group-address* [**mask** { *mask* | *mask-length* }] [*source-address* [**mask** { *mask* | *mask-length* }]] } }
清除依据 IGMP SSM Mapping 规则创建的组播组信息	**reset igmp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **ssm-mapping group** { **all** | **interface** *interface-type interface-number* { **all** | *group-address* [**mask** { *mask* | *mask-length* }] [*source-address* [**mask** { *mask* | *mask-length* }]] } }

说明

通过 **reset igmp group** 命令不能够清除以静态方式加入的 IGMP 组播组的信息。

注意

执行 **reset igmp group** 命令可能导致接收者中断组播信息的接收。

1.8 IGMP 典型配置举例

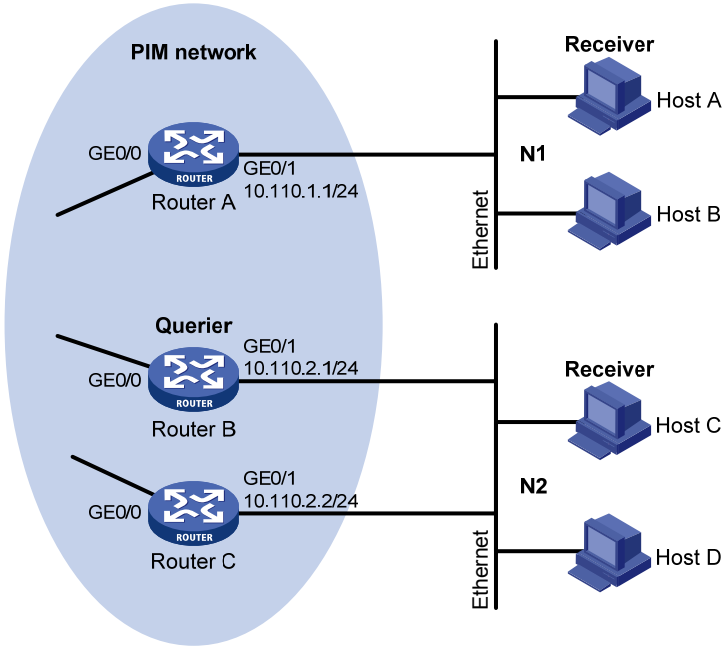
1.8.1 IGMP 基本功能配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者组成末梢网络 N1 和 N2，HostA 与 HostC 分别为 N1 和 N2 中的组播信息接收者。
- PIM 网络中的 Router A 连接 N1，Router B 与 Router C 共同连接 N2。
- Router A 通过 GigabitEthernet0/1 连接 N1，通过 GigabitEthernet0/0 连接 PIM 网络中的其它设备。
- Router B 与 Router C 分别通过各自的 GigabitEthernet0/1 连接 N2，并分别通过各自的 GigabitEthernet0/0 连接 PIM 网络中的其它设备。
- Router A 与 N1 之间运行 IGMPv2；Router B、Router C 与 N2 之间也分别运行 IGMPv2，且由于 Router B 的接口 IP 地址较小，因此在 N2 中通常由其来充当 IGMP 查询器。

2. 组网图

图1-5 IGMP 基本功能配置组网图



3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-5配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM 网络内的各路由器之间采用 OSPF 协议进行互连，确保 PIM 网络内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-DM 和 IGMP

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在主机侧接口 GigabitEthernet0/1 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] igmp enable
[RouterA-GigabitEthernet0/1] pim dm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/0
[RouterA-GigabitEthernet0/0] pim dm
[RouterA-GigabitEthernet0/0] quit
```

在 Router B 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在主机侧接口 GigabitEthernet0/1 上使能 IGMP。

```
<RouterB> system-view
[RouterB] multicast routing-enable
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] igmp enable
[RouterB-GigabitEthernet0/1] pim dm
[RouterB-GigabitEthernet0/1] quit
[RouterB] interface GigabitEthernet 0/0
[RouterB-GigabitEthernet0/0] pim dm
[RouterB-GigabitEthernet0/0] quit
```

在 Router C 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在主机侧接口 GigabitEthernet0/1 上使能 IGMP。

```
<RouterC> system-view
[RouterC] multicast routing-enable
[RouterC] interface GigabitEthernet 0/1
[RouterC-GigabitEthernet0/1] igmp enable
[RouterC-GigabitEthernet0/1] pim dm
[RouterC-GigabitEthernet0/1] quit
[RouterC] interface GigabitEthernet 0/0
[RouterC-GigabitEthernet0/0] pim dm
[RouterC-GigabitEthernet0/0] quit
```

(3) 检验配置效果

通过使用 **display igmp interface** 命令可以查看各路由器接口上 IGMP 的配置和运行情况。例如：

查看 Router B 在 GigabitEthernet0/1 上的 IGMP 信息。

```
[RouterB] display igmp interface GigabitEthernet 0/1
 GigabitEthernet0/1(10.110.2.1):
   IGMP is enabled
   Current IGMP version is 2
   Value of query interval for IGMP(in seconds): 60
   Value of other querier present interval for IGMP(in seconds): 125
   Value of maximum query response time for IGMP(in seconds): 10
   Querier for IGMP: 10.110.2.1 (this Router)
  Total 1 IGMP Group reported
```

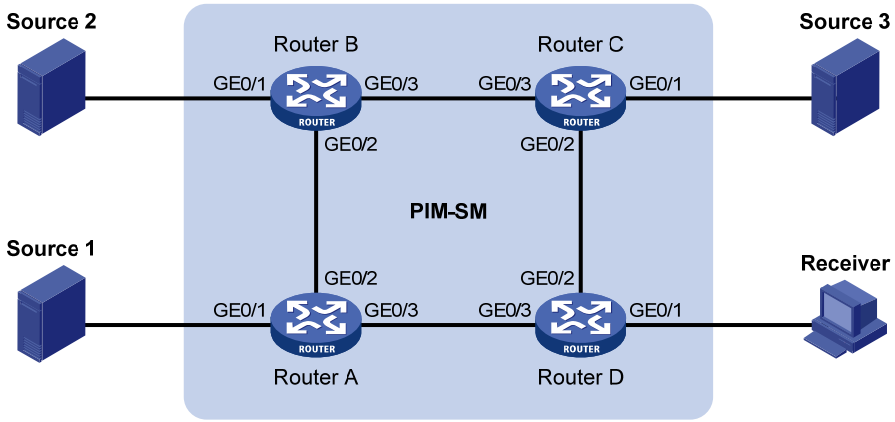
1.8.2 IGMP SSM Mapping 功能配置举例

1. 组网需求

- PIM-SM 网络中同时采用 ASM 和 SSM 方式提供组播服务，将 Router D 的接口 Ethernet1/3 配置为 C-BSR 和 C-RP，SSM 组播组的范围为 232.1.1.0/24。
- Router D 的接口 GigabitEthernet1/1 上运行 IGMPv3，接收者主机上运行 IGMPv2，且不能升级至 IGMPv3，因此该主机在加入组播组时无法指定组播源。
- Source 1、Source 2 和 Source 3 都向 SSM 组范围内的组播组发送组播数据，要求通过在 Router D 上配置 IGMP SSM Mapping 功能，使接收者主机只能接收来自 Source 1 和 Source 3 的组播数据。

2. 组网图

图1-6 IGMP SSM Mapping 功能配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source 1	-	133.133.1.1/24	Source 3	-	133.133.3.1/24
Source 2	-	133.133.2.1/24	Receiver	-	133.133.4.1/24
Router A	GE0/1	133.133.1.2/24	Router C	GE0/1	133.133.3.2/24
	GE0/2	192.168.1.1/24		GE0/2	192.168.3.1/24
	GE0/3	192.168.4.2/24		GE0/3	192.168.2.2/24
Router B	GE0/1	133.133.2.2/24	Router D	GE0/1	133.133.4.2/24
	GE0/2	192.168.1.2/24		GE0/2	192.168.3.2/24
	GE0/3	192.168.2.1/24		GE0/3	192.168.4.1/24

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-6配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-SM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-SM、IGMP 和 IGMP SSM Mapping 功能

在 Router D 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet0/1 上使能 IGMP 和 IGMP SSM Mapping 功能，配置 IGMP 版本为 3。

```
<RouterD> system-view
[RouterD] multicast routing-enable
[RouterD] interface GigabitEthernet 0/1
[RouterD-GigabitEthernet0/1] igmp enable
```

```
[RouterD-GigabitEthernet0/1] igmp version 3
[RouterD-GigabitEthernet0/1] igmp ssm-mapping enable
[RouterD-GigabitEthernet0/1] pim sm
[RouterD-GigabitEthernet0/1] quit
[RouterD] interface GigabitEthernet 0/2
[RouterD-GigabitEthernet0/2] pim sm
[RouterD-GigabitEthernet0/2] quit
[RouterD] interface GigabitEthernet 0/3
[RouterD-GigabitEthernet0/3] pim sm
[RouterD-GigabitEthernet0/3] quit
```

在 Router A 上使能 IP 组播路由，并在各接口上使能 PIM-SM。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] pim sm
[RouterA-GigabitEthernet0/2] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] pim sm
[RouterA-GigabitEthernet0/3] quit
```

Router B 和 Router C 的配置与 Router A 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP

在 Router D 上配置 C-BSR 和 C-RP 的位置。

```
[RouterD] pim
[RouterD-pim] c-bsr GigabitEthernet 0/3
[RouterD-pim] c-rp GigabitEthernet 0/3
[RouterD-pim] quit
```

(4) 配置 SSM 组播组的地址范围

在 Router D 上配置 SSM 组播组的地址范围为 232.1.1.0/24。

```
[RouterD] acl number 2000
[RouterD-acl-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[RouterD-acl-basic-2000] quit
[RouterD] pim
[RouterD-pim] ssm-policy 2000
[RouterD-pim] quit
```

Router A、Router B 和 Router C 的配置与 Router D 相似，配置过程略。

(5) 配置 IGMP SSM Mapping 规则

在 Router D 上配置 IGMP SSM Mapping 规则。

```
[RouterD] igmp
[RouterD-igmp] ssm-mapping 232.1.1.0 24 133.133.1.1
[RouterD-igmp] ssm-mapping 232.1.1.0 24 133.133.3.1
[RouterD-igmp] quit
```

(6) 检验配置效果

通过使用 **display igmp ssm-mapping** 命令可以查看路由器上的 IGMP SSM Mapping 规则。例如：

查看 Router D 上公网实例中组播组 232.1.1.1 的 IGMP SSM Mapping 规则。

```
[RouterD] display igmp ssm-mapping 232.1.1.1
Vpn-Instance: public net
Group: 232.1.1.1
Source list:
       133.133.1.1
       133.133.3.1
```

通过使用 **display igmp ssm-mapping group** 命令可以查看路由器上依据 IGMP SSM Mapping 规则创建的组播组信息。例如：

查看 Router D 上公网实例中依据 IGMP SSM Mapping 规则创建的组播组信息。

```
[RouterD] display igmp ssm-mapping group
Total 1 IGMP SSM-mapping Group(s).
Interface group report information of VPN-Instance: public net
 GigabitEthernet0/1(133.133.4.2):
  Total 1 IGMP SSM-mapping Group reported
   Group Address       Last Reporter   Uptime      Expires
   232.1.1.1           133.133.4.1     00:02:04    off
```

通过使用 **display pim routing-table** 命令可以查看路由器的 PIM 路由表信息。例如：

查看 Router D 上公网实例中的 PIM 路由表信息。

```
[RouterD] display pim routing-table
Vpn-instance: public net
Total 0 (*, G) entry; 2 (S, G) entry

 (133.133.1.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: GigabitEthernet0/3
         Upstream neighbor: 192.168.4.2
         RPF prime neighbor: 192.168.4.2
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: GigabitEthernet0/1
                 Protocol: igmp, UpTime: 00:13:25, Expires: -

 (133.133.3.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: GigabitEthernet0/2
         Upstream neighbor: 192.168.3.1
         RPF prime neighbor: 192.168.3.1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: GigabitEthernet0/1
```

```
Protocol: igmp, UpTime: 00:13:25, Expires: -
```

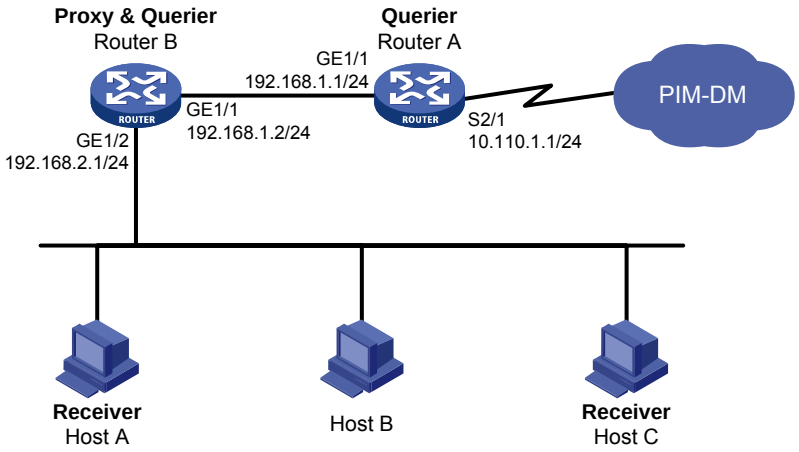
1.8.3 IGMP Proxying 功能配置举例

1. 组网需求

- 核心网络中运行 PIM-DM，末梢网络中的接收者 Host A 和 Host C 通过组播组 224.1.1.1 点播视频节目。
- 要求通过在 Router B 上配置 IGMP Proxying 功能，使其在不运行 PIM-DM 的情况下实现组成员关系的维护和组播数据的正常转发。

2. 组网图

图1-7 IGMP Proxying 功能配置组网图



3. 配置步骤

(1) 配置 IP 地址

请按照 图 1-7配置各接口的IP地址和子网掩码，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-DM、IGMP 和 IGMP Proxying

在 Router A 上使能 IP 组播路由，在接口 Serial2/1 上使能 PIM-DM，并在接口 GigabitEthernet1/1 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface serial 2/1
[RouterA-Serial2/1] pim dm
[RouterA-Serial2/1] quit
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] igmp enable
[RouterA-Ethernet1/1] pim dm
[RouterA-GigabitEthernet1/1] quit
```

在 Router B 上使能 IP 组播路由，在接口 GigabitEthernet1/1 上使能 IGMP Proxying，并在接口 GigabitEthernet1/2 上使能 IGMP。

```
<RouterB> system-view
[RouterB] multicast routing-enable
```

```
[RouterB] interface gigabitethernet 1/1
[RouterB-GigabitEthernet1/1] igmp proxying enable
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface gigabitethernet 1/2
[RouterB-GigabitEthernet1/2] igmp enable
[RouterB-GigabitEthernet1/2] quit
```

(3) 检验配置效果

通过使用 **display igmp interface** 命令可以查看各路由器接口上 IGMP 的配置和运行情况。例如：

查看 Router B 在 GigabitEthernet1/1 上 IGMP 配置和运行的详细信息。

```
[RouterB] display igmp interface gigabitethernet 1/1 verbose
 GigabitEthernet1/1(192.168.1.2):
   IGMP proxy is enabled
   Current IGMP version is 2
   Multicast routing on this interface: enabled
   Require-router-alert: disabled
   Version1-querier-present-timer-expiry: 00:00:20
```

通过使用 **display igmp group** 命令可以查看 IGMP 组播组的信息。例如：

查看 Router A 上的 IGMP 组播组信息。

```
[RouterA] display igmp group
Total 1 IGMP Group(s).
Interface group report information of VPN-Instance: public net
 GigabitEthernet1/1(192.168.1.1):
  Total 1 IGMP Groups reported
   Group Address   Last Reporter   Uptime      Expires
   224.1.1.1       192.168.1.2     00:02:04    00:01:15
```

由此可见，主机的 IGMP 报告通过 Router B 的代理接口 GigabitEthernet1/1 发给 Router A。

1.9 常见配置错误举例

1.9.1 接收者侧路由器上无组成员信息

1. 故障现象

当某主机发送了加入组播组 G 的报文后，离该主机最近的路由器上却没有组播组 G 的组成员信息。

2. 故障分析

- 组网、接口连线的正确与否以及接口的协议层是否 up 将直接影响组播组成员信息的生成；
- 在路由器上必须使能组播路由，在连接主机的接口上必须使能 IGMP；
- 如果路由器接口上运行的 IGMP 版本比主机的低，那么路由器将无法识别主机发来的较高版本的 IGMP 报告报文；
- 如果在接口上使用命令 **igmp group-policy** 对加入组播组 G 进行了限制后，该接口将不再接收未通过过滤的那些要求加入组播组 G 的报文。

3. 处理过程

(1) 检查组网是否正确，接口间的连线是否正确，以及接口状态是否正常，是否配置了正确的 IP 地址。通过命令 **display igmp interface** 查看接口信息。若无接口信息输出，说明接口状态异常，原因通常是接口上配置了 **shutdown** 命令，或者接口连线不正确，或者接口上没有配置正确的 IP 地址。

(2) 检查是否使能了组播路由。通过命令 **display current-configuration** 查看是否配置了命令 **multicast routing-enable**。若缺少该配置，则需要在系统视图下执行命令 **multicast routing-enable** 使能 IP 组播路由，同时也需要在相应接口上使能 IGMP。

(3) 检查接口上运行的 IGMP 版本。通过命令 **display igmp interface** 来检查接口上运行的 IGMP 版本是否低于主机所使用的版本。

(4) 检查接口上是否配置了 ACL 规则来限制主机加入组播组 G。通过命令 **display current-configuration interface** 观察是否配置了 **igmp group-policy** 命令。如果配置的 ACL 规则对加入组播组 G 进行了限制，则需要修改该 ACL 规则，允许接受组播组 G 的报告报文。

1.9.2 同一网段各路由器上组成员关系不一致

1. 故障现象

在同一网段的不同 IGMP 路由器上，各自维护的组成员关系不一致。

2. 故障分析

- 运行 IGMP 的路由器为每个接口维护多个参数，各参数之间相互影响，非常复杂。如果同一网段路由器的 IGMP 接口参数配置不一致，必然导致组成员关系的混乱；
- 另外，IGMP 目前有 3 个版本，版本不同的 IGMP 路由器与主机之间虽然可以兼容，但是连接在同一网段的所有路由器必须运行相同版本的 IGMP。如果同一网段路由器的 IGMP 版本不一致，也将导致 IGMP 组成员关系的混乱。

3. 处理过程

(1) 检查 IGMP 配置。通过命令 **display current-configuration** 观察接口上 IGMP 的配置信息。

(2) 在同一网段的所有路由器上执行命令 **display igmp interface** 来检查 IGMP 相关定时器的参数，确保配置一致。

(3) 通过命令 **display igmp interface** 来检查各路由器上运行的 IGMP 版本是否一致。

目 录

1 MSDP配置 1-1
1.1 MSDP简介 1-1
1.1.1 MSDP概述 1-1
1.1.2 MSDP原理 1-1
1.1.3 多实例的MSDP 1-6
1.1.4 协议规范 1-6
1.2 MSDP配置任务简介 1-7
1.3 配置MSDP基本功能 1-7
1.3.1 配置准备 1-7
1.3.2 使能MSDP 1-7
1.3.3 创建MSDP对等体连接 1-8
1.3.4 配置静态RPF对等体 1-9
1.4 配置MSDP对等体连接 1-9
1.4.1 配置准备 1-9
1.4.2 配置MSDP对等体描述信息 1-9
1.4.3 配置MSDP全连接组 1-9
1.4.4 配置MSDP对等体连接控制 1-10
1.5 配置SA消息 1-11
1.5.1 配置准备 1-11
1.5.2 配置SA消息内容 1-11
1.5.3 配置SA请求消息 1-11
1.5.4 配置SA消息过滤规则 1-12
1.5.5 配置SA消息缓存 1-13
1.6 MSDP显示和维护 1-13
1.7 MSDP典型配置举例 1-14
1.7.1 借助BGP路由的AS间组播配置举例 1-14
1.7.2 借助静态RPF对等体的AS间组播配置举例 1-19
1.7.3 Anycast RP应用配置举例 1-22
1.7.4 SA消息过滤机制配置举例 1-26
1.8 常见配置错误举例 1-30
1.8.1 MSDP对等体一直处于down状态 1-30
1.8.2 路由器SA缓存中没有SA表项 1-30
1.8.3 Anycast RP应用中RP间互通信息异常 1-31

1 MSDP 配置

说明

- 本文所涉及的路由器和路由器图标，代表了一般意义下的路由器或运行了 MSDP 协议的三层交换机。
- 本文中所提到的 DR（Designated Router，指定路由器）、BSR（BootStrap Router，自举路由器）、C-BSR（Candidate-BSR，候选 BSR）、RP（Rendezvous Point，汇集点）、C-RP（Candidate RP，候选 RP）、SPT（Shortest Path Tree，最短路径树）和 RPT（Rendezvous Point Tree，共享树或汇集树）等概念的详细介绍，请参见“IP 组播分册”中的“PIM 配置”。

1.1 MSDP 简介

1.1.1 MSDP 概述

MSDP 是 Multicast Source Discovery Protocol（组播源发现协议）的简称，是为了解决多个 PIM-SM（Protocol Independent Multicast Sparse Mode，协议无关组播—稀疏模式）域之间的互连而开发的一种域间组播解决方案，用来发现其它 PIM-SM 域内的组播源信息。

在基本的 PIM-SM 模式下，组播源只向本 PIM-SM 域内的 RP 注册，且各域的组播源信息是相互隔离的，因此 RP 仅知道本域内的组播源信息，只能在本域内建立组播分发树，将本域内组播源发出的组播数据分发给本地用户。如果能够有一种机制，将其它域内的组播源信息传递给本域内的 RP，则本域内的 RP 就可以向其它域内的组播源发起加入过程并建立组播分发树，从而实现组播数据的跨域传输。

基于这一设想，MSDP 通过在网络中选取适当的路由器建立 MSDP 对等体关系，以连通各 PIM-SM 域的 RP。通过在各 MSDP 对等体之间交互 SA（Source Active，信源有效）消息来共享组播源信息。

注意

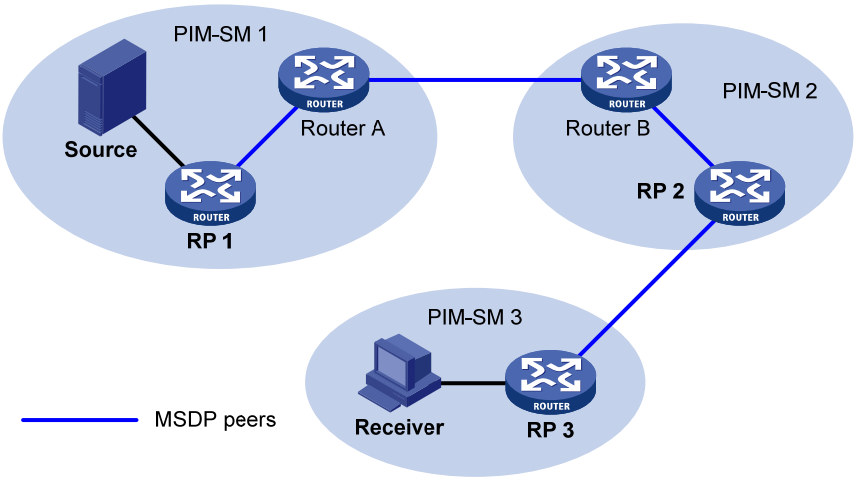
- MSDP 的适用前提：域内组播路由协议必须是 PIM-SM。
- MSDP 仅对 ASM（Any-Source Multicast，任意信源组播）模型有意义。

1.1.2 MSDP 原理

1. MSDP 对等体

通过在网络中配置一对或多对 MSDP 对等体，形成彼此相连的一张“MSDP 连通图”，以连通各个 PIM-SM 域的 RP。通过这些 MSDP 对等体之间的接力，可以把某 RP 发出的 SA 消息传递给其它所有的 RP。

图1-1 MSDP 对等体的位置



如 图 1-1所示，MSDP对等体可以创建在任意的PIM-SM路由器上，在不同角色的PIM-SM路由器上所创建的MSDP对等体的功能有所不同：

(1) 在 RP 上创建的 MSDP 对等体

- 源端 MSDP 对等体：即离组播源（Source）最近的 MSDP 对等体（通常也就是源端 RP，如 RP 1）。源端 RP 创建 SA 消息并发送给远端 MSDP 对等体，通告在本 RP 上注册的组播源信息。源端 MSDP 对等体必须配置在 RP 上，否则将无法向外发布组播源信息。
- 接收者端 MSDP 对等体：即离接收者（Receiver）最近的 MSDP 对等体（如 RP 3）。接收者端 MSDP 对等体在收到 SA 消息后，根据该消息中所包含的组播源信息，跨域加入以该组播源为根的 SPT；当来自该组播源的组播数据到达后，再沿 RPT 向本地接收者转发。
- 中间 MSDP 对等体：即拥有多个远端 MSDP 对等体的 MSDP 对等体（如 RP 2）。中间 MSDP 对等体把从一个远端 MSDP 对等体收到的 SA 消息转发给其它远端 MSDP 对等体，其作用相当于传输组播源信息的中转站。

(2) 在普通的 PIM-SM 路由器（非 RP）上创建的 MSDP 对等体

如 Router A 和 Router B，其作用仅限于将收到的 SA 消息转发出去。

说明

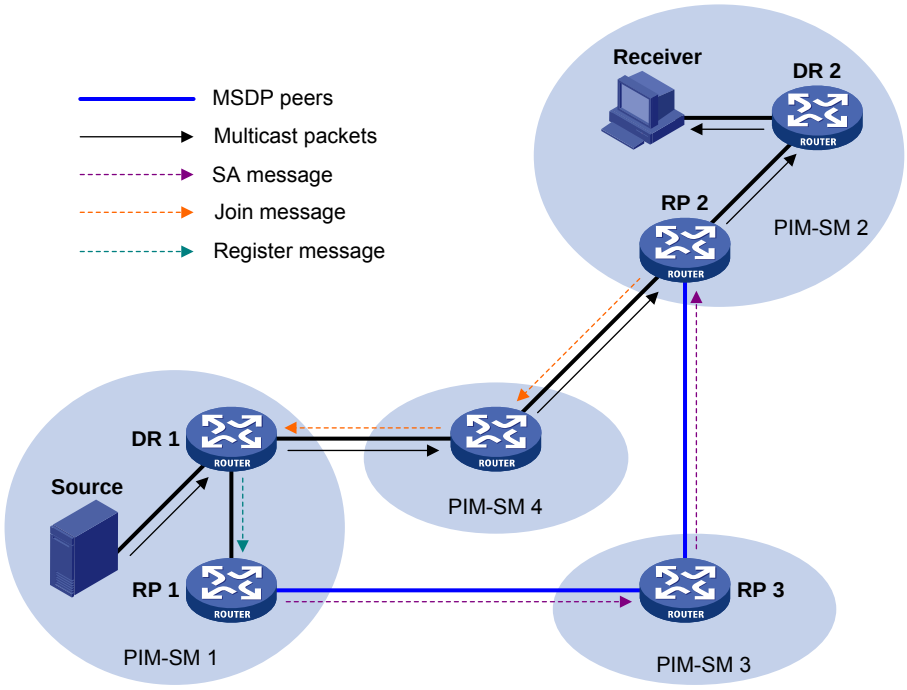
对于通过 BSR 机制动态选举 RP 的 PIM-SM 网络来说，RP 是由 C-RP 选举产生的。为了增强其网络的健壮性，一个 PIM-SM 域内往往存在不止一个 C-RP。由于无法预计 RP 选举的结果，为了保证选举获胜的 C-RP 能始终位于“MSDP 连通图”上，需要在所有的 C-RP 之间建立 MSDP 对等体关系。而选举落败的 C-RP 在“MSDP 连通图”上所担当的角色相当于普通的 PIM-SM 路由器。

---

2. 借助 MSDP 对等体实现域间组播

如 图 1-2所示，PIM-SM 1 域内存在激活的组播源（Source），RP 1 通过组播源注册过程了解到了该组播源的存在。如果PIM-SM 2 和PIM-SM 3 域也希望知道该组播源的具体位置，进而能够从该组播源获取组播数据，则需要在RP 1 与RP 3、RP 2 与RP 3 之间分别建立MSDP对等体关系。

图1-2 MSDP 对等体示意图



借助 MSDP 对等体进行域间组播的工作过程如下：

(1) 当 PIM-SM 1 域内的组播源向组播组 G 发送第一个组播数据包时，DR 1 将该组播数据封装在注册消息（Register Message）中，并发给 RP 1。RP 1 因此获知了该组播源的相关信息。

(2) RP 1 作为源端 RP，创建 SA 消息，并周期性地向其它 MSDP 对等体发送。SA 消息中包含组播源的地址 S、组播组的地址 G 以及创建该 SA 消息的源端 RP（即 RP 1）的地址。

(3) MSDP 对等体对收到的 SA 消息进行 RPF（Reverse Path Forwarding，逆向路径转发）检查，以及各种转发策略的过滤，从而只接受和转发来自正确路径并通过过滤的 SA 消息，以避免 SA 消息传递环路；另外，可以在 MSDP 对等体之间配置 MSDP 全连接组（Mesh Group），以避免 SA 消息在 MSDP 对等体之间的泛滥。

(4) SA 消息在 MSDP 对等体之间转发，最终该组播源的相关信息将传遍所有建立了 MSDP 对等体关系的 PIM-SM 域（即 PIM-SM 2 和 PIM-SM 3）。

(5) PIM-SM 2 中的 RP 2 在收到该 SA 消息后，检查本域内是否有组播组 G 的接收者（Receiver）存在：

- 如果有接收者，RP 2 与接收者之间维护组播组 G 的 RPT。RP 2 创建（S，G）表项，向源端的 DR 1 逐跳发送（S，G）加入消息（Join Message），从而跨越各 PIM-SM 域直接加入以该组播源为根的 SPT。组播数据沿 SPT 到达 RP 2 后，再沿 RPT 向接收者转发。当接收者端的 DR 2 收到来自组播源的组播数据后，可以自行决定是否发起从 RPT 向 SPT 的切换；
- 如果没有接收者，RP 2 不会创建（S，G）表项，也不加入以该组播源为根的 SPT。

说明

- MSDP 全连接组：要求所有组成员之间两两建立 MSDP 对等体关系，且所有组成员均使用相同的组名称。
- 在使用 MSDP 进行域间组播时，RP 在收到组播源的信息后就不再需要依赖其它 PIM-SM 域内的 RP，此时接收者可以跨越各 PIM-SM 域内的 RP，而直接加入基于组播源的 SPT。

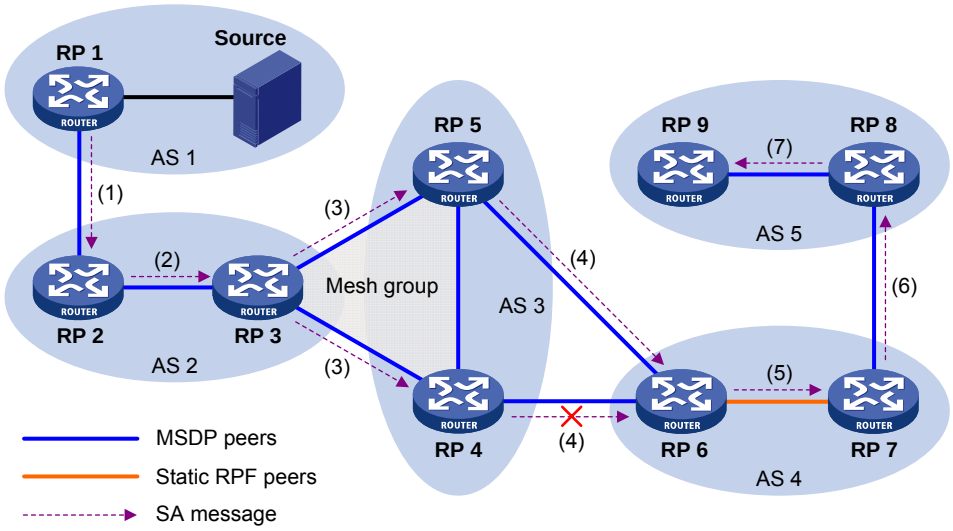
3. SA 消息的 RPF 检查规则

如图 1-3所示，网络中有五个自治系统AS 1～AS 5，AS内部使用IGP互联，AS之间使用BGP或MBGP互联。每个AS中包含至少一个PIM-SM域，且每个PIM-SM域中包含至少一个RP。各RP之间建立起MSDP对等体关系，其中RP 3、RP 4 和RP 5 之间建立MSDP全连接组，并在RP 7 上将RP 6 配置为其静态RPF对等体。

说明

当PIM-SM域内只存在一个MSDP对等体时，该域又称为STUB域（如 图 1-3中的AS 4）。STUB域内的MSDP对等体可以同时拥有多个远端MSDP对等体，用户可以从中选取其中一个或多个配置为静态RPF对等体。对于来自静态RPF对等体的SA消息不进行RPF检查，直接接受并向其它对等体转发。

图1-3 SA 消息的 RPF 检查规则



对照 图 1-3，这些MSDP对等体将按照如下RPF检查规则处理收到的SA消息：

(1) 当 RP 2 收到 RP 1 发来的 SA 消息时

由于 SA 消息中所携带的源端 RP 的地址与 MSDP 对等体的地址相同，说明发出 SA 消息的 MSDP 对等体就是创建该 SA 消息的 RP，于是 RP 2 接受该 SA 消息并向其它对等体（RP 3）转发。

(2) 当 RP 3 收到 RP 2 发来的 SA 消息时

由于 SA 消息来自同一个 AS 的 MSDP 对等体（RP 2），且该对等体是到源端 RP 最佳路径上的下一跳，于是 RP 3 接受该 SA 消息并向其它对等体（RP 4 和 RP 5）转发。

(3) 当 RP 4 和 RP 5 分别收到 RP 3 发来的 SA 消息时

由于 SA 消息来自同一个全连接组的 MSDP 对等体（RP 3），于是 RP 4 和 RP 5 均接受该 SA 消息并不再向本组其它成员转发，而只向本组之外的其它 MSDP 对等体（RP 6）转发。

(4) 当 RP 6 收到 RP 4 和 RP 5（假设 RP 5 的 IP 地址较大）发来的 SA 消息时

尽管同处 AS 3 的 RP 4 和 RP 5 都与 RP 6 建立了 MSDP 对等体关系，但由于 RP 5 的 IP 地址较大，于是 RP 6 只接受 IP 地址较高的 MSDP 对等体（RP 5）发来的 SA 消息。

(5) 当 RP 7 收到 RP 6 发来的 SA 消息时

由于 SA 消息来自其静态 RPF 对等体（RP 6），于是 RP 7 接受该 SA 消息并向其它对等体（RP 8）转发。

(6) 当 RP 8 收到 RP 7 发来的 SA 消息时

属于不同 AS 的 MSDP 对等体之间存在 BGP 或 MBGP 路由。由于 SA 消息来自不同 AS 的 MSDP 对等体（RP 7），且该对等体是到源端 RP 的 BGP 或 MBGP 路由的下一跳，于是 RP 8 接受该 SA 消息并向其它对等体（RP 9）转发。

(7) 当 RP 9 收到 RP 8 发来的 SA 消息时

由于只有一个 MSDP 对等体（RP 8），于是 RP 9 接受该 SA 消息。

对于由其它路径到来的 SA 消息，MSDP 对等体将不接受也不转发。

4. 借助 MSDP 对等体实现域内 Anycast RP

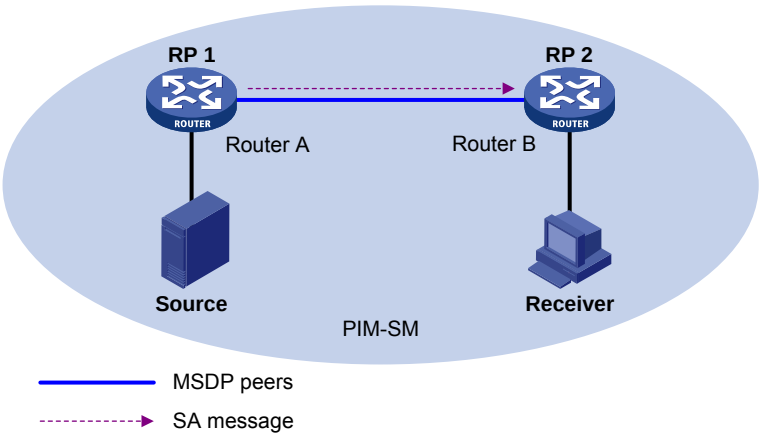
Anycast RP（任播 RP）是指在同一个 PIM-SM 域内设置两个或多个具有相同地址的 RP，并在这些 RP 之间建立 MSDP 对等体关系，以实现域内各 RP 之间的负载分担和冗余备份。

如 图 1-4所示，在一个PIM-SM域内，组播源（Source）向组播组G发送组播数据，接收者（Receiver）是组播组G的成员。分别在Router A和Router B上配置相同的IP地址（称为Anycast RP地址，通常使用私有地址），同时将这些接口配置为C-RP，并在Router A和Router B之间建立MSDP对等体关系。

说明

通常在设备的逻辑接口（如 Loopback 接口）上配置 Anycast RP 地址。

图1-4 Anycast RP 典型组网图



Anycast RP 的工作过程如下：

(1) 组播源选择距离最近的 RP 进行注册。如：Source 向 RP 1 注册，注册消息中封装有 Source 发出的组播数据。当该注册消息到达 RP 1 后，进行解封装。

(2) 接收者向距离最近的 RP 发送加入消息，加入以该 RP 为根的 RPT。如：Receiver 加入以 RP 2 为根的 RPT。

(3) RP 之间通过发送 SA 消息，共享注册的组播源信息。如：RP 1 创建一个 SA 消息，发送给 RP 2，该 SA 消息中封装有 Source 发出的组播数据。当该 SA 消息到达 RP 2 后，进行解封装。

(4) 接收者沿 RPT 收到组播数据后，直接加入以该组播源为根的 SPT。如：RP 2 沿 RPT 将组播数据向下转发。当 Receiver 收到来自 Source 的组播数据后，直接加入以 Source 为根的 SPT。

Anycast RP 的意义如下：

- RP 路径最优：组播源向距离最近的 RP 进行注册，建立路径最优的 SPT；接收者向距离最近的 RP 发起加入，建立路径最优的 RPT。
- RP 间的负载分担：每个 RP 上只需维护 PIM-SM 域内的部分源/组信息、转发部分的组播数据，从而实现了 RP 间的负载分担。
- RP 间的冗余备份：当某 RP 失效后，原先在该 RP 上注册或加入的组播源或接收者会自动选择就近的 RP 进行注册或加入操作，从而实现了 RP 间的冗余备份。

注意

- 必须为 Anycast RP 地址配置 32 位的子网掩码（即 255.255.255.255），也即将其配置为一个主机地址。
- MSDP 对等体的地址不能与 Anycast RP 地址相同。

1.1.3 多实例的 MSDP

属于同一实例的组播路由器各接口之间可以建立 MSDP 对等体。通过在 MSDP 对等体之间交互 SA 消息，可以实现跨域的 VPN 组播。

应用多实例的组播路由器，为其所支持的每一个实例都独立维护了一套 MSDP 机制，包括：SA 缓存、对等体连接、定时器、发送缓存和 PIM 交互的缓冲区。同时，保证不同实例之间信息隔离。所以，只有属于同一实例的 MSDP 和 PIM-SM 信息才可以交互。

1.1.4 协议规范

与 MSDP 相关的协议规范有：

- RFC 3618：Multicast Source Discovery Protocol (MSDP)
- RFC 3446：Anycast Rendevous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)

1.2 MSDP 配置任务简介

表1-1 MSDP 配置任务简介

配置任务		说明	详细配置
配置 MSDP 基本功能	使能 MSDP	必选	1.3.2
	创建 MSDP 对等体连接	必选	1.3.3
	配置静态 RPF 对等体	可选	1.3.4
配置 MSDP 对等体连接	配置 MSDP 对等体描述信息	可选	1.4.2
	配置 MSDP 全连接组	可选	1.4.3
	配置 MSDP 对等体连接控制	可选	1.4.4
配置 SA 消息	配置 SA 消息内容	可选	1.5.2
	配置 SA 请求消息	可选	1.5.3
	配置 SA 消息过滤规则	可选	1.5.4
	配置 SA 消息缓存	可选	1.5.5

1.3 配置 MSDP 基本功能

说明

本节的所有配置都是在 PIM-SM 域内的 RP 上进行的，这些 RP 将成为 MSDP 对等体的一端。

1.3.1 配置准备

在配置 MSDP 基本功能之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 PIM-SM，实现域内组播

在配置 MSDP 基本功能之前，需准备以下数据：

- MSDP 对等体的 IP 地址
- 表示 RP 地址过滤策略的地址前缀列表

1.3.2 使能 MSDP

在配置 MSDP 各功能之前，必须先使能 MSDP。

1. 使能公网实例中的 MSDP

表1-2 使能公网实例中的 MSDP

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态

操作	命令	说明
使能 MSDP，并进入公网实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，MSDP 处于关闭状态

2. 使能 VPN 实例中的 MSDP

表1-3 使能 VPN 实例中的 MSDP

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
使能 MSDP，并进入 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	必选 缺省情况下，MSDP 处于关闭状态

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.3.3 创建 MSDP 对等体连接

MSDP 对等体使用地址对来标识，即本端 MSDP 对等体地址和远端 MSDP 对等体地址。需要在互为对等体的两端都创建 MSDP 对等体连接。

表1-4 创建 MSDP 对等体连接

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
创建 MSDP 对等体连接	**peer** *peer-address* **connect-interface** *interface-type interface-number*	必选 缺省情况下，没有创建 MSDP 对等体连接

说明

如果某接口同时作为 MSDP 对等体和 BGP/MBGP 对等体中的一端，则建议为 MSDP 对等体配置与 BGP 或 MBGP 对等体相同的 IP 地址。

1.3.4 配置静态 RPF 对等体

通过配置静态 RPF 对等体可以免除对收到的 SA 消息进行 RPF 检查。

表1-5 配置静态 RPF 对等体

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
配置静态 RPF 对等体	**static-rpf-peer** *peer-address* [**rp-policy** *ip-prefix-name*]	必选 缺省情况下，没有配置静态 RPF 对等体

说明

如果在一台路由器上只配置了一个 MSDP 对等体，则该 MSDP 对等体将被当作静态 RPF 对等体。

1.4 配置 MSDP 对等体连接

1.4.1 配置准备

在配置 MSDP 对等体连接之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 MSDP 基本功能

在配置 MSDP 对等体连接之前，需准备以下数据：

- MSDP 对等体的描述信息
- MSDP 全连接组的名称
- 建立 MSDP 对等体连接的重试周期

1.4.2 配置 MSDP 对等体描述信息

管理员可以通过 MSDP 对等体的描述信息方便地区分不同的 MSDP 对等体，从而更好地管理 MSDP 对等体。

表1-6 配置 MSDP 对等体描述信息

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
配置 MSDP 对等体的描述信息	**peer** *peer-address* **description** *text*	必选 缺省情况下，MSDP 对等体没有描述信息

1.4.3 配置 MSDP 全连接组

一个自治系统内可能包含多个 MSDP 对等体，为了避免这些 MSDP 对等体之间泛滥 SA 消息，可

以使用 MSDP 全连接组来优化数据流量。

构成全连接组的 MSDP 对等体，一方面将来自组外并通过了 RPF 检查的 SA 消息转发给组内的其它成员；另一方面，对来自组内成员的 SA 消息不经 RPF 检查就接受，也不在组内进行重复转发。这种操作既避免了 SA 消息的泛滥，同时还由于不需要在 MSDP 对等体之间运行 BGP 或 MBGP，所以也就简化了对等体 RPF 检查机制。

通过为多个 MSDP 对等体配置相同的全连接组名称，可以建立 MSDP 全连接组。

表1-7 配置 MSDP 全连接组

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
把 MSDP 对等体加入全连接组	**peer** *peer-address* **mesh-group** *name*	必选 缺省情况下，MSDP 对等体不属于任何全连接组

说明

- 在配置 MSDP 全连接组之前，应使各路由器之间保持两两互连。
- 如果在同一 MSDP 对等体上多次配置加入全连接组，最后一个配置有效。

1.4.4 配置 MSDP 对等体连接控制

MSDP 对等体之间使用 TCP 进行连接（端口号为 639），用户可以手工关闭或重建 MSDP 对等体连接，灵活控制 MSDP 对等体之间的会话。当关闭了 MSDP 对等体连接后，MSDP 对等体之间不再传递 SA 消息，TCP 连接关闭，并不再重试建立连接，但配置信息会被保留。

当新创建了 MSDP 对等体、或重新启动了被关闭的 MSDP 对等体连接、或发生故障的 MSDP 对等体尝试恢复工作时，需要在 MSDP 对等体之间建立 TCP 连接。用户可以灵活地调整建立 MSDP 对等体连接的重试周期。

表1-8 配置 MSDP 对等体连接控制

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
手工关闭 MSDP 对等体连接	**shutdown** *peer-address*	可选 缺省情况下，MSDP 对等体处于连接状态
配置建立 MSDP 对等体连接的重试周期	**timer retry** *interval*	可选 缺省情况下，建立 MSDP 对等体连接的重试周期为 30 秒

1.5 配置 SA 消息

1.5.1 配置准备

在配置 SA 消息传递之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 MSDP 基本功能

在配置 SA 消息传递之前，需准备以下数据：

- 表示 SA 请求消息过滤规则的 ACL 规则
- 表示 SA 消息创建规则的 ACL 规则
- 表示接收或转发 SA 消息的过滤规则的 ACL 规则
- 封装在 SA 消息中组播数据报文的 TTL（Time to live，生存时间）阈值
- 可缓存从指定 MSDP 对等体学到的（S，G）表项的最大数量

1.5.2 配置 SA 消息内容

某些组播源发送组播数据的时间间隔较长，超出了（S，G）表项的超时时间。在这种情况下，源端 DR 只能将组播数据逐个封装在注册消息中，发送给源端 RP。源端 RP 使用 SA 消息将（S，G）信息传输给远端 RP。然后，远端 RP 向源端 DR 发起加入过程，并创建 SPT。由于（S，G）表项已超时，远端用户将永远无法收到该组播源发出的组播数据。

当在源端 RP 上使能了在 SA 消息中封装注册消息的功能后，源端 RP 会将注册消息封装在 SA 消息中发送出去。远端 RP 收到该 SA 消息后解封装，将注册消息中包含的组播数据沿 RPT 传输给本端用户。

MSDP 对等体之间传递 SA 消息，当路由器对收到的 SA 消息进行 RPF 检查时，如果发现对端 RP 的地址与本地 RP 的地址相同，就会丢弃该 SA 消息。但在 Anycast RP 应用中，要求在同一个 PIM-SM 域内的两台或多台路由器上配置 IP 地址相同的 RP，并在这些路由器之间建立 MSDP 对等体关系，因此必须为 SA 消息指定一个与实际 RP 的地址不同的逻辑 RP 地址（即逻辑接口上的 RP 地址），以通过 RPF 检查。

表1-9 配置 SA 消息内容

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
使能在 SA 消息中封装注册消息	**encap-data-enable**	可选 缺省情况下，在 SA 消息中只包含（S，G）表项，不封装注册消息
配置接口地址为 SA 消息的 RP 地址	**originating-rp** *interface-type interface-number*	可选 缺省情况下，SA 消息的 RP 地址为 PIM 的 RP 地址

1.5.3 配置 SA 请求消息

缺省情况下，当一个新接收者加入时，路由器不会主动向其 MSDP 对等体发送 SA 请求消息，而是等待其 MSDP 对等体在下一个周期发来的 SA 消息，这将延迟接收者获取组播源信息的时间。为了

尽快让新接收者了解到当前活跃的组播源信息，需要主动向 MSDP 对等体发送 SA 请求消息。

表1-10 配置 SA 请求消息

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
使能发送 SA 请求消息	**peer** *peer-address* **request-sa-enable**	可选 缺省情况下，路由器收到新的组加入消息时，不向其 MSDP 对等体发送 SA 请求消息，而是等待下一周期 SA 消息的到来
配置 SA 请求消息的过滤规则	**peer** *peer-address* **sa-request-policy** [**acl** *acl-number*]	可选 缺省情况下，不对 SA 请求消息进行过滤

注意

在使能发送 SA 请求消息功能之前，必须首先关闭 SA 消息缓存机制，否则设备不会向外发送 SA 请求消息。

1.5.4 配置 SA 消息过滤规则

通过配置 SA 消息的创建规则，路由器可以在创建 SA 消息时，对其通告的（S，G）表项进行过滤，从而实现在创建 SA 消息时对组播源消息传播的控制。

通过配置接收或转发 SA 消息的过滤规则，路由器可以在接收或转发 SA 消息时，对其通告的（S，G）转发项进行过滤，从而实现在接收和转发 SA 消息时，对组播源消息传播的控制。

通过配置封装在 SA 消息中组播数据报文的 TTL 阈值，可以对组播数据报文在 SA 消息中的封装以及传输范围进行限制：

- 路由器在创建封装有组播数据报文的 SA 消息之前，先检查该组播数据报文 IP 头的 TTL 值：如果小于阈值，则不创建该 SA 消息；如果大于或等于阈值，则将组播数据报文封装在 SA 消息中并转发出去。
- 路由器在收到封装有组播数据报文的 SA 消息之后，先将该组播数据报文 IP 头的 TTL 值减 1，再检查此时的 TTL 值：如果小于阈值，则不再向其指定的 MSDP 对等体转发；如果大于或等于阈值，则重新将组播数据报文封装在 SA 消息中并转发出去。

表1-11 配置 SA 消息过滤规则

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
配置 SA 消息的创建规则	**import-source** [**acl** *acl-number*]	必选 缺省情况下，在创建 SA 消息时，对其通告的（S，G）项不作限制

操作	命令	说明
配置接收或转发 SA 消息的过滤规则	**peer** *peer-address* **sa-policy** { **import** | **export** } [**acl** *acl-number*]	必选 缺省情况下，不对接收或转发的 SA 消息进行过滤
配置封装在 SA 消息中组播数据报文的 TTL 阈值	**peer** *peer-address* **minimum-ttl** *ttl-value*	可选 缺省情况下，封装在 SA 消息中组播数据报文的 TTL 阈值为 0

1.5.5 配置 SA 消息缓存

为了减少获取组播信息的延迟时间，可以在路由器上使能 SA 消息缓存机制，即在本地缓存 SA 消息中所包含的（S，G）表项。缓存的（S，G）表项越多，所占用的内存空间越大。

在使能了 SA 消息缓存机制后，当收到一个新的组加入消息（*，G）时，路由器首先查找 SA 缓存：

- 如果缓存中没有对应的（S，G），便等候其 MSDP 对等体在下一个周期发来的 SA 消息；
- 如果缓存中有对应的（S，G），则直接加入以 S 为根的 SPT。

为了有效防止路由器受到 DoS（Denial of Service，拒绝服务）攻击，可以在路由器上配置可缓存（S，G）表项的最大数量。

表1-12 配置 SA 消息缓存

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 MSDP 视图	**msdp** [**vpn-instance** *vpn-instance-name*]	-
使能 SA 消息缓存机制	**cache-sa-enable**	可选 在缺省情况下，SA 消息缓存机制处于使能状态，即在设备收到 SA 消息后缓存其中包含的（S，G）表项
配置可缓存从指定 MSDP 对等体学到的（S，G）表项的最大数量	**peer** *peer-address* **sa-cache-maximum** *sa-limit*	可选 缺省情况下，可缓存从任一 MSDP 对等体学到的（S，G）表项的最大数量为 8192

1.6 MSDP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MSDP 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 MSDP 的统计信息。

表1-13 MSDP 显示和维护

操作	命令
查看 MSDP 对等体的简要信息	**display msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **brief** [**state** { **connect** | **down** | **listen** | **shutdown** | **up** }]
查看 MSDP 对等体的详细状态信息	**display msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **peer-status** [*peer-address*]

操作	命令
查看 SA 缓存中的（S，G）表项信息	**display msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **sa-cache** [*group-address* | *source-address* | *as-number*] *
查看 SA 缓存中（S，G）表项的数量	**display msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **sa-count** [*as-number*]
重置与 MSDP 对等体的 TCP 连接	**reset msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **peer** [*peer-address*]
清除 SA 缓存中的（S，G）表项	**reset msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **sa-cache** [*group-address*]
清除 MSDP 对等体的统计信息	**reset msdp** [**vpn-instance** *vpn-instance-name* | **all-instance**] **statistics** [*peer-address*]

1.7 MSDP 典型配置举例

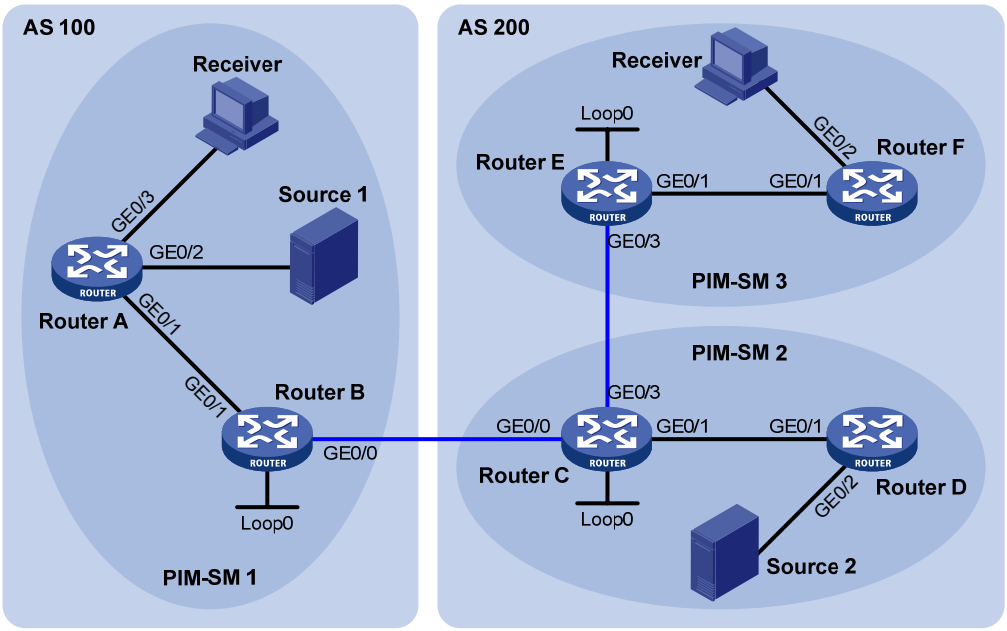
1.7.1 借助 BGP 路由的 AS 间组播配置举例

1. 组网需求

- 网络中存在两个自治系统：PIM-SM 1 属于 AS 100，PIM-SM 2 和 PIM-SM 3 属于 AS 200。各 AS 内部采用 OSPF 进行互联，AS 之间采用 BGP 交换路由信息；
- 每个 PIM-SM 域分别拥有 0 或 1 个组播源和接收者，域内运行 OSPF 协议以提供单播路由；
- 将 Router B、Router C 和 Router E 各自的 Loopback0 接口分别配置为各自 PIM-SM 域的 C-BSR 和 C-RP；
- 在 Router B 与 Router C 之间通过 EBGP 建立 MSDP 对等体关系，在 Router C 与 Router E 之间通过 IBGP 建立 MSDP 对等体关系。

2. 组网图

图1-5 借助 BGP 路由的 AS 间组播配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/1	10.110.1.2/24	Router D	GE0/1	10.110.4.2/24
	GE0/2	10.110.2.1/24		GE0/2	10.110.5.1/24
	GE0/3	10.110.3.1/24	Router E	GE0/1	10.110.6.1/24
Router B	GE0/1	10.110.1.1/24		GE0/3	192.168.3.2/24
	GE0/0	192.168.1.1/24		Loop0	3.3.3.3/32
	Loop0	1.1.1.1/32	Router F	GE0/1	10.110.6.2/24
Router C	GE0/1	10.110.4.1/24		GE0/2	10.110.7.1/24
	GE0/3	192.168.3.1/24	Source 1	-	10.110.2.100/24
	GE0/0	192.168.1.2/24	Source 2	-	10.110.5.100/24
	Loop0	2.2.2.2/32			

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-5配置各接口的IP地址和子网掩码，具体配置过程略。

配置 AS 内的各路由器之间采用 OSPF 协议进行互连，确保各 AS 内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，使能 PIM-SM 和 IGMP，并配置 BSR 的服务边界

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet 0/3 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] pim sm
[RouterA-GigabitEthernet0/2] quit
```

```
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] igmp enable
[RouterA-GigabitEthernet0/3] pim sm
[RouterA-GigabitEthernet0/3] quit
```

Router B、Router C、Router D、Router E 和 Router F 上的配置与 Router A 相似，配置过程略。

在 Router B 上配置 BSR 的服务边界。

```
[RouterB] interface GigabitEthernet 0/0
[RouterB-GigabitEthernet0/0] pim bsr-boundary
[RouterB-GigabitEthernet0/0] quit
```

Router C 和 Router E 上的配置与 Router B 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP 的位置

在 Router B 上将 Loopback0 接口配置为 C-BSR 和 C-RP。

```
[RouterB] pim
[RouterB-pim] c-bsr loopback 0
[RouterB-pim] c-rp loopback 0
[RouterB-pim] quit
```

Router C 和 Router E 上的配置与 Router B 相似，配置过程略。

(4) 配置 BGP 协议，将 BGP 与 OSPF 互相引入

在 Router B 上配置 EBGP，并引入 OSPF 路由。

```
[RouterB] bgp 100
[RouterB-bgp] Router-id 1.1.1.1
[RouterB-bgp] peer 192.168.1.2 as-number 200
[RouterB-bgp] import-route ospf 1
[RouterB-bgp] quit
```

在 Router C 上配置 IBGP 和 EBGP，并引入 OSPF 路由。

```
[RouterC] bgp 200
[RouterC-bgp] Router-id 2.2.2.2
[RouterC-bgp] peer 192.168.1.1 as-number 100
[RouterC-bgp] peer 192.168.3.2 as-number 200
[RouterC-bgp] import-route ospf 1
[RouterC-bgp] quit
```

在 Router E 上配置 IBGP，并引入 OSPF 路由。

```
[RouterE] bgp 200
[RouterE-bgp] Router-id 3.3.3.3
[RouterE-bgp] peer 192.168.3.1 as-number 200
[RouterE-bgp] import-route ospf 1
[RouterE-bgp] quit
```

在 Router B 的 OSPF 中引入 BGP。

```
[RouterB] ospf 1
[RouterB-ospf-1] import-route bgp
[RouterB-ospf-1] quit
```

Router C 和 Router E 上的配置与 Router B 相似，配置过程略。

(5) 配置 MSDP 对等体

在 Router B 上配置 MSDP 对等体。

```
[RouterB] msdp
[RouterB-msdp] peer 192.168.1.2 connect-interface GigabitEthernet 0/0
[RouterB-msdp] quit
```

在 Router C 上配置 MSDP 对等体。

```
[RouterC] msdp
[RouterC-msdp] peer 192.168.1.1 connect-interface GigabitEthernet 0/0
[RouterC-msdp] peer 192.168.3.2 connect-interface GigabitEthernet 0/3
[RouterC-msdp] quit
```

在 Router E 上配置 MSDP 对等体。

```
[RouterE] msdp
[RouterE-msdp] peer 192.168.3.1 connect-interface GigabitEthernet 0/3
[RouterE-msdp] quit
```

(6) 检验配置效果

通过使用 **display bgp peer** 命令可以查看路由器之间 BGP 对等体的关系。例如：

查看 Router B 上 BGP 对等体关系的信息。

```
[RouterB] display bgp peer

 BGP local Router ID : 1.1.1.1
 Local AS number : 100
 Total number of peers : 1                 Peers in established state : 1

  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.2  4 200       24       21     0       6 00:13:09 Established
```

查看 Router C 上 BGP 对等体关系的信息。

```
[RouterC] display bgp peer

 BGP local Router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 2                 Peers in established state : 2

  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.1  4 100       18       16     0       1 00:12:04 Established
  192.168.3.2  4 200       21       20     0       6 00:12:05 Established
```

查看 Router E 上 BGP 对等体关系的信息。

```
[RouterE] display bgp peer

BGP local Router ID : 3.3.3.3
 Local AS number : 200
 Total number of peers : 1                 Peers in established state : 1
```

```
  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.3.1  4 200       16       14     0       1 00:10:58 Established
```

通过使用 **display bgp routing-table** 命令可以查看路由器上的 BGP 路由表。例如：

查看 Router C 上 BGP 路由表的信息。

```
[RouterC] display bgp routing-table

 Total Number of Routes: 13

 BGP Local Router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   1.1.1.1/32        192.168.1.1   0                  0       100?
 *>i  2.2.2.2/32        192.168.3.2   0          100     0       ?
 *>   3.3.3.3/32        0.0.0.0       0                  0       ?
 *>   192.168.1.0       0.0.0.0       0                  0       ?
 *                      192.168.1.1   0                  0       100?
 *>   192.168.1.1/32    0.0.0.0       0                  0       ?
 *>   192.168.1.2/32    0.0.0.0       0                  0       ?
 *                      192.168.1.1   0                  0       100?
 *>   192.168.3.0       0.0.0.0       0                  0       ?
 * i                    192.168.3.2   0          100     0       ?
 *>   192.168.3.1/32    0.0.0.0       0                  0       ?
 *>   192.168.3.2/32    0.0.0.0       0                  0       ?
 * i                    192.168.3.2   0          100     0       ?
```

当 PIM-SM 1 和 PIM-SM 2 域内的组播源 Source 1 和 Source 2 发送组播信息时，PIM-SM 1 和 PIM-SM 3 域内的接收者（Receiver）能收到该组播信息。通过使用 **display msdp brief** 命令可以查看路由器之间 MSDP 对等体建立情况。例如：

查看 Router B 上 MSDP 对等体建立情况的简要信息。

```
[RouterB] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.1.2       Up       00:12:27        200    13         0
```

查看 Router C 上 MSDP 对等体建立情况的简要信息。

```
[RouterC] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
```

```
  192.168.3.2       Up       00:15:32        200    8          0
  192.168.1.1       UP       00:06:39        100    13         0
```

查看 Router E 上 MSDP 对等体建立情况的简要信息。

```
[RouterE] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.1       Up       01:07:08        200    8          0
```

查看 Router B 上 MSDP 对等体的详细信息。

```
[RouterB] display msdp peer-status
MSDP Peer Information of VPN-Instance: public net
  MSDP Peer 192.168.1.2, AS 200
  Description:
  Information about connection status:
    State: Up
    Up/down time: 00:15:47
    Resets: 0
    Connection interface: GE0/0 (192.168.1.1)
    Number of sent/received messages: 16/16
    Number of discarded output messages: 0
    Elapsed time since last connection or counters clear: 00:17:51
  Information about (Source, Group)-based SA filtering policy:
    Import policy: none
    Export policy: none
  Information about SA-Requests:
    Policy to accept SA-Request messages: none
    Sending SA-Requests status: disable
  Minimum TTL to forward SA with encapsulated data: 0
  SAs learned from this peer: 0, SA-cache maximum for the peer: none
  Input queue size: 0, Output queue size: 0
  Counters for MSDP message:
    Count of RPF check failure: 0
    Incoming/outgoing SA messages: 0/0
    Incoming/outgoing SA requests: 0/0
    Incoming/outgoing SA responses: 0/0
    Incoming/outgoing data packets: 0/0
```

1.7.2 借助静态 RPF 对等体的 AS 间组播配置举例

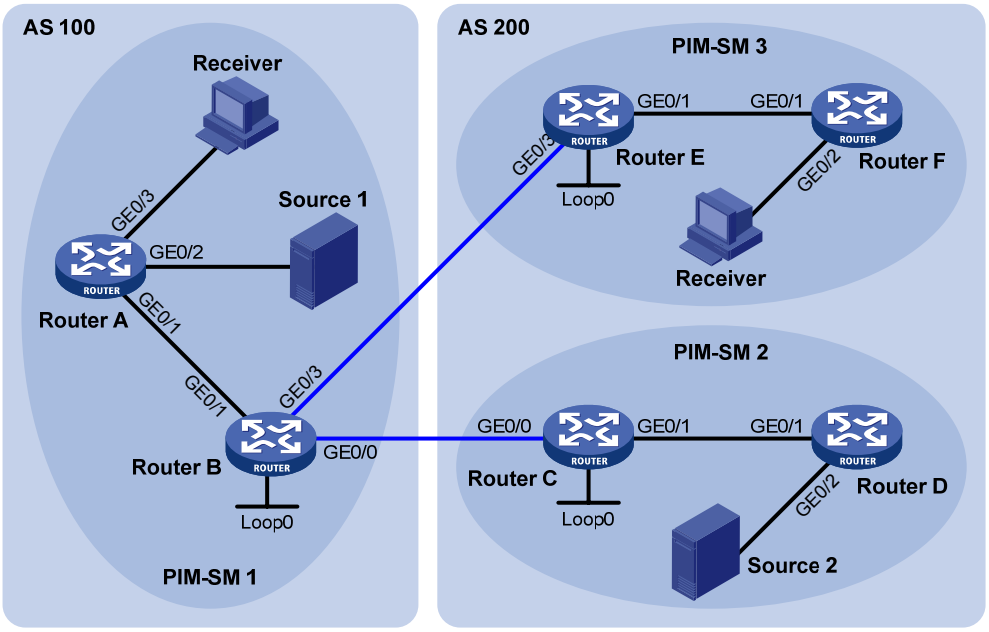
1. 组网需求

- 网络中存在两个自治系统：PIM-SM 1 属于 AS 100，PIM-SM 2 和 PIM-SM 3 属于 AS 200。各 AS 内部采用 OSPF 进行互联，AS 之间采用 BGP 交换路由信息；
- 每个 PIM-SM 域分别拥有 0 或 1 个组播源和接收者，域内运行 OSPF 以提供单播路由；

- PIM-SM 2 和 PIM-SM 3 都是 STUB 区域，这两个区域与 PIM-SM 1 之间无需运行 BGP 或 MBGP，而是与之建立静态 RPF 对等体，从而避免对 SA 消息进行 RPF 检查；
- 将 Router B、Router C 和 Router E 各自的 Loopback0 接口分别配置为各自 PIM-SM 域的 C-BSR 和 C-RP；
- Router B 的静态 RPF 对等体为 Router C 和 Router E，Router C 和 Router E 也分别只有 Router B 这一个静态 RPF 对等体，任何路由器都能够接收来自其静态 RPF 对等体的且被相应过滤策略允许的 SA 消息。

2. 组网图

图1-6 借助静态 RPF 对等体的 AS 间组播配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/1	10.110.1.2/24	Router D	GE0/1	10.110.4.2/24
	GE0/2	10.110.2.1/24		GE0/2	10.110.5.1/24
	GE0/3	10.110.3.1/24	Router E	GE0/1	10.110.6.1/24
Router B	GE0/1	10.110.1.1/24		GE0/3	192.168.3.2/24
	GE0/0	192.168.1.1/24		Loop0	3.3.3.3/32
	GE0/3	192.168.3.1/24	Router F	GE0/1	10.110.6.2/24
	Loop0	1.1.1.1/32		GE0/2	10.110.7.1/24
Router C	GE0/0	192.168.1.2/24	Source 1	-	10.110.2.100/24
	GE0/1	10.110.4.1/24	Source 2	-	10.110.5.100/24
	Loop0	2.2.2.2/32			

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-6配置各接口的IP地址和掩码，具体配置过程略。

配置 AS 内的各路由器之间采用 OSPF 协议进行互连，确保 AS 内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，使能 PIM-SM 和 IGMP，并配置 BSR 的服务边界

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet 0/3 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface GigabitEthernet 0/2
[RouterA-GigabitEthernet0/2] pim sm
[RouterA-GigabitEthernet0/2] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] igmp enable
[RouterA-GigabitEthernet0/3] pim sm
[RouterA-GigabitEthernet0/3] quit
```

Router B、Router C、Router D、Router E 和 Router F 上的配置与 Router A 相似，配置过程略。

在 Router B 上配置 BSR 的服务边界。

```
[RouterB] interface GigabitEthernet 0/3
[RouterB-GigabitEthernet0/3] pim bsr-boundary
[RouterB-GigabitEthernet0/3] quit
[RouterB] interface GigabitEthernet 0/0
[RouterB-GigabitEthernet0/0] pim bsr-boundary
[RouterB-GigabitEthernet0/0] quit
```

Router C 和 Router E 上的配置与 Router B 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP 的位置

在 Router B 上将 Loopback0 接口配置为 C-BSR 和 C-RP。

```
[RouterB] pim
[RouterB-pim] c-bsr loopback 0
[RouterB-pim] c-rp loopback 0
[RouterB-pim] quit
```

Router C 和 Router E 上的配置与 Router B 相似，配置过程略。

(4) 配置静态 RPF 对等体

配置 Router C 和 Router E 作为 Router B 的 MSDP 对等体及静态 RPF 对等体。

```
[RouterB] ip ip-prefix list-df permit 192.168.0.0 16 greater-equal 16 less-equal 32
[RouterB] msdp
[RouterB-msdp] peer 192.168.3.2 connect-interface GigabitEthernet 0/3
[RouterB-msdp] peer 192.168.1.2 connect-interface GigabitEthernet 0/0
[RouterB-msdp] static-rpf-peer 192.168.3.2 rp-policy list-df
[RouterB-msdp] static-rpf-peer 192.168.1.2 rp-policy list-df
[RouterB-msdp] quit
```

配置 Router B 作为 Router C 的 MSDP 对等体及静态 RPF 对等体。

```
[RouterC] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16 less-equal 32
[RouterC] msdp
[RouterC-msdp] peer 192.168.1.1 connect-interface GigabitEthernet 0/0
```

```
[RouterC-msdp] static-rpf-peer 192.168.1.1 rp-policy list-c
[RouterC-msdp] quit
```

配置 Router B 作为 Router E 的 MSDP 对等体及静态 RPF 对等体。

```
[RouterE] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16 less-equal 32
[RouterE] msdp
[RouterE-msdp] peer 192.168.3.1 connect-interface GigabitEthernet 0/3
[RouterE-msdp] static-rpf-peer 192.168.3.1 rp-policy list-c
[RouterE-msdp] quit
```

(5) 检验配置效果

通过使用 **display bgp peer** 命令可以查看路由器之间 BGP 对等体建立情况，如果无任何信息输出，说明路由器之间未建立 BGP 对等体关系。

当 PIM-SM 1 和 PIM-SM 2 域内的组播源 Source 1 和 Source 2 发送组播信息时，PIM-SM 1 和 PIM-SM 3 域内的接收者（Receiver）能收到该组播信息，通过使用 **display msdp brief** 命令可以查看路由器之间 MSDP 对等体的建立情况。例如：

查看 Router B 上 MSDP 对等体的简要信息。

```
[RouterB] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.2       Up       01:07:08        ?      8          0
  192.168.1.2       Up       00:16:39        ?      13         0
```

查看 Router C 上 MSDP 对等体的简要信息。

```
[RouterC] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.1.1       Up       01:07:09        ?      8          0
```

查看 Router E 上 MSDP 对等体的简要信息。

```
[RouterE] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.1       Up       00:16:40        ?      13         0
```

1.7.3 Anycast RP 应用配置举例

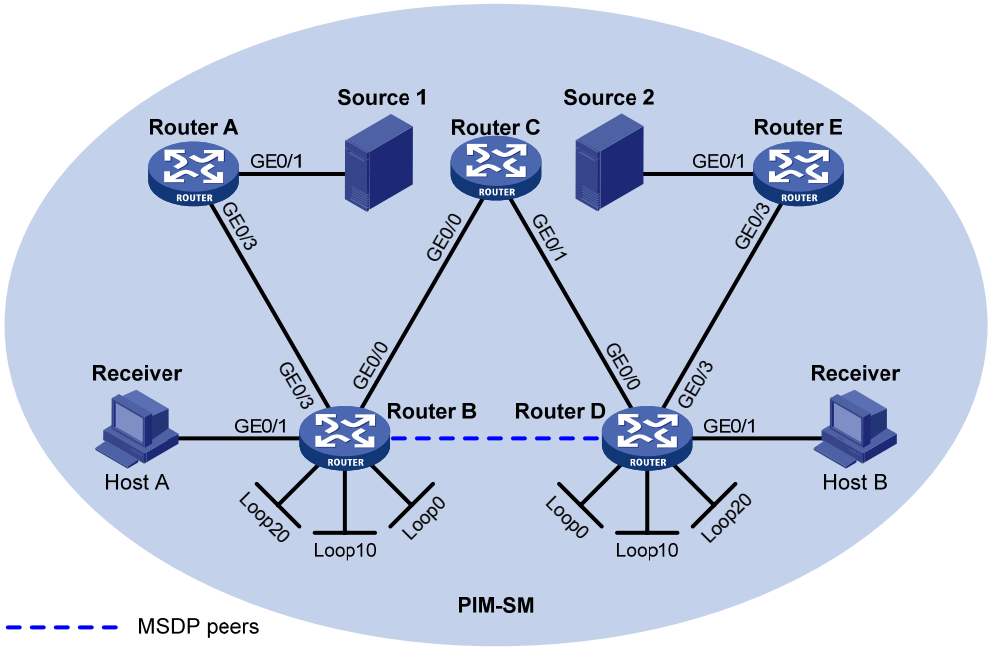
1. 组网需求

- PIM-SM 域内拥有多个组播源和接收者，并在域内运行 OSPF 协议以提供单播路由；

- 通过配置 Anycast RP，使接收者侧 DR 能够向拓扑距离最近的 RP 发起加入，组播源侧 DR 也向拓扑距离最近的 RP 发起注册；
- 将 Router B 和 Router D 各自的 Loopback10 接口配置为 C-BSR、Loopback20 接口配置为 C-RP；
- Router B 的 Router ID 为 1.1.1.1，Router D 的 Router ID 为 2.2.2.2，在 Router B 和 Router D 之间建立 MSDP 对等体关系。

2. 组网图

图1-7 Anycast RP 应用配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source 1	-	10.110.5.100/24	Router C	GE0/0	192.168.1.2/24
Source 2	-	10.110.6.100/24		GE0/1	192.168.2.2/24
Router A	GE0/1	10.110.5.1/24	Router D	GE0/1	10.110.3.1/24
	GE0/3	10.110.2.2/24		GE0/3	10.110.4.1/24
Router B	GE0/1	10.110.1.1/24		GE0/0	192.168.2.1/24
	GE0/3	10.110.2.1/24		Loop0	2.2.2.2/32
	GE0/0	192.168.1.1/24		Loop10	4.4.4.4/32
	Loop0	1.1.1.1/32		Loop20	10.1.1.1/32
	Loop10	3.3.3.3/32	Router E	GE0/1	10.110.6.1/24
	Loop20	10.1.1.1/32		GE0/1	10.110.4.2/24

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-7配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-SM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-SM 和 IGMP

在 Router B 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet 0/1 上使能 IGMP。

```
<RouterB> system-view
```

```
[RouterB] multicast routing-enable
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] igmp enable
[RouterB-GigabitEthernet0/1] pim sm
[RouterB-GigabitEthernet0/1] quit
[RouterB] interface GigabitEthernet 0/3
[RouterB-GigabitEthernet0/3] pim sm
[RouterB-GigabitEthernet0/3] quit
[RouterB] interface GigabitEthernet 0/0
[RouterB-GigabitEthernet0/0] pim sm
[RouterB-GigabitEthernet0/0] quit
[RouterB] interface loopback 0
[RouterB-LoopBack0] pim sm
[RouterB-LoopBack0] quit
[RouterB] interface loopback 10
[RouterB-LoopBack10] pim sm
[RouterB-LoopBack10] quit
[RouterB] interface loopback 20
[RouterB-LoopBack20] pim sm
[RouterB-LoopBack20] quit
```

Router A、Router C、Router D 和 Router E 上的配置与 Router B 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP 的位置

在 RouterB 上将 Loopback10 配置为 C-BSR，将 Loopback20 配置为 C-RP。

```
[RouterB] pim
[RouterB-pim] c-bsr loopback 10
[RouterB-pim] c-rp loopback 20
[RouterB-pim] quit
```

Router D 上的配置与 Router B 相似，配置过程略。

(4) 配置 MSDP 对等体

在 Router B 的 Loopback0 接口上配置 MSDP 对等体。

```
[RouterB] msdp
[RouterB-msdp] originating-rp loopback 0
[RouterB-msdp] peer 2.2.2.2 connect-interface loopback 0
[RouterB-msdp] quit
```

在 Router D 的 Loopback0 接口上配置 MSDP 对等体。

```
[RouterD] msdp
[RouterD-msdp] originating-rp loopback 0
[RouterD-msdp] peer 1.1.1.1 connect-interface loopback 0
[RouterD-msdp] quit
```

(5) 检验配置效果

通过使用 **display msdp brief** 命令可以查看路由器之间 MSDP 对等体建立情况。

查看 Router B 上 MSDP 对等体的简要信息。

```
[RouterB] display msdp brief
```

```
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  2.2.2.2           Up       00:10:17        ?      0          0
```

查看 Router D 上 MSDP 对等体的简要信息。

```
[RouterD] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  1.1.1.1           Up       00:10:18        ?      0          0
```

通过使用 **display pim routing-table** 命令可以查看路由器上的 PIM 路由。

当 Source 1（10.110.5.100/24）开始向组播组 G（225.1.1.1）发送组播信息时，Host A 加入组播组 G。通过比较 Router B 与 Router D 上 PIM 路由的显示信息，可知当前的有效 RP 为 Router B：Source 1 向 Router B 注册，Host A 向 Router B 加入。

查看 Router B 上的 PIM 路由的信息。

```
[RouterB] display pim routing-table
 Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:15:04
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet0/1
             Protocol: igmp, UpTime: 00:15:04, Expires: -

 (10.110.5.100, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT 2MSDP ACT
     UpTime: 00:46:28
     Upstream interface: GigabitEthernet0/3
         Upstream neighbor: 10.110.2.2
         RPF prime neighbor: 10.110.2.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet0/1
             Protocol: pim-sm, UpTime:  - , Expires:  -
```

查看 Router D 上的 PIM 路由的信息。

```
[RouterD] display pim routing-table
```

Router D 上没有信息输出。

Host A 离开组播组 G，Source 1 也停止向组播组 G 发送组播数据。当 Source 2（10.110.6.100/24）开始向组播组 G 发送组播信息时，Host B 加入组播组 G。通过比较 Router B 与 Router D 上 PIM 路由的显示信息，可知当前的有效 RP 为 Router D：Source 2 向 Router D 注册，Host B 向 Router D 加入。

查看 Router B 上的 PIM 路由的信息。

```
[RouterB] display pim routing-table
```

Router B 上没有信息输出。

查看 Router D 上的 PIM 路由的信息。

```
[RouterD] display pim routing-table
 Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:12:07
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet01/1
             Protocol: igmp, UpTime: 00:12:07, Expires: -

 (10.110.6.100, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT 2MSDP ACT
     UpTime: 00:40:22
     Upstream interface: GigabitEthernet0/3
         Upstream neighbor: 10.110.4.2
         RPF prime neighbor: 10.110.4.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet01/1
             Protocol: pim-sm, UpTime:  - , Expires:  -
```

1.7.4 SA 消息过滤机制配置举例

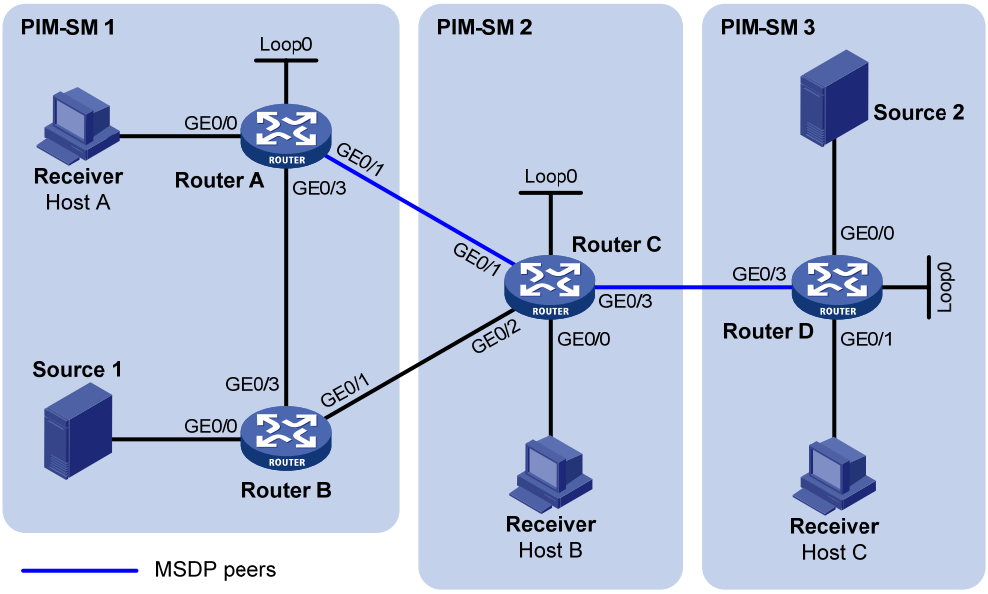
1. 组网需求

- 网络中存在三个 PIM-SM 域，各域内部以及域之间均运行 OSPF 协议以提供单播路由；
- 将 Router A、Router C 和 Router D 各自的 Loopback0 接口分别配置为各自 PIM-SM 域的 C-BSR 和 C-RP；

- 分别在 Router A 与 Router C、Router C 与 Router D 之间建立 MSDP 对等体关系；
- 组播源 Source 1 向组播组 225.1.1.0/30 和 226.1.1.0/30 发送组播数据，组播源 Source 2 向组播组 227.1.1.0/30 发送组播数据；
- 通过配置 SA 消息过滤规则，使接收者 Host A 和 Host B 只能接收发往组播组 225.1.1.0/30 和 226.1.1.0/30 的组播数据，而 Host C 则只能接收发往组播组 226.1.1.0/30 和 227.1.1.0/30 的组播数据。

2. 组网图

图1-8 SA 消息过滤机制配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source 1	-	10.110.3.100/24	Router C	GE0/0	10.110.4.1/24
Source 2	-	10.110.6.100/24		GE0/3	10.110.5.1/24
Router A	GE0/0	10.110.1.1/24		GE0/1	192.168.1.2/24
	GE0/3	10.110.2.1/24		GE0/2	192.168.2.2/24
	GE0/1	192.168.1.1/24		Loop0	2.2.2.2/32
	Loop0	1.1.1.1/32	Router D	GE0/0	10.110.6.1/24
Router B	GE0/0	10.110.3.1/24		GE0/1	10.110.7.1/24
	GE0/3	10.110.2.2/24		GE0/3	10.110.5.2/24
	GE0/1	192.168.2.1/24		Loop0	3.3.3.3/32

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-8配置各接口的IP地址和掩码，具体配置过程略。

配置各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部以及各 PIM-SM 域之间在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，使能 PIM-SM 和 IGMP，并配置 BSR 的服务边界

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet 0/1 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/0
```

```
[RouterA-GigabitEthernet0/0] igmp enable
[RouterA-GigabitEthernet0/0] pim sm
[RouterA-GigabitEthernet0/0] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet 0/3] pim sm
[RouterA-GigabitEthernet 0/3] quit
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
[RouterA] interface loopback 0
[RouterA-LoopBack0] pim sm
[RouterA-LoopBack0] quit
```

Router B、Router C 和 Router D 上的配置与 Router A 相似，配置过程略。

在 Router C 上配置 BSR 的服务边界。

```
[RouterC] interface GigabitEthernet 0/1
[RouterC-GigabitEthernet0/1] pim bsr-boundary
[RouterC-GigabitEthernet0/1] quit
[RouterC] interface GigabitEthernet 0/2
[RouterC-GigabitEthernet0/2] pim bsr-boundary
[RouterC-GigabitEthernet0/2] quit
[RouterC] interface GigabitEthernet 0/3
[RouterC-GigabitEthernet 0/3] pim bsr-boundary
[RouterC-GigabitEthernet 0/3] quit
```

Router A、Router B 和 Router D 上的配置与 Router C 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP 的位置

在 Router A 上将 Loopback0 接口配置为 C-BSR 和 C-RP。

```
[RouterA] pim
[RouterA-pim] c-bsr loopback 0
[RouterA-pim] c-rp loopback 0
[RouterA-pim] quit
```

Router C 和 Router D 上的配置与 Router A 相似，配置过程略。

(4) 配置 MSDP 对等体

在 Router A 上配置 MSDP 对等体。

```
[RouterA] msdp
[RouterA-msdp] peer 192.168.1.2 connect-interface GigabitEthernet 0/1
[RouterA-msdp] quit
```

在 Router C 上配置 MSDP 对等体。

```
[RouterC] msdp
[RouterC-msdp] peer 192.168.1.1 connect-interface GigabitEthernet 0/1
[RouterC-msdp] peer 10.110.5.2 connect-interface GigabitEthernet 0/3
[RouterC-msdp] quit
```

在 Router D 上配置 MSDP 对等体。

```
[RouterD] msdp
```

```
[RouterD-msdp] peer 10.110.5.1 connect-interface GigabitEthernet 0/3
[RouterD-msdp] quit
```

(5) 配置 SA 消息过滤规则

在 Router C 上配置不向 Router D 转发有关（Source 1，225.1.1.0/30）的 SA 消息。

```
[RouterC] acl number 3001
[RouterC-acl-adv-3001] rule deny ip source 10.110.3.100 0 destination 225.1.1.0 0.0.0.3
[RouterC-acl-adv-3001] rule permit ip source any destination any
[RouterC-acl-adv-3001] quit
[RouterC] msdp
[RouterC-msdp] peer 10.110.5.2 sa-policy export acl 3001
[RouterC-msdp] quit
```

在 Router D 上配置不创建有关 Source 2 的 SA 消息。

```
[RouterD] acl number 2001
[RouterD-acl-basic-2001] rule deny source 10.110.6.100 0
[RouterD-acl-basic-2001] quit
[RouterD] msdp
[RouterD-msdp] import-source acl 2001
[RouterD-msdp] quit
```

(6) 检验配置效果

通过使用 **display msdp sa-cache** 命令可以查看路由器上 SA 缓存中的（S，G）表项信息。例如：

查看 Router C 上 SA 缓存中的（S，G）表项信息。

```
[RouterC] display msdp sa-cache
MSDP Source-Active Cache Information of VPN-Instance: public net
 MSDP Total Source-Active Cache - 8 entries
 MSDP matched 8 entries

(Source, Group)              Origin RP       Pro  AS     Uptime   Expires
(10.110.3.100, 225.1.1.0)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.1)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.2)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.3)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.0)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.1)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.2)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.3)    1.1.1.1         ?    ?      02:03:30 00:05:31
```

查看 Router D 上 SA 缓存中的（S，G）表项信息。

```
[RouterD] display msdp sa-cache
MSDP Source-Active Cache Information of VPN-Instance: public net
 MSDP Total Source-Active Cache - 4 entries
 MSDP matched 4 entries

(Source, Group)              Origin RP       Pro  AS     Uptime   Expires
(10.110.3.100, 226.1.1.0)    1.1.1.1         ?    ?      00:32:53 00:05:07
(10.110.3.100, 226.1.1.1)    1.1.1.1         ?    ?      00:32:53 00:05:07
(10.110.3.100, 226.1.1.2)    1.1.1.1         ?    ?      00:32:53 00:05:07
```

```
(10.110.3.100, 226.1.1.3)    1.1.1.1         ?    ?      00:32:53 00:05:07
```

1.8 常见配置错误举例

1.8.1 MSDP 对等体一直处于 down 状态

1. 故障现象

配置了 MSDP 对等体，但其状态一直显示为 down。

2. 分析

- 所配置的本地接口地址与 MSDP 对等体地址之间，建立起基于 TCP 连接的 MSDP 对等体关系；
- 如果本地接口地址与对端路由器上所配置的 MSDP 对等体地址不一致，TCP 连接就不会建立起来；
- 如果两个 MSDP 对等体之间没有路由，TCP 连接也不会建立起来。

3. 处理过程

(1) 检查各路由器之间的路由是否通达。通过命令 **display ip routing-table** 查看各路由器之间单播路由是否正确。

(2) 检查将成为 MSDP 对等体的两个路由器间是否存在到达对方的单播路由。

(3) 检查 MSDP 对等体之间的接口地址是否匹配。通过命令 **display current-configuration** 查看本地接口地址是否与对端 MSDP 对等体地址一致，应确保一致。

1.8.2 路由器 SA 缓存中没有 SA 表项

1. 故障现象

MSDP 没有将（S，G）转发项通过 SA 消息发送出去。

2. 分析

- 命令 **import-source** 用于控制将本域的（S，G）项通过 SA 消息发送给 MSDP 对等体。如果没有指定参数 *acl-number* 则表示默认过滤掉所有的（S，G）项，即不通告本域的所有（S，G）项；
- 未配置 **import-source** 命令时，系统将发送本域的所有（S，G）信息。如果 MSDP 没有将本域的（S，G）项通过 SA 消息发送出去，应检查 **import-source** 命令的配置是否正确。

3. 处理过程

(1) 检查各路由器之间的路由是否通达。通过命令 **display ip routing-table** 查看各路由器之间单播路由是否正确。

(2) 检查将成为 MSDP 对等体的两个路由器间是否存在到达对方的单播路由。

(3) 检查命令 **import-source** 及其参数 *acl-number* 的配置情况，确保 ACL 规则能够过滤合适的（S，G）信息。

1.8.3 Anycast RP 应用中 RP 间互通信息异常

1. 故障现象

在 Anycast RP 的应用中，各 RP 之间没有互相交换其本地注册的（S，G）信息。

2. 分析

- 在 Anycast RP 应用中，通过将同一 PIM-SM 域内的 RP 配置为 MSDP 对等体，可以实现 RP 间的负载分担；
- MSDP 对等体地址不能与 Anycast RP 地址相同，且 C-BSR 和 C-RP 必须配置在不同的设备或接口上；
- 在使用 **originating-rp** 命令进行配置后，MSDP 将利用该命令所指定的接口地址替换 SA 消息中的 RP 地址；
- 当 MSDP 对等体对收到的 SA 消息进行 RPF 检查时，如果发现 RP 地址与本地配置的 RP 地址相同，将拒绝接收 SA 消息。

3. 处理过程

(1) 检查各路由器之间的路由是否通达。通过命令 **display ip routing-table** 查看各路由器之间单播路由是否正确。

(2) 检查将成为 MSDP 对等体的两个路由器间是否存在到达对方的单播路由。

(3) 检查 **originating-rp** 命令的配置情况。在 Anycast RP 的应用环境中，一定要配置 **originating-rp** 命令，而且 **originating-rp** 命令所指定的接口地址要与建立 MSDP 对等体连接的本地接口地址相同。

(4) 检查所配置的 C-BSR 地址是否与 Anycast RP 的地址不同，应确保两个地址不同。

目 录

1 PIM配置 1-1
1.1 PIM简介 1-1
1.1.1 PIM-DM简介 1-1
1.1.2 PIM-DM工作机制 1-1
1.1.3 PIM-SM简介 1-4
1.1.4 PIM-SM工作机制 1-5
1.1.5 PIM-SM管理域机制介绍 1-10
1.1.6 SSM模型在PIM中的实现 1-11
1.1.7 多实例的PIM 1-13
1.1.8 协议规范 1-13
1.2 配置PIM-DM 1-13
1.2.1 PIM-DM配置任务简介 1-13
1.2.2 配置准备 1-13
1.2.3 使能PIM-DM 1-14
1.2.4 使能状态刷新能力 1-15
1.2.5 配置状态刷新参数 1-15
1.2.6 配置PIM-DM定时器 1-16
1.3 配置PIM-SM 1-16
1.3.1 PIM-SM配置任务简介 1-16
1.3.2 配置准备 1-17
1.3.3 使能PIM-SM 1-17
1.3.4 配置RP 1-18
1.3.5 配置BSR 1-20
1.3.6 配置管理域 1-23
1.3.7 配置组播源注册 1-25
1.4 配置PIM-SSM 1-26
1.4.1 PIM-SSM配置任务简介 1-26
1.4.2 配置准备 1-26
1.4.3 使能PIM-SM 1-26
1.4.4 配置SSM组播组范围 1-27
1.5 配置PIM公共特性 1-28
1.5.1 PIM公共特性配置任务简介 1-28
1.5.2 配置准备 1-28
1.5.3 配置组播数据过滤器 1-29
1.5.4 配置Hello报文过滤器 1-29
1.5.5 配置Hello报文选项 1-30
1.5.6 配置PIM公共定时器 1-31
1.5.7 配置加入/剪枝报文规格 1-33
1.6 PIM显示和维护 1-33
1.7 PIM典型配置举例 1-34
1.7.1 PIM-DM典型配置举例 1-34
1.7.2 PIM-SM非管理域配置举例 1-37

1.7.3 PIM-SM管理域配置举例 1-43
1.7.4 PIM-SSM典型配置举例 1-49
1.8 常见配置错误举例 1-52
1.8.1 无法正确建立组播分发树 1-52
1.8.2 组播数据异常终止在中间路由器 1-53
1.8.3 PIM-SM中RP无法加入SPT 1-53
1.8.4 PIM-SM中无法建立RPT或无法进行源注册 1-54

1 PIM 配置

说明

本文所涉及的路由器和路由器图标，代表了一般意义下的路由器或运行了 PIM 协议的三层交换机。

1.1 PIM 简介

PIM 是 Protocol Independent Multicast（协议无关组播）的简称，表示可以利用静态路由或者任意单播路由协议（包括 RIP、OSPF、IS-IS、BGP 等）所生成的单播路由表为 IP 组播提供路由。组播路由与所采用的单播路由协议无关，只要能够通过单播路由协议产生相应的组播路由表项即可。PIM 借助 RPF（Reverse Path Forwarding，逆向路径转发）机制实现对组播报文的转发。当组播报文到达本地设备时，首先对其进行 RPF 检查：若 RPF 检查通过，则创建相应的组播路由表项，从而进行组播报文的转发；若 RPF 检查失败，则丢弃该报文。有关 RPF 的详细介绍，请参见“IP 组播分册”中的“组播路由与转发配置”。

根据实现机制的不同，PIM 分为以下两种模式：

- PIM-DM（Protocol Independent Multicast-Dense Mode，协议无关组播—密集模式）
- PIM-SM（Protocol Independent Multicast-Sparse Mode，协议无关组播—稀疏模式）

说明

为了描述的方便，本文中把由支持 PIM 协议的组播路由器所组成的网络简称为“PIM 域”。

1.1.1 PIM-DM 简介

PIM-DM 属于密集模式的组播路由协议，使用“推（Push）模式”传送组播数据，通常适用于组播组成员相对比较密集的小型网络。

PIM-DM 的基本原理如下：

- PIM-DM 假设网络中的每个子网都存在至少一个组播组成员，因此组播数据将被扩散（Flooding）到网络中的所有节点。然后，PIM-DM 对没有组播数据转发的分支进行剪枝（Prune），只保留包含接收者的分支。这种“扩散—剪枝”现象周期性地发生，被剪枝的分支也可以周期性地恢复成转发状态。
- 当被剪枝分支的节点上出现了组播组的成员时，为了减少该节点恢复成转发状态所需的时间，PIM-DM 使用嫁接（Graft）机制主动恢复其对组播数据的转发。

一般说来，密集模式下数据包的转发路径是有源树（Source Tree，即以组播源为“根”、组播组成员为“枝叶”的一棵转发树）。由于有源树使用的是从组播源到接收者的最短路径，因此也称为最短路径树（Shortest Path Tree，SPT）。

1.1.2 PIM-DM 工作机制

PIM-DM 的工作机制可以概括如下：

- 邻居发现
- 构建 SPT
- 嫁接
- 断言

1. 邻居发现

在 PIM 域中，路由器通过周期性地的向所有 PIM 路由器（224.0.0.13）以组播方式发送 PIM Hello 报文（以下简称 Hello 报文），以发现 PIM 邻居，维护各路由器之间的 PIM 邻居关系，从而构建和维护 SPT。

 说明

路由器每个运行了 PIM 协议的接口都会周期性地发送 Hello 报文，从而了解与该接口相关的 PIM 邻居信息。

---

2. 构建 SPT

构建 SPT 的过程也就是“扩散—剪枝”的过程：

(1) 在 PIM-DM 域中，组播源 S 向组播组 G 发送组播报文时，首先对组播报文进行扩散：路由器对该报文的 RPF 检查通过后，便创建一个（S，G）表项，并将该报文向网络中的所有下游节点转发。经过扩散，PIM-DM 域内的每个路由器上都会创建（S，G）表项。

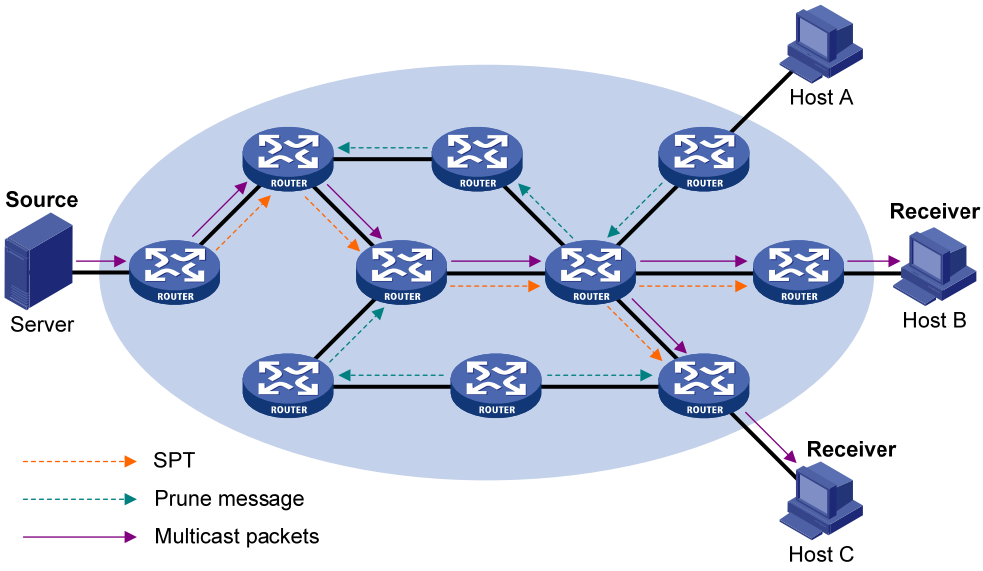
(2) 然后对那些下游没有接收者的节点进行剪枝：由没有接收者的下游节点向上游节点发剪枝报文（Prune Message），以通知上游节点将相应的接口从其组播转发表项（S，G）所对应的出接口列表中删除，并不再转发该组播组的报文至该节点。

 说明

- （S，G）表项包括组播源的地址 S、组播组的地址 G、出接口列表和入接口等。
- 路由器上收到组播数据的接口称为“上游”，转发组播数据的接口称为“下游”。

剪枝过程最先由叶子路由器发起，如 图 1-1所示，没有接收者（Receiver）的路由器（如与Host A直连的路由器）主动发起剪枝，并一直持续到PIM-DM域中只剩下必要的分支，这些分支共同构成了SPT。

图1-1 PIM-DM 中构建 SPT 示意图



“扩散—剪枝”的过程是周期性发生的。各个被剪枝的节点提供超时机制，当剪枝超时后便重新开始这一过程。

说明

剪枝在 PIM-SM 中有着相似的应用。

---

3. 嫁接

当被剪枝的节点上出现了组播组的成员时，为了减少该节点恢复成转发状态所需的时间，PIM-DM 使用嫁接机制主动恢复其对组播数据的转发，过程如下：

(1) 需要恢复接收组播数据的节点向其上游节点发送嫁接报文（Graft Message）以申请重新加入到 SPT 中；

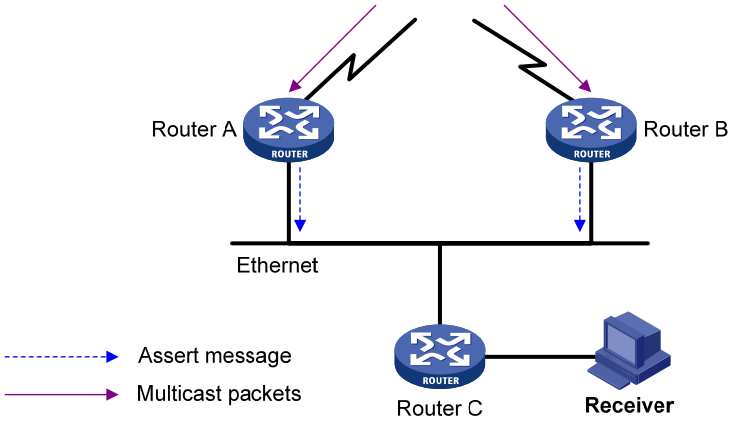
(2) 当上游节点收到该报文后恢复该下游节点的转发状态，并向其回应一个嫁接应答报文（Graft-Ack Message）以进行确认；

(3) 如果发送嫁接报文的下游节点没有收到来自其上游节点的嫁接应答报文，将重新发送嫁接报文直到被确认为止。

4. 断言

在一个网段内如果存在多台组播路由器，则相同的组播报文可能会被重复发送到该网段。为了避免出现这种情况，就需要通过断言（Assert）机制来选定唯一的组播数据转发者。

图1-2 Assert 机制示意图



如 图 1-2所示，当Router A和Router B从上游节点收到（S，G）组播报文后，都会向本地网段转发该报文，于是处于下游的节点Router C就会收到两份相同的组播报文，Router A和Router B也会从各自的本地接口收到对方转发来的该组播报文。此时，Router A和Router B会通过本地接口向所有PIM路由器（224.0.0.13）以组播方式发送断言报文（Assert Message），该报文中携带有以下信息：组播源地址S、组播组地址G、到组播源的单播路由的优先级和度量值。通过一定的规则对这些参数进行比较后，Router A和Router B中的获胜者将成为（S，G）组播报文在本网段的转发者，比较规则如下：

(1) 到组播源的单播路由的优先级较高者获胜；

(2) 如果到组播源的单播路由的优先级相等，那么到组播源的度量值较小者获胜；

(3) 如果到组播源的度量值也相等，则本地接口 IP 地址较大者获胜。

1.1.3 PIM-SM 简介

PIM-DM 使用以“扩散—剪枝”方式构建的 SPT 来传送组播数据。尽管 SPT 的路径最短，但是其建立的过程效率较低，并不适合大中型网络。

PIM-SM 属于稀疏模式的组播路由协议，使用“拉（Pull）模式”传送组播数据，通常适用于组播组成员分布相对分散、范围较广的大中型网络。

PIM-SM 的基本原理如下：

- PIM-SM 假设所有主机都不需要接收组播数据，只向明确提出需要组播数据的主机转发。PIM-SM 实现组播转发的核心任务就是构造并维护 RPT（Rendezvous Point Tree，共享树或汇集树），RPT 选择 PIM 域中某台路由器作为公用的根节点 RP（Rendezvous Point，汇集点），组播数据通过 RP 沿着 RPT 转发给接收者；
- 连接接收者的路由器向某组播组对应的 RP 发送加入报文（Join Message），该报文被逐跳送达 RP，所经过的路径就形成了 RPT 的分支；
- 组播源如果要向某组播组发送组播数据，首先由组播源侧 DR（Designated Router，指定路由器）负责向 RP 进行注册，把注册报文（Register Message）通过单播方式发送给 RP，该报文到达 RP 后触发建立 SPT。之后组播源把组播数据沿着 SPT 发向 RP，当组播数据到达 RP 后，被复制并沿着 RPT 发送给接收者。

说明

复制仅发生在分发树的分支处，这个过程能够自动重复直到数据包最终到达接收者。

1.1.4 PIM-SM 工作机制

PIM-SM 的工作机制可以概括如下：

- 邻居发现
- DR 选举
- RP 发现
- 构建 RPT
- 组播源注册
- 断言

1. 邻居发现

PIM-SM使用与PIM-DM类似的邻居发现机制，具体请参见“1.1.2 1. 邻居发现”一节。

2. DR 选举

借助 Hello 报文还可以为共享网络（如 Ethernet）选举 DR，DR 将作为该共享网络中组播数据的唯一转发者。

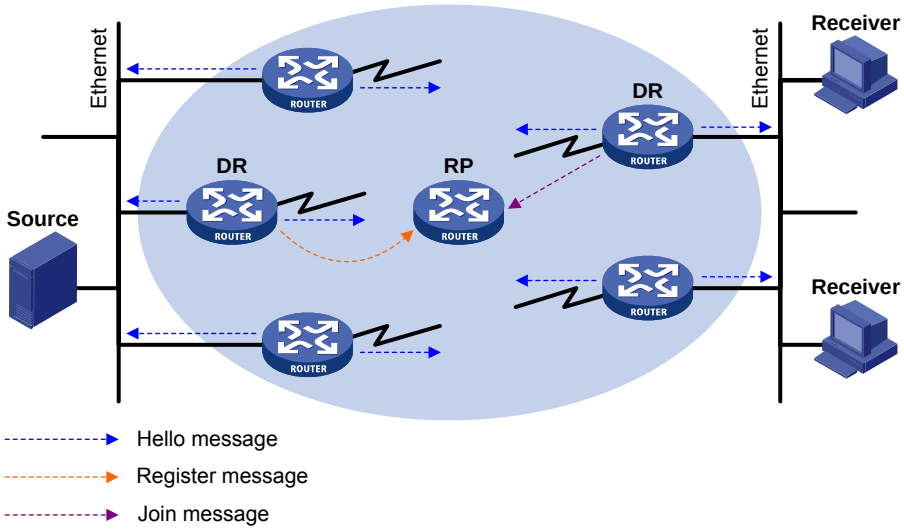
无论是与组播源相连的网络，还是与接收者相连的网络，都需要选举 DR。接收者侧的 DR 负责向 RP 发送加入报文；组播源侧的 DR 负责向 RP 发送注册报文。

说明

- 各路由器之间通过比较 Hello 报文中所携带的优先级和 IP 地址，可以为多路由器网段选举 DR。选举出的 DR 对于 PIM-SM 有实际的意义；而对于 PIM-DM 来说，其本身其实并不需要 DR，但如果 PIM-DM 域中的共享网络上运行了 IGMPv1，则需要选举出 DR 来充当共享网络上的 IGMPv1 查询器。
- 在充当接收者侧 DR 的设备上必须使能 IGMP，否则连接在该 DR 上的接收者将不能通过该 DR 加入组播组。

有关 IGMP 的介绍，请参见“IP 组播分册”中的“IGMP 配置”。

图1-3 DR 选举示意图



如 图 1-3所示，DR的选举过程如下：

(1) 共享网络上的各路由器相互之间发送 Hello 报文（携带有竞选 DR 优先级的参数），拥有最高优先级的路由器将成为 DR；

(2) 如果优先级相同，或者网络中至少有一台路由器不支持在 Hello 报文中携带竞选 DR 优先级的参数，则根据各路由器的 IP 地址大小来竞选 DR，IP 地址最大的路由器将成为 DR。

当 DR 出现故障时，其余路由器在超时后仍没有收到来自 DR 的 Hello 报文，则会触发新的 DR 选举过程。

3. RP 发现

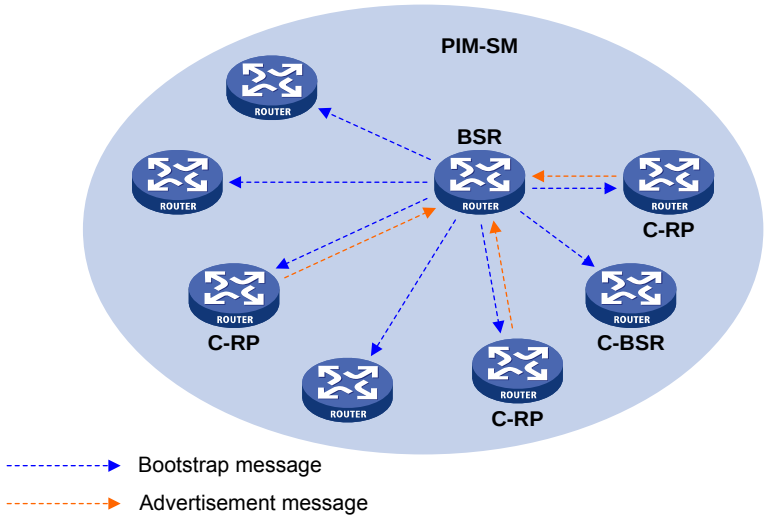
RP 是 PIM-SM 域中的核心设备。在结构简单的小型网络中，组播信息量少，整个网络仅依靠一个 RP 进行组播信息的转发即可，此时可以在 PIM-SM 域中的各路由器上静态指定 RP 的位置；但是在更多的情况下，PIM-SM 域的规模都很大，通过 RP 转发的组播信息量巨大。为了缓解 RP 的负担并优化 RPT 的拓扑结构，可以在 PIM-SM 域中配置多个 C-RP（Candidate-RP，候选 RP），通过自举机制来动态选举 RP，使不同的 RP 服务于不同的组播组，此时需要配置 BSR（BootStrap Router，自举路由器）。BSR 是 PIM-SM 域的管理核心，一个 PIM-SM 域内只能有一个 BSR，但可以配置多个 C-BSR（Candidate-BSR，候选 BSR）。这样，一旦 BSR 发生故障，其余 C-BSR 能够通过自动选举产生新的 BSR，从而确保业务免受中断。

说明

- 一个 RP 可以同时服务于多个组播组，但一个组播组只能唯一对应一个 RP。
- 一台设备可以同时充当 C-RP 和 C-BSR。

如 图 1-4所示，BSR负责收集网络中由C-RP发来的宣告报文（Advertisement Message），该报文中携带有C-RP的地址和优先级以及其服务的组范围，BSR将这些信息汇总为RP-Set（RP集，即组播组与RP的映射关系数据库），封装在自举报文（Bootstrap Message）中并发布到整个PIM-SM域。

图1-4 RP 与 BSR 信息交互示意图



网络中的各路由器将依据 RP-Set 提供的信息，使用相同的规则从众多 C-RP 中为特定组播组选择其对应的 RP，具体规则如下：

(1) 首先比较 C-RP 的优先级，优先级较高者获胜。

(2) 若优先级相同，则使用哈希（Hash）函数计算哈希值，该值较大者获胜。

(3) 若优先级和哈希值都相同，则 C-RP 地址较大者获胜。

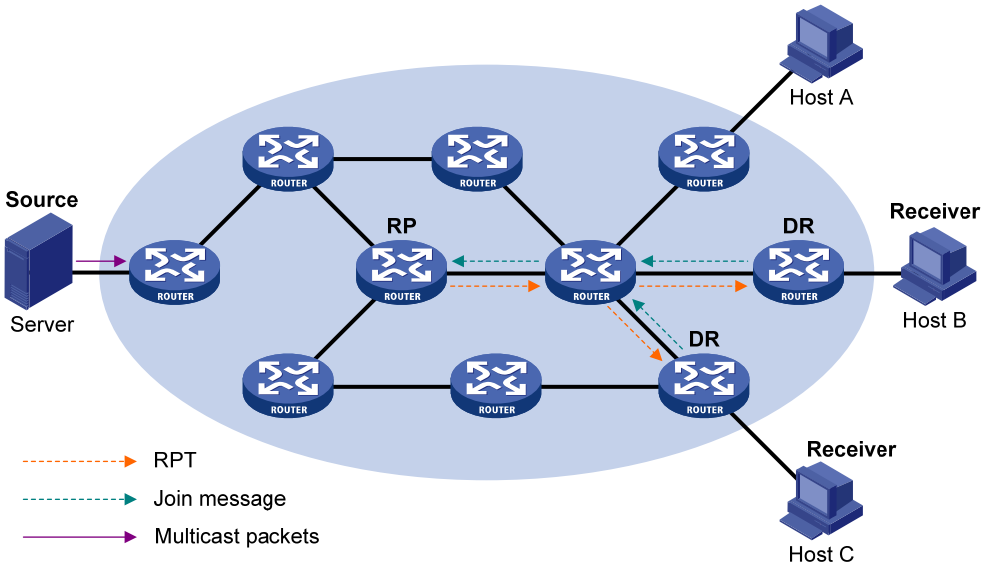
哈希函数的表达式为：Value (G, M, C_i) = (1103515245 * ((1103515245 * (G & M) + 12345) XOR C_i) + 12345) mod 2^{31}，其中各符号的含义如 表 1-1所示。

表1-1 哈希函数中各符号含义

符号	含义
Value	哈希值
G	IP 组播组的地址
M	哈希掩码长度（Hash Mask Length）
C_i	C-RP 的 IP 地址
&	逻辑运算符，表示与运算
XOR	逻辑运算符，表示异或运算
mod	算术运算符，表示整除取余

4. 构建 RPT

图1-5 PIM-SM 中构建 RPT 示意图



如 图 1-5所示，RPT的构建过程如下：

(1) 当接收者加入一个组播组 G 时，先通过 IGMP 报文通知与其直连的 DR；

(2) DR 掌握了组播组 G 的接收者的信息后，向该组所对应的 RP 方向逐跳发送加入报文；

(3) 从 DR 到 RP 所经过的路由器就形成了 RPT 的分支，这些路由器都在其转发表中生成了（*，G）表项，这里的“*”表示来自任意组播源。RPT 以 RP 为根，以 DR 为叶子。

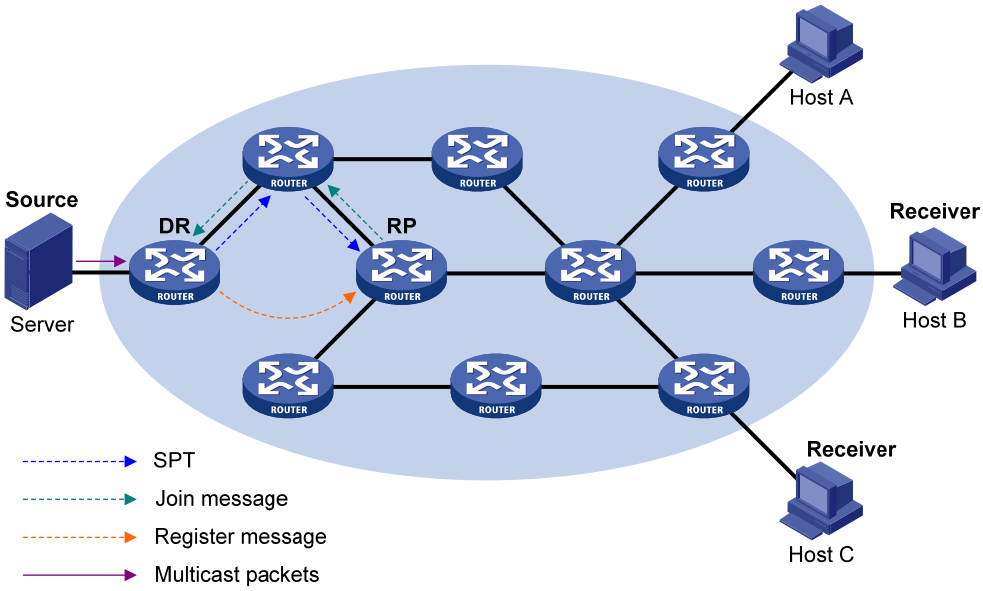
当发往组播组 G 的组播数据流经 RP 时，数据就会沿着已建立好的 RPT 到达 DR，进而到达接收者。

当某接收者对组播组 G 的信息不再感兴趣时，与其直连的 DR 会逆着 RPT 向该组的 RP 方向逐跳发送剪枝报文；上游节点收到该报文后在其出接口列表中删除与下游节点相连的接口，并检查自己是否拥有该组播组的接收者，如果没有则继续向其上游转发该剪枝报文。

5. 组播源注册

组播源注册的目的是向 RP 通知组播源的存在。

图1-6 组播源注册示意图



如 图 1-6所示，组播源向RP注册的过程如下：

(1) 当组播源 S 向组播组 G 发送了一个组播报文时，与组播源直连的 DR 在收到该报文后，就将其封装成注册报文，并通过单播方式发送给相应的 RP；

(2) 当 RP 收到该报文后，一方面解封装注册报文并将封装在其中的组播报文沿着 RPT 转发给接收者，另一方面向组播源方向逐跳发送（S，G）加入报文。这样，从 RP 到组播源所经过的路由器就形成了 SPT 的分支，这些路由器都在其转发表中生成了（S，G）表项。SPT 以组播源侧的 DR 为根，以 RP 为叶子。

(3) 组播源发出的组播数据沿着已建立好的 SPT 到达 RP，然后由 RP 把组播数据沿着 RPT 向接收者进行转发。当 RP 收到沿着 SPT 转发来的组播数据后，通过单播方式向与组播源直连的 DR 发送注册停止报文（Register-Stop Message），组播源注册过程结束。

6. SPT 切换

在 PIM-SM 域中，一个组播组唯一对应一个 RP 和一棵 RPT。在 SPT 切换前，所有发往该组的组播报文都必须先由组播源侧 DR 封装在注册报文中发往 RP，由 RP 解封装后再沿 RPT 分发给接收者侧的 DR，RP 是所有组播数据必经的中转站。

(1) RP 发起的 SPT 切换

RP 周期性地检测组播数据（S，G）的转发速率，缺省情况下，收到组播数据后，立即向组播源方向发送（S，G）加入报文，在组播源侧 DR 与 RP 之间建立起 SPT 分支，后续的组播报文都直接沿该分支到达 RP。

说明

由RP发起的SPT切换的详细过程，请参见“1.1.4 5. 组播源注册”一节。

---

(2) 接收者侧 DR 发起的 SPT 切换

接收者侧 DR 周期性地检测组播数据（S，G）的转发速率，缺省情况下，收到组播数据后，立即发起 SPT 切换，过程如下：

- 首先，接收者侧 DR 向组播源方向发送（S，G）加入报文，并最终送达组播源侧 DR，沿途经过的所有路由器在其转发表中都生成了（S，G）表项，从而建立了 SPT 分支；
- 随后，当接收者侧 DR 收到沿 SPT 发来的组播数据后，开始丢弃沿 RPT 发来的组播数据，并同时向 RP 逐跳发送包含 RP 位的剪枝报文，RP 收到该报文后会向组播源方向继续发送剪枝报文（假设此时只有这一个接收者），从而最终实现从 RPT 向 SPT 的切换；
- 最终，组播数据将沿 SPT 从组播源发送到接收者。

通过 SPT 切换，PIM-SM 能够以比 PIM-DM 更经济的方式建立 SPT。

7. 断言

PIM-SM使用与PIM-DM类似的断言机制，具体请参见“1.1.2 4. 断言”一节。

1.1.5 PIM-SM 管理域机制介绍

1. PIM-SM 域两种机制的划分

一般情况下，在一个 PIM-SM 域内只能有一个 BSR，并由该 BSR 负责在整个 PIM-SM 域内宣告 RP-Set 信息，所有组播组的信息都在此 BSR 管理的网络范围内进行转发，我们称之为非管理域机制。

考虑到管理的精细化，可以将整个 PIM-SM 域划分为一个 Global 域（Global-scope Zone）和多个管理域（Admin-scope Zone），一方面可以有效分担单一 BSR 的管理压力，另一方面可以使用私有组地址为特定区域提供专门的服务。相应地，我们称之为管理域机制。

管理域与组播组相对应，针对不同组播组划分相应的管理域。管理域的边界由 ZBR（Zone Border Router，区域边界路由器）构成，每个管理域各维护一个 BSR，为特定范围的组播组服务，属于此范围的组播协议报文（如断言报文、BSR 自举报文等）无法通过管理域边界。不同管理域所服务的组播组范围可以重叠，该组播组只在本管理域内有效，相当于私有组地址。不属于任何管理域的组播组则一律属于 Global 域的服务范围，Global 域中维护一个 BSR，为剩余的所有组播组服务。

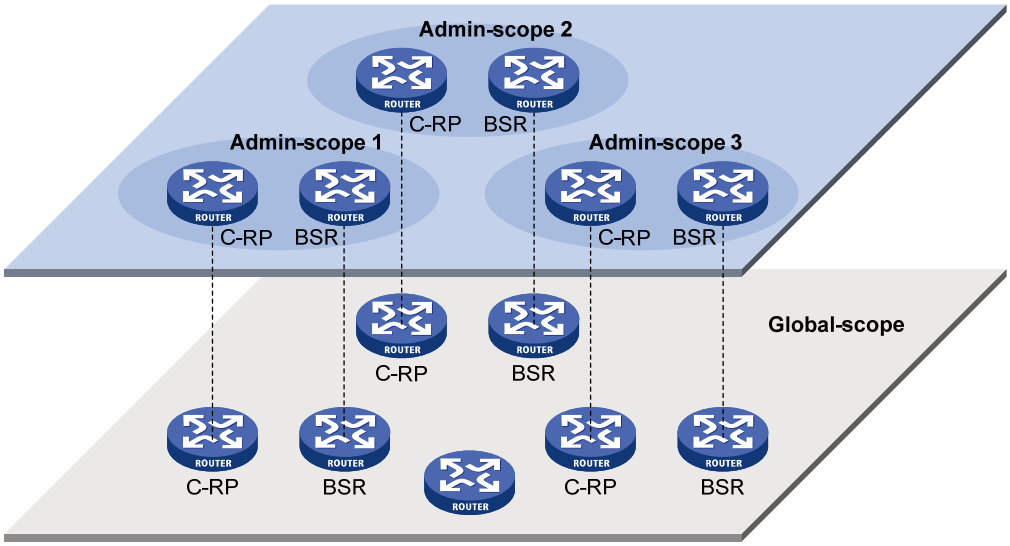
2. 管理域与 Global 域的关系

每个管理域和 Global 域都有独立的 C-RP 和 BSR 设备，这些设备仅在其所属域有效，也就是说 BSR 机制与 RP 选举在各管理域之间是隔离的；每个管理域都有自己的边界，各管理域所服务组播组范围内的组播信息不能进、出该边界。为了更清晰地理解管理域和 Global 域，可以从以下两个角度进行考虑：

(1) 地域空间角度

管理域是针对特定组播组的逻辑管理区域，各管理域在地域上必须相互独立，如 图 1-7所示。

图1-7 地域空间上管理域与 Global 域的关系

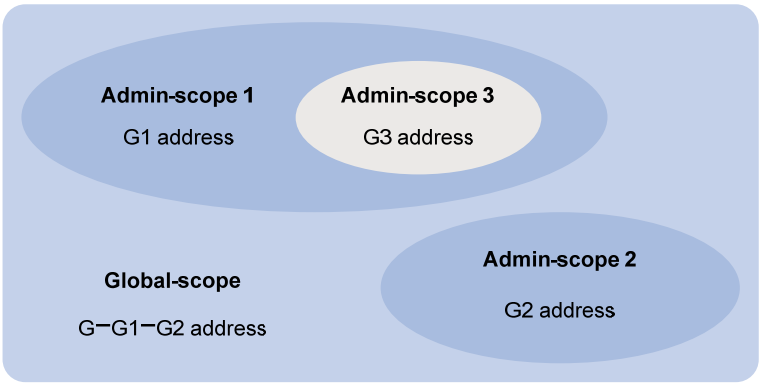


在地域空间上，各管理域之间相互隔离，即同一路由器不能从属于多个管理域。换句话说，各管理域所包含的路由器互不相同。而 Global 域则包含了 PIM-SM 域内的所有路由器。

(2) 组地址范围角度

每个管理域为特定的组播组提供服务，这些组播组地址之间通常没有交集，但是也可能存在相互交叉和重叠关系。

图1-8 组地址范围上管理域与 Global 域的关系



如 图 1-8所示，管理域 1 与管理域 2 所对应的组地址范围无交集，而管理域 3 的组地址是管理域 1 组地址的子集；Global域所对应的组地址范围是除各管理域组地址外的其它所有组地址，即 G−G1−G2。也就是说，Global域和所有管理域之间就组地址范围来说是互补关系。

1.1.6 SSM 模型在 PIM 中的实现

SSM（Source-Specific Multicast，指定信源组播）模型和 ASM（Any-Source Multicast，任意信源组播）模型是两个完全对等的模型。目前，ASM 模型包括 PIM-DM 和 PIM-SM 两种模式，SSM 模型能够借助 PIM-SM 的部分技术来实现。

SSM 模型为指定源组播提供了解决方案，通过 IGMPv3 来维护主机与路由器之间的关系。在实际应用中，通常采用 PIM-SM 模式的一部分技术来实现 SSM 模型。由于接收者已经通过其它渠道（如广告咨询等）知道了组播源的具体位置，因此在 SSM 模型中无需 RP，无需构建 RPT，无需组播

源注册过程，也无需通过 MSDP（Multicast Source Discovery Protocol，组播源发现协议）来发现其它 PIM 域内的组播源。

与 ASM 模型相比，SSM 模型仅需要 IGMPv3 和 PIM-SM 部分子集的支持，SSM 模型在 PIM 域内的工作机制可以简单概括如下：

- 邻居发现
- DR 选举
- 构建 SPT

1. 邻居发现

PIM-SSM使用与PIM-SM完全相同的邻居发现机制，具体请参见“1.1.4 1. 邻居发现”一节。

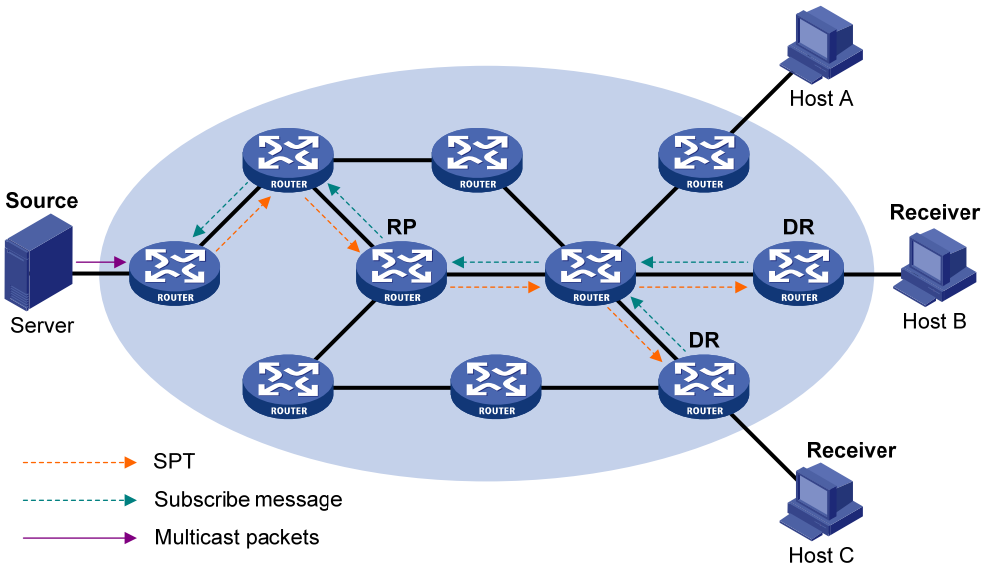
2. DR 选举

PIM-SSM使用与PIM-SM完全相同的DR选举机制，具体请参见 “1.1.4 2. DR选举”一节。

3. 构建 SPT

构建为 PIM-SM 服务的 RPT，还是构建为 PIM-SSM 服务的 SPT，关键在于接收者准备加入的组播组是否属于 SSM 组地址范围（IANA 保留的 SSM 组地址范围为 232.0.0.0/8）。

图1-9 PIM-SSM 中构建 SPT 示意图



如 图 1-9所示，Host B和Host C为组播信息的接收者（Receiver），由其借助IGMPv3 的报告报文向DR报告自己对来自组播源S、发往组播组G的信息感兴趣。收到该报告报文的DR先判断该报文中的组地址是否在SSM组地址范围内：

- 如果在 SSM 组地址范围内，则构建 PIM-SSM，并向组播源 S 逐跳发送通道（Channel）的订阅报文（Subscribe Message）。沿途所有路由器上都创建（S，G）表项，从而在网络内构建了一棵以组播源 S 为根、以接收者为叶子的 SPT，该 SPT 就是 PIM-SSM 中的传输通道；
- 如果不在 SSM 组地址范围内，则仍旧按照 PIM-SM 的流程进行后续处理，此时 DR 需要向 RP 发送（*，G）加入报文，同时需要进行组播源的注册。

说明

在 PIM-SSM 中，借助“通道”的概念表示组播组，借助“订阅报文”的概念表示加入报文。

1.1.7 多实例的 PIM

在多实例应用中，组播路由器需要针对不同的 VPN 实例分别维护 PIM 邻居表、组播路由表、BSR 信息和 RP-Set 信息，并保持多个 VPN 实例间上述信息的相互独立。

当组播路由器收到组播数据报文时，需要区分出该数据报文所属的 VPN 实例，并根据该 VPN 实例对应的组播路由表将其转发，或创建与该 VPN 实例的 PIM 相关的组播路由表项。

1.1.8 协议规范

与 PIM 相关的协议规范有：

- RFC 4601：Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)
- RFC 3973：Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification (Revised)
- RFC 4607：Source-Specific Multicast for IP
- RFC 5059：Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- draft-ietf-ssm-overview-05：An Overview of Source-Specific Multicast (SSM)

1.2 配置 PIM-DM

1.2.1 PIM-DM 配置任务简介

表1-2 PIM-DM 配置任务简介

配置任务	说明	详细配置
使能 PIM-DM	必选	1.2.3
使能状态刷新能力	可选	1.2.4
配置状态刷新参数	可选	1.2.5
配置 PIM-DM 定时器	可选	1.2.6
配置 PIM 公共特性	可选	1.5

1.2.2 配置准备

在配置 PIM-DM 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通

在配置 PIM-DM 之前，需准备以下数据：

- 发送状态刷新报文的时间间隔
- 等待接收新状态刷新报文的最小时间
- 状态刷新报文的 TTL 值

- 嫁接报文的重传时间

1.2.3 使能 PIM-DM

在接口上使能了 PIM-DM 后，路由器会定期发送 Hello 报文以发现 PIM 邻居，并对收到的来自 PIM 邻居的报文进行处理。在部署 PIM-DM 域时，建议在其所有非边界接口上均使能 PIM-DM。

1. 使能公网实例中的 PIM-DM

表1-3 使能公网实例中的 PIM-DM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 PIM-DM	**pim dm**	必选 缺省情况下，PIM-DM 处于关闭状态

2. 使能 VPN 实例中的 PIM-DM

表1-4 使能 VPN 实例中的 PIM-DM

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 PIM-DM	**pim dm**	必选 缺省情况下，PIM-DM 处于关闭状态

注意

- 在同一台设备的同一个 VPN 实例中，所有接口上所启用的 PIM 模式必须相同。
- PIM-DM 不能与处于 SSM 组地址范围内的组播组同时使用。

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.2.4 使能状态刷新能力

只有当一个共享网段内的所有 PIM 路由器上都使能了状态刷新能力时，该共享网段才具备状态刷新能力。

请在 PIM-DM 域内的所有路由器上进行如下配置。

表1-5 使能状态刷新能力

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能状态刷新能力	**pim state-refresh-capable**	可选 缺省情况下，状态刷新能力处于使能状态

1.2.5 配置状态刷新参数

为了避免各路由器上被剪枝的接口因为超时而恢复转发，与组播源直连的路由器会周期性地发送（S，G）状态刷新报文，该报文沿着 PIM-DM 域最初的扩散路径逐跳进行转发，从而刷新沿途所有路由器上的剪枝定时器的状态。

路由器可能在短时间内收到多个状态刷新报文，而其中有些报文可能是重复的。为了避免接收这些重复的报文，可以配置接收新状态刷新报文的等待时间：路由器将丢弃在该时间内收到的状态刷新报文；当该时间超时后，路由器将正常接收新的状态刷新报文，并更新自己的 PIM-DM 状态，同时重置该等待时间。

在收到状态刷新报文时，路由器会将该报文的 TTL 值减 1 后转发给其下游，直至该报文的 TTL 值减为 0，当网络规模很小时，状态刷新报文将在网络中循环传递。因此，为了有效控制刷新报文的传递范围，需要根据网络规模大小配置合适的 TTL 值。

请在 PIM-DM 域内的所有路由器上进行如下配置。

表1-6 配置状态刷新参数

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置发送状态刷新报文的时间间隔	**state-refresh-interval** *interval*	可选 缺省情况下，发送状态刷新报文的时间间隔为 60 秒
配置接收新状态刷新报文的等待时间	**state-refresh-rate-limit** *interval*	可选 缺省情况下，接收新状态刷新报文的等待时间为 30 秒

操作	命令	说明
配置状态刷新报文的 TTL 值	**state-refresh-ttl** *ttl-value*	可选 缺省情况下，状态刷新报文的 TTL 值为 255

1.2.6 配置 PIM-DM 定时器

嫁接报文是 PIM-DM 中唯一使用确认机制的报文。在 PIM-DM 域中，下游路由器发出嫁接报文后，如果在指定时间内没有收到来自其上游路由器的嫁接应答报文，则会重发嫁接报文，直到被确认。

表1-7 配置 PIM-DM 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置嫁接报文的重传时间	**pim timer graft-retry** *interval*	可选 缺省情况下，嫁接报文的重传时间为 3 秒



说明

有关PIM-DM其它定时器的相关配置，请参见“1.5.6 配置PIM公共定时器”。

1.3 配置 PIM-SM

1.3.1 PIM-SM 配置任务简介

表1-8 PIM-SM 配置任务简介

配置任务		说明	详细配置
使能 PIM-SM		必选	1.3.3
配置 RP	配置静态 RP	可选	1.3.4 1.
	配置 C-RP	可选	1.3.4 2.
	使能自动 RP 侦听	可选	1.3.4 3.
	全局配置 C-RP 定时器	可选	1.3.4 4.
配置 BSR	配置 C-BSR	可选	1.3.5 1.
	配置 BSR 服务边界	可选	1.3.5 2.
	全局配置 C-BSR 参数	可选	1.3.5 3.
	配置 C-BSR 定时器	可选	1.3.5 4.
配置管理域	使能管理域机制	可选	1.3.6 1.
	配置管理域边界	可选	1.3.6 2.
	配置各域的 C-BSR	可选	1.3.6 3.

配置任务	说明	详细配置
配置组播源注册	可选	1.3.7
配置 PIM 公共特性	可选	1.5

1.3.2 配置准备

在配置 PIM-SM 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通

在配置 PIM-SM 之前，需准备以下数据：

- 静态 RP 的 IP 地址及表示其所服务组范围的 ACL 规则
- C-RP 的优先级及表示其所服务组范围的 ACL 规则
- 合法 C-RP 的地址范围及表示其所服务组范围的 ACL 规则
- 发送宣告报文的时间间隔
- C-RP 超时时间
- C-BSR 的优先级
- 哈希掩码长度
- 表示合法 BSR 地址范围的 ACL 规则
- 自举时间间隔
- 自举超时时间
- 表示注册报文过滤规则的 ACL 规则
- 注册抑制时间
- 注册探测时间

1.3.3 使能 PIM-SM

在接口上使能了 PIM-SM 后，路由器会定期发送 Hello 报文以发现 PIM 邻居，并对收到的来自 PIM 邻居的报文进行处理。在部署 PIM-SM 域时，建议在其所有非边界接口上均使能 PIM-SM。

1. 使能公网实例中的 PIM-SM

表1-9 使能公网实例中的 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 PIM-SM	**pim sm**	必选 缺省情况下，PIM-SM 处于关闭状态

2. 使能 VPN 实例中的 PIM-SM

表1-10 使能 VPN 实例中的 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
创建VPN实例，并进入VPN实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 PIM-SM	**pim sm**	必选 缺省情况下，PIM-SM 处于关闭状态

注意

在同一台设备的同一个 VPN 实例中，所有接口上所启用的 PIM 模式必须相同。

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.3.4 配置 RP

RP 可以通过手工方式静态配置，也可以通过 BSR 机制动态选举。对于大型 PIM 网络，配置静态 RP 将会非常繁琐。所以，通常静态 RP 是作为动态选举 RP 机制的备份手段，以提高网络的健壮性，增强组播网络的运营管理能力。

1. 配置静态 RP

当网络内仅有一个动态 RP 时，通过手工配置静态 RP 可以避免因单一节点故障而引起的通信中断，同时也可以避免 C-RP 与 BSR 之间频繁的信息交互而占用带宽。

请在 PIM-SM 域内的所有路由器上进行如下配置。

表1-11 配置静态 RP

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-

操作	命令	说明
配置静态 RP	**static-rp** *rp-address* [*acl-number*] [**preferred**]	必选 缺省情况下，没有配置静态 RP

注意

为了让静态 RP 功能正常发挥作用，必须在所有路由器上指定相同的静态 RP 地址。

2. 配置 C-RP

在 PIM-SM 域中，可以把有意成为 RP 的路由器配置为 C-RP。BSR 通过接收来自 C-RP 的 C-RP 信息，或者接收来自其它路由器的自动 RP 宣告，收集 C-RP 信息并将其汇总为 RP-Set 信息，然后在全网内扩散。之后，网络内的其它路由器根据 RP-Set 信息计算出特定组播组范围所对应的 RP。建议在骨干网路由器上配置 C-RP。

为了防止 C-RP 欺骗，需要在 BSR 上配置合法的 C-RP 地址范围及其所服务的组播组范围。同时由于每个 C-BSR 都可能成为 BSR，因此必须在 PIM-SM 域内的所有 C-BSR 上都配置相同的过滤策略。

表1-12 配置 C-RP

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置某接口为 C-RP	**c-rp** *interface-type interface-number* [**group-policy** *acl-number* | **priority** *priority* | **holdtime** *hold-interval* | **advertisement-interval** *adv-interval*] *	必选 缺省情况下，没有配置 C-RP
配置合法 C-RP 的地址范围及其所服务的组播组范围	**crp-policy** *acl-number*	可选 缺省情况下，C-RP 的地址范围及其所服务的组播组范围不受任何限制

说明

- 在配置 C-RP 时，应在 C-RP 与 PIM-SM 域中的其它设备之间保留较大的通信带宽。
- 一个 RP 可以为多个组播组服务，也可以为所有组播组服务。每个组播组在任意时刻，只能由唯一的一个 RP 为其转发数据，而不能由多个 RP 转发数据。

3. 使能自动 RP 侦听

自动 RP 宣告（Announce）和发现（Discovery）报文的目的地址分别为组播组地址 224.0.1.39 和 224.0.1.40。在设备上使能了自动 RP 侦听功能后，该设备便能够接收这两种报文，并记录报文中所携带的 RP 信息。

表1-13 使能自动 RP 侦听

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
使能自动 RP 侦听功能	**auto-rp enable**	必选 缺省情况下，自动 RP 侦听功能处于关闭状态

4. 全局配置 C-RP 定时器

为了使 BSR 能够在 PIM-SM 域内分发 RP-Set 信息，C-RP 必须周期性地向 BSR 发送宣告报文，BSR 从该报文中学习 RP-Set 信息，并将该信息与自己的 IP 地址一起封装在自举报文中向域中的所有 PIM 路由器（224.0.0.13）进行宣告。

C-RP 在其宣告报文中封装一个保持时间，BSR 在收到该报文后，从中获得该时间值并启动 C-RP 超时定时器，如果超时后 BSR 仍没有收到来自 C-RP 后续的宣告报文，则认为目前网络中的 C-RP 失效或不可达。

请在已配置为 C-RP 的路由器上进行如下配置。

表1-14 全局配置 C-RP 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置发送宣告报文的时间间隔	**c-rp advertisement-interval** *interval*	可选 缺省情况下，发送宣告报文的时间间隔为 60 秒
配置 C-RP 超时时间	**c-rp holdtime** *interval*	可选 缺省情况下，C-RP 的超时时间为 150 秒



说明

有关PIM-SM其它定时器的相关配置，请参见“1.5.6 配置PIM公共定时器”。

1.3.5 配置 BSR

在一个 PIM-SM 域中只能有一个 BSR，但需要配置至少一个 C-BSR。任意一台路由器都可以被配置为 C-BSR。在 C-BSR 之间通过自动选举产生 BSR，BSR 负责在 PIM-SM 域中收集并发布 RP 信息。

1. 配置 C-BSR

C-BSR 应配置在骨干网的路由器上，在将路由器配置为 C-BSR 时，必须同时指定一个使能了 PIM-SM 的接口。C-BSR 间的自动选举机制简单描述如下：

- 最初，每个 C-BSR 都认为自己是本 PIM-SM 域的 BSR，并使用接口的 IP 地址作为 BSR 地址，发送自举报文；

- 当某 C-BSR 收到其它 C-BSR 发来的自举报文时，首先比较自己与后者的优先级，优先级较高者获胜；在优先级相同的情况下，再比较自己与后者的 BSR 地址，拥有较大 IP 地址者获胜。如果后者获胜，则用后者的 BSR 地址替换自己的 BSR 地址，并不再认为自己是 BSR；否则，保留自己的 BSR 地址，并继续认为自己是 BSR。

通过在路由器上配置合法 BSR 的地址范围，可以对收到的 BSR 报文按照地址范围进行过滤，从而防止某些恶意主机非法伪装成 BSR，以避免合法的 BSR 被恶意取代。必须在 PIM-SM 域内的所有路由器上进行相同的配置。通常针对以下两类情况实施预防措施：

(1) 某些恶意主机通过伪造 BSR 报文以欺骗路由器，试图更改 RP 映射关系。这种攻击通常发生在边缘路由器上，由于 BSR 处于网络内部，主机在网络外部，因此边缘路由器通过对收到的 BSR 报文进行邻居检查和 RPF 检查，丢弃不符合要求的报文，就可以避免外部网络用户对内部网络 BSR 的攻击；

(2) 网络中某台路由器被攻击者控制，或者有非法接入的路由器时，攻击者可以将这样的路由器配置为 C-BSR，并使其在竞争中获胜，从而控制网络中 RP 信息的发布权。由于在被配置为 C-BSR 后，路由器会自动向整个网络扩散 BSR 报文，而 BSR 报文是 TTL 值为 1 的组播报文，所以只要其邻居路由器不接收该 BSR 报文，就不会影响整个网络。因此，通过在整个网络的所有路由器上都配置合法 BSR 的地址范围，从而丢弃合法范围之外的 BSR 报文，就可以防止此类攻击。

以上两种预防策略可以部分地保护网络中 BSR 的安全。但是如果某台合法的 BSR 路由器被攻击者控制，还是可能导致问题。

表1-15 配置 C-BSR

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或VPN实例PIM视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置某接口为 C-BSR	**c-bsr** *interface-type interface-number* [*hash-length* [*priority*]]	必选 缺省情况下，没有配置 C-BSR
配置合法的BSR地址范围	**bsr-policy** *acl-number*	可选 缺省情况下，BSR 的地址范围不受任何限制

说明

- 由于 BSR 与 PIM-SM 域中的其它设备需要交换大量信息，因此应在 C-BSR 与 PIM-SM 域中的其它设备之间保留较大的通信带宽。
- 如果 C-BSR 之间通过 GRE（Generic Routing Encapsulation，通用路由封装）隧道相连接，则需要配置组播静态路由，以保证到 C-BSR 下一跳的是 GRE 口。有关组播静态路由的相关配置，请参见“IP 组播分册”中的“组播路由与转发配置”。

2. 配置 BSR 服务边界

BSR 作为 PIM-SM 域中的管理核心，负责将收集到的 RP-Set 信息以自举报文的形式发向 PIM-SM 域中的所有路由器。

BSR 的服务边界，即 PIM-SM 域的边界。BSR 是针对特定的服务范围而言的，众多的 BSR 服务边界接口将网络划分成不同的 PIM-SM 域，自举报文无法通过 PIM-SM 域的边界，BSR 服务边界之外的路由器也不能参与本 PIM-SM 域内的组播转发。

请在有待成为 BSR 服务边界的路由器上进行如下配置。

表1-16 配置 BSR 服务边界

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 BSR 的服务边界	**pim bsr-boundary**	必选 缺省情况下，没有配置 BSR 的服务边界

3. 全局配置 C-BSR 参数

在一个 PIM-SM 域中，从众多 C-BSR 中选举出唯一的 BSR。PIM-SM 域内的 C-RP 向 BSR 发送宣告报文，由 BSR 汇总为 RP-Set，并向本 PIM-SM 域内的所有路由器进行宣告。所有路由器都使用统一的哈希算法，得到特定组播组所对应 RP 的地址。

请在已配置为 C-BSR 的路由器上进行如下配置。

表1-17 全局配置 C-BSR 参数

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置哈希掩码长度	**c-bsr hash-length** *hash-length*	可选 缺省情况下，哈希掩码长度为 30
配置 C-BSR 的优先级	**c-bsr priority** *priority*	可选 缺省情况下，C-BSR 的优先级为 0



说明

关于哈希掩码长度和 C-BSR 的优先级这两个值：

- 可以在三种域范围内进行配置：全局、管理域和 Global 域；
- 如果在管理域或 Global 域下进行了配置，则取其配置值；
- 如果未在管理域或 Global 域下进行配置，则取相应的全局值。

有关管理域和Global域C-BSR参数的相关配置，请参见“1.3.6 3. 配置各域的C-BSR”。

4. 配置 C-BSR 定时器

当某 C-BSR 竞选成为 BSR 后，它会通过自举报文向其所服务的区域以组播方式发送自己的 IP 地址和 RP-Set 信息。BSR 以自举时间间隔周期性地向网络发送自举报文，收到该报文的 C-BSR 会在自举超时时间内将其保持，此时 BSR 选举过程暂停；当自举超时时间超时后，C-BSR 之间会触发新一轮的 BSR 选举过程。

请在已配置为 C-BSR 的路由器上进行如下配置。

表1-18 配置 C-BSR 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置自举时间间隔	**c-bsr interval** *interval*	可选 缺省取值请参见 表 1-18后的相关说明
配置自举超时时间	**c-bsr holdtime** *interval*	可选 缺省取值请参见 表 1-18后的相关说明



说明

关于自举时间间隔的取值：

- 如果没有进行手工配置，则其取值由如下公式决定：自举时间间隔 =（自举超时时间 − 10）÷ 2。缺省情况下，自举超时时间为 130 秒，则自举时间间隔的缺省值 =（130 − 10）÷ 2 = 60（秒）。
- 如果进行了手工配置，则取配置值。

关于自举超时时间的取值：

- 如果没有进行手工配置，则其取值由如下公式决定：自举超时时间 = 自举时间间隔 × 2 + 10。缺省情况下，自举时间间隔为 60 秒，则自举超时时间的缺省值 = 60 × 2 + 10 = 130（秒）。
- 如果进行了手工配置，则取配置值。



注意

配置时，应保证自举时间间隔小于自举超时时间。

1.3.6 配置管理域

在非管理域机制下，一个 PIM-SM 域中只有唯一的 BSR，整个网络都在该 BSR 的管理范围之内。为了更有针对性地管理，可以将整个 PIM-SM 域划分为多个管理域：每个管理域中各维护一个 BSR，服务于特定范围的组播组；而 Gloab 域中也维护一个 BSR，为所有剩余的组播组提供服务。

1. 使能管理域机制

在配置管理域各项功能之前，必须先使能管理域机制。

请在 PIM-SM 域内的所有路由器上进行如下配置。

表1-19 使能管理域机制

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或VPN实例PIM视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
使能管理域机制	**c-bsr admin-scope**	必选 缺省情况下，管理域机制处于关闭状态

2. 配置管理域边界

在管理域机制中，各管理域的边界由ZBR构成，每个管理域各维护一个BSR，服务于特定范围的组播组，属于此范围的组播协议报文（如断言报文、BSR自举报文等）无法通过管理域边界。

请在有待成为ZBR的路由器上进行如下配置。

表1-20 配置管理域边界

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置组播转发边界	**multicast boundary** *group-address* { *mask* | *mask-length* }	必选 缺省情况下，没有配置组播转发边界

说明

有关 **multicast boundary** 命令的详细介绍，请参见“IP组播分册”中的“组播路由与转发命令”。

3. 配置各域的C-BSR

在应用了管理域机制的网络中，从众多C-BSR中选举出针对不同组播组的BSR。网络内的C-RP只向对应的BSR发送宣告报文，由BSR汇总为RP-Set，并向其所服务管理域内的所有路由器进行宣告。所有路由器都使用统一的哈希算法，得到特定组播组所对应RP的地址。

请在有待成为管理域或Global域C-BSR的路由器上进行如下配置。

表1-21 配置各域的C-BSR

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或VPN实例PIM视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置管理域的C-BSR	**c-bsr group** *group-address* { *mask* | *mask-length* } [**hash-length** *hash-length* | **priority** *priority*] *	必选 缺省情况下，没有配置管理域的C-BSR
配置Global域的C-BSR	**c-bsr global** [**hash-length** *hash-length* | **priority** *priority*] *	必选 缺省情况下，没有配置Global域的C-BSR

说明

关于哈希掩码长度和 C-BSR 的优先级这两个值：

- 可以在三种域范围内进行配置：全局、管理域和 Global 域；
- 如果在管理域或 Global 域下进行了配置，则取其配置值；
- 如果未在管理域或 Global 域下进行配置，则取相应的全局值。

有关全局C-BSR参数的相关配置，请参见“1.3.5 3. 全局配置C-BSR参数”。

1.3.7 配置组播源注册

在 PIM-SM 域内，组播源侧 DR 向 RP 发送注册报文，而这些注册报文拥有不同的组播源或组播组地址。为了让 RP 服务于特定的组播组，可以对注册报文进行过滤。如果某个（S，G）表项被过滤规则拒绝，或者过滤规则中没有定义对它的操作，RP 都会向 DR 发送注册停止报文，以停止该组播数据的注册过程。

出于对注册报文在传递过程中完整性的考虑，可以配置根据整个报文来计算校验和。但为了减少往注册报文中封装数据报文的工作量并考虑到互通性，一般情况下不建议配置根据注册报文的全部内容来计算校验和的方式。

当接收者不再通过 RP 接收发往某组播组的数据（即 RP 不再服务于该组播组），或 RP 正式开始接收来自组播源的组播数据时，RP 将向组播源侧 DR 发送注册停止报文，DR 收到该报文后将停止发送封装有组播数据的注册报文并启动注册停止定时器（Register-Stop Timer）。当注册停止定时器超时后，DR 会向 RP 发送一个空注册报文（Null-Register Message，即不封装组播数据的注册报文）：如果 DR 在注册探测时间（Register_Probe_Time）内收到了来自 RP 的注册停止报文，DR 将刷新其注册停止定时器；否则，DR 将重新开始发送封装有组播数据的注册报文。

注册停止定时器的超时时间是一个随机值，由其它两个时间值决定：注册抑制时间（Register_Suppression_Time）和注册探测时间。其具体取值范围如下：（0.5×注册抑止时间，1.5×注册抑止时间）－注册探测时间。

请在所有已配置为 C-RP 的路由器上配置注册报文的过滤规则和根据注册报文的全部内容来计算校验和；请在所有可能成为组播源侧 DR 的路由器上配置注册抑制时间和注册探测时间。

说明

SR6600 路由器只有第一个注册报文可以携带组播数据。

表1-22 配置组播源注册

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置注册报文的过滤规则	**register-policy** *acl-number*	可选 缺省情况下，没有配置注册报文的过滤规则
配置根据注册报文的全部内容来计算校验和	**register-whole-checksum**	可选 缺省情况下，仅根据注册报文头来计算校验和

操作	命令	说明
配置注册抑制时间	**register-suppression-timeout** *interval*	可选 缺省情况下，注册抑制时间为 60 秒
配置注册探测时间	**probe-interval** *interval*	可选 缺省情况下，注册探测时间为 5 秒

1.4 配置 PIM-SSM

说明

PIM-SSM 模型需要 IGMPv3 的支持，因此应确保连接有接收者的 PIM 路由器上使能了 IGMPv3。

1.4.1 PIM-SSM 配置任务简介

表1-23 PIM-SSM 配置任务简介

配置任务	说明	详细配置
使能 PIM-SM	必选	1.4.3
配置 SSM 组播组范围	可选	1.4.4
配置 PIM 公共特性	可选	1.5

1.4.2 配置准备

在配置 PIM-SSM 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通

在配置 PIM-SSM 之前，需准备以下数据：

- SSM 组播组范围

1.4.3 使能 PIM-SM

由于 SSM 模型是通过 PIM-SM 的部分子集功能实现的，因此系统在使能了 PIM-SM 的同时也具备了 PIM-SSM 能力。在部署 PIM-SM 域时，建议在其所有非边界接口上均使能 PIM-SM。

1. 使能公网实例中的 PIM-SM

表1-24 使能公网实例中的 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能 PIM-SM	**pim sm**	必选 缺省情况下，PIM-SM 处于关闭状态

2. 使能 VPN 实例中的 PIM-SM

表1-25 使能 VPN 实例中的 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 PIM-SM	**pim sm**	必选 缺省情况下，PIM-SM 处于关闭状态

注意

在同一台设备的同一个 VPN 实例中，所有接口上所启用的 PIM 模式必须相同。

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.4.4 配置 SSM 组播组范围

在把来自组播源的信息传递给接收者的过程中，是采用 PIM-SSM 模型还是 PIM-SM 模型，这取决于接收者订阅通道（S，G）中的组播组是否在 SSM 组播组范围之内，所有使能了 PIM-SM 的接口将会认为属于该范围内的组播组采用了 PIM-SSM 模型。

请在 PIM-SM 域内的所有路由器上进行如下配置。

表1-26 配置 SSM 组播组范围

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-

操作	命令	说明
配置 SSM 组播组的范围	**ssm-policy** *acl-number*	可选 缺省情况下，SSM 组播组的范围为 232.0.0.0/8

注意

- 应确保域内所有路由器上配置的 SSM 组播组地址范围都一致，否则组播信息将无法通过 SSM 模型进行传输。
- 如果某组播组属于 SSM 组播组范围，但该组成员使用 IGMPv1 或 IGMPv2 发送加入报文，则设备不会触发（*，G）加入报文。

1.5 配置 PIM 公共特性

说明

对于本节中那些既可在 PIM 视图、又可在接口视图下进行配置的功能或参数来说：

- 在 PIM 视图下所作的配置对所有接口生效，在接口视图下所作的配置仅对当前接口生效；
- 如果在这两个视图下对某功能或参数均进行了配置，则不论配置先后，接口视图下的配置将被优先采用。

1.5.1 PIM 公共特性配置任务简介

表1-27 PIM 公共特性配置任务简介

配置任务	说明	详细配置
配置组播数据过滤器	可选	1.5.3
配置 Hello 报文选项	可选	1.5.5
配置 PIM 公共定时器	可选	1.5.6
配置加入/剪枝报文规格	可选	1.5.7

1.5.2 配置准备

在配置 PIM 公共特性之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 PIM-DM（或 PIM-SM、PIM-SSM）

在配置 PIM 公共特性之前，需准备以下数据：

- 表示组播数据过滤规则的 ACL 规则
- 竞选 DR 的优先级（全局值/接口值）
- 保持 PIM 邻居可达状态的时间（全局值/接口值）
- 剪枝延迟时间（全局值/接口值）
- 剪枝否决时间（全局值/接口值）

- 发送 Hello 报文的时间间隔（全局值/接口值）
- 发送 Hello 报文的最大延迟时间（接口值）
- 保持断言状态的时间（全局值/接口值）
- 发送加入/剪枝报文的时间间隔（全局值/接口值）
- 保持加入/剪枝状态的时间（全局值/接口值）
- 组播源生存时间
- 加入/剪枝报文的最大长度
- 加入/剪枝报文中（S，G）表项的最大数量

1.5.3 配置组播数据过滤器

无论在 PIM-DM 还是 PIM-SM 域内，各路由器都可以对流经自己的组播数据进行检查，通过比较是否符合过滤规则，从而决定是否继续转发组播数据。也就是说 PIM 域内的路由器能够成为组播数据的过滤器。过滤器的存在一方面有助于实现信息流量控制，另一方面可以在安全性方面限定下游接收者能够获得的信息。

表1-28 配置组播数据过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim [vpn-instance** *vpn-instance-name*]	-
配置组播数据过滤器	**source-policy** *acl-number*	必选 缺省情况下，没有配置组播数据过滤器

说明

- 通常，过滤器的位置离组播源距离越近，过滤影响越明显。
- 过滤器不仅过滤独立的组播数据，还过滤封装在注册报文中的组播数据。

1.5.4 配置 Hello 报文过滤器

随着 PIM 协议的推广和应用，对其安全性的要求也越来越高。建立正确的 PIM 邻居是 PIM 协议安全应用的前提。如果在接口上指定了合法 Hello 报文的源地址范围，便能够保证 PIM 邻居的正确建立，从而有效防止各种 PIM 协议报文攻击，提高设备对 PIM 协议报文处理的安全性。

表1-29 配置 Hello 报文过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置合法 Hello 报文的源地址范围	**pim neighbor-policy** *acl-number*	必选 缺省情况下，Hello 报文的源地址范围不受任何限制

说明

当 Hello 报文过滤器的配置生效后，对于之前已建立的 PIM 邻居，若由于其 Hello 报文被过滤而导致无法收到后续的 Hello 报文，将会在老化超时后被自动删除。

1.5.5 配置 Hello 报文选项

无论在 PIM-DM 域还是在 PIM-SM 域内，各路由器之间发送的 Hello 报文都包含很多可供配置的选项，对各选项的介绍如下：

- DR_Priority（仅用于 PIM-SM）：表示竞选 DR 的优先级，优先级高的设备被选举为 DR。可以在与组播源或接收者直连的共享网络中的所有路由器上都配置此参数。
- Holdtime：表示保持 PIM 邻居可达状态的时间，若超时后仍没有收到 Hello 报文，则认为 PIM 邻居失效或不可达。
- LAN_Prune_Delay：表示在共享网络上传递剪枝报文的延迟时间，该选项由三部分组成：Lan-delay（剪枝延迟时间）、Override-interval（剪枝否决时间）和禁止加入报文抑制能力。当共享网段内各 PIM 路由器的剪枝延迟或否决时间值不同时，取其中最大的值；当要禁止加入报文抑制能力时，须在共享网段内的所有 PIM 路由器上都禁止该能力。

剪枝延迟时间将使上游路由器延迟处理收到的剪枝报文。如果配置的剪枝延迟时间过小，可能导致下游路由器还未否决剪枝动作，上游路由器就已经停止转发组播数据，因此配置时需要慎重。

剪枝否决时间表示允许下游路由器否决剪枝动作的时间。路由器在收到下游路由器发来的剪枝报文后并不立即执行剪枝动作，而是仍然将当前的转发状态保持剪枝延迟时间＋剪枝否决时间。如果下游路由器需要继续接收组播数据，则必须在剪枝否决时间内否决剪枝动作；否则上游路由器会在剪枝延迟时间＋剪枝否决时间超时后执行剪枝动作。

PIM 路由器发送 Hello 报文时，会生成一个随机数 Generation ID 并携带在该报文中。一台 PIM 路由器的 Generation ID 一般不会变化，除非该路由器的状态发生了改变（如接口刚使能了 PIM 或设备进行了重启），此时，当其开始或重新开始发送 Hello 报文时，会生成一个新的 Generation ID。这样，如果 PIM 路由器发现其上游邻居发来的 Hello 报文中 Generation ID 发生了改变，就会认为该邻居的状态已经丢失或其上游邻居已经改变，从而触发发送加入报文以进行状态刷新。

在禁止加入报文抑制能力（即使能邻居跟踪）时，应在共享网段的所有 PIM 路由器上都禁止该能力，否则上游路由器无法跟踪每台下游路由器的加入报文。

1. 全局配置 Hello 报文选项

表1-30 全局配置 Hello 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置竞选 DR 的优先级	**hello-option dr-priority** *priority*	可选 缺省情况下，竞选 DR 的优先级为 1
配置保持 PIM 邻居可达状态的时间	**hello-option holdtime** *interval*	可选 缺省情况下，保持 PIM 邻居可达状态的时间为 105 秒

操作	命令	说明
配置剪枝延迟时间	**hello-option lan-delay** *interval*	可选 缺省情况下，剪枝延迟时间为 500 毫秒
配置剪枝否决时间	**hello-option override-interval** *interval*	可选 缺省情况下，剪枝否决时间为 2500 毫秒
禁止加入报文抑制能力	**hello-option neighbor-tracking**	必选 缺省情况下，加入报文抑制能力处于使能状态

2. 在接口上配置 Hello 报文选项

表1-31 在接口上配置 Hello 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置竞选 DR 的优先级	**pim hello-option dr-priority** *priority*	可选 缺省情况下，竞选 DR 的优先级为 1
配置保持 PIM 邻居可达状态的时间	**pim hello-option holdtime** *interval*	可选 缺省情况下，保持 PIM 邻居可达状态的时间为 105 秒
配置剪枝延迟时间	**pim hello-option lan-delay** *interval*	可选 缺省情况下，剪枝延迟时间为 500 毫秒
配置剪枝否决时间	**pim hello-option override-interval** *interval*	可选 缺省情况下，剪枝否决时间为 2500 毫秒
禁止加入报文抑制能力	**pim hello-option neighbor-tracking**	必选 缺省情况下，加入报文抑制能力处于使能状态
配置不接受无 Generation ID 的 Hello 报文	**pim require-genid**	必选 缺省情况下，接受无 Generation ID 的 Hello 报文

1.5.6 配置 PIM 公共定时器

PIM 路由器通过周期性地发送 Hello 报文，以发现 PIM 邻居，并维护各路由器之间的 PIM 邻居关系。

为了避免多个 PIM 路由器同时发送 Hello 报文而导致冲突，当 PIM 路由器在收到 Hello 报文时，将延迟一段时间再发送 Hello 报文，该时间值为小于“触发 Hello 报文的最大延迟时间”的一个随机值。

PIM 路由器通过周期性地向其上游路由器发送加入/剪枝报文以更新状态，在该报文中携带有保持时间，上游路由器为被剪枝的下游接口设置保持加入/剪枝状态定时器。

在断言中落选的路由器将会剪掉其下游的转发接口，并把这种断言状态保持一段时间。超时后，落选的路由器会重新恢复转发组播数据。

当路由器没有收到来自组播源 S 的后续组播数据时，不会立即删除（S，G）表项，而是将其维持一段时间后再删除，这段时间就称为组播源的生存时间。

1. 全局配置 PIM 公共定时器

表1-32 全局配置 PIM 公共定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置发送 Hello 报文的时间间隔	**timer hello** *interval*	可选 缺省情况下，发送 Hello 报文的时间间隔为 30 秒
配置发送加入/剪枝报文的时间间隔	**timer join-prune** *interval*	可选 缺省情况下，发送加入/剪枝报文的时间间隔为 60 秒
配置保持加入/剪枝状态的时间	**holdtime join-prune** *interval*	可选 缺省情况下，保持加入/剪枝状态的时间为 210 秒
配置保持断言状态的时间	**holdtime assert** *interval*	可选 缺省情况下，保持断言状态的时间为 180 秒
配置组播源生存时间	**source-lifetime** *interval*	可选 缺省情况下，组播源的生存时间为 210 秒

2. 在接口上配置 PIM 公共定时器

表1-33 在接口上配置 PIM 公共定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置发送 Hello 报文的时间间隔	**pim timer hello** *interval*	可选 缺省情况下，发送 Hello 报文的时间间隔为 30 秒
配置触发 Hello 报文的最大延迟时间	**pim triggered-hello-delay** *interval*	可选 缺省情况下，触发 Hello 报文的最大延迟时间为 5 秒
配置发送加入/剪枝报文的时间间隔	**pim timer join-prune** *interval*	可选 缺省情况下，发送加入/剪枝报文的时间间隔为 60 秒
配置保持加入/剪枝状态的时间	**pim holdtime join-prune** *interval*	可选 缺省情况下，保持加入/剪枝状态的时间为 210 秒
配置保持断言状态的时间	**pim holdtime assert** *interval*	可选 缺省情况下，保持断言状态的时间为 180 秒



说明

如果网络没有特殊要求，建议采用缺省值。

1.5.7 配置加入/剪枝报文规格

如果加入/剪枝报文的尺寸较大，则丢失一个报文将导致较多信息的遗失；如果加入/剪枝报文的尺寸较小，则单个报文的丢失所产生的影响也将降低。

通过控制加入/剪枝报文中（S，G）表项的数目，可以有效减少单位时间内发送的（S，G）表项数量。

表1-34 配置加入/剪枝报文规格

操作	命令	说明
进入系统视图	**system-view**	-
进入公网实例或 VPN 实例 PIM 视图	**pim** [**vpn-instance** *vpn-instance-name*]	-
配置加入/剪枝报文的最大长度	**jp-pkt-size** *packet-size*	可选 缺省情况下，加入/剪枝报文的最大长度为 8100 字节
配置加入/剪枝报文中（S，G）表项的最大数量	**jp-queue-size** *queue-size*	可选 缺省情况下，加入/剪枝报文中（S，G）表项的最大数量为 1020 个

1.6 PIM 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 PIM 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 PIM 统计信息。

表1-35 PIM 显示和维护

操作	命令
查看 PIM-SM 域中的 BSR 信息，以及本地配置并生效的 C-RP 信息	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **bsr-info**
查看 PIM 所使用的单播路由信息	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **claimed-route** [*source-address*]
查看 PIM 控制报文的数量	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **control-message counters** [**message-type** { **probe** | **register** | **register-stop** } | [**interface** *interface-type interface-number* | **message-type** { **assert** | **bsr** | **crp** | **graft** | **graft-ack** | **hello** | **join-prune** | **state-refresh** }] *]
查看尚未被确认的嫁接信息	**display pim** [**all-instance** | **vpn-instance** *vpn-instance-name*] **grafts**
查看接口上的 PIM 信息	**display pim** [**all-instance** | **vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*] [**verbose**]
查看待发送的加入/剪枝报文信息	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **join-prune mode** { **sm** [**flags** *flag-value*] | **ssm** } [**interface** *interface-type interface-number* | **neighbor** *neighbor-address*] * [**verbose**]
查看 PIM 邻居信息	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **neighbor** [**interface** *interface-type interface-number* | *neighbor-address* | **verbose**] *

操作	命令
查看 PIM 路由表的内容	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **routing-table** [*group-address* [**mask** { *mask-length* | *mask* }] | *source-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** [*interface-type interface-number* | **register**] | **outgoing-interface** { **include** | **exclude** | **match** } { *interface-type interface-number* | **register** } | **mode** *mode-type* | **flags** *flag-value* | **fsm**] *
查看 RP 的信息	**display pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **rp-info** [*group-address*]
重置 PIM 控制报文计数器	**reset pim** [**vpn-instance** *vpn-instance-name* | **all-instance**] **control-message counters** [**interface** *interface-type interface-number*]

1.7 PIM 典型配置举例

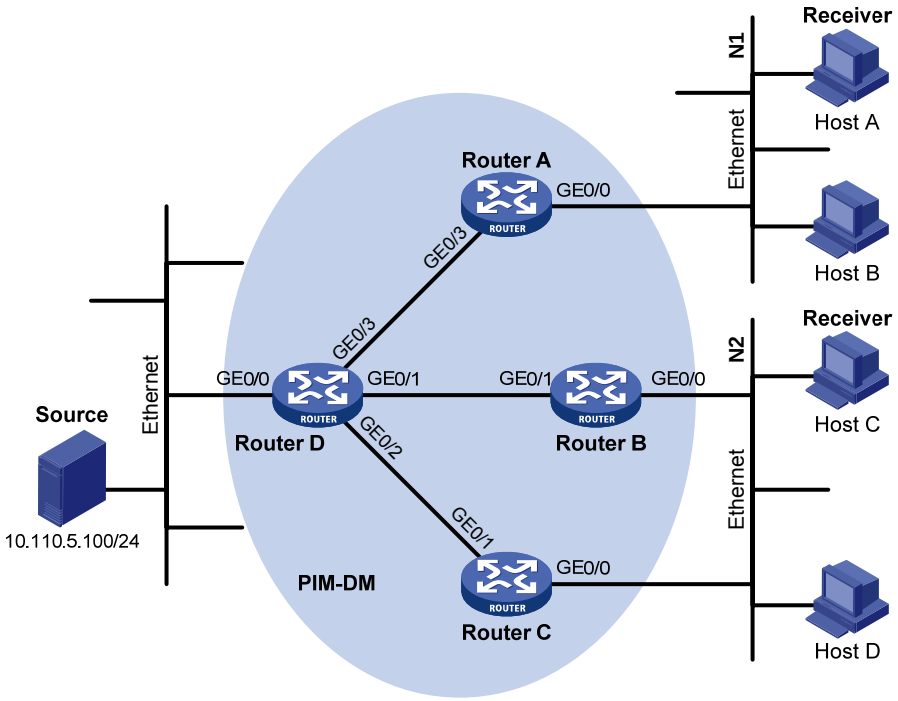
1.7.1 PIM-DM 典型配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 PIM 域采用 DM 方式。
- Host A 和 Host C 为两个末梢网络中的组播信息接收者；Router D 通过 GigabitEthernet0/0 接口与组播源（Source）所在的网络连接；Router A 通过 GigabitEthernet0/0 接口连接末梢网络 N1，通过 GE0/3 接口连接 Router D；Router B 和 Router C 分别通过各自的 GigabitEthernet0/0 接口连接末梢网络 N2，并分别通过各自的 GigabitEthernet0/1 接口连接 Router D。
- Router A 与末梢网络 N1 之间运行 IGMPv2；Router B 和 Router C 与末梢网络 N2 之间也运行 IGMPv2。

2. 组网图

图1-10 PIM-DM 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/0	10.110.1.1/24	Router D	GE0/0	10.110.5.1/24
	GE0/3	192.168.1.1/24		GE0/3	192.168.1.2/24
Router B	GE0/0	10.110.2.1/24		GE0/1	192.168.2.2/24
	GE0/1	192.168.2.1/24		GE0/2	192.168.3.2/24
Router C	GE0/0	10.110.2.2/24			
	GE0/1	192.168.3.1/24			

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-10配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-DM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-DM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-DM 和 IGMP

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-DM，并在其连接末梢网络的接口 GigabitEthernet0/0 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/0
[RouterA-GigabitEthernet0/0] igmp enable
[RouterA-GigabitEthernet0/0] pim dm
[RouterA-GigabitEthernet0/0] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] pim dm
[RouterA-GigabitEthernet0/3] quit
```

Router B 和 Router C 的配置与 Router A 相似，配置过程略。

在 Router D 上使能 IP 组播路由，并在各接口上使能 PIM-DM。

```
<RouterD> system-view
[RouterD] multicast routing-enable
[RouterD] interface GigabitEthernet 0/0
[RouterD-GigabitEthernet0/0] pim dm
[RouterD-GigabitEthernet0/0] quit
[RouterD] interface GigabitEthernet 0/3
[RouterD-GigabitEthernet0/3] pim dm
[RouterD-GigabitEthernet0/3] quit
[RouterD] interface GigabitEthernet 0/1
[RouterD-GigabitEthernet0/1] pim dm
[RouterD-GigabitEthernet0/1] quit
[RouterD] interface GigabitEthernet 0/2
[RouterD-GigabitEthernet0/2] pim dm
[RouterD-GigabitEthernet0/2] quit
```

(3) 检验配置效果

通过使用 **display pim interface** 命令可以查看路由器接口上 PIM 的配置和运行情况。例如：

查看 Router D 上 PIM 的配置信息。

```
[RouterD] display pim interface
 VPN-Instance: public net
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
GE0/0              0      30         1          10.110.5.1     (local)
GE0/3              1      30         1          192.168.1.2    (local)
GE0/1              1      30         1          192.168.2.2    (local)
GE0/2              1      30         1          192.168.3.2    (local)
```

通过使用 **display pim neighbor** 命令可以查看路由器之间的 PIM 邻居关系。例如：

查看 Router D 上 PIM 的邻居关系信息。

```
[RouterD] display pim neighbor
 VPN-Instance: public net
 Total Number of Neighbors = 3

 Neighbor         Interface           Uptime   Expires  Dr-Priority
 192.168.1.1      GE0/3              00:02:22 00:01:27 1
 192.168.2.1      GE0/1              00:00:22 00:01:29 3
 192.168.3.1      GE0/2              00:00:23 00:01:31 5
```

假如 Host A 需要接收组播组 G（225.1.1.1）的信息，当组播源 S（10.110.5.100/24）向组播组 G 发送组播数据时，通过扩散生成 SPT，SPT 路径中各路由器（Router A 和 Router D）上都存在（S，G）表项，Host A 向 Router A 发送 IGMP 报告以加入组播组 G，在 Router A 上生成（*，G）表项。

通过使用 **display pim routing-table** 命令可以查看路由器的 PIM 路由表信息。例如：

查看 Router A 上的 PIM 路由表信息。

```
[RouterA] display pim routing-table
 VPN-Instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry
```

```
 (*, 225.1.1.1)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:04:25
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet0/0
             Protocol: igmp, UpTime: 00:04:25, Expires: never

 (10.110.5.100, 225.1.1.1)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:06:14
     Upstream interface: GE0/3,
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet0/0
             Protocol: pim-dm, UpTime: 00:04:25, Expires: never
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 PIM 路由表信息。

```
[RouterD] display pim routing-table
 VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.1)
     Protocol: pim-dm, Flag: LOC ACT
     UpTime: 00:03:27
     Upstream interface: GigabitEthernet0/0
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 3
         1: GE0/3
             Protocol: pim-dm, UpTime: 00:03:27, Expires: never
         2: GE0/1
             Protocol: pim-dm, UpTime: 00:03:27, Expires: never
         3: GE0/2
             Protocol: pim-dm, UpTime: 00:03:27, Expires: never
```

1.7.2 PIM-SM 非管理域配置举例

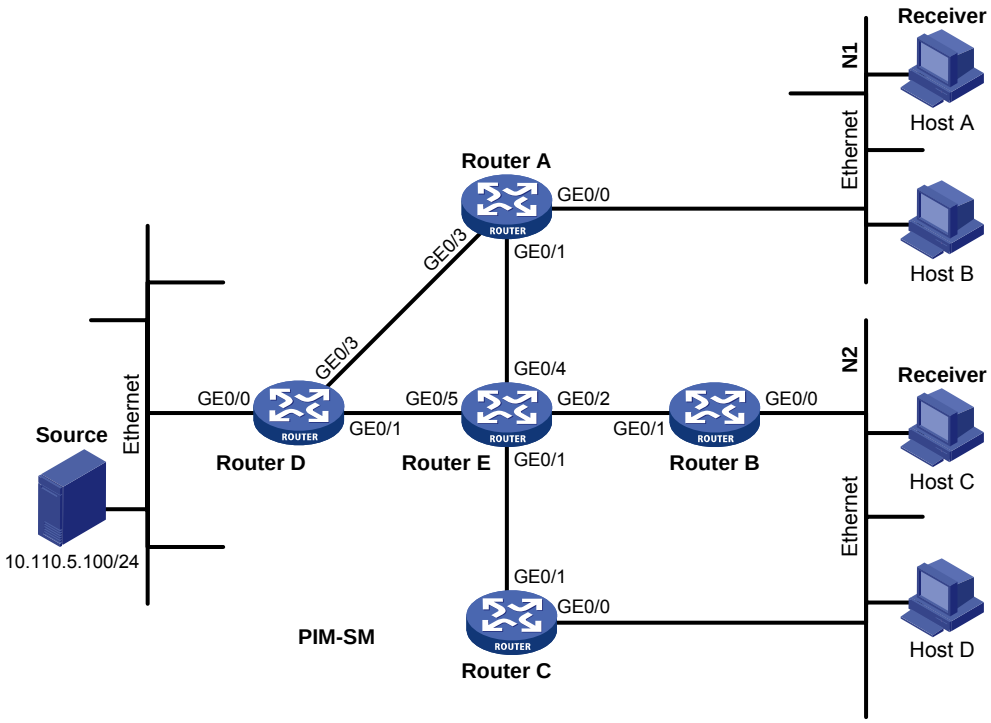
1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 PIM 域采用 SM 非管理域方式。

- Host A 和 Host C 为两个末梢网络中的组播信息接收者；Router D 通过 GigabitEthernet0/0 接口与组播源（Source）所在网络连接；Router A 通过 GigabitEthernet0/0 接口连接末梢网络 N1，通过 GE0/3 接口和 GE0/1 接口分别连接 Router D 和 Router E；Router B 和 Router C 分别通过各自的 GigabitEthernet0/0 接口连接末梢网络 N2，并分别通过各自的 GE0/1 接口连接 Router E。
- 将 Router D 的 GE0/1 接口和 Router E 的 GE0/4 接口都配置为 C-BSR 和 C-RP，其中 Router E 上 C-BSR 的优先级较高；C-RP 所服务的组播组范围为 225.1.1.0/24，通过改变哈希掩码长度使此范围内的组地址间隔映射到这两个 C-RP 上。
- Router A 与末梢网络 N1 之间运行 IGMPv2；Router B 和 Router C 与末梢网络 N2 之间也运行 IGMPv2。

2. 组网图

图1-11 PIM-SM 非管理域配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/0	10.110.1.1/24	Router D	GE0/0	10.110.5.1/24
	GE0/3	192.168.1.1/24		GE0/3	192.168.1.2/24
	GE0/1	192.168.9.1/24		GE0/1	192.168.4.2/24
Router B	GE0/0	10.110.2.1/24	Router E	GE0/1	192.168.3.2/24
	GE0/1	192.168.2.1/24		GE0/2	192.168.2.2/24
Router C	GE0/0	10.110.2.2/24		GE0/4	192.168.9.2/24
	GE0/1	192.168.3.1/24		GE0/5	192.168.4.1/24

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-11配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-SM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-SM 和 IGMP

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在其连接末梢网络的接口 GigabitEthernet0/0 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/0
[RouterA-GigabitEthernet0/0] igmp enable
[RouterA-GigabitEthernet0/0] pim sm
[RouterA-GigabitEthernet0/0] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] pim sm
[RouterA-GigabitEthernet0/3] quit
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
```

Router B 和 Router C 的配置与 Router A 相似，Router D 和 Router E 除了不需要在相应接口上使能 IGMP 外，其它的配置也与 Router A 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP

在 Router D 上配置 RP 通告的服务范围，以及 C-BSR 和 C-RP 的位置，并指定哈希掩码长度为 32，C-BSR 的优先级为 10。

```
<RouterD> system-view
[RouterD] acl number 2005
[RouterD-acl-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[RouterD-acl-basic-2005] quit
[RouterD] pim
[RouterD-pim] c-bsr GigabitEthernet 0/1 32 10
[RouterD-pim] c-rp GigabitEthernet 0/1 group-policy 2005
[RouterD-pim] quit
```

在 Router E 上配置 RP 通告的服务范围，以及 C-BSR 和 C-RP 的位置，并指定哈希掩码长度为 32，C-BSR 的优先级为 20。

```
<RouterE> system-view
[RouterE] acl number 2005
[RouterE-acl-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[RouterE-acl-basic-2005] quit
[RouterE] pim
[RouterE-pim] c-bsr GigabitEthernet 0/4 32 20
[RouterE-pim] c-rp GigabitEthernet 0/4 group-policy 2005
[RouterE-pim] quit
```

(4) 检验配置效果

通过使用 **display pim interface** 命令可以查看路由器接口上 PIM 的配置和运行情况。例如：

查看 Router A 上 PIM 的配置信息。

```
[RouterA] display pim interface
 VPN-Instance: public net
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
```

```
 GE0/0              0      30         1          10.110.1.1     (local)
 GE0/3              1      30         1          192.168.1.2
 GE0/1              1      30         1          192.168.9.2
```

通过使用 **display pim bsr-info** 命令可以查看路由器上 BSR 选举的信息，以及本地配置并生效的 C-RP 信息。例如：

查看 Router A 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterA] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Accept Preferred
     Scope: Not scoped
     Uptime: 00:40:40
     Expires: 00:01:42
```

查看 Router D 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterD] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Accept Preferred
     Scope: Not scoped
     Uptime: 00:05:26
     Expires: 00:01:45
 Candidate BSR Address: 192.168.4.2
     Priority: 10
     Hash mask length: 32
     State: Candidate
     Scope: Not scoped

 Candidate RP: 192.168.4.2(GE0/1)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:34
```

查看 Router E 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterE] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Elected
     Scope: Not scoped
     Uptime: 00:01:18
     Next BSR message scheduled at: 00:01:52
```

```
 Candidate BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Elected
     Scope: Not scoped

 Candidate RP: 192.168.9.2(GE0/4)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:48
```

通过使用 **display pim rp-info** 命令可以查看路由器上获取的 RP 信息。例如：

查看 Router A 上的 RP 信息。

```
[RouterA] display pim rp-info
 VPN-Instance: public net
 PIM-SM BSR RP information:
 Group/MaskLen: 225.1.1.0/24
     RP: 192.168.4.2
     Priority: 0
     HoldTime: 150
     Uptime: 00:51:45
     Expires: 00:02:22

     RP: 192.168.9.2
     Priority: 0
     HoldTime: 150
     Uptime: 00:51:45
     Expires: 00:02:22
```

假如 Host A 需要接收组播组 G（225.1.1.0）的信息，由于根据哈希算法得出 G 对应的 RP 为 Router E，因此 Router A 和 Router E 之间会生成 RPT。当组播源 S（10.110.5.100/24）向 RP 发起注册后，Router D 和 Router E 之间会生成 SPT。当 Router A 收到组播数据后立即执行从 RPT 到 SPT 的切换。RPT 路径中的路由器（Router A 和 Router E）上存在（*，G）表项，而 SPT 路径中的路由器（Router A 和 Router D）上存在（S，G）表项，通过使用 **display pim routing-table** 命令可以查看路由器的 PIM 路由表信息。例如：

查看 Router A 上的 PIM 路由表信息。

```
[RouterA] display pim routing-table
 VPN-Instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.0)
     RP: 192.168.9.2
     Protocol: pim-sm, Flag: WC
     UpTime: 00:13:46
     Upstream interface: GE0/1
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
```

```
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: GigabitEthernet0/0
            Protocol: igmp, UpTime: 00:13:46, Expires: 00:03:06

 (10.110.5.100, 225.1.1.0)
    RP: 192.168.9.2
    Protocol: pim-sm, Flag: SPT ACT
    UpTime: 00:00:42
    Upstream interface: GE0/3
        Upstream neighbor: 192.168.9.2
        RPF prime neighbor: 192.168.9.2
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: GigabitEthernet0/0
            Protocol: pim-sm, UpTime: 00:00:42, Expires: 00:03:06
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 PIM 路由表信息。

```
[RouterD] display pim routing-table
 VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.0)
    RP: 192.168.9.2
    Protocol: pim-sm, Flag: SPT LOC ACT
    UpTime: 00:00:42
    Upstream interface: GigabitEthernet0/0
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface(s) information:
    Total number of downstreams: 1
        1: GE0/1
            Protocol: pim-sm, UpTime: 00:00:42, Expires: 00:02:26
```

查看 Router E 上的 PIM 路由表信息。

```
[RouterE] display pim routing-table
 VPN-Instance: public net
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, 225.1.1.0)
    RP: 192.168.9.2 (local)
    Protocol: pim-sm, Flag: WC
    UpTime: 00:13:16
    Upstream interface: Register
        Upstream neighbor: 192.168.4.2
        RPF prime neighbor: 192.168.4.2
    Downstream interface(s) information:
    Total number of downstreams: 1
```

```
1: GE0/4
    Protocol: pim-sm, UpTime: 00:13:16, Expires: 00:03:22
```

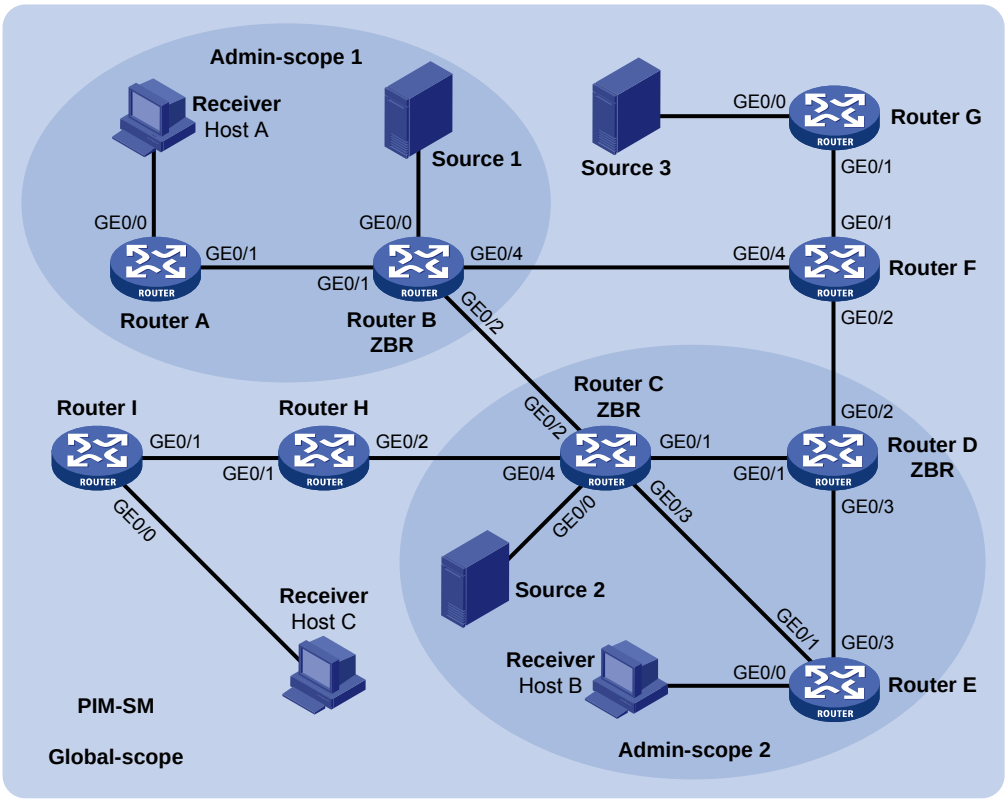
1.7.3 PIM-SM 管理域配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，整个 PIM 域采用 SM 管理域方式，划分为管理域 1、管理域 2 和 Global 域，Router B、Router C 和 Router D 为各管理域的 ZBR。
- Source 1 和 Source 2 分别向组播组 239.1.1.1 发送内容不同的组播信息，Host A 和 Host B 则分别只接收来自 Source 1 和 Source 2 的组播信息；Source 3 向组播组 224.1.1.1 发送组播信息，Host C 为其接收者。
- Router B 的 GigabitEthernet0/1 接口为管理域 1 的 C-BSR 和 C-RP，服务于 239.0.0.0/8；Router D 的 GigabitEthernet0/1 接口为管理域 2 的 C-BSR 和 C-RP，服务于 239.0.0.0/8；Router F 的 GigabitEthernet0/1 接口为 Global 域的 C-BSR 和 C-RP，服务于 239.0.0.0/8 以外的所有组播组。
- Router A、Router E 和 Router I 分别与各自所连接的接收者之间运行 IGMPv2。

2. 组网图

图1-12 PIM-SM 管理域配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/0	192.168.1.1/24	Router D	GE0/1	10.110.4.2/24
	GE0/1	10.110.1.1/24		GE0/3	10.110.7.1/24
Router B	GE0/0	192.168.2.1/24		GE0/2	10.110.8.1/24
	GE0/1	10.110.1.2/24	Router E	GE0/0	192.168.4.1/24
	GE0/2	10.110.2.1/24		GE0/1	10.110.5.2/24
	GE0/4	10.110.3.1/24		GE0/3	10.110.7.2/24
Router C	GE0/0	192.168.3.1/24	Router F	GE0/1	10.110.9.1/24
	GE0/1	10.110.4.1/24		GE0/2	10.110.8.2/24
	GE0/3	10.110.5.1/24		GE0/4	10.110.3.2/24
	GE0/2	10.110.2.2/24	Router G	GE0/0	192.168.5.1/24
	GE0/4	10.110.6.1/24		GE0/1	10.110.9.2/24
Router H	GE0/1	10.110.10.1/24	Source 1	-	192.168.2.10/24
	GE0/2	10.110.6.2/24	Source 2	-	192.168.3.10/24
Router I	GE0/0	192.168.6.1/24	Source 3	-	192.168.5.10/24
	GE0/1	10.110.10.2/24			

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-12配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-SM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由和管理域机制，并使能 PIM-SM 和 IGMP

在 Router A 上使能 IP 组播路由和管理域机制，在各接口上使能 PIM-SM，并在其连接有接收者的接口 GigabitEthernet0/0 上使能 IGMP。

```
<RouterA> system-view
[RouterA] multicast routing-enable
```

```
[RouterA] pim
[RouterA-pim] c-bsr admin-scope
[RouterA-pim] quit
[RouterA] interface GigabitEthernet 0/0
[RouterA-GigabitEthernet0/0] igmp enable
[RouterA-GigabitEthernet0/0] pim sm
[RouterA-GigabitEthernet0/0] quit
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
```

Router E 和 Router I 的配置与 Router A 相似，配置过程略。

在 Router B 上使能 IP 组播路由和管理域机制，并在各接口上使能 PIM-SM。

```
<RouterB> system-view
[RouterB] multicast routing-enable
[RouterB] pim
[RouterB-pim] c-bsr admin-scope
[RouterB-pim] quit
[RouterB] interface GigabitEthernet 0/0
[RouterB-GigabitEthernet0/0] pim sm
[RouterB-GigabitEthernet0/0] quit
[RouterB] interface GigabitEthernet 0/1
[RouterB-GigabitEthernet0/1] pim sm
[RouterB-GigabitEthernet0/1] quit
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] pim sm
[RouterB-GigabitEthernet0/2] quit
[RouterB] interface GigabitEthernet 0/4
[RouterB-GE0/4] pim sm
[RouterB-GE0/4] quit
```

Router C、Router D、Router F、Router G 和 Router H 的配置与 Router B 相似，配置过程略。

(3) 配置管理域边界

在 Router B 上将接口 GE0/2 和 GE0/4 配置为管理域 1 的边界。

```
[RouterB] interface GigabitEthernet 0/2
[RouterB-GigabitEthernet0/2] multicast boundary 239.0.0.0 8
[RouterB-GigabitEthernet0/2] quit
[RouterB] interface GigabitEthernet 0/4
[RouterB-GE0/4] multicast boundary 239.0.0.0 8
[RouterB-GE0/4] quit
```

在 Router C 上将接口 GE0/2 和 GE0/4 配置为管理域 2 的边界。

```
<RouterC> system-view
[RouterC] interface GigabitEthernet 0/2
[RouterC-GigabitEthernet0/2] multicast boundary 239.0.0.0 8
[RouterC-GigabitEthernet0/2] quit
[RouterC] interface GigabitEthernet 0/4
[RouterC-GE0/4] multicast boundary 239.0.0.0 8
```

```
[RouterC-GE0/4] quit
```

在 Router D 上将接口 GE0/2 配置为管理域 2 的边界。

```
<RouterD> system-view
[RouterD] interface GigabitEthernet 0/2
[RouterD-GigabitEthernet0/2] multicast boundary 239.0.0.0 8
[RouterD-GigabitEthernet0/2] quit
```

(4) 配置 C-BSR 和 C-RP

在 Router B 上配置 RP 通告的服务范围，并将接口 GigabitEthernet0/1 配置为管理域 1 的 C-BSR 和 C-RP。

```
[RouterB] acl number 2001
[RouterB-acl-basic-2001] rule permit source 239.0.0.0 0.255.255.255
[RouterB-acl-basic-2001] quit
[RouterB] pim
[RouterB-pim] c-bsr group 239.0.0.0 8
[RouterB-pim] c-bsr GigabitEthernet 0/1
[RouterB-pim] c-rp GigabitEthernet 0/1 group-policy 2001
[RouterB-pim] quit
```

在 Router D 上配置 RP 通告的服务范围，并将接口 GigabitEthernet0/1 配置为管理域 2 的 C-BSR 和 C-RP。

```
[RouterD] acl number 2001
[RouterD-acl-basic-2001] rule permit source 239.0.0.0 0.255.255.255
[RouterD-acl-basic-2001] quit
[RouterD] pim
[RouterD-pim] c-bsr group 239.0.0.0 8
[RouterD-pim] c-bsr GigabitEthernet 0/1
[RouterD-pim] c-rp GigabitEthernet 0/1 group-policy 2001
[RouterD-pim] quit
```

在 Router F 上将接口 GigabitEthernet0/1 配置为 Global 域的 C-BSR 和 C-RP。

```
<RouterF> system-view
[RouterF] pim
[RouterF-pim] c-bsr global
[RouterF-pim] c-bsr GigabitEthernet 0/1
[RouterF-pim] c-rp GigabitEthernet 0/1
[RouterF-pim] quit
```

(5) 检验配置效果

通过使用 **display pim bsr-info** 命令可以查看路由器上 BSR 选举的信息，以及本地配置并生效的 C-RP 信息。例如：

查看 Router B 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterB] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Accept Preferred
```

```
    Scope: Global
    Uptime: 00:01:45
    Expires: 00:01:25
 Elected BSR Address: 10.110.1.2
    Priority: 0
    Hash mask length: 30
    State: Elected
    Scope: 239.0.0.0/8
    Uptime: 00:04:54
    Next BSR message scheduled at: 00:00:06
 Candidate BSR Address: 10.110.1.2
    Priority: 0
    Hash mask length: 30
    State: Elected
    Scope: 239.0.0.0/8

 Candidate RP: 10.110.1.2(GigabitEthernet0/1)
    Priority: 0
    HoldTime: 150
    Advertisement Interval: 60
    Next advertisement scheduled at: 00:00:15
```

查看 Router D 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterD] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 10.110.9.1
    Priority: 0
    Hash mask length: 30
    State: Accept Preferred
    Scope: Global
    Uptime: 00:01:45
    Expires: 00:01:25
 Elected BSR Address: 10.110.4.2
    Priority: 0
    Hash mask length: 30
    State: Elected
    Scope: 239.0.0.0/8
    Uptime: 00:03:48
    Next BSR message scheduled at: 00:01:12
 Candidate BSR Address: 10.110.4.2
    Priority: 0
    Hash mask length: 30
    State: Elected
    Scope: 239.0.0.0/8

 Candidate RP: 10.110.4.2(GigabitEthernet0/1)
    Priority: 0
    HoldTime: 150
    Advertisement Interval: 60
```

```
     Next advertisement scheduled at: 00:00:10
```

查看 Router F 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterF] display pim bsr-info
 VPN-Instance: public net
 Elected BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: Global
     Uptime: 00:11:11
     Next BSR message scheduled at: 00:00:49
 Candidate BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: Global

 Candidate RP: 10.110.9.1(GigabitEthernet0/1)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:55
```

通过使用 **display pim rp-info** 命令可以查看路由器上获取的 RP 信息。例如：

查看 Router B 上的 RP 信息。

```
[RouterB] display pim rp-info
 VPN-Instance: public net
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
     RP: 10.110.9.1
     Priority: 0
     HoldTime: 150
     Uptime: 00:03:39
     Expires: 00:01:51

Group/MaskLen: 239.0.0.0/8
     RP: 10.110.1.2 (local)
     Priority: 0
     HoldTime: 150
     Uptime: 00:07:44
     Expires: 00:01:51
```

查看 Router D 上的 RP 信息。

```
[RouterD] display pim rp-info
 VPN-Instance: public net
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
     RP: 10.110.9.1
```

```
    Priority: 0
    HoldTime: 150
    Uptime: 00:03:42
    Expires: 00:01:48

Group/MaskLen: 239.0.0.0/8
    RP: 10.110.4.2 (local)
    Priority: 0
    HoldTime: 150
    Uptime: 00:06:54
    Expires: 00:02:41
```

查看 Router F 上的 RP 信息。

```
[RouterF] display pim rp-info
 VPN-Instance: public net
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
    RP: 10.110.9.1 (local)
    Priority: 0
    HoldTime: 150
    Uptime: 00:00:32
    Expires: 00:01:58
```

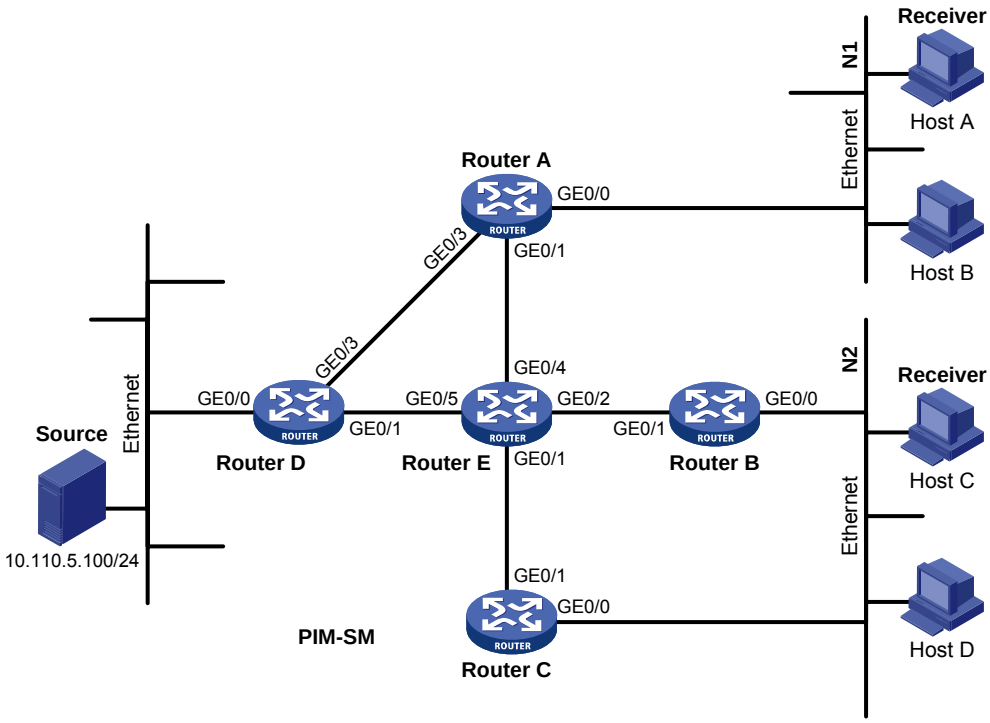
1.7.4 PIM-SSM 典型配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 PIM 域采用 SSM 方式。
- Host A 和 Host C 为两个末梢网络中的组播信息接收者；Router D 通过 GigabitEthernet0/0 接口与组播源（Source）所在网络连接；Router A 通过 GigabitEthernet0/0 接口连接末梢网络 N1，通过 GE0/3 接口和 GE0/1 接口分别连接 Router D 和 Router E；Router B 和 Router C 分别通过各自的 GigabitEthernet0/0 接口连接末梢网络 N2，并分别通过各自的 GE0/1 接口连接 Router E；Router E 连接 Router A、Router B、Router C 和 Router D。
- SSM 组播组的范围是 232.1.1.0/24。
- Router A 与末梢网络 N1 之间运行 IGMPv3；Router B 和 Router C 与末梢网络 N2 之间也运行 IGMPv3。

2. 组网图

图1-13 PIM-SSM 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE0/0	10.110.1.1/24	Router D	GE0/0	10.110.5.1/24
	GE0/3	192.168.1.1/24		GE0/3	192.168.1.2/24
	GE0/1	192.168.9.1/24		GE0/1	192.168.4.2/24
Router B	GE0/0	10.110.2.1/24	Router E	GE0/1	192.168.3.2/24
	GE0/1	192.168.2.1/24		GE0/2	192.168.2.2/24
Router C	GE0/0	10.110.2.2/24		GE0/4	192.168.9.2/24
	GE0/1	192.168.3.1/24		GE0/5	192.168.4.1/24

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 1-13配置各接口的IP地址和子网掩码，具体配置过程略。

配置 PIM-SM 域内的各路由器之间采用 OSPF 协议进行互连，确保 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IP 组播路由，并使能 PIM-SM 和 IGMP

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在其连接末梢网络的接口 GigabitEthernet0/0 上使能 IGMP，且配置其版本为 3。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 0/0
[RouterA-GigabitEthernet0/0] igmp enable
[RouterA-GigabitEthernet0/0] igmp version 3
[RouterA-GigabitEthernet0/0] pim sm
[RouterA-GigabitEthernet0/0] quit
[RouterA] interface GigabitEthernet 0/3
[RouterA-GigabitEthernet0/3] pim sm
```

```
[RouterA-GigabitEthernet0/3] quit
[RouterA] interface GigabitEthernet 0/1
[RouterA-GigabitEthernet0/1] pim sm
[RouterA-GigabitEthernet0/1] quit
```

Router B 和 Router C 的配置与 Router A 相似，Router D 和 Router E 除了不需要在相应接口上使能 IGMP 外，其它的配置也与 Router A 相似，配置过程略。

(3) 配置 SSM 组播组的地址范围

在 Router A 上配置 SSM 组播组地址范围为 232.1.1.0/24。

```
[RouterA] acl number 2000
[RouterA-acl-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[RouterA-acl-basic-2000] quit
[RouterA] pim
[RouterA-pim] ssm-policy 2000
[RouterA-pim] quit
```

Router B、Router C、Router D 和 Router E 的配置与 Router A 相似，配置过程略。

(4) 检验配置效果

通过使用 **display pim interface** 命令可以查看路由器接口上 PIM 的配置和运行情况。例如：

查看 Router A 上 PIM 的配置信息。

```
[RouterA] display pim interface
 VPN-Instance: public net
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
 GE0/0               0      30         1          10.110.1.1     (local)
 GE0/3               1      30         1          192.168.1.2
 GE0/1               1      30         1          192.168.9.2
```

假如 Host A 需要接收指定组播源 S（10.110.5.100/24）发往组播组 G（232.1.1.1）的信息，Router A 会向组播源方向构造 SPT，SPT 路径中的路由器（Router A 和 Router D）上生成（S，G）表项，而 SPT 路径之外的路由器（Router E）上没有组播路由项，通过使用 **display pim routing-table** 命令可以查看路由器的 PIM 路由表信息。例如：

查看 Router A 上的 PIM 路由表信息。

```
[RouterA] display pim routing-table
 VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: GE0/3
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet0/0
             Protocol: igmp, UpTime: 00:13:25, Expires: 00:03:25
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 PIM 路由表信息。

```
[RouterD] display pim routing-table
 VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol: pim-ssm, Flag: LOC
     UpTime: 00:12:05
     Upstream interface: GigabitEthernet0/0
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GE0/3
             Protocol:  pim-ssm, UpTime: 00:12:05, Expires: 00:03:25
```

1.8 常见配置错误举例

1.8.1 无法正确建立组播分发树

1. 故障现象

网络中各路由器（包括直连组播源或接收者的路由器）上都没有组播转发项，也就是说无法正确建立组播分发树，客户端无法接收组播数据。

2. 分析

- 当全网运行 PIM-DM 时，组播数据由直连组播源的第一跳路由器扩散到直连客户端的最后一跳路由器。无论组播数据扩散到哪一台路由器，只有该路由器存在到达组播源的路由，才会创建（S，G）项。反之，如果没有到达组播源的路由或者到达组播源的 RPF 接口没有使能 PIM-DM，则该路由器无法创建（S，G）项。
- 当全网运行 PIM-SM 时，路由器在准备加入 SPT 时，只有存在到达组播源的路由，才会创建（S，G）项。反之，如果没有到达组播源的路由或者到达组播源的 RPF 接口没有使能 PIM-SM，则该路由器无法创建（S，G）项。
- 对于某个 RPF 检查对象，在现存的单播路由表中查找到达该对象的最优路由，该路由的出接口作为 RPF 接口，下一跳作为 RPF 邻居。RPF 接口完全依赖于现存的单播路由，并且与 PIM 本身无关。RPF 接口必须使能 PIM，而且 RPF 邻居也必须是 PIM 邻居。如果 RPF 接口或 RPF 邻居所在路由器上没有使能 PIM，必然导致组播分发树无法正确建立，导致组播数据转发异常。
- Hello 报文并不携带 PIM 的模式信息，所以运行 PIM 的路由器无法掌握自己的 PIM 邻居运行的是何种模式的 PIM。如果 RPF 接口和 RPF 邻居所在路由器的对应接口没有使能相同模式的 PIM，必然导致组播分发树无法正确建立，导致组播数据转发异常。
- 全网必须运行相同模式的 PIM。否则，组播分发树必然无法正确建立，导致组播数据转发异常。

3. 处理过程

(1) 检查单播路由。使用命令 **display ip routing-table** 检查从接收者主机是否有到达组播源的单播路由项。

(2) 检查接口上是否使能 PIM，尤其是 RPF 接口上是否使能 PIM。通过命令 **display pim interface** 观察接口上的 PIM 信息。若接口上没有使能 PIM，请使用 **pim dm** 或 **pim sm** 命令使能 PIM-DM 或者 PIM-SM。

(3) 检查 RPF 邻居是否是 PIM 邻居。通过命令 **display pim neighbor** 观察 PIM 邻居信息。

(4) 检查直连组播源或接收者的路由器接口上是否使能了 PIM 和 IGMP。

(5) 检查 PIM 模式是否一致。通过命令 **display pim interface verbose** 检查 RPF 接口和 RPF 邻居所在路由器的对应接口上是否使能了相同模式的 PIM。

(6) 检查全部网络中各路由器上的 PIM 模式是否一致。确保全网所有路由器配置相同模式的 PIM，即要么全部配置为 PIM-SM，要么全部配置为 PIM-DM。如果配置为 PIM-SM，则还需要检查 BSR 以及 RP 的配置是否正确。

1.8.2 组播数据异常终止在中间路由器

1. 故障现象

组播数据可以到达中间路由器，但无法到达最后一跳路由器。中间路由器某接口上收到组播数据，但 PIM 路由表中没有创建相应的（S，G）项。

2. 分析

- 路由器在转发组播数据时，将收到组播数据报文的 TTL 值减 1，重新计算校验值，然后向所有的出接口转发。如果出接口上配置了命令 **multicast minimum-ttl**，组播数据报文的 TTL 值必须大于配置的最小 TTL 值才会被转发，否则将会被丢弃。
- 命令 **multicast boundary** 用于在接口上设置组播转发边界，如果组播数据无法通过该边界，PIM 是无法创建路由项的。
- 此外，**source-policy** 命令用于过滤接收到的组播数据报文。如果组播数据报文无法通过该命令的 ACL 规则，PIM 也是无法创建路由项的。

3. 处理过程

(1) 检查组播转发最小 TTL 值。通过命令 **display current-configuration** 查看组播转发的最小 TTL 值，应调高 TTL 值。

(2) 检查组播转发边界的配置。通过命令 **display current-configuration** 查看组播转发边界上的设置，使用 **multicast boundary** 命令更改组播转发边界的设置，使组播数据能够通过该边界。

(3) 检查组播过滤器配置。通过命令 **display current-configuration** 查看组播过滤器的配置，更改 **source-policy** 命令的 ACL 规则，使组播数据的源/组地址通过 ACL 过滤。

1.8.3 PIM-SM 中 RP 无法加入 SPT

1. 故障现象

共享树无法正确建立，或者 RP 无法加入到达组播源的最短路径树。

2. 分析

- RP 是 PIM-SM 的核心，为特定的组服务。网络中可以同时存在多个 RP。必须保证所有路由器的 RP 信息完全一致，并且对于某个特定的组映射到相同的 RP。否则必然导致组播数据转发异常。
- 如果使用静态 RP 机制，必须在全网所有路由器上配置完全相同的静态 RP 命令。否则必然导致组播数据转发异常。

3. 处理过程

(1) 检查是否有到达 RP 的路由。通过命令 **display ip routing-table** 查看各路由器上是否有到达 RP 的路由。

(2) 检查动态 RP 信息。通过命令 **display pim rp-info** 查看各路由器上的 RP 信息是否一致。

(3) 检查静态 RP 的配置。通过命令 **display pim rp-info** 查看全网所有路由器上是否配置了完全相同的静态 RP。

1.8.4 PIM-SM 中无法建立 RPT 或无法进行源注册

1. 故障现象

C-RP 无法向 BSR 单播通告报文，BSR 没有发布包含 C-RP 的自举报文，BSR 上没有到达各 C-RP 的单播路由，共享树无法正确建立，或者 DR 无法向 RP 进行源注册。

2. 分析

- 由于 C-RP 周期性地向 BSR 单播宣告报文，如果 C-RP 没有到达 BSR 的单播路由就无法发送宣告报文，BSR 就收不到 C-RP 宣告报文，也就不会发布包含该 C-RP 的自举报文。
- 另外，如果 BSR 没有到达 C-RP 的单播路由，就会丢弃该宣告报文，也就不会发布包含该 C-RP 的自举报文。
- RP 是 PIM-SM 域的核心。必须保证全网所有路由器的 RP 信息完全一致，并且对于某个特定的组 G 映射到相同的 RP，并且存在到达 RP 的单播路由。

3. 处理过程

(1) 检查是否有到达各 C-RP、BSR 的路由。通过命令 **display ip routing-table** 查看各路由器上是否有到达 RP 和 BSR 的路由，及 RP 和 BSR 之间的路由。确保 C-RP 上存在到达 BSR 的单播路由，BSR 上存在到达 C-RP 的单播路由，全网络所有路由器上存在到达 RP 的单播路由。

(2) 检查 RP 和 BSR 信息。PIM-SM 协议需要有 RP 和 BSR 的支持，首先使用命令 **display pim bsr-info** 查看是否有 BSR 信息，使用 **display pim rp-info** 命令查看 RP 信息是否正确。

(3) 检查 PIM 邻居关系。通过命令 **display pim neighbor** 来查看是否正确建立了邻居关系。

目 录

1 组播VPN配置 1-1
1.1 组播VPN简介 1-1
1.1.1 MPLS L3VPN概述 1-1
1.1.2 组播VPN概述 1-2
1.1.3 MD VPN概述 1-3
1.1.4 协议规范 1-6
1.2 MD VPN实现原理 1-6
1.2.1 创建Share-MDT 1-6
1.2.2 基于Share-MDT的传输 1-9
1.2.3 Switch-MDT切换 1-11
1.2.4 跨AS的MD VPN 1-12
1.3 组播VPN配置任务简介 1-14
1.4 配置MD VPN 1-14
1.4.1 配置准备 1-14
1.4.2 使能VPN实例中的IP组播路由 1-14
1.4.3 配置Share-Group并绑定MTI 1-15
1.4.4 配置Switch-MDT切换参数 1-15
1.4.5 打开Switch-Group重用日志输出开关 1-16
1.5 配置BGP MDT 1-16
1.5.1 配置准备 1-16
1.5.2 配置BGP MDT对等体/对等体组 1-17
1.5.3 配置BGP MDT路由反射器 1-17
1.6 组播VPN显示和维护 1-18
1.7 组播VPN典型配置举例 1-18
1.7.1 单AS内MD VPN配置举例 1-18
1.7.2 跨AS的MD VPN配置举例 1-31
1.8 常见配置错误举例 1-44
1.8.1 无法建立Share-MDT 1-44
1.8.2 VPN实例无法正确建立组播路由表 1-44

1 组播 VPN 配置

1.1 组播 VPN 简介

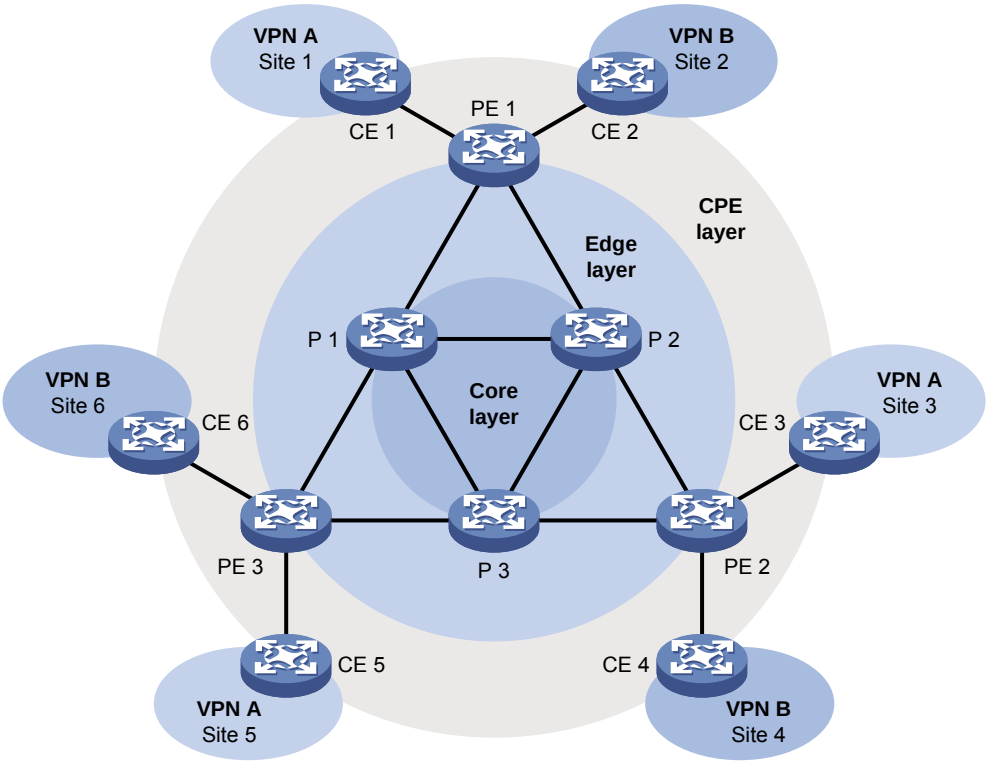
说明

- 有关 MPLS L3VPN 的详细介绍和相关配置，请参见“MPLS 分册”中的“MPLS L3VPN 配置”。
- 有关 BGP 的详细介绍和相关配置，请参见“IP 路由分册”中的“BGP 配置”。

1.1.1 MPLS L3VPN 概述

组播 VPN 是基于 MPLS L3VPN 网络来实现组播传输的技术。MPLS L3VPN 是一种基于 BGP（Border Gateway Protocol，边界网关协议）和 MPLS（Multiprotocol Label Switching，多协议标签交换）扩展技术实现的 VPN（Virtual Private Network，虚拟专用网），由运营商的骨干网和用户的各个 Site（站点）组成，各 Site 之间彼此相互孤立，只有借助骨干网才能实现互通。VPN 可以看作是一组策略，控制着各 Site 之间的连接。

图1-1 MPLS L3VPN 典型应用



如 图 1-1所示，由Site 1、Site 3 和Site 5 组成VPN A网络，由Site 2、Site 4 和Site 6 组成VPN B网络，其中包括以下三种类型的设备：

- P（Provider）：骨干网核心设备，不与 CE 直接相连，负责 MPLS 转发。
- PE（Provider Edge）：骨干网边缘设备，与 CE 直接相连，负责 VPN 路由的处理，是 MPLS L3VPN 的主要实现者。

- CE（Customer Edge）：用户网边缘设备，可以是路由器或交换机，也可以是一台主机，负责用户网络路由的发布。

在 MPLS L3VPN 中，属于同一 VPN 的两个 Site 之间使用两层标签来穿越公网传输报文。公网入口 PE 为报文打上内外两层标签：

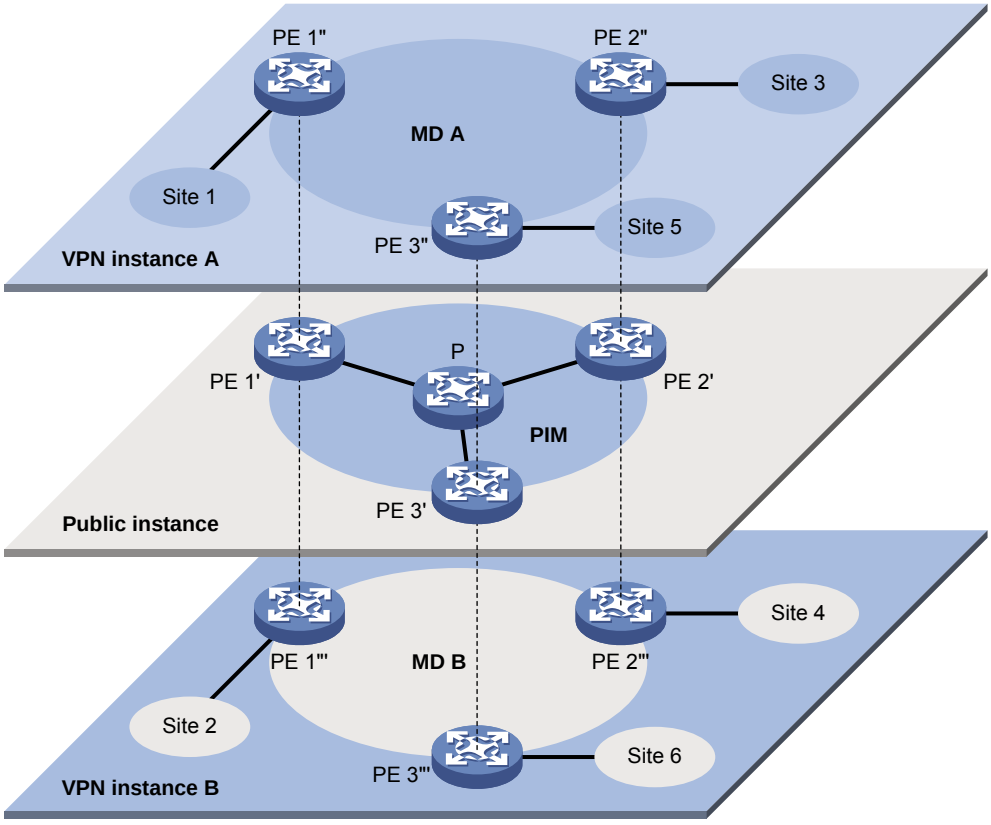
- 外层标签：用于在骨干网内部进行交换，代表从本端 PE 到对端 PE 的一条 LSP（Label Switched Path，标签交换路径）。借助这层标签，报文可以沿着 LSP 到达对端 PE。
- 内层标签：代表通过骨干网相连的两个 CE 之间的一条 LSP。主要标识报文属于哪个 Site，PE 根据内层标签将报文转发到相应的 CE。

1.1.2 组播 VPN 概述

如 图 1-2所示，网络中同时承载着三个相互独立的组播业务：公网实例、VPN实例A和VPN实例B。公共网络边缘的PE组播设备支持多实例，相当于多台独立运行的组播设备。各实例之间形成彼此隔离的平面，每个实例对应一个平面。

举例来说，图 1-1中的PE 1 上运行着三个实例：公网实例、VPN实例A和VPN实例B，可以把这三个不同的实例想象成三台独立运行的虚拟设备，分别是PE 1’、PE 1”和PE 1”’，每台虚拟设备则分别对应着一个平面，对应关系如 图 1-2所示。

图1-2 VPN 中基于多实例的组播



以 VPN 实例 A 为例，组播 VPN 是指：当 VPN A 中的组播源向某组播组发送组播数据时，在网络中所有可能的接收者中，只有属于 VPN A（即 Site 1、Site 3 或 Site 5 中）的组播组成员才能收到该组播源发来的组播数据。组播数据在各 Site 以及公网中均以组播方式进行传输。

由此分析，实现组播 VPN 所需具备的网络条件如下：

(1) 在每个 Site 内支持基于 VPN 实例的组播。

(2) 在公共网络内支持基于公网实例的组播。

(3) PE 设备支持多实例组播：

- 通过 VPN 实例连接 Site，支持基于 VPN 实例的组播；
- 通过公网实例连接公共网络，支持基于公网实例的组播；
- 支持公网实例与 VPN 实例之间的信息交流和数据转换。

1.1.3 MD VPN 概述

说明

有关 PIM（Protocol Independent Multicast，协议无关组播）、BSR（BootStrap Router，自举路由器）、C-BSR（Candidate-BSR，候选 BSR）、RP（Rendezvous Point，汇集点）、C-RP（Candidate RP，候选 RP）、SPT（Shortest Path Tree，最短路径树）和 RPT（Rendezvous Point Tree，共享树或汇集树）等概念的介绍，请参见“IP 组播分册”中的“PIM 配置”。

系统利用 MD（Multicast Domain，组播域）方案来实现组播 VPN，简称为 MD VPN。

MD 方案最大的优点是仅需要 PE 设备支持多实例，而无需升级 CE 和 P 设备，并且无需修改 CE 和 P 设备上原有的 PIM 配置。也就是说，MD 方案对于 CE 和 P 是透明的。

1. MD VPN 基本概念

MD VPN中所涉及到的基本概念如 表 1-1所示。

表1-1 MD VPN 中的基本概念

概念	全称及中文解释	说明
MD	Multicast Domain，组播域	由一组相互连通的 MVRF 所组成的集合称为 MD。每个 MD 都唯一对应着一个具备组播能力的 VPN，连接该 VPN 的所有 PE 都属于这个 MD
MDT	Multicast Distribution Tree，组播分发树	是建立在属于同一 VPN 所有 PE 之间的组播分发树，包括 Share-MDT 和 Switch-MDT
MT	Multicast Tunnel，组播隧道	在 MD 内将各 MVRF 连接到一起的通道称为 MT，用来传输 MD 内部的私网数据
MTI	Multicast Tunnel Interface，组播隧道接口	MTI 是 MT 的入/出口，相当于 MD 的入/出口，它只收发组播报文，而不收发单播报文。MTI 在 VPN 实例配置了 Share-Group 并绑定 MTI 后自动创建。MVRF 通过 MTI 访问 MD，在 MVRF 看来，MTI 就像一个 LAN 接口，而 MD 就像一个 LAN 网络，所有 PE 都连接在这个 LAN 网络里
MVRF	Multicast VPN Routing and Forwarding，组播 VPN 路由转发	使能了三层组播的 VPN 实例同时维护单播和组播的路由转发表，我们称之为 MVRF。不同 PE 上的 MVRF 加入到同一个 MD 中，并通过 MD 内自动建立的 MT 连接起来，实现了不同 Site 间组播业务的互通，从而形成了一个组播 VPN 网络
Share-Group	共享组	每个 MD 在公网上分配一个独立的组播地址，称为 Share-Group。它是 MD 在公网上的唯一标志，用来在公网上建立 MD 所对应的 Share-MDT
Share-MDT	Share-Multicast Distribution Tree，共享组播分发树	以 Share-Group 为组地址的 MDT，称为 Share-MDT。VPN 使用 Share-Group 唯一标识一棵 Share-MDT。Share-MDT 是在配置完成后自动生成的，在公网中将会一直存在，而不论公网或私网中有没有实际的组播业务

概念	全称及中文解释	说明
Switch-MDT	Switch-Multicast Distribution Tree，切换组播分发树	以 Switch-Group 为组地址的 MDT，称为 Switch-MDT。下游存在接收者的 PE 加入 Switch-Group，形成一棵 Switch-MDT，入口 PE 使用 Switch-MDT 在公网中转发封装后的私网组播数据
Switch-Group	切换组	当某私网组播数据的流量达到或超过域值时，入口 PE 会为其分配一个独立的组播地址，称为 Switch-Group，并通知其它 PE 使用该地址在公网内转发该组播数据流量，从而实现 Switch-MDT 切换
Switch-Group-Pool	切换组地址池	在进行 Switch-MDT 切换时，从 Switch-Group-Pool 中选取一个地址（即 Switch-Group），从 PE 进入公网的私网组播报文将使用该地址进行封装。不同 VPN 的 Switch-Group-Pool 不能重叠，某 VPN 的 Switch-Group-Pool 中也不能含有其它 VPN 的 Share-Group
VRF	VPN Routing and Forwarding，VPN 路由转发	VPN 实例独立维护的单播路由转发表

2. MD VPN 实现机制

MD VPN 实现机制的要点列举如下：

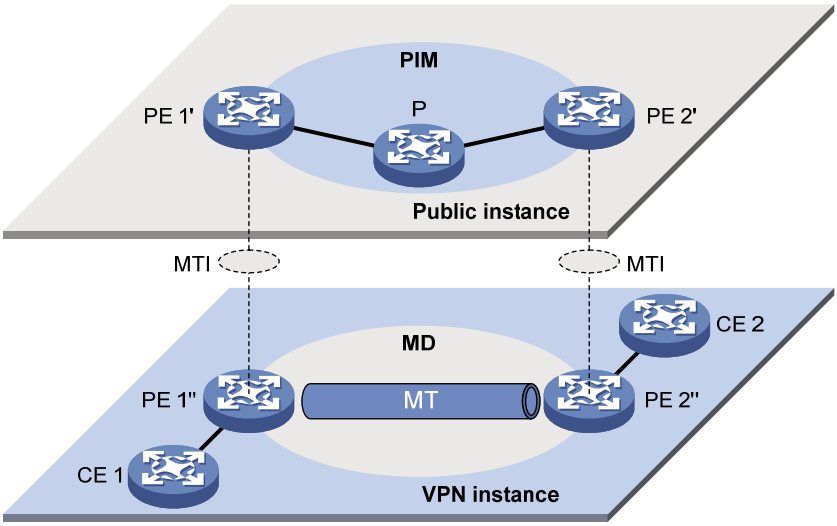
(1) 运营商构建的公共网络支持组播功能。PE 同时支持公网实例和多个 VPN 实例，每个实例各自运行相互间独立的 PIM。PE 与 CE 之间通过 VPN 实例进行私网组播传输；PE 与 P 之间则通过公网实例进行公网组播传输。

(2) MD在逻辑上表示某一特定VPN的私网组播数据在公网中的传播范围，在实际中则标识了网络中支持该VPN实例的所有PE设备。不同的VPN实例对应不同的MD。如 图 1-2所示，其中每个VPN实例平面的中央椭圆区域表示一个MD，服务于某个特定的VPN，在该VPN中传输的所有私网组播数据，都在此MD内传输。

(3) 在 MD 内部，私网数据通过 MT 进行传输。MT 传输过程为：本地 PE 将私网组播报文封装成公网组播数据报文，并在公网内进行组播转发，远端 PE 收到该报文后通过解封装将其还原成私网报文。

(4) 本地PE将私网数据通过MTI发出，而远端PE则从MTI接收私网数据。如 图 1-3所示，可以将MD比作一个私网数据的传输池，MTI则是MD的入/出口。本地PE将私网数据从入口（MTI）投入传输池，传输池自动将私网数据复制并传输到MD的所有出口（MTI），任何有需求的远端PE都可以从各自的出口（MTI）“打捞”私网数据。

图1-3 公网实例 PIM 与 VPN 实例 MD 对应关系示意图



(5) 一个 VPN 实例唯一指定一个 Share-Group 地址。私网数据信息对于公网来说是透明的，不论私网组播报文属于哪个组播组、是协议报文还是数据报文，PE 都统一将其封装为普通的公网组播数据报文，并以 Share-Group 作为其所属的公网组播组。之后，PE 将封装好的公网组播数据报文发送到公网中。

(6) 一个 Share-Group 唯一对应一个 MD，并利用公网资源唯一创建一棵 Share-MDT 以进行数据转发。在该 VPN 中传输的所有私网组播报文，无论从哪个 PE 进入公网，都经由此 Share-MDT 转发。

(7) 一个 Share-Group 唯一确定一个 Switch-Group-Pool 以备进行 Switch-MDT 切换。在进行 Switch-MDT 切换时，从 Switch-Group-Pool 中选取一个空闲的地址（即 Switch-Group），从 PE 进入公网的、流量达到或超过切换阈值的私网组播报文将使用该地址进行封装。

(8) 网络中所有 PE 都在监测 Share-MDT 的转发速率。当从某 PE 进入公网的数据转发速率超过一定阈值时，该 PE 将作为源端沿 Share-MDT 向其下游发出切换消息，使用 Switch-Group 在该 PE 和有接收需求的远端 PE 之间新建一棵 Switch-MDT。之后，进行 Switch-MDT 切换：即所有从该 PE 进入公网的私网组播数据流，不再使用 Share-Group 地址进行封装，而是被封装成公网的 Switch-Group 组播报文，从 Share-MDT 切换到新创建的 Switch-MDT 上。

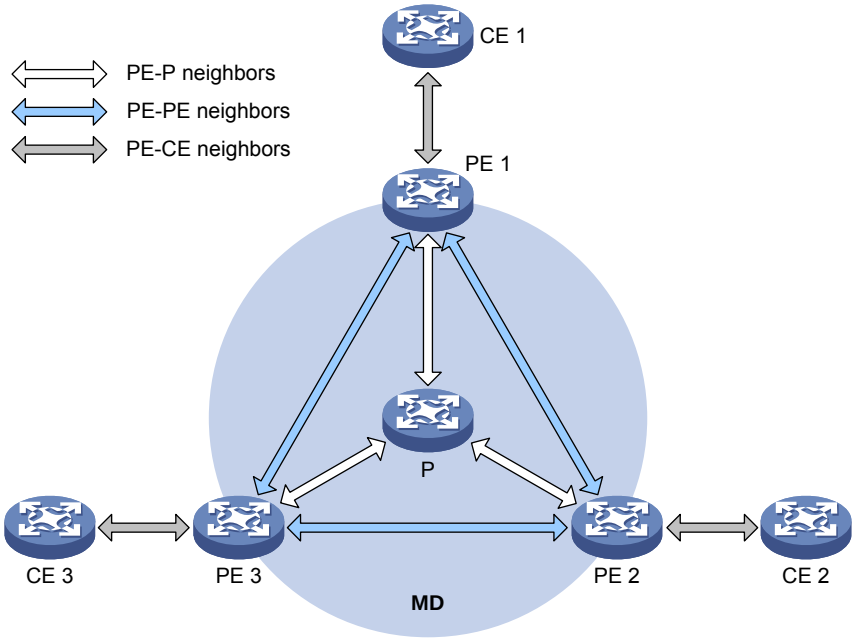
说明

- VPN、MD、MTI、Share-Group 与 Switch-Group-Pool 两两之间分别一一对应。
- 有关Switch-MDT切换的详细介绍，请参见“1.2.3 Switch-MDT切换”。

---

3. MD VPN 中的 PIM 邻居关系

图1-4 MD VPN 中的 PIM 邻居关系示意图



PIM邻居关系建立在直接相连且属于同一网段的两台或多台设备之间。如 图 1-4所示，在MD VPN中存在如下三种PIM邻居关系：

- PE-P 邻居关系：指 PE 上公网实例接口与链路对端 P 上的接口之间所建立的 PIM 邻居关系。
- PE-PE 邻居关系：指 PE 上的 VPN 实例通过 MTI 收到远端 PE 上的 VPN 实例发来的 PIM Hello 报文后所建立的邻居关系。
- PE-CE 邻居关系：指 PE 上绑定 VPN 实例的接口与链路对端 CE 上的接口之间所建立的 PIM 邻居关系。

1.1.4 协议规范

与组播 VPN 相关的协议规范有：

- RFC 4684：Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)
- draft-rosen-vpn-mcast-08：Multicast in MPLS/BGP IP VPNs

1.2 MD VPN 实现原理

本节介绍 MD VPN 技术的实现原理，包括组播分发树的构建和工作过程，以及跨 AS 的实现方式。

对于 VPN 实例来说，公网传输是透明的，私网数据的传输在 PE 上的 MTI 处完成了无缝连接：VPN 实例只知道将私网数据从 MTI 发出，然后远端就能从 MTI 接收。其实中间经历了复杂的公网传输过程，即 MDT 传输过程。

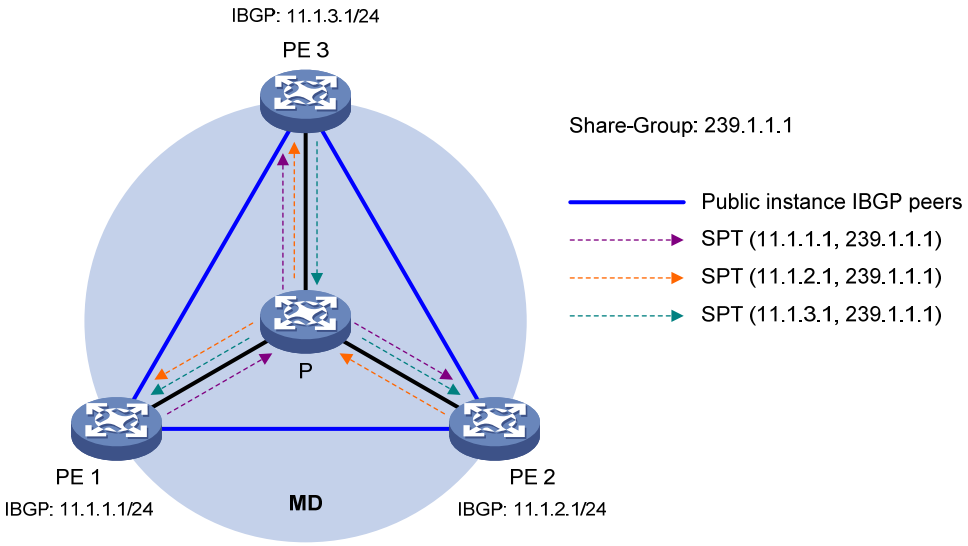
1.2.1 创建 Share-MDT

以 Share-Group 为组地址的 MDT，称为 Share-MDT。公网传输是透明的，私网数据传输在 PE 上的 MTI 处完成了无缝连接：VPN 实例只知道将私网数据从 MTI 发出，然后远端就能从 MTI 接收。其实中间经历了复杂的公网传输过程，即 MDT 传输过程。

公网中运行的组播路由协议可以是 PIM-DM、PIM-SM 或 PIM-SSM 中的一种。在这三种情况下，创建 Share-MDT 的过程是有区别的。

1. 在 PIM-DM 网络中创建 Share-MDT

图1-5 在 PIM-DM 网络中创建 Share-MDT



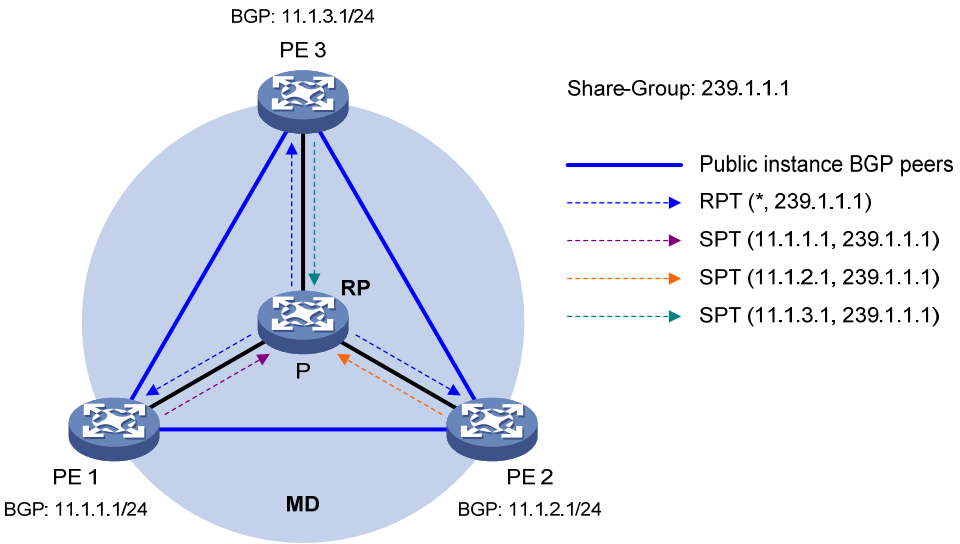
如 图 1-5所示，公网中运行PIM-DM，且PE 1、PE 2 和PE 3 都支持VPN实例A。Share-MDT的创建过程如下：PE 1 的公网实例以IBGP接口地址（即建立BGP对等体时所使用的接口地址）为组播源地址、Share-Group地址为组播组地址、其它支持VPN实例A的PE为组播组成员，在整个公网范围内发起扩散—剪枝过程，在公网沿途的各设备上分别创建（11.1.1.1，239.1.1.1）表项，从而形成一棵以PE 1 为根，PE 2 和PE 3 为叶的SPT。

同时 PE 2 和 PE 3 也各自发起类似的扩散—剪枝过程，最终在 MD 中形成三棵相互独立的 SPT。

在 PIM-DM 网络中，由这三棵相互独立的 SPT 共同组成了一棵 Share-MDT。

2. 在 PIM-SM 网络中创建 Share-MDT

图1-6 在 PIM-SM 网络中创建 Share-MDT



如 图 1-6所示，公网中运行PIM-SM，且PE 1、PE 2 和PE 3 都支持VPN实例A。Share-MDT的创建过程如下：

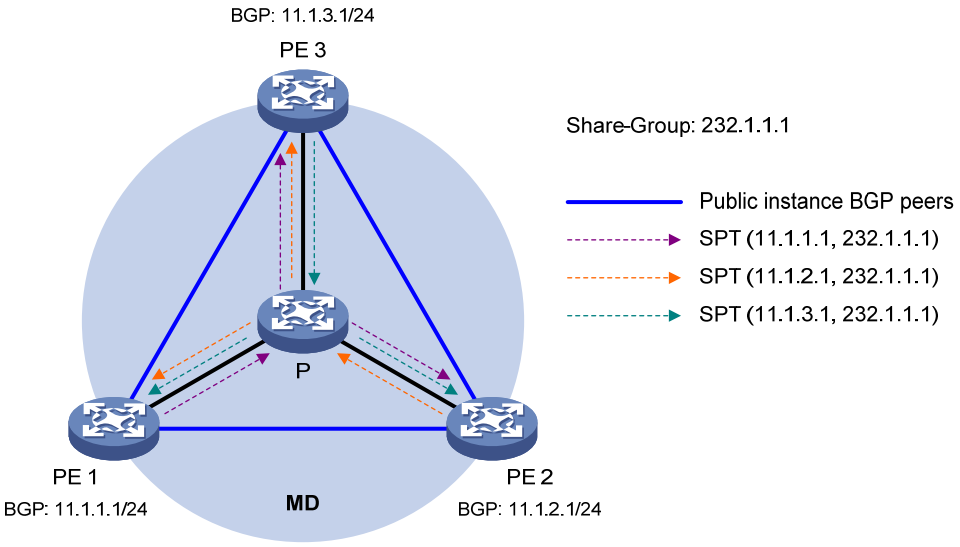
(1) PE 1 的公网实例向公网 RP 发起加入（Join），以 Share-Group 地址为组播组地址，在公网中沿途的各设备上分别创建（*，239.1.1.1）表项。同时，PE 2 和 PE 3 也各自发起类似的加入过程，最终在 MD 中形成一棵以公网 RP 为根，PE 1、PE 2 和 PE 3 为叶的 RPT。

(2) PE 1 的公网实例向公网 RP 发起注册（Register），以 BGP 接口地址为组播源地址、Share-Group 地址为组播组地址，在公网沿途的各设备上分别创建（11.1.1.1，239.1.1.1）表项。同时，PE 2 和 PE 3 也各自发起类似的注册过程，最终在 MD 中形成三棵相互独立的连接 PE 与 RP 的 SPT。

在 PIM-SM 网络中，由 RPT（*，239.1.1.1）和这三棵相互独立的 SPT 共同组成了一棵 Share-MDT。

3. 在 PIM-SSM 网络中创建 Share-MDT

图1-7 在 PIM-SSM 网络中创建 Share-MDT



如 图 1-7所示，公网中运行PIM-SSM，且PE 1、PE 2 和PE 3 都支持VPN实例A。Share-MDT的创建过程如下：PE 1 的公网实例分别向PE 2 和PE 3 的公网实例通知其本地的BGP MDT信息（包含BGP接口地址和Share-Group地址等信息），并互换BGP MDT路由信息；PE 2 和PE 3 收到来自PE 1 的BGP MDT信息后，分别向PE 1 的BGP接口地址逐跳发送订阅报文，在公网沿途的各设备上分别创建（11.1.1.1，232.1.1.1）表项，从而形成一棵以PE 1 为根、PE 2 和PE 3 为叶的SPT。

同时，分别以 PE 2 和 PE 3 为根的 SPT 形成过程与此类似，最终在 MD 中形成三棵相互独立的 SPT。

在 PIM-SSM 网络中，由这三棵相互独立的 SPT 共同组成了一棵 Share-MDT。

4. Share-MDT 的特点

综前，无论公网中运行的是 PIM-DM、PIM-SM 还是 PIM-SSM，Share-MDT 都具有以下特点：

- 网络中所有支持 VPN 实例 A 的 PE 都加入该 Share-MDT。
- 所有属于 VPN A 的私网组播报文（包括协议报文和数据报文）在进入公网后，均沿该 Share-MDT 向各 PE 转发，无论 PE 所连接的 Site 中是否存在接收者。

1.2.2 基于 Share-MDT 的传输

Share-MDT 构建完成后，就可以进行组播报文的传输。其中，组播协议报文和组播数据报文的传输过程是不同的，下面分别进行介绍。

1. 组播协议报文的传输

当私网组播接收者和组播源位于 VPN 网络中的不同 Site 时，组播协议报文必须跨越公网进行传输。协议报文在本地 PE 上被封装为普通的公网组播数据并沿 Share-MDT 进行传输，在远端 PE 上被解封装，从而继续进行正常的协议交互过程，最终建立一棵跨越公网的组播分发树。

注意

在属于同一 VPN 的所有接口（包括 PE 上绑定 VPN 实例的接口和 MTI）上必须运行相同模式的 PIM 协议。

---

(1) 当 VPN 网络中运行 PIM-DM 或 PIM-SSM 时，需要跨越公网创建 SPT。

(2) 当 VPN 网络中运行 PIM-SM 时：

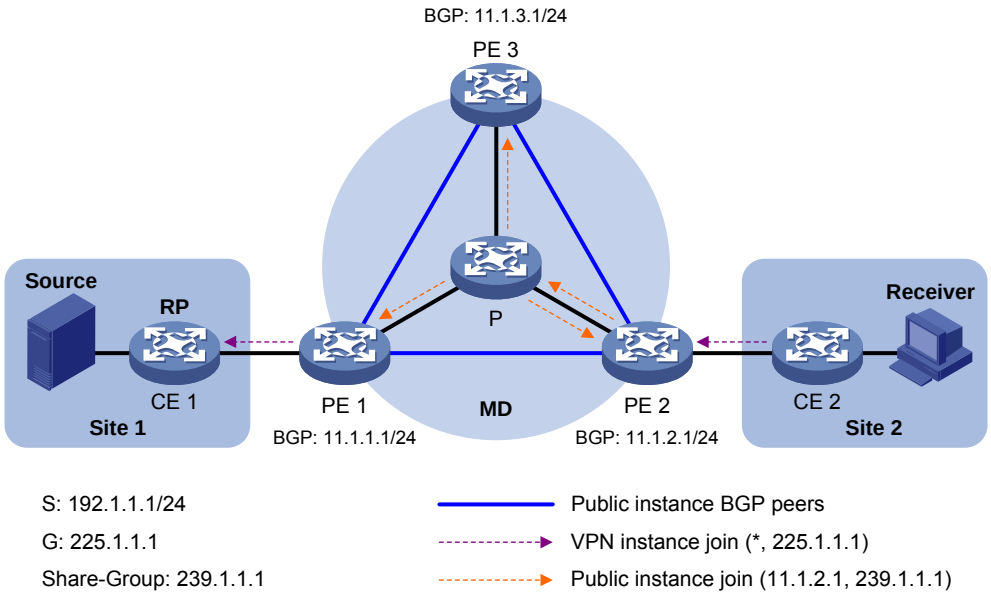
- 如果接收者与私网 RP 属于不同的 Site，需要跨越公网发起加入以创建 RPT。
- 如果组播源与私网 RP 属于不同的 Site，需要跨越公网发起注册以创建 SPT。

说明

下面以公网和 VPN 网络中均运行 PIM-SM、私网接收者跨越公网发起加入为例，介绍基于 Share-MDT 的组播协议报文的传输过程。

如 图 1-8所示，公网和VPN网络中分别运行PIM-SM，属于Site 2 的私网组播组G（225.1.1.1）的接收者（Receiver）与CE 2 相连；属于Site 1 的CE 1 作为G的RP；用于公网组播数据转发的Share-Group地址为 239.1.1.1。

图1-8 组播协议报文的传输过程



组播协议报文的交互过程如下：

(1) Receiver 向 CE 2 发送 IGMP 报告以加入组播组 G。CE 2 在本地创建（*，225.1.1.1）表项，同时向私网 RP（CE 1）发起加入。

(2) PE 2 上的 VPN 实例收到 CE 2 发来的加入消息（Join Message）后，在本地创建（*，225.1.1.1）表项，并指定上游接口为 MTI，然后 PE 2 将对该加入消息做进一步处理。这时，PE 2 上的 VPN 实例将认为加入消息已从 MTI 发出。

(3) PE 2 对加入消息以 GRE（Generic Routing Encapsulation，通用路由封装）方式进行封装，以 PE 2 的 BGP 接口地址为组播源地址、Share-Group 地址为组播组地址，转换成普通的公网组播数据报文（11.1.2.1，239.1.1.1），然后交由 PE 2 上的公网实例向公网进行转发。

(4) 组播数据报文（11.1.2.1，239.1.1.1）沿 Share-MDT 传输给各 PE 上的公网实例。各 PE 对其进行解封装，还原为发往私网 RP 的加入消息。然后，各 PE 分别检查该加入消息，如果发现私网 RP 在与其直连的 Site 中，则交由其上的 VPN 实例处理，否则丢弃该加入消息。

(5) PE 1 上的 VPN 实例收到加入消息后，认为是从 MTI 获得的。在本地创建（*，225.1.1.1）表项，并指定下游接口为 MTI、上游接口为朝向 CE 1 的接口。同时向私网 RP 发送加入消息。

(6) CE 1 收到 PE 1 上的 VPN 实例发来的加入消息后，在本地更新或创建（*，225.1.1.1）表项。至此跨越公网的 RPT 创建完成。


说明

有关 GRE 的详细介绍，请参见“VPN 分册”中的“GRE 配置”。

---

2. 组播数据报文的传输

当组播分发树创建完成后，私网组播数据沿组播分发树传输给各 Site 中的接收者。私网组播数据在本地 PE 上被封装为普通的公网组播数据并沿 Share-MDT 进行传输，在远端 PE 上被解封装，从而继续在私网内进行组播数据的传输。

(1) 当 VPN 网络中运行 PIM-DM 或 PIM-SSM 时，沿 SPT 跨越公网进行私网组播数据的传输。

(2) 当 VPN 网络中运行 PIM-SM 时（以下仅指 RPT 向 SPT 切换前的情形）：

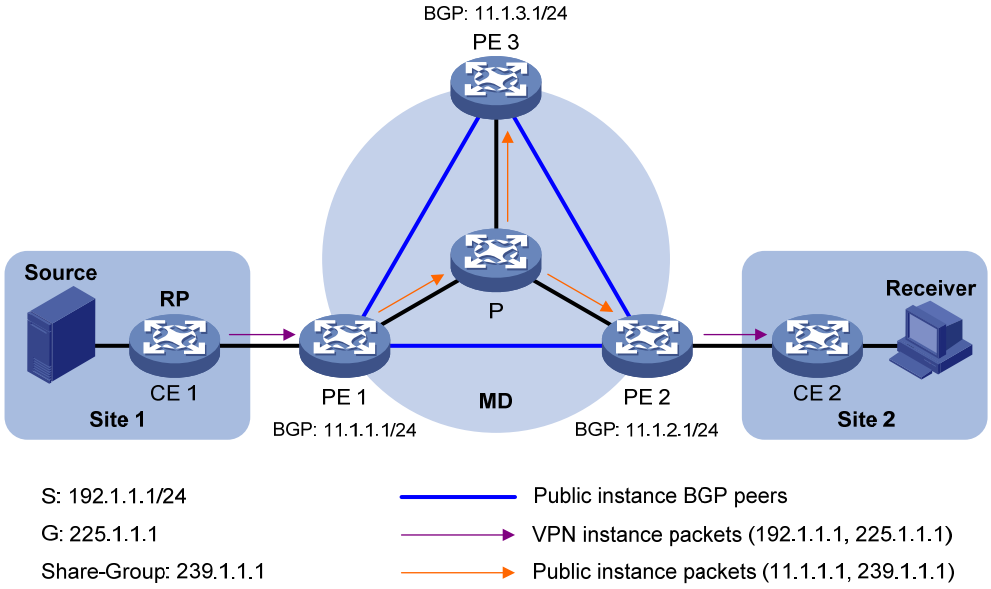
- 如果接收者与私网 RP 属于不同的 Site，沿 RPT 跨越公网进行私网组播数据的传输。
- 如果组播源与私网 RP 属于不同的 Site，沿 SPT 跨越公网进行私网组播数据的传输。


说明

- 有关 RPT 向 SPT 切换的详细介绍，请参见“IP 组播分册”中的“PIM 配置”。
- 下面以公网和 VPN 网络中均运行 PIM-DM、沿 SPT 跨越公网进行私网组播数据的传输为例，介绍基于 Share-MDT 的组播数据报文的传输过程。

如 图 1-9所示，公网和VPN网络中分别运行PIM-DM，属于Site 2 的私网组播组G（225.1.1.1）的接收者（Receiver）与CE 2 相连；属于Site 1 的组播源（Source）向G发送组播数据；用于公网组播数据转发的Share-Group地址为 239.1.1.1。

图1-9 组播数据报文的传输过程



私网组播数据跨越公网进行传输的过程如下：

(1) Source 发送私网组播数据（192.1.1.1，225.1.1.1）到 CE 1。

(2) CE 1 沿 SPT 将私网组播数据转发给 PE 1，PE 1 上的 VPN 实例查找转发表项。如果对应转发表项的出接口列表中包含 MTI，则 PE 1 将对该私网组播数据做进一步处理。这时，PE 1 上的 VPN 实例将认为私网组播数据已从 MTI 发出。

(3) PE 1 对该私网组播数据以 GRE 方式进行封装，以 PE 1 的 BGP 接口地址为组播源地址、Share-Group 地址为组播组地址，转换成普通的公网组播数据报文（11.1.1.1，239.1.1.1），然后交由 PE 1 上的公网实例向公网进行转发。

(4) 组播数据报文（11.1.1.1，239.1.1.1）沿 Share-MDT 传输给各 PE 上的公网实例。各 PE 对其进行解封装，还原为私网组播数据，然后交由相应的 VPN 实例处理。如果该 PE 上存在 SPT 的下游接口，则沿 SPT 转发该私网组播数据，否则将其丢弃。

(5) PE 2 上的 VPN 实例查找转发表项，最终将私网组播数据送达 Receiver。至此跨越公网的私网组播数据传输完成。

1.2.3 Switch-MDT 切换

1. 由 Share-MDT 向 Switch-MDT 切换

在公网中通过 Share-MDT 传送组播数据时，组播报文被传输到支持同一 VPN 实例的所有 PE 上，无论该 PE 所连接的 Site 内是否存在接收者。当私网中组播数据的传输速率比较大时，可能在公网中造成数据的泛滥。这样即浪费网络带宽，又增加了 PE 的处理负担。

为了解决上述问题，MD 方案对此进行了优化：为进入公网的大流量私网组播数据，在连接有私网接收者和私网组播源的各 PE 之间，建立起专用的 Switch-MDT。然后将该组播数据流从 Share-MDT 切换到 Switch-MDT，从而实现按需进行组播。

从 Share-MDT 向 Switch-MDT 切换的过程如下：

(1) 源端 PE（如 PE 1）周期性地检测私网组播数据的转发速率。发起从 Share-MDT 向 Switch-MDT 的切换要求私网组播数据通过了由 Share-MDT 向 Switch-MDT 切换的 ACL 规则的过滤，否则仍沿 Share-MDT 转发；

(2) PE 1 从 Switch-Group-Pool 中分配一个空闲的 Switch-Group 地址，沿 Share-MDT 向所有下游 PE 发送切换消息。该消息中包括私网组播源地址、私网组播组地址和 Switch-Group 地址。

(3) 其它 PE 收到该消息后，检查自己是否连接有该私网组播数据的接收者：如果有，则加入以 PE 1 为根的 Switch-MDT；如果没有，则将该消息缓存起来，等待有接收者时再加入 Switch-MDT。

(4) 当 PE 1 发送切换消息一定时间后，PE 1 停止使用 Share-Group 地址对私网组播数据进行封装，并改用 Switch-Group 地址进行封装，组播数据沿 Switch-MDT 向下分发。

(5) 当 Share-MDT 切换到 Switch-MDT 之后，PE 1 会周期性地发送切换消息，以便后续有 PE 加入 Switch-MDT。当某下游 PE 不再连接有接收者时，可以退出 Switch-MDT。



说明

Switch-MDT 和 Share-MDT 都是同一个 MD 中的转发隧道。Share-MDT 由 Share-Group 地址唯一确定；Switch-MDT 则由 Switch-Group 地址唯一确定。每个 Share-Group 地址关联一组 Switch-Group 地址（即 Switch-Group-Pool）。

---

2. 由 Switch-MDT 向 Share-MDT 反向切换

当私网组播数据切换到 Switch-MDT 之后，由于情况变化而导致了不满足切换条件，PE 1 就会把此私网组播数据从 Switch-MDT 反向切换回 Share-MDT，反向切换的触发条件包括（满足其一即可）：

- 当更改了 Switch-Group-Pool 后，用于私网组播数据封装的 Switch-Group 地址被排除在新地址池之外。
- 当控制私网组播数据由 Share-MDT 向 Switch-MDT 切换的 ACL 规则发生了变化，从而导致私网组播数据不能通过新 ACL 规则的过滤。

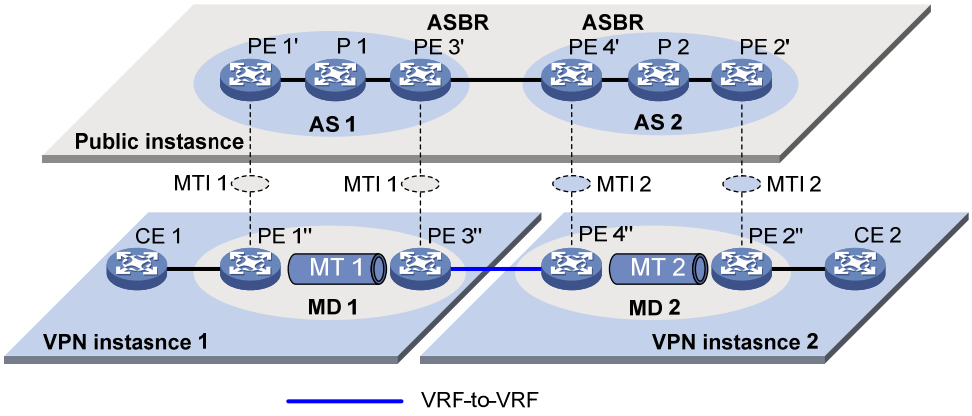
1.2.4 跨 AS 的 MD VPN

当整个 VPN 跨越多个 AS（Autonomous System，自治系统）时，需要连通分布在不同 AS 内的各 VPN 节点。跨 AS 的 MD VPN 有以下两种实现方式：

1. VRF-to-VRF 连接方式

如 图 1-10所示，VPN跨越了AS 1 和AS 2 两个自治系统，PE 3 和PE 4 分别是AS 1 和AS 2 的ASBR。PE 3 和PE 4 通过各自的VPN实例相连，并互把对方视为CE设备。

图1-10 VRF-to-VRF 连接方式示意图



采用 VRF-to-VRF 方式时，需在每个 AS 内各建立一个独立的 MD，在各 MD 之间实现私网组播数据跨 AS 的传输。

说明

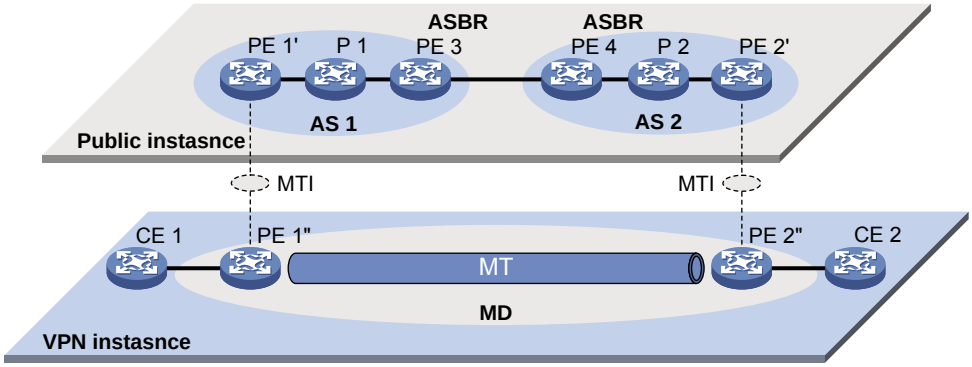
由于 ASBR 之间转发的只是私网组播数据，因此各 AS 内部运行的公网 PIM 模式可以不同，但属于同一 VPN 的所有接口（包括 ASBR 上绑定 VPN 实例的接口）上必须运行统一的 PIM 模式（PIM-DM、PIM-SM 或 PIM-SSM）。

---

2. Multi-hop EBGP 连接方式

如 图 1-11所示，VPN跨越了AS 1 和AS 2 两个自治系统，PE 3 和PE 4 分别是AS 1 和AS 2 的ASBR。PE 3 和PE 4 通过各自的公网实例相连，并互把对方视为P设备。

图1-11 Multi-hop EBGP 连接方式



采用 Multi-hop EBGP 方式时，只需在所有 AS 内统一建立一个 MD 即可，在该 MD 内部实现公网组播数据跨 AS 的传输。

1.3 组播 VPN 配置任务简介

表1-2 组播 VPN 配置任务简介

配置任务		说明	详细配置
配置 MD VPN	使能 VPN 实例中的 IP 组播路由	必选	1.4.2
	配置 Share-Group 并绑定 MTI	必选	1.4.3
	配置 Switch-MDT 切换参数	可选	1.4.4
	打开 Switch-Group 重用日志输出开关	可选	1.4.5
配置 BGP MDT	配置 BGP MDT 对等体/对等体组	必选	1.5.2
	配置 BGP MDT 路由反射器	可选	1.5.3

1.4 配置 MD VPN

1.4.1 配置准备

在配置 MD VPN 之前，需完成以下任务：

- 配置任一单播路由协议，实现域内网络层互通
- 配置 MPLS L3VPN
- 配置 PIM-DM、PIM-SM 或 PIM-SSM

在配置 MD VPN 之前，需准备以下数据：

- VPN 实例的名称和 RD（Route Distinguisher，路由标识符）
- Share-Group 组地址和 MTI 接口编号
- Switch-Group-Pool 的范围和切换条件 ACL
- 由 Share-MDT 向 Switch-MDT 切换的延迟时间
- 由 Switch-MDT 向 Share-MDT 反向切换的延迟时间

1.4.2 使能 VPN 实例中的 IP 组播路由

在配置 MD VPN 的各项功能之前，必须先使能 VPN 实例中的 IP 组播路由。

表1-3 使能 VPN 实例中的 IP 组播路由

操作	命令	说明
进入系统视图	**system-view**	-
创建 VPN 实例，并进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 VPN 实例的 RD	**route-distinguisher** *route-distinguisher*	必选 缺省情况下，VPN 实例没有 RD
使能 IP 组播路由	**multicast routing-enable**	必选 缺省情况下，IP 组播路由处于关闭状态

说明

- 有关 **ip vpn-instance** 和 **route-distinguisher** 命令的详细介绍，请参见“MPLS 分册”中的“MPLS L3VPN 命令”。
- 有关 **multicast routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“组播路由与转发命令”。

1.4.3 配置 Share-Group 并绑定 MTI

通过在 PE 上启动多实例，可以为多个 VPN 服务。不同 PE 上相同的 VPN 实例需要配置相同的 Share-Group 地址。配置 Share-Group 并绑定 MTI 时，系统会自动创建 MTI、将 Share-Group 绑定到 MTI，并将 MTI 绑定到本 VPN 实例。

表1-4 配置 Share-Group 并绑定 MTI

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
指定 Share-Group 并为本 VPN 实例绑定 MTI	**multicast-domain share-group** *group-address* **binding mtunnel** *mtunnel-number*	必选 缺省情况下，没有指定 Share-Group，也没有为本 VPN 实例绑定 MTI

说明

- 当使用命令 **peer connect-interface** 配置了 BGP 对等体之后，MTI 接口将自动获取 **connect-interface** 的地址为其 IP 地址，且该地址不能在 VPN 中再被使用，否则 MTI 接口将无法获取 IP 地址；在一台设备上配置多个 BGP 对等体时应指定相同的 **connect-interface**，否则 MTI 接口也无法获取 IP 地址。有关 **peer connect-interface** 命令的详细介绍，请参见“IP 路由分册”中的“BGP 命令”。
- 只有当 VPN 实例内至少一个接口上使能了 PIM 协议，MTI 接口上的 PIM 协议才能被使能；当 VPN 实例内的所有接口上都关闭了 PIM 协议后，MTI 接口上的 PIM 协议也将被同时关闭。

1.4.4 配置 Switch-MDT 切换参数

在某些情况下，私网组播数据的转发速率会在切换阈值上下振荡。为了避免组播数据流在 Share-MDT 与 Switch-MDT 之间进行频繁切换：

- 当转发速率高于阈值后并不立即切换，而是等待 Switch-Delay 时间（即由 Share-MDT 向 Switch-MDT 切换的延迟时间）。在这段时间内如果转发速率始终高于阈值，则切换至 Switch-MDT，否则继续使用 Share-MDT 进行转发；
- 当转发速率低于阈值后也不立即切换，而是等待 Switch-Holddown 时间（即由 Switch-MDT 向 Share-MDT 反向切换的延迟时间）。在这段时间内如果转发速率始终低于阈值，则切换回 Share-MDT，否则继续使用 Switch-MDT 进行转发。

表1-5 配置 Switch-MDT 切换参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
配置 Switch-Group-Pool 的范围和切换条件	**multicast-domain switch-group-pool** *switch-group-pool* { *mask* | *mask-length* } [**acl** *acl-number*] *	必选 缺省情况下，没有指定 Switch-Group-Pool，也永不向 Switch-MDT 切换
配置由 Share-MDT 向 Switch-MDT 切换的延迟时间	**multicast-domain switch-delay** *switch-delay*	可选 缺省情况下，由 Share-MDT 向 Switch-MDT 切换的延迟时间为 5 秒
配置由 Switch-MDT 向 Share-MDT 反向切换的延迟时间	**multicast-domain holddown-time** *interval*	可选 缺省情况下，由 Switch-MDT 向 Share-MDT 反向切换的延迟时间为 60 秒

1.4.5 打开 Switch-Group 重用日志输出开关

在源 PE 上的 VPN 实例中，如果需要切换的私网组播数据流的个数超过了 Switch-Group-Pool 中组地址的个数时，可以重复使用该地址池中的组地址。通过打开 Switch-Group 重用日志输出开关可以记录组地址重用的日志信息。

表1-6 打开 Switch-Group 重用日志输出开关

操作	命令	说明
进入系统视图	**system-view**	-
进入 VPN 实例视图	**ip vpn-instance** *vpn-instance-name*	-
打开 Switch-Group 重用日志输出开关	**multicast-domain log switch-group-reuse**	必选 缺省情况下，Switch-Group 重用日志输出开关处于关闭状态

说明

- 组地址重用日志信息的级别为 informational，隶属于 MD 模块。有关日志信息的详细介绍，请参见“系统分册”中的“信息中心配置”。
- 打开 Switch-Group 重用日志输出开关后，会生成组地址重用日志信息并发往信息中心，该日志信息的输出规则可在信息中心中进行配置。有关信息中心的具体配置，请参见“系统分册”中的“信息中心配置”。

1.5 配置 BGP MDT

当在公网中运行 PIM-SSM 时，需要进行 BGP MDT 的配置。

1.5.1 配置准备

在配置 BGP MDT 之前，需完成以下任务：

- 配置 MPLS L3VPN

- 配置 BGP 基本功能

在配置 BGP MDT 之前，需准备以下数据：

- 路由反射器的集群 ID

1.5.2 配置 BGP MDT 对等体/对等体组

通过配置 BGP MDT 对等体/对等体组，使各 PE 间可通过交换 BGP MDT 路由信息以获取远端 PE 的地址来建立 Share-MDT。

表1-7 配置 BGP MDT 对等体/对等体组

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-MDT 子地址族视图	**ipv4-family mdt**	-
激活 BGP MDT 对等体/对等体组	**peer** { *group-name* | *ip-address* } **enable**	必选 缺省情况下，BGP MDT 对等体/对等体组处于未激活状态
向 BGP MDT 对等体组中添加 BGP MDT 对等体	**peer** *ip-address* **group** *group-name*	可选 缺省情况下， BGP MDT 对等体也不属于任何 BGP MDT 对等体组

说明

BGP MDT 对等体/对等体组是指在 BGP-MDT 子地址族视图下创建的对等体/对等体组。

1.5.3 配置 BGP MDT 路由反射器

为保证位于同一 AS 内的 BGP MDT 对等体间的连通性，需在对等体之间建立全连接关系。而当对等体的数目很多时，建立全连接的开销很大，使用路由反射器则可以解决这个问题。在配置了路由反射器之后，由其它路由器作为客户机与路由反射器之间建立 BGP MDT 连接，路由反射器在客户机之间传递（反射）BGP MDT 路由信息，从而使各客户机之间无需建立全连接。但如果客户机之间已经是全连接的，则可以通过禁止客户机间的路由反射以减少开销。

路由反射器及其客户机共同组成了一个集群。通常，一个集群中只有一个路由反射器，并通过其 Router ID 来识别该集群。为了增强网络的可靠性，可在一个集群中配置多个路由反射器，此时应为每个路由反射器配置相同的集群 ID，以避免产生路由环路。

表1-8 配置 BGP MDT 路由反射器

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP-MDT 子地址族视图	**ipv4-family mdt**	-

操作	命令	说明
配置本设备作为路由反射器，并指定其路由反射客户	**peer** { *group-name* | *ip-address* } **reflect-client**	必选 缺省情况下，没有配置路由反射器及其客户
禁止客户到客户的路由反射	**undo reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
配置路由反射器的集群 ID	**reflector cluster-id** { *cluster-id* | *ip-address* }	可选 缺省情况下，路由反射器的 Router ID 就是其集群 ID

1.6 组播 VPN 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后组播 VPN 的运行情况，通过查看显示信息验证配置的效果。

在用户试图下执行 **reset** 命令可以复位 BGP MDT 的连接。

表1-9 组播 VPN 显示和维护

操作	命令
查看 MD 中指定 VPN 实例的 Share-Group 组信息	**display multicast-domain vpn-instance** *vpn-instance-name* **share-group** { **local** | **remote** }
查看 MD 中指定 VPN 实例收到的 Switch-Group 信息	**display multicast-domain vpn-instance** *vpn-instance-name* **switch-group receive** [**brief** | [**active** | **group** *group-address* | **sender** *source-address* | *vpn-source-address* [**mask** { *mask-length* | *mask* }] | *vpn-group-address* [**mask** { *mask-length* | *mask* }]] *]
查看 MD 中指定 VPN 实例发送的 Switch-Group 信息	**display multicast-domain vpn-instance** *vpn-instance-name* **switch-group send** [**group** *group-address* | **reuse** *interval* | *vpn-source-address* [**mask** { *mask-length* | *mask* }] | *vpn-group-address* [**mask** { *mask-length* | *mask* }]] *
查看 BGP MDT 对等体组的信息	**display bgp mdt group** [*group-name*]
查看 BGP MDT 对等体的信息	**display bgp mdt peer** [[*ip-address*] **verbose**]
查看 BGP MDT 路由信息	**display bgp mdt** { **all** | **route-distinguisher** *route-distinguisher* } **routing-table** [*ip-address* [*mask* | *mask-length*]]
复位 BGP MDT 连接	**reset bgp mdt** { *as-number* | *ip-address* | **all** | **external** | **group** *group-name* | **internal** }

1.7 组播 VPN 典型配置举例

1.7.1 单 AS 内 MD VPN 配置举例

1. 组网需求

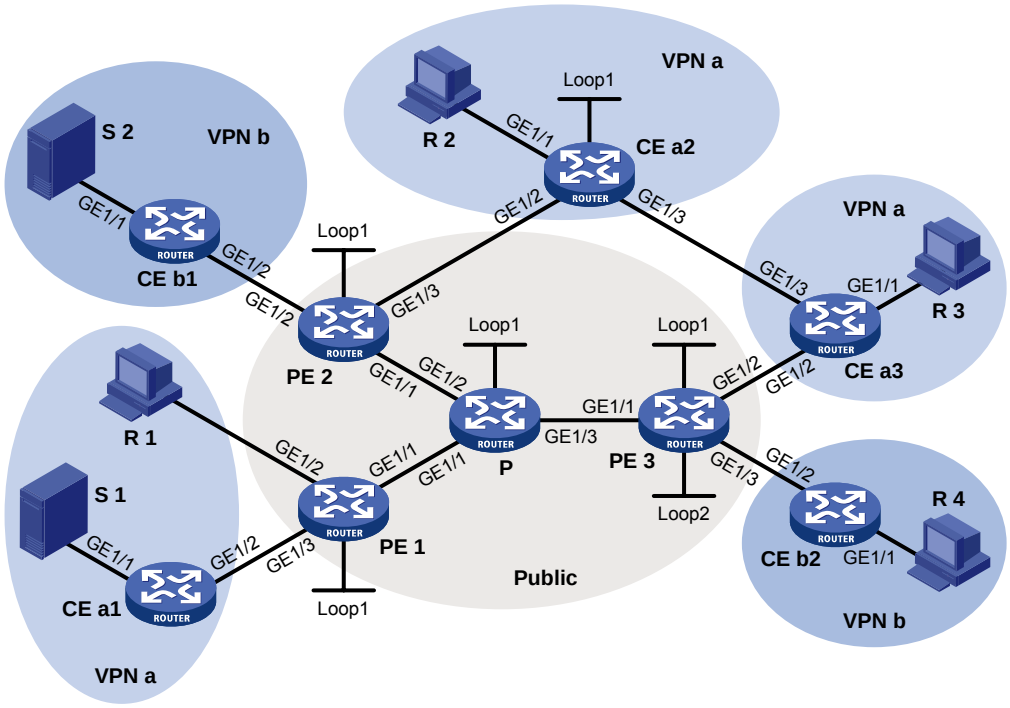
组网需求如 表 1-10所示。

表1-10 单 AS 内 MD VPN 配置组网需求

项目	组网需求
组播源和接收者	● VPN a 中的组播源为 S 1，接收者为 R 1、R 2 和 R 3 ● VPN b 中的组播源为 S 2，接收者为 R 4 ● VPN a 中的 Share-Group 地址为 239.1.1.1，Switch-Group-Pool 范围为 225.2.2.0～225.2.2.15 ● VPN b 中的 Share-Group 地址为 239.2.2.2，Switch-Group-Pool 范围为 225.4.4.0～225.4.4.15
PE 上各接口所属的 VPN 实例	● PE 1：GigabitEthernet1/2 和 GigabitEthernet1/3 属于 VPN 实例 a，GigabitEthernet1/1 和 Loopback1 属于公网实例 ● PE 2：GigabitEthernet1/2 属于 VPN 实例 b，GigabitEthernet1/3 属于 VPN 实例 a，GigabitEthernet1/1 和 Loopback1 属于公网实例 ● PE 3：GigabitEthernet1/2 属于 VPN 实例 a，GigabitEthernet1/3 和 Loopback2 属于 VPN 实例 b，GigabitEthernet1/1 和 Loopback1 接口属于公网实例
单播路由协议和 MPLS	● 在公网中配置 OSPF，在各 PE 与 CE 之间配置 RIP ● 在 PE 1、PE 2 和 PE 3 各自的 Loopback1 接口之间建立 BGP 对等体连接并传递所有私网路由 ● 在公网中配置 MPLS
IP 组播路由功能	● 在 P 上使能 IP 组播路由 ● 在 PE 1、PE 2 和 PE 3 的公网实例中均使能 IP 组播路由 ● 在 PE 1、PE 2 和 PE 3 的 VPN 实例 a 中均使能 IP 组播路由 ● 在 PE 2 和 PE 3 的 VPN 实例 b 中均使能 IP 组播路由 ● 在 CE a1、CE a2、CE a3、CE b1 和 CE b2 上均使能 IP 组播路由
IGMP 功能	● 在 PE 1 的 GigabitEthernet1/2 接口上使能 IGMPv2 ● 在 CE a2、CE a3 和 CE b2 各自的 GigabitEthernet1/1 接口上均使能 IGMPv2
PIM 功能	● 在 P 的所有接口上均使能 PIM-SM ● 在 PE 1、PE 2 和 PE 3 的所有公网和私网接口上均使能 PIM-SM ● 在 CE a1、CE a2、CE a3、CE b1 和 CE b2 的所有接口上均使能 PIM-SM ● P 的 Loopback1 接口为公网的 C-BSR 和 C-RP（服务于所有组播组） ● CE a2 的 Loopback1 接口为 VPN a 的 C-BSR 和 C-RP（服务于所有组播组） ● PE 3 的 Loopback2 接口为 VPN b 的 C-BSR 和 C-RP（服务于所有组播组）

2. 组网图

图1-12 单 AS 内 MD VPN 配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
S 1	-	10.110.7.2/24	PE 3	GE1/1	192.168.8.1/24
S 2	-	10.110.8.2/24		GE1/2	10.110.5.1/24
R 1	-	10.110.1.2/24		GE1/3	10.110.6.1/24
R 2	-	10.110.9.2/24		Loop1	1.1.1.3/32
R 3	-	10.110.10.2/24		Loop2	33.33.33.33/32
R 4	-	10.110.11.2/24	CE a1	GE1/1	10.110.7.1/24
P	GE1/1	192.168.6.2/24		GE1/2	10.110.2.2/24
	GE1/2	192.168.7.2/24	CE a2	GE1/1	10.110.9.1/24
	GE1/3	192.168.8.2/24		GE1/2	10.110.4.2/24
	Loop1	2.2.2.2/32		GE1/3	10.110.12.1/24
PE 1	GE1/1	192.168.6.1/24		Loop1	22.22.22.22/32
	GE1/2	10.110.1.1/24	CE a3	GE1/1	10.110.10.1/24
	GE1/3	10.110.2.1/24		GE1/2	10.110.5.2/24
	Loop1	1.1.1.1/32		GE1/3	10.110.12.2/24
PE 2	GE1/1	192.168.7.1/24	CE b1	GE1/1	10.110.8.1/24
	GE1/2	10.110.3.1/24		GE1/2	10.110.3.2/24
	GE1/3	10.110.4.1/24	CE b2	GE1/1	10.110.11.1/24
	Loop1	1.1.1.2/32		GE1/2	10.110.6.2/24

3. 配置步骤

(1) 配置 PE 1

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE1> system-view
[PE1] router id 1.1.1.1
[PE1] multicast routing-enable
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

创建 VPN 实例 a，为其配置 RD，并创建其出、入口路由。

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
```

使能 VPN 实例 a 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE1-vpn-instance-a] multicast routing-enable
[PE1-vpn-instance-a] multicast-domain share-group 239.1.1.1 binding mtunnel 0
[PE1-vpn-instance-a] multicast-domain switch-group-pool 225.2.2.0 28
[PE1-vpn-instance-a] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE1] interface GigabitEthernet 1/1
[PE1-GigabitEthernet1/1] ip address 192.168.6.1 24
[PE1-GigabitEthernet1/1] pim sm
[PE1-GigabitEthernet1/1] mpls
[PE1-GigabitEthernet1/1] mpls ldp
[PE1-GigabitEthernet1/1] quit
```

把接口 GigabitEthernet1/2 绑定到 VPN 实例 a，配置 IP 地址，并使能 IGMP 和 PIM-SM。

```
[PE1] interface GigabitEthernet 1/2
[PE1-GigabitEthernet1/2] ip binding vpn-instance a
[PE1-GigabitEthernet1/2] ip address 10.110.1.1 24
[PE1-GigabitEthernet1/2] igmp enable
[PE1-GigabitEthernet1/2] pim sm
[PE1-GigabitEthernet1/2] quit
```

把接口 GigabitEthernet1/3 绑定到 VPN 实例 a，配置 IP 地址，并使能 PIM-SM。

```
[PE1] interface GigabitEthernet 1/3
[PE1-GigabitEthernet1/3] ip binding vpn-instance a
[PE1-GigabitEthernet1/3] ip address 10.110.2.1 24
[PE1-GigabitEthernet1/3] pim sm
[PE1-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

配置 BGP 协议。

```
[PE1] bgp 100
[PE1-bgp] group vpn-g internal
[PE1-bgp] peer vpn-g connect-interface loopback 1
[PE1-bgp] peer 1.1.1.2 group vpn-g
[PE1-bgp] peer 1.1.1.3 group vpn-g
[PE1–bgp] ipv4-family vpn-instance a
[PE1-bgp-a] import-route rip 2
```

```
[PE1-bgp-a] import-route direct
[PE1-bgp-a] quit
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer vpn-g enable
[PE1-bgp-af-vpnv4] peer 1.1.1.2 group vpn-g
[PE1-bgp-af-vpnv4] peer 1.1.1.3 group vpn-g
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

当 PE 1 上配置了 BGP 对等体之后，MTI0 接口将自动获得 IP 地址，该地址与配置 BGP 对等体时所指定的 Loopback 接口的 IP 地址相同；MTI0 接口上运行的 PIM 协议类型也与其所属的 VPN 实例 a 中运行的 PIM 协议类型相同。

配置 OSPF 协议。

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.255.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

配置 RIP 协议。

```
[PE1] rip 2 vpn-instance a
[PE1-rip-2] network 10.0.0.0
[PE1-rip-2] import-route bgp
[PE1-rip-2] return
```

(2) 配置 PE 2

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE2> system-view
[PE2] router id 1.1.1.2
[PE2] multicast routing-enable
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

创建 VPN 实例 b，为其配置 RD，并创建其出、入口路由。

```
[PE2] ip vpn-instance b
[PE2-vpn-instance-b] route-distinguisher 200:1
[PE2-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE2-vpn-instance-b] vpn-target 200:1 import-extcommunity
```

使能 VPN 实例 b 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE2-vpn-instance-b] multicast routing-enable
[PE2-vpn-instance-b] multicast-domain share-group 239.2.2.2 binding mtunnel 1
[PE2-vpn-instance-b] multicast-domain switch-group-pool 225.4.4.0 28
[PE2-vpn-instance-b] quit
```

创建 VPN 实例 a，为其配置 RD，并创建其出、入口路由。

```
[PE2] ip vpn-instance a
[PE2-vpn-instance-a] route-distinguisher 100:1
[PE2-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE2-vpn-instance-a] vpn-target 100:1 import-extcommunity
```

使能 VPN 实例 a 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE2-vpn-instance-a] multicast routing-enable
[PE2-vpn-instance-a] multicast-domain share-group 239.1.1.1 binding mtunnel 0
[PE2-vpn-instance-a] multicast-domain switch-group-pool 225.2.2.0 28
[PE2-vpn-instance-a] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE2] interface GigabitEthernet 1/1
[PE2-GigabitEthernet1/1] ip address 192.168.7.1 24
[PE2-GigabitEthernet1/1] pim sm
[PE2-GigabitEthernet1/1] mpls
[PE2-GigabitEthernet1/1] mpls ldp
[PE2-GigabitEthernet1/1] quit
```

把接口 GigabitEthernet1/2 绑定到 VPN 实例 b，配置 IP 地址，并使能 PIM-SM。

```
[PE2] interface GigabitEthernet 1/2
[PE2-GigabitEthernet1/2] ip binding vpn-instance b
[PE2-GigabitEthernet1/2] ip address 10.110.3.1 24
[PE2-GigabitEthernet1/2] pim sm
[PE2-GigabitEthernet1/2] quit
```

把接口 GigabitEthernet1/3 绑定到 VPN 实例 a，配置 IP 地址，并使能 PIM-SM。

```
[PE2] interface GigabitEthernet 1/3
[PE2-GigabitEthernet1/3] ip binding vpn-instance a
[PE2-GigabitEthernet1/3] ip address 10.110.4.1 24
[PE2-GigabitEthernet1/3] pim sm
[PE2-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 1.1.1.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

配置 BGP 协议。

```
[PE2] bgp 100
[PE2-bgp] group vpn-g internal
[PE2-bgp] peer vpn-g connect-interface loopback 1
[PE2-bgp] peer 1.1.1.1 group vpn-g
[PE2-bgp] peer 1.1.1.3 group vpn-g
[PE2–bgp] ipv4-family vpn-instance a
[PE2-bgp-a] import-route rip 2
[PE2-bgp-a] import-route direct
```

```
[PE2-bgp-a] quit
[PE2–bgp] ipv4-family vpn-instance b
[PE2-bgp-b] import-route rip 3
[PE2-bgp-b] import-route direct
[PE2-bgp-b] quit
[PE2–bgp] ipv4-family vpnv4
[PE2–bgp-af-vpnv4] peer vpn-g enable
[PE2-bgp-af-vpnv4] peer 1.1.1.1 group vpn-g
[PE2–bgp-af-vpnv4] peer 1.1.1.3 group vpn-g
[PE2–bgp-af-vpnv4] quit
[PE2–bgp] quit
```

当 PE 2 上配置了 BGP 对等体之后，MTI0 和 MTI1 接口将自动获得各自的 IP 地址，该地址与配置 BGP 对等体时所指定的 Loopback 接口的 IP 地址相同；MTI0 和 MTI1 接口上运行的 PIM 协议类型也分别与其所属的 VPN 实例 a 和 b 中运行的 PIM 协议类型相同。

配置 OSPF 协议。

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.255.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

配置 RIP 协议。

```
[PE2] rip 2 vpn-instance a
[PE2-rip-2] network 10.0.0.0
[PE2-rip-2] import-route bgp
[PE2-rip-2] quit
[PE2] rip 3 vpn-instance b
[PE2-rip-3] network 10.0.0.0
[PE2-rip-3] import-route bgp
[PE2-rip-3] return
```

(3) 配置 PE 3

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE3> system-view
[PE3] router id 1.1.1.3
[PE3] multicast routing-enable
[PE3] mpls lsr-id 1.1.1.3
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
```

创建 VPN 实例 a，为其配置 RD，并创建其出、入口路由。

```
[PE3] ip vpn-instance a
[PE3-vpn-instance-a] route-distinguisher 100:1
[PE3-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE3-vpn-instance-a] vpn-target 100:1 import-extcommunity
```

使能 VPN 实例 a 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE3-vpn-instance-a] multicast routing-enable
[PE3-vpn-instance-a] multicast-domain share-group 239.1.1.1 binding mtunnel 0
[PE3-vpn-instance-a] multicast-domain switch-group-pool 225.2.2.0 28
[PE3-vpn-instance-a] quit
```

创建 VPN 实例 b，为其配置 RD，并创建其出、入口路由。

```
[PE3] ip vpn-instance b
[PE3-vpn-instance-b] route-distinguisher 200:1
[PE3-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE3-vpn-instance-b] vpn-target 200:1 import-extcommunity
```

使能 VPN 实例 b 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE3-vpn-instance-b] multicast routing-enable
[PE3-vpn-instance-b] multicast-domain share-group 239.2.2.2 binding mtunnel 1
[PE3-vpn-instance-b] multicast-domain switch-group-pool 225.4.4.0 28
[PE3-vpn-instance-b] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE3] interface GigabitEthernet 1/1
[PE3-GigabitEthernet1/1] ip address 192.168.8.1 24
[PE3-GigabitEthernet1/1] pim sm
[PE3-GigabitEthernet1/1] mpls
[PE3-GigabitEthernet1/1] mpls ldp
[PE3-GigabitEthernet1/1] quit
```

把接口 GigabitEthernet1/2 绑定到 VPN 实例 a，配置 IP 地址，并使能 PIM-SM。

```
[PE3] interface GigabitEthernet 1/2
[PE3-GigabitEthernet1/2] ip binding vpn-instance a
[PE3-GigabitEthernet1/2] ip address 10.110.5.1 24
[PE3-GigabitEthernet1/2] pim sm
[PE3-GigabitEthernet1/2] quit
```

把接口 GigabitEthernet1/3 绑定到 VPN 实例 b，配置 IP 地址，并使能 PIM-SM。

```
[PE3] interface GigabitEthernet 1/3
[PE3-GigabitEthernet1/3] ip binding vpn-instance b
[PE3-GigabitEthernet1/3] ip address 10.110.6.1 24
[PE3-GigabitEthernet1/3] pim sm
[PE3-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE3] interface loopback 1
[PE3-LoopBack1] ip address 1.1.1.3 32
[PE3-LoopBack1] pim sm
[PE3-LoopBack1] quit
```

把接口 Loopback2 绑定到 VPN 实例 b，配置 IP 地址，并使能 PIM-SM。

```
[PE3] interface loopback 2
```

```
[PE3-LoopBack2] ip binding vpn-instance b
[PE3-LoopBack2] ip address 33.33.33.33 32
[PE3-LoopBack2] pim sm
[PE3-LoopBack2] quit
```

配置 Loopback2 接口为 VPN b 的 C-BSR 和 C-RP。

```
[PE3] pim vpn-instance b
[PE3-pim-b] c-bsr loopback 2
[PE3-pim-b] c-rp loopback 2
[PE3-pim-b] quit
```

配置 BGP 协议。

```
[PE3] bgp 100
[PE3-bgp] group vpn-g internal
[PE3-bgp] peer vpn-g connect-interface loopback 1
[PE3-bgp] peer 1.1.1.1 group vpn-g
[PE3-bgp] peer 1.1.1.2 group vpn-g
[PE3–bgp] ipv4-family vpn-instance a
[PE3-bgp-a] import-route rip 2
[PE3-bgp-a] import-route direct
[PE3-bgp-a] quit
[PE3–bgp] ipv4-family vpn-instance b
[PE3-bgp-b] import-route rip 3
[PE3-bgp-b] import-route direct
[PE3-bgp-b] quit
[PE3–bgp] ipv4-family vpnv4
[PE3–bgp-af-vpnv4] peer vpn-g enable
[PE3-bgp-af-vpnv4] peer 1.1.1.1 group vpn-g
[PE3–bgp-af-vpnv4] peer 1.1.1.2 group vpn-g
[PE3–bgp-af-vpnv4] quit
[PE3–bgp] quit
```

当 PE 3 上配置了 BGP 对等体之后，MTI0 和 MTI1 接口将自动获得各自的 IP 地址，该地址与配置 BGP 对等体时所指定的 Loopback 接口的 IP 地址相同；MTI0 和 MTI1 接口上运行的 PIM 协议类型也分别与其所属的 VPN 实例 a 和 b 中运行的 PIM 协议类型相同。

配置 OSPF 协议。

```
[PE3] ospf 1
[PE3-ospf-1] area 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 1.1.1.3 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.255.255
[PE3-ospf-1-area-0.0.0.0] quit
[PE3-ospf-1] quit
```

配置 RIP 协议。

```
[PE3] rip 2 vpn-instance a
[PE3-rip-2] network 10.0.0.0
[PE3-rip-2] import-route bgp
[PE3-rip-2] quit
[PE3] rip 3 vpn-instance b
```

```
[PE3-rip-3] network 10.0.0.0
[PE3-rip-3] network 33.0.0.0
[PE3-rip-3] import-route bgp
[PE3-rip-3] return
```

(4) 配置 P

使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<P> system-view
[P] multicast routing-enable
[P] mpls lsr-id 2.2.2.2
[P] mpls
[P-mpls] quit
[P] mpls ldp
[P-mpls-ldp] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[P] interface GigabitEthernet 1/1
[P-GigabitEthernet1/1] ip address 192.168.6.2 24
[P-GigabitEthernet1/1] pim sm
[P-GigabitEthernet1/1] mpls
[P-GigabitEthernet1/1] mpls ldp
[P-GigabitEthernet1/1] quit
```

在公网接口 GigabitEthernet1/2 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[P] interface GigabitEthernet 1/2
[P-GigabitEthernet1/2] ip address 192.168.7.2 24
[P-GigabitEthernet1/2] pim sm
[P-GigabitEthernet1/2] mpls
[P-GigabitEthernet1/2] mpls ldp
[P-GigabitEthernet1/2] quit
```

在公网接口 GigabitEthernet1/3 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[P] interface GigabitEthernet 1/3
[P-GigabitEthernet1/3] ip address 192.168.8.2 24
[P-GigabitEthernet1/3] pim sm
[P-GigabitEthernet1/3] mpls
[P-GigabitEthernet1/3] mpls ldp
[P-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.2 32
[P-LoopBack1] pim sm
[P-LoopBack1] quit
```

配置 Loopback1 接口为公网实例的 C-BSR 和 C-RP。

```
[P] pim
[P-pim] c-bsr loopback 1
[P-pim] c-rp loopback 1
[P-pim] quit
```

配置 OSPF 协议。

```
[P] ospf 1
[P-ospf-1] area 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.255.255
```

(5) 配置 CE a1

使能 IP 组播路由。

```
<CEa1> system-view
[CEa1] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，并使能 PIM-SM。

```
[CEa1] interface GigabitEthernet 1/1
[CEa1-GigabitEthernet1/1] ip address 10.110.7.1 24
[CEa1-GigabitEthernet1/1] pim sm
[CEa1-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEa1] interface GigabitEthernet 1/2
[CEa1-GigabitEthernet1/2] ip address 10.110.2.2 24
[CEa1-GigabitEthernet1/2] pim sm
[CEa1-GigabitEthernet1/2] quit
```

配置 RIP 协议。

```
[CEa1] rip 2
[CEa1-rip-2] network 10.0.0.0
```

(6) 配置 CE b1

使能 IP 组播路由。

```
<CEb1> system-view
[CEb1] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，并使能 PIM-SM。

```
[CEb1] interface GigabitEthernet 1/1
[CEb1-GigabitEthernet1/1] ip address 10.110.8.1 24
[CEb1-GigabitEthernet1/1] pim sm
[CEb1-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEb1] interface GigabitEthernet 1/2
[CEb1-GigabitEthernet1/2] ip address 10.110.3.2 24
[CEb1-GigabitEthernet1/2] pim sm
[CEb1-GigabitEthernet1/2] quit
```

配置 RIP 协议。

```
[CEb1] rip 3
[CEb1-rip-3] network 10.0.0.0
```

(7) 配置 CE a2

使能 IP 组播路由。

```
<CEa2> system-view
```

```
[CEa2] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，使能 IGMP 和 PIM-SM。

```
[CEa2] interface GigabitEthernet 1/1
[CEa2-GigabitEthernet1/1] ip address 10.110.9.1 24
[CEa2-GigabitEthernet1/1] igmp enable
[CEa2-GigabitEthernet1/1] pim sm
[CEa2-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEa2] interface GigabitEthernet 1/2
[CEa2-GigabitEthernet1/2] ip address 10.110.4.2 24
[CEa2-GigabitEthernet1/2] pim sm
[CEa2-GigabitEthernet1/2] quit
```

在接口 GigabitEthernet1/3 上配置 IP 地址，并使能 PIM-SM。

```
[CEa2] interface GigabitEthernet 1/3
[CEa2-GigabitEthernet1/3] ip address 10.110.12.1 24
[CEa2-GigabitEthernet1/3] pim sm
[CEa2-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[CEa2] interface loopback 1
[CEa2-LoopBack1] ip address 22.22.22.22 32
[CEa2-LoopBack1] pim sm
[CEa2-LoopBack1] quit
```

配置 Loopback1 接口为 VPN a 的 BSR 和 RP。

```
[CEa2] pim
[CEa2-pim] c-bsr loopback 1
[CEa2-pim] c-rp loopback 1
[CEa2-pim] quit
```

配置 RIP 协议。

```
[CEa2] rip 2
[CEa2-rip-2] network 10.0.0.0
[CEa2-rip-2] network 22.0.0.0
```

(8) 配置 CE a3

使能 IP 组播路由。

```
<CEa3> system-view
[CEa3] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，使能 IGMP 和 PIM-SM。

```
[CEa3] interface GigabitEthernet 1/1
[CEa3-GigabitEthernet1/1] ip address 10.110.10.1 24
[CEa3-GigabitEthernet1/1] igmp enable
[CEa3-GigabitEthernet1/1] pim sm
[CEa3-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEa3] interface GigabitEthernet 1/2
```

```
[CEa3-GigabitEthernet1/2] ip address 10.110.5.2 24
[CEa3-GigabitEthernet1/2] pim sm
[CEa3-GigabitEthernet1/2] quit
```

在接口 GigabitEthernet1/3 上配置 IP 地址，并使能 PIM-SM。

```
[CEa3] interface GigabitEthernet 1/3
[CEa3-GigabitEthernet1/3] ip address 10.110.12.2 24
[CEa3-GigabitEthernet1/3] pim sm
[CEa3-GigabitEthernet1/3] quit
```

配置 RIP 协议。

```
[CEa3] rip 2
[CEa3-rip-2] network 10.0.0.0
```

(9) 配置 CE b2

使能 IP 组播路由。

```
<CEb2> system-view
[CEb2] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，使能 IGMP 和 PIM-SM。

```
[CEb2] interface GigabitEthernet 1/1
[CEb2-GigabitEthernet1/1] ip address 10.110.11.1 24
[CEb2-GigabitEthernet1/1] igmp enable
[CEb2-GigabitEthernet1/1] pim sm
[CEb2-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEb2] interface GigabitEthernet 1/2
[CEb2-GigabitEthernet1/2] ip address 10.110.6.2 24
[CEb2-GigabitEthernet1/2] pim sm
[CEb2-GigabitEthernet1/2] quit
```

配置 RIP 协议。

```
[CEb2] rip 3
[CEb2-rip-3] network 10.0.0.0
```

(10) 检验配置效果

通过使用 **display multicast-domain vpn-instance share-group** 命令可以查看指定 VPN 实例的 Share-Group 组信息。

查看 PE 1 上 VPN 实例 a 中的本地 Share-Group 组信息。

```
<PE1> display multicast-domain vpn-instance a share-group local
MD local share-group information for VPN-Instance: a
  Share-group: 239.1.1.1
  MTunnel address: 1.1.1.1
```

查看 PE 2 上 VPN 实例 a 中的本地 Share-Group 组信息。

```
<PE2> display multicast-domain vpn-instance a share-group local
MD local share-group information for VPN-Instance: a
  Share-group: 239.1.1.1
  MTunnel address: 1.1.1.2
```

查看 PE 2 上 VPN 实例 b 中的本地 Share-Group 组信息。

```
<PE2> display multicast-domain vpn-instance b share-group local
MD local share-group information for VPN-Instance: b
  Share-group: 239.2.2.2
  MTunnel address: 1.1.1.2
```

查看 PE 3 上 VPN 实例 a 中的本地 Share-Group 组信息。

```
<PE3> display multicast-domain vpn-instance a share-group local
MD local share-group information for VPN-Instance: a
  Share-group: 239.1.1.1
  MTunnel address: 1.1.1.3
```

查看 PE 3 上 VPN 实例 b 中的本地 Share-Group 组信息。

```
<PE3> display multicast-domain vpn-instance b share-group local
MD local share-group information for VPN-Instance: b
  Share-group: 239.2.2.2
  MTunnel address: 1.1.1.3
```

1.7.2 跨 AS 的 MD VPN 配置举例

1. 组网需求

组网需求如 表 1-11所示。

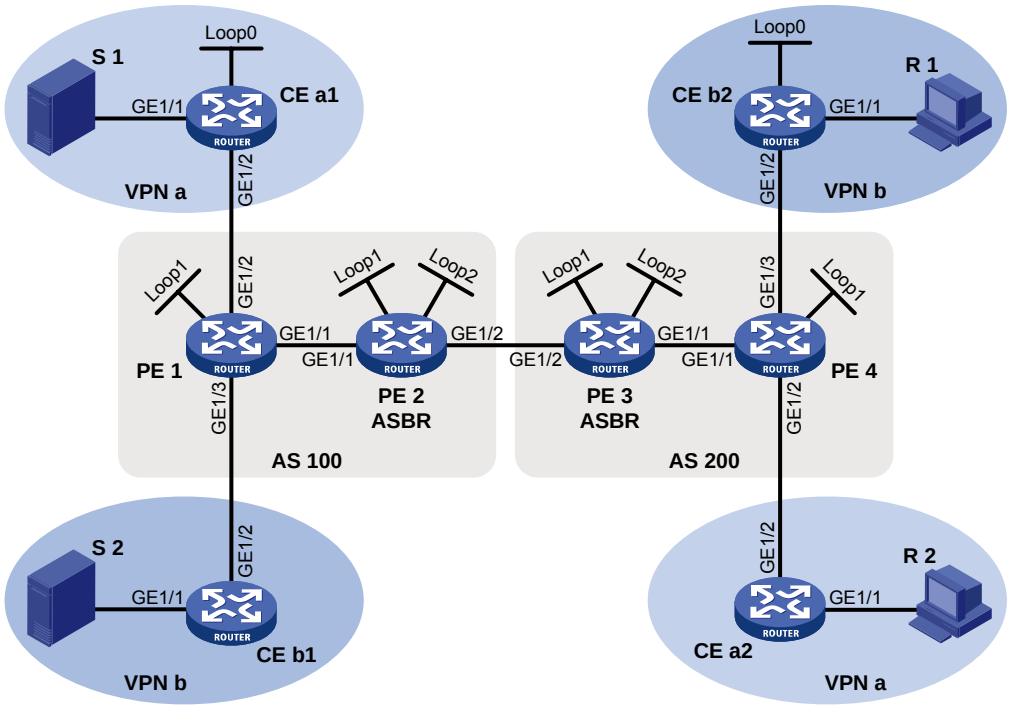
表1-11 跨 AS 的 MD VPN 配置组网需求

项目	组网需求
组播源和接收者	• VPN a 的组播源为 S 1，接收者为 R 2 • VPN b 的组播源为 S 2，接收者为 R 1 • VPN a 中的 Share-Group 地址为 239.1.1.1，Switch-Group 地址池为 225.1.1.0～225.1.1.15 • VPN b 中的 Share-Group 地址为 239.4.4.4，Switch-Group 地址池为 225.4.4.0～225.4.4.15
PE 上各接口所属的 VPN 实例	• PE 1：GigabitEthernet1/2 属于 VPN 实例 a，GigabitEthernet1/3 属于 VPN 实例 b，GigabitEthernet1/1 和 Loopback1 属于公网实例 • PE 2：GigabitEthernet1/1、GigabitEthernet1/2、Loopback1 和 Loopback2 属于公网实例 • PE 3：GigabitEthernet1/1、GigabitEthernet1/2、Loopback1 和 Loopback2 属于公网实例 • PE 4：GigabitEthernet1/2 属于 VPN 实例 a，GigabitEthernet1/3 属于 VPN 实例 b，GigabitEthernet1/1 和 Loopback1 属于公网实例
单播路由协议和 MPLS	• 在 AS 100 和 AS 200 中分别配置 OSPF，在各 PE 与 CE 之间也配置 OSPF • 在 PE 1、PE 2、PE 3 和 PE 4 各自的 Loopback1 接口之间建立 BGP 对等体连接并传递所有私网路由 • 在 AS 100 和 AS 200 中分别配置 MPLS
IP 组播路由功能	• 在 PE 1、PE 2、PE 3 和 PE 4 的公网实例中均使能 IP 组播路由 • 在 PE 1 和 PE 4 的 VPN 实例 a 中均使能 IP 组播路由 • 在 PE 1 和 PE 4 的 VPN 实例 b 中均使能 IP 组播路由 • 在 CE a1、CE a2、CE b1、CE b2 上均使能 IP 组播路由
IGMP 功能	• 在 CE a2 的 GigabitEthernet1/1 接口上使能 IGMPv2 • 在 CE b2 的 GigabitEthernet1/1 接口上使能 IGMPv2

项目	组网需求
PIM 功能	• 在 PE 2 和 PE 3 的所有公网接口上均使能 PIM-SM • 在 PE 1 和 PE 4 的所有公网和私网接口上均使能 PIM-SM • 在 CE a1、CE a2、CE b1 和 CE b2 的所有接口上均使能 PIM-SM • PE 2 和 PE 3 的 Loopback2 接口为各自所在 AS 的 C-BSR 和 C-RP（服务于所有组播组） • CE a1 的 Loopback0 接口为 VPN a 的 C-BSR 和 C-RP（服务于所有组播组） • CE b1 的 Loopback0 接口为 VPN b 的 C-BSR 和 C-RP（服务于所有组播组）
MSDP 功能	• 在 PE 2 和 PE 3 的 Loopback1 接口之间建立 MSDP 对等体

2. 组网图

图1-13 跨 AS 的 MD VPN 配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
S 1	-	10.11.5.2/24	R 1	-	10.11.8.2/24
S 2	-	10.11.6.2/24	R 2	-	10.11.7.2/24
PE 1	GE1/1	10.10.1.1/24	PE 3	GE1/1	10.10.2.1/24
	GE1/2	10.11.1.1/24		GE1/2	192.168.1.2/24
	GE1/3	10.11.2.1/24		Loop1	1.1.1.3/32
	Loop1	1.1.1.1/32		Loop2	22.22.22.22/32
PE 2	GE1/1	10.10.1.2/24	PE 4	GE1/1	10.10.2.2/24
	GE1/2	192.168.1.1/24		GE1/2	10.11.3.1/24
	Loop1	1.1.1.2/32		GE1/3	10.11.4.1/32
	Loop2	11.11.11.11/32		Loop2	1.1.1.4/32
CE a1	GE1/1	10.11.5.1/24	CE b1	GE1/1	10.11.6.1/24
	GE1/2	10.11.1.2/24		GE1/2	10.11.2.2/24
	Loop0	2.2.2.2/32	CE b2	GE1/1	10.11.8.1/24
CE a2	GE1/1	10.11.7.1/24		GE1/2	10.11.4.2/24
	GE1/2	10.11.3.2/24		Loop0	3.3.3.3/32

3. 配置步骤

(1) 配置 PE 1

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE1> system-view
[PE1] router id 1.1.1.1
[PE1] multicast routing-enable
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
```

创建 VPN 实例 a，为其配置 RD，并创建其出、入口路由；使能 VPN 实例 a 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE1-vpn-instance-a] multicast routing-enable
[PE1-vpn-instance-a] multicast-domain share-group 239.1.1.1 binding mtunnel 0
[PE1-vpn-instance-a] multicast-domain switch-group-pool 225.1.1.0 28
[PE1-vpn-instance-a] quit
```

创建 VPN 实例 b，为其配置 RD，并创建其出、入口路由；使能 VPN 实例 b 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE1] ip vpn-instance b
[PE1-vpn-instance-b] route-distinguisher 200:1
[PE1-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE1-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE1-vpn-instance-b] multicast routing-enable
[PE1-vpn-instance-b] multicast-domain share-group 239.4.4.4 binding mtunnel 1
[PE1-vpn-instance-b] multicast-domain switch-group-pool 225.4.4.0 28
[PE1-vpn-instance-b] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE1] interface GigabitEthernet 1/1
[PE1-GigabitEthernet1/1] ip address 10.10.1.1 24
[PE1-GigabitEthernet1/1] pim sm
[PE1-GigabitEthernet1/1] mpls
[PE1-GigabitEthernet1/1] mpls ldp
[PE1-GigabitEthernet1/1] quit
```

把接口 GigabitEthernet1/2 绑定到 VPN 实例 a，配置 IP 地址，并使能 PIM-SM。

```
[PE1] interface GigabitEthernet 1/2
[PE1-GigabitEthernet1/2] ip binding vpn-instance a
[PE1-GigabitEthernet1/2] ip address 10.11.1.1 24
[PE1-GigabitEthernet1/2] pim sm
[PE1-GigabitEthernet1/2] quit
```

把接口 GigabitEthernet1/3 绑定到 VPN 实例 b，配置 IP 地址，并使能 PIM-SM。

```
[PE1] interface GigabitEthernet 1/3
[PE1-GigabitEthernet1/3] ip binding vpn-instance b
[PE1-GigabitEthernet1/3] ip address 10.11.2.1 24
```

```
[PE1-GigabitEthernet1/3] pim sm
[PE1-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

配置 BGP 协议。

```
[PE1] bgp 100
[PE1-bgp] group pe1-pe2 internal
[PE1-bgp] peer pe1-pe2 label-route-capability
[PE1-bgp] peer pe1-pe2 connect-interface loopback 1
[PE1-bgp] peer 1.1.1.2 group pe1-pe2
[PE1-bgp] group pe1-pe4 external
[PE1-bgp] peer pe1-pe4 as-number 200
[PE1-bgp] peer pe1-pe4 ebgp-max-hop 255
[PE1-bgp] peer 1.1.1.4 group pe1-pe4
[PE1-bgp] peer 1.1.1.4 connect-interface loopback 1
[PE1–bgp] ipv4-family vpn-instance a
[PE1-bgp-a] import-route ospf 2
[PE1-bgp-a] import-route direct
[PE1-bgp-a] quit
[PE1–bgp] ipv4-family vpn-instance b
[PE1-bgp-b] import-route ospf 3
[PE1-bgp-b] import-route direct
[PE1-bgp-b] quit
[PE1–bgp] ipv4-family vpnv4
[PE1–bgp-af-vpnv4] peer 1.1.1.4 enable
[PE1–bgp-af-vpnv4] quit
[PE1–bgp] quit
```

当 PE 1 上配置了 BGP 对等体之后，MTI0 和 MTI1 接口将分别自动获得其 IP 地址，该地址与配置 BGP 对等体时所指定的 Loopback 接口的 IP 地址相同；MTI0 接口上运行的 PIM 协议类型与 VPN 实例 a 中运行的 PIM 协议类型相同；MTI1 接口上运行的 PIM 协议类型与 VPN 实例 b 中运行的 PIM 协议类型相同。

配置 OSPF 协议。

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 10.10.0.0 0.0.255.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
[PE1] ospf 2 vpn-instance a
[PE1-ospf-2] import-route bgp
[PE1-ospf-2] area 0.0.0.0
[PE1-ospf-2-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[PE1-ospf-2-area-0.0.0.0] quit
```

```
[PE1-ospf-2] quit
[PE1] ospf 3 vpn-instance b
[PE1-ospf-3] import-route bgp
[PE1-ospf-3] area 0.0.0.0
[PE1-ospf-3-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[PE1-ospf-3-area-0.0.0.0] quit
[PE1-ospf-3] quit
```

(2) 配置 PE 2

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE2> system-view
[PE2] router id 1.1.1.2
[PE2] multicast routing-enable
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE2] interface GigabitEthernet 1/1
[PE2-GigabitEthernet1/1] ip address 10.10.1.2 24
[PE2-GigabitEthernet1/1] pim sm
[PE2-GigabitEthernet1/1] mpls
[PE2-GigabitEthernet1/1] mpls ldp
[PE2-GigabitEthernet1/1] quit
```

在公网接口 GigabitEthernet1/2 上配置 IP 地址，使能 PIM-SM 和 MPLS。

```
[PE2] interface GigabitEthernet 1/2
[PE2-GigabitEthernet1/2] ip address 192.168.1.1 24
[PE2-GigabitEthernet1/2] pim sm
[PE2-GigabitEthernet1/2] mpls
[PE2-GigabitEthernet1/2] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 1.1.1.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

配置 Loopback2 接口的 IP 地址，并使能 PIM-SM。

```
[PE2] interface loopback 2
[PE2-LoopBack2] ip address 11.11.11.11 32
[PE2-LoopBack2] pim sm
[PE2-LoopBack2] quit
```

配置 Loopback2 接口为公网实例的 C-BSR 和 C-RP。

```
[PE2] pim
[PE2-pim] c-bsr loopback 2
[PE2-pim] c-rp loopback 2
```

```
[PE2-pim] quit
```

配置 BSR 的服务边界。

```
[PE2] interface GigabitEthernet 1/2
[PE2-GigabitEthernet1/2] pim bsr-boundary
[PE2-GigabitEthernet1/2] quit
```

配置 MSDP 对等体。

```
[PE2] msdp
[PE2-msdp] encap-data-enable
[PE2-msdp] peer 1.1.1.3 connect-interface loopback 1
```

配置静态路由。

```
[PE2] ip route-static 1.1.1.3 32 GigabitEthernet 1/2 192.168.1.2
```

配置 BGP 协议。

```
[PE2] bgp 100
[PE2-bgp] import-route ospf 1
[PE2-bgp] group pe2-pe1 internal
[PE2-bgp] peer pe2-pe1 route-policy map2 export
[PE2-bgp] peer pe2-pe1 label-route-capability
[PE2-bgp] peer pe2-pe1 connect-interface loopback 1
[PE2-bgp] peer 1.1.1.1 group pe2-pe1
[PE2-bgp] group pe2-pe3 external
[PE2-bgp] peer pe2-pe3 as-number 200
[PE2-bgp] peer pe2-pe3 ebgp-max-hop 255
[PE2-bgp] peer pe2-pe3 route-policy map1 export
[PE2-bgp] peer pe2-pe3 label-route-capability
[PE2-bgp] peer pe2-pe3 connect-interface loopback 1
[PE2-bgp] peer 1.1.1.3 group pe2-pe3
[PE2–bgp] quit
```

配置 OSPF 协议。

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 11.11.11.11 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.10.0.0 0.0.255.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

配置路由策略。

```
[PE2] route-policy map1 permit node 10
[PE2-route-policy] apply mpls-label
[PE2-route-policy] quit
[PE2] route-policy map2 permit node 10
[PE2-route-policy] if-match mpls-label
[PE2-route-policy] apply mpls-label
[PE2-route-policy] quit
```

(3) 配置 PE 3

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE3> system-view
[PE3] router id 1.1.1.3
[PE3] multicast routing-enable
[PE3] mpls lsr-id 1.1.1.3
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE3] interface GigabitEthernet 1/1
[PE3-GigabitEthernet1/1] ip address 10.10.2.1 24
[PE3-GigabitEthernet1/1] pim sm
[PE3-GigabitEthernet1/1] mpls
[PE3-GigabitEthernet1/1] mpls ldp
[PE3-GigabitEthernet1/1] quit
```

在公网接口 GigabitEthernet1/2 上配置 IP 地址，使能 PIM-SM 和 MPLS。

```
[PE3] interface GigabitEthernet 1/2
[PE3-GigabitEthernet1/2] ip address 192.168.1.2 24
[PE3-GigabitEthernet1/2] pim sm
[PE3-GigabitEthernet1/2] mpls
[PE3-GigabitEthernet1/2] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE3] interface loopback 1
[PE3-LoopBack1] ip address 1.1.1.3 32
[PE3-LoopBack1] pim sm
[PE3-LoopBack1] quit
```

配置 Loopback2 接口的 IP 地址，并使能 PIM-SM。

```
[PE3] interface loopback 2
[PE3-LoopBack2] ip address 22.22.22.22 32
[PE3-LoopBack2] pim sm
[PE3-LoopBack2] quit
```

配置 Loopback2 接口为公网实例的 C-BSR 和 C-RP。

```
[PE3] pim
[PE3-pim] c-bsr loopback 2
[PE3-pim] c-rp loopback 2
[PE3-pim] quit
```

配置 BSR 的服务边界。

```
[PE3] interface GigabitEthernet 1/2
[PE3-GigabitEthernet1/2] pim bsr-boundary
[PE3-GigabitEthernet1/2] quit
```

配置 MSDP 对等体。

```
[PE3] msdp
[PE3-msdp] encap-data-enable
```

```
[PE3-msdp] peer 1.1.1.2 connect-interface loopback 1
```

配置静态路由。

```
[PE3] ip route-static 1.1.1.2 32 GigabitEthernet 1/2 192.168.1.1
```

配置 BGP 协议。

```
[PE3] bgp 200
[PE3-bgp] import-route ospf 1
[PE3-bgp] group pe3-pe4 internal
[PE3-bgp] peer pe3-pe4 route-policy map2 export
[PE3-bgp] peer pe3-pe4 label-route-capability
[PE3-bgp] peer pe3-pe4 connect-interface loopback 1
[PE3-bgp] peer 1.1.1.4 group pe3-pe4
[PE3-bgp] group pe3-pe4 external
[PE3-bgp] peer pe3-pe4 as-number 100
[PE3-bgp] peer pe3-pe4 ebgp-max-hop 255
[PE3-bgp] peer pe3-pe4 route-policy map1 export
[PE3-bgp] peer pe3-pe4 label-route-capability
[PE3-bgp] peer pe3-pe4 connect-interface loopback 1
[PE3-bgp] peer 1.1.1.2 group pe3-pe4
[PE3–bgp] quit
```

配置 OSPF 协议。

```
[PE3] ospf 1
[PE3-ospf-1] area 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 1.1.1.3 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 22.22.22.22 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 10.10.0.0 0.0.255.255
[PE3-ospf-1-area-0.0.0.0] quit
[PE3-ospf-1] quit
```

配置路由策略。

```
[PE3] route-policy map1 permit node 10
[PE3-route-policy] apply mpls-label
[PE3-route-policy] quit
[PE3] route-policy map2 permit node 10
[PE3-route-policy] if-match mpls-label
[PE3-route-policy] apply mpls-label
[PE3-route-policy] quit
```

(4) 配置 PE 4

配置 Router ID，使能公网实例的 IP 组播路由，配置 MPLS LSR ID，并使能 LDP 能力。

```
<PE4> system-view
[PE4] router id 1.1.1.4
[PE4] multicast routing-enable
[PE4] mpls lsr-id 1.1.1.4
[PE4] mpls
[PE4-mpls] quit
[PE4] mpls ldp
[PE4-mpls-ldp] quit
```

创建 VPN 实例 a，为其配置 RD，并创建其出、入口路由；使能 VPN 实例 a 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE4] ip vpn-instance a
[PE4-vpn-instance-a] route-distinguisher 100:1
[PE4-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE4-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE4-vpn-instance-a] multicast routing-enable
[PE4-vpn-instance-a] multicast-domain share-group 239.1.1.1 binding mtunnel 0
[PE4-vpn-instance-a] multicast-domain switch-group-pool 225.1.1.0 28
[PE4-vpn-instance-a] quit
```

创建 VPN 实例 b，为其配置 RD，并创建其出、入口路由；使能 VPN 实例 b 中的 IP 组播路由，配置 Share-Group 地址，绑定 MTI 接口并指定 Switch-Group-Pool 的范围。

```
[PE4] ip vpn-instance b
[PE4-vpn-instance-b] route-distinguisher 200:1
[PE4-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE4-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE4-vpn-instance-b] multicast routing-enable
[PE4-vpn-instance-b] multicast-domain share-group 239.4.4.4 binding mtunnel 1
[PE4-vpn-instance-b] multicast-domain switch-group-pool 225.4.4.0 28
[PE4-vpn-instance-b] quit
```

在公网接口 GigabitEthernet1/1 上配置 IP 地址，使能 PIM-SM 和 LDP 能力。

```
[PE4] interface GigabitEthernet 1/1
[PE4-GigabitEthernet1/1] ip address 10.10.2.2 24
[PE4-GigabitEthernet1/1] pim sm
[PE4-GigabitEthernet1/1] mpls
[PE4-GigabitEthernet1/1] mpls ldp
[PE4-GigabitEthernet1/1] quit
```

把接口 GigabitEthernet1/2 绑定到 VPN 实例 a，配置 IP 地址，并使能 PIM-SM。

```
[PE4] interface GigabitEthernet 1/2
[PE4-GigabitEthernet1/2] ip binding vpn-instance a
[PE4-GigabitEthernet1/2] ip address 10.11.3.1 24
[PE4-GigabitEthernet1/2] pim sm
[PE4-GigabitEthernet1/2] quit
```

把接口 GigabitEthernet1/3 绑定到 VPN 实例 b，配置 IP 地址，并使能 PIM-SM。

```
[PE4] interface GigabitEthernet 1/3
[PE4-GigabitEthernet1/3] ip binding vpn-instance b
[PE4-GigabitEthernet1/3] ip address 10.11.4.1 24
[PE4-GigabitEthernet1/3] pim sm
[PE4-GigabitEthernet1/3] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[PE4] interface loopback 1
[PE4-LoopBack1] ip address 1.1.1.4 32
[PE4-LoopBack1] pim sm
[PE4-LoopBack1] quit
```

配置 BGP 协议。

```
[PE4] bgp 100
[PE4-bgp] group pe4-pe3 internal
[PE4-bgp] peer pe4-pe3 label-route-capability
[PE4-bgp] peer pe4-pe3 connect-interface loopback 1
[PE4-bgp] peer 1.1.1.3 group pe4-pe3
[PE4-bgp] group pe4-pe1 external
[PE4-bgp] peer pe4-pe1 as-number 100
[PE4-bgp] peer pe4-pe1 ebgp-max-hop 255
[PE4-bgp] peer 1.1.1.1 group pe4-pe1
[PE4-bgp] peer 1.1.1.1 connect-interface loopback 1
[PE4–bgp] ipv4-family vpn-instance a
[PE4-bgp-a] import-route ospf 2
[PE4-bgp-a] import-route direct
[PE4-bgp-a] quit
[PE4–bgp] ipv4-family vpn-instance b
[PE4-bgp-b] import-route ospf 3
[PE4-bgp-b] import-route direct
[PE4-bgp-b] quit
[PE4–bgp] ipv4-family vpnv4
[PE4–bgp-af-vpnv4] peer 1.1.1.1 enable
[PE4–bgp-af-vpnv4] quit
[PE4–bgp] quit
```

当 PE 4 上配置了 BGP 对等体之后，MTI0 和 MTI1 接口将分别自动获得其 IP 地址，该地址与配置 BGP 对等体时所指定的 Loopback 接口的 IP 地址相同；MTI0 接口上运行的 PIM 协议类型与 VPN 实例 a 中运行的 PIM 协议类型相同；MTI1 接口上运行的 PIM 协议类型与 VPN 实例 b 中运行的 PIM 协议类型相同。

配置 OSPF 协议。

```
[PE4] ospf 1
[PE4-ospf-1] area 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 1.1.1.4 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 10.10.0.0 0.0.255.255
[PE4-ospf-1-area-0.0.0.0] quit
[PE4-ospf-1] quit
[PE4] ospf 2 vpn-instance a
[PE4-ospf-2] import-route bgp
[PE4-ospf-2] area 0.0.0.0
[PE4-ospf-2-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[PE4-ospf-2-area-0.0.0.0] quit
[PE4-ospf-2] quit
[PE4] ospf 3 vpn-instance b
[PE4-ospf-3] import-route bgp
[PE4-ospf-3] area 0.0.0.0
[PE4-ospf-3-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[PE4-ospf-3-area-0.0.0.0] quit
[PE4-ospf-3] quit
```

(5) 配置 CE a1

使能 IP 组播路由。

```
<CEa1> system-view
[CEa1] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，并使能 PIM-SM。

```
[CEa1] interface GigabitEthernet 1/1
[CEa1-GigabitEthernet1/1] ip address 10.11.5.1 24
[CEa1-GigabitEthernet1/1] pim sm
[CEa1-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEa1] interface GigabitEthernet 1/2
[CEa1-GigabitEthernet1/2] ip address 10.11.1.2 24
[CEa1-GigabitEthernet1/2] pim sm
[CEa1-GigabitEthernet1/2] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[CEa1] interface loopback 1
[CEa1-LoopBack1] ip address 2.2.2.2 32
[CEa1-LoopBack1] pim sm
[CEa1-LoopBack1] quit
```

配置 Loopback1 接口为 VPN a 的 C-BSR 和 C-RP。

```
[CEa1] pim vpn-instance a
[CEa1-pim-a] c-bsr loopback 1
[CEa1-pim-a] c-rp loopback 1
[CEa1-pim-a] quit
```

配置 OSPF 协议。

```
[CEa1] ospf 1
[CEa1-ospf-1] area 0.0.0.0
[CEa1-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[CEa1-ospf-1-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[CEa1-ospf-1-area-0.0.0.0] quit
[CEa1-ospf-1] quit
```

(6) 配置 CE b1

使能 IP 组播路由。

```
<CEb1> system-view
[CEb1] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，并使能 PIM-SM。

```
[CEb1] interface GigabitEthernet 1/1
[CEb1-GigabitEthernet1/1] ip address 10.11.6.1 24
[CEb1-GigabitEthernet1/1] pim sm
[CEb1-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEb1] interface GigabitEthernet 1/2
```

```
[CEb1-GigabitEthernet1/2] ip address 10.11.2.2 24
[CEb1-GigabitEthernet1/2] pim sm
[CEb1-GigabitEthernet1/2] quit
```

配置 OSPF 协议。

```
[CEb1] ospf 1
[CEb1-ospf-1] area 0.0.0.0
[CEb1-ospf-1-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[CEb1-ospf-1-area-0.0.0.0] quit
[CEb1-ospf-1] quit
```

(7) 配置 CE a2

使能 IP 组播路由。

```
<CEa2> system-view
[CEa2] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，使能 IGMP 和 PIM-SM。

```
[CEa2] interface GigabitEthernet 1/1
[CEa2-GigabitEthernet1/1] ip address 10.11.7.1 24
[CEa2-GigabitEthernet1/1] igmp enable
[CEa2-GigabitEthernet1/1] pim sm
[CEa2-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEa2] interface GigabitEthernet 1/2
[CEa2-GigabitEthernet1/2] ip address 10.11.3.2 24
[CEa2-GigabitEthernet1/2] pim sm
[CEa2-GigabitEthernet1/2] quit
```

配置 OSPF 协议。

```
[CEa2] ospf 1
[CEa2-ospf-1] area 0.0.0.0
[CEa2-ospf-1-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[CEa2-ospf-1-area-0.0.0.0] quit
[CEa2-ospf-1] quit
```

(8) 配置 CE b2

使能 IP 组播路由。

```
<CEb2> system-view
[CEb2] multicast routing-enable
```

在接口 GigabitEthernet1/1 上配置 IP 地址，使能 IGMP 和 PIM-SM。

```
[CEb2] interface GigabitEthernet 1/1
[CEb2-GigabitEthernet1/1] ip address 10.11.8.1 24
[CEb2-GigabitEthernet1/1] igmp enable
[CEb2-GigabitEthernet1/1] pim sm
[CEb2-GigabitEthernet1/1] quit
```

在接口 GigabitEthernet1/2 上配置 IP 地址，并使能 PIM-SM。

```
[CEb2] interface GigabitEthernet 1/2
[CEb2-GigabitEthernet1/2] ip address 10.11.4.2 24
```

```
[CEb2-GigabitEthernet1/2] pim sm
[CEb2-GigabitEthernet1/2] quit
```

配置 Loopback1 接口的 IP 地址，并使能 PIM-SM。

```
[CEb2] interface loopback 1
[CEb2-LoopBack1] ip address 3.3.3.3 32
[CEb2-LoopBack1] pim sm
[CEb2-LoopBack1] quit
```

配置 Loopback1 接口为 VPN b 的 C-BSR 和 C-RP。

```
[CEb2] pim vpn-instance b
[CEb2-pim-b] c-bsr loopback 1
[CEb2-pim-b] c-rp loopback 1
[CEb2-pim-b] quit
```

配置 OSPF 协议。

```
[CEb2] ospf 1
[CEb2-ospf-1] area 0.0.0.0
[CEb2-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[CEb2-ospf-1-area-0.0.0.0] network 10.11.0.0 0.0.255.255
[CEb2-ospf-1-area-0.0.0.0] quit
[CEb2-ospf-1] quit
```

(9) 检验配置效果

通过使用 **display multicast-domain vpn-instance share-group** 命令可以查看指定 VPN 实例的 Share-Group 组信息。

查看 PE 1 上 VPN 实例 a 中的本地 Share-Group 组信息。

```
<PE1> display multicast-domain vpn-instance a share-group local
MD local share-group information for VPN-Instance: a
  Share-group: 239.1.1.1
  MTunnel address: 1.1.1.1
```

查看 PE 1 上 VPN 实例 b 中的本地 Share-Group 组信息。

```
<PE1> display multicast-domain vpn-instance b share-group local
MD local share-group information for VPN-Instance: b
  Share-group: 239.4.4.4
  MTunnel address: 1.1.1.1
```

查看 PE 4 上 VPN 实例 a 中的本地 Share-Group 组信息。

```
<PE4> display multicast-domain vpn-instance a share-group local
MD local share-group information for VPN-Instance: a
  Share-group: 239.1.1.1
  MTunnel address: 1.1.1.4
```

查看 PE 4 上 VPN 实例 b 中的本地 Share-Group 组信息。

```
<PE4> display multicast-domain vpn-instance b share-group local
MD local share-group information for VPN-Instance: b
  Share-group: 239.4.4.4
  MTunnel address: 1.1.1.4
```

1.8 常见配置错误举例

1.8.1 无法建立 Share-MDT

1. 故障现象

无法正确建立 Share-MDT，不同 PE 设备上相同的 VPN 实例之间无法建立起 PIM 邻居关系。

2. 分析

- 在不同的 PE 设备上，相同的 VPN 实例需要配置相同的 Share-Group 地址，每个 Share-Group 地址唯一标识一个 Share-MDT，如果不同 PE 设备上的相同 VPN 实例没有配置相同的 Share-Group 地址，则该 VPN 实例在不同的 PE 设备上无法建立起 Share-MDT。
- 在不同的 PE 设备上，相同 VPN 实例的各接口必须使能相同类型的 PIM 协议，P 设备上的所有接口必须使能相同类型的 PIM 协议，这样才能正确地建立 Share-MDT，本地 PE 设备和远端 PE 设备的相同 VPN 实例上才能建立起 PIM 邻居关系。否则无法建立 Share-MDT。
- 只有配置了 BGP 和单播路由，MTI 接口才能自动获得 IP 地址；只有 VPN 实例内至少有一个接口上使能了 PIM 协议，MTI 接口上的 PIM 协议才能被使能，从而使不同 PE 设备的相同 VPN 实例之间建立起 PIM 邻居。否则无法建立起 PIM 邻居关系。

3. 处理过程

(1) 检查 Share-Group 地址。使用命令 **display multicast-domain vpn-instance share-group** 检查不同 PE 设备上相同的 VPN 实例是否配置了相同的 Share-Group 地址。

(2) 检查各设备 VPN 实例内是否有至少一个接口上使能了 PIM 协议，不同 PE 设备上属于同一 VPN 实例的各接口上是否使能了相同类型的 PIM 协议，以及 P 设备的各接口上是否使能了相同类型的 PIM 协议。使用命令 **display pim interface verbose** 查看各接口上的 PIM 信息。

(3) 检查单播路由。使用命令 **display ip routing-table** 检查本地 PE 设备的 VPN 实例是否有到达远端 PE 设备的相同 VPN 实例的单播路由项。

(4) 检查是否配置 BGP 对等体。使用命令 **display bgp peer** 查看配置的 BGP 对等体信息。

1.8.2 VPN 实例无法正确建立组播路由表

1. 故障现象

VPN 实例无法正确建立起组播路由表。

2. 分析

- 如果 VPN 实例使能的是 PIM-SM，需要有该 VPN 实例的 BSR 信息，否则不能正确建立该 VPN 实例的组播路由表。
- 如果 VPN 实例使能的是 PIM-SM，需要有该 VPN 实例的 RP 信息，如果没有通向 RP 的单播路由，公网实例和 VPN 实例没有正确建立起 PIM 邻居关系，VPN 实例就无法正确建立组播路由表。
- 私网 DR 需要有到达私网 RP 的路由。

3. 处理过程

(1) 使用 **display pim bsr-info** 命令查看公网实例和 VPN 实例是否有 BSR 信息。如果不存在 BSR 信息，则需要查看是否有通向 BSR 的单播路由。

(2) 使用 **display pim rp-info** 命令查看 RP 信息是否正确。如果没有 RP 信息，则需要检查是否有通向 RP 的单播路由。使用 **display pim neighbor** 命令查看公网和私网上是否正确建立起了邻居关系。

(3) 使用 **ping** 命令检查私网 DR 与私网 RP 之间是否通达。

目 录

1 MBGP配置 1-1
1.1 MBGP简介 1-1
1.2 协议规范 1-1
1.3 MBGP配置任务简介 1-2
1.4 配置MBGP基本功能 1-2
1.4.1 配置准备 1-2
1.4.2 配置MBGP基本功能 1-2
1.5 控制路由信息的发布与接收 1-3
1.5.1 配置准备 1-3
1.5.2 配置MBGP引入其他路由 1-3
1.5.3 配置MBGP引入缺省路由 1-3
1.5.4 配置MBGP路由聚合 1-4
1.5.5 配置向IPv4 MBGP对等体/IPv4 MBGP对等体组发送缺省路由 1-4
1.5.6 配置MBGP路由信息的发布策略 1-5
1.5.7 配置MBGP路由信息的接收策略 1-6
1.5.8 配置MBGP路由衰减 1-6
1.6 配置MBGP的路由属性 1-7
1.6.1 配置准备 1-7
1.6.2 配置MBGP路由管理优先级 1-7
1.6.3 配置本地优先级缺省值 1-7
1.6.4 配置MED属性 1-8
1.6.5 配置发布路由时将自身地址作为下一跳 1-8
1.6.6 配置AS_PATH属性 1-8
1.7 调整和优化MBGP网络 1-9
1.7.1 配置准备 1-9
1.7.2 配置MBGP软复位 1-9
1.7.3 配置MBGP负载分担路由条数 1-10
1.8 配置大型MBGP网络 1-10
1.8.1 配置准备 1-10
1.8.2 配置IPv4 MBGP对等体组 1-10
1.8.3 配置MBGP团体 1-11
1.8.4 配置MBGP路由反射器 1-12
1.9 MBGP显示和维护 1-12
1.9.1 MBGP配置显示 1-12
1.9.2 复位MBGP连接 1-13
1.9.3 清除MBGP信息 1-14
1.10 MBGP典型配置举例 1-14

1 MBGP 配置

说明

在以下路由协议的介绍中所指的路由器及路由器图标，代表了一般意义下的路由器以及运行了路由协议的以太网交换机。

1.1 MBGP 简介

为了提供对多种网络层协议的支持，IETF（Internet Engineering Task Force，因特网工程任务组）对 BGP-4 进行了扩展，形成 MP-BGP（Multi-Protocol BGP，多协议 BGP），使 BGP 能够为多种路由应用提供路由信息。

组播的网络拓扑和单播拓扑有可能不同，需要通过 MP-BGP 扩展使得 BGP 能够将单播 NLRI（Network Layer Reachability Information，网络层可达性信息）和组播 NLRI 分开运载，其中组播 NLRI 专用于执行 RPF（Reverse Path Forwarding，逆向路径转发）功能。这使得在单播路由表和组播路由表中，对于相同的前缀有不同的路径选择，维护了一致的单播转发并使域间组播正常工作。

目前的 MP-BGP 标准是 RFC 2858（Multiprotocol Extensions for BGP-4，BGP-4 的多协议扩展）。

MP-BGP 在组播上的应用简称为 MBGP（Multicast BGP，组播 BGP）。

说明

- 本章主要介绍 MP-BGP 应用于组播的配置事项，即 MBGP 配置事项，有关 BGP 的详细内容以及配置事项，请参见“IP 路由分册”中的“BGP 配置”。
- 有关 RPF 检查的详细内容以及配置事项，请参见“IP 组播分册”中的“组播路由与转发配置”。

1.2 协议规范

与 MBGP 相关的协议规范有：

- RFC 2858：Multiprotocol Extensions for BGP-4
- RFC 3392：Capabilities Advertisement with BGP-4
- draft-ietf-idmr-bgp-mcast-attr-00：BGP Attributes for Multicast Tree Construction

1.3 MBGP 配置任务简介

表1-1 MBGP 配置任务简介

配置任务		说明	详细配置
配置 MBGP 基本功能		必选	1.4.2
控制路由信息的发布与接收	配置 MBGP 引入其他路由	必选	1.5.2
	配置 MBGP 引入缺省路由	可选	1.5.3
	配置 MBGP 路由聚合	可选	1.5.4
	配置向 IPv4 MBGP 对等体/IPv4 MBGP 对等体组发送缺省路由	可选	1.5.5
	配置 MBGP 路由信息的发布策略	可选	1.5.6
	配置 MBGP 路由信息的接收策略	可选	1.5.7
	配置 MBGP 路由衰减	可选	1.5.8
配置 MBGP 的路由属性	配置 MBGP 路由管理优先级	可选	1.6.2
	配置本地优先级缺省值	可选	1.6.3
	配置 MED 属性	可选	1.6.4
	配置发布路由时将自身地址做为下一跳	可选	1.6.5
	配置 AS_PATH 属性	可选	1.6.6
调整和优化 MBGP 网络	配置 MBGP 软复位	可选	1.7.2
	配置 MBGP 负载分担路由条数	可选	1.7.3
配置大型 MBGP 网络	配置 IPv4 MBGP 对等体组	可选	1.8.2
	配置 MBGP 团体	可选	1.8.3
	配置 MBGP 路由反射器	可选	1.8.4

1.4 配置 MBGP 基本功能

1.4.1 配置准备

在配置 MBGP 之前，需完成以下任务：相邻节点的网络层互通。

1.4.2 配置 MBGP 基本功能

表1-2 配置 MBGP 基本功能

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
指定对等体/对等体组的 AS 号	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	必选 缺省情况下，对等体/对等体组无 AS 号
创建并进入 MBGP 地址族视图	**ipv4-family multicast**	必选

操作	命令	说明
激活在 IPv4 单播视图下创建的对等体/对等体组	**peer** { *group-name* | *ip-address* } **enable**	必选 缺省情况下，对等体/对等体组在 BGP IPv4 组播地址族视图下处于非激活状态
为从 IPv4 MBGP 对等体/IPv4 MBGP 对等体组接收的路由分配首选值	**peer** { *group-name* | *ip-address* } **preferred-value** *value*	可选 缺省情况下，从 IPv4 MBGP 对等体/IPv4 MBGP 对等体组接收的路由的首选值为 0

1.5 控制路由信息的发布与接收

1.5.1 配置准备

在控制 MBGP 路由信息的发布与接收之前，需完成以下任务：配置 MBGP 基本功能。

1.5.2 配置 MBGP 引入其他路由

MBGP 可以向邻居 AS 发送本地 AS 内部网络的路由信息，但 MBGP 不是自己去发现 AS 内部的路由信息，而是引入 IGP 的路由信息到 MBGP 路由表中，并发布给 MBGP 对等体。在引入 IGP 路由时，还可以针对不同的路由协议来对路由信息进行过滤。

表1-3 配置 MBGP 引入其他路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
引入其它协议路由信息并通告	**import-route** *protocol* [*process-id* [**med** *med-value* | **route-policy** *route-policy-name*] *]	二者必选其一 缺省情况下，MBGP 不引入、不通告其它协议的路由，不发布任何网段路由
将网段路由发布到 MBGP 路由表中	**network** *ip-address* [*mask* | *mask-length*] [**short-cut** | **route-policy** *route-policy-name*]	

说明

- 通过 **import-route** 命令引入到 MBGP 路由表中的路由的 ORIGIN 属性为 Incomplete。
- 使用 **network** 命令发布到 MBGP 路由表中的网段路由的 ORIGIN 属性为 IGP。
- 要发布的网段路由必须存在于本地的 IP 路由表中，使用路由策略可以更为灵活的控制所发布的路由。

1.5.3 配置 MBGP 引入缺省路由

MBGP 不能通过 **import-route** 命令从其它协议引入缺省路由，如果要引入其它协议的缺省路由，必须要配置下列命令。

表1-4 配置 MBGP 引入其它协议的缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
引入其它协议路由信息并通告	**import-route** *protocol* [*process-id* [**med** *med-value* | **route-policy** *route-policy-name*] *]	必选 缺省情况下，MBGP 不引入其它协议的路由
允许将其它协议的缺省路由引入到 MBGP 路由表中	**default-route imported**	必选 缺省情况下，MBGP 不引入其它协议的缺省路由

1.5.4 配置 MBGP 路由聚合

在中型或大型 MBGP 网络中，在向组播对等体发布路由信息时，需要配置路由聚合，减小对等体路由表中的路由数量。MBGP 支持以下两种聚合方式：

自动聚合：对 MBGP 引入的 IGP 子网路由进行聚合。配置自动聚合后，MBGP 将不再发布从 IGP 引入的子网路由，而是发布聚合后的自然网段的路由。缺省路由和用 **network** 命令引入的路由不能进行自动聚合。

手动聚合：对 MBGP 本地路由进行聚合。手动聚合的优先级高于自动聚合的优先级。

表1-5 配置 MBGP 路由聚合

操作		命令	说明
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	-
进入 MBGP 地址族视图		**ipv4-family multicast**	-
配置 MBGP 路由聚合	配置对引入的子网路由进行自动聚合	**summary automatic**	二者必选其一 缺省情况下，不进行路由聚合 可以根据需求选择路由聚合方式；当二者同时配置时，手动路由聚合生效
	配置手动路由聚合	**aggregate** *ip-address* { *mask* | *mask-length* } [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	

1.5.5 配置向 IPv4 MBGP 对等体/IPv4 MBGP 对等体组发送缺省路由

表1-6 配置向 IPv4 MBGP/IPv4 MBGP 组发送缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-

操作	命令	说明
向 IPv4 MBGP 对等体/IPv4 MBGP 对等体组发送缺省路由	**peer** { *group-name* | *ip-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	必选 缺省情况下，不向 IPv4 MBGP 对等体/IPv4 MBGP 对等体组发送缺省路由

说明

执行 **peer default-route-advertise** 命令后，不论本地路由表中是否存在缺省路由，都将向指定 IPv4 MBGP 对等体/IPv4 MBGP 对等体组发布一条下一跳地址为本地地址的缺省路由。

1.5.6 配置 MBGP 路由信息的发布策略

用户可以根据需求选择过滤策略，同时配置几种过滤策略时，按照如下顺序执行：

- **filter-policy export**
- **peer filter-policy export**
- **peer as-path-acl export**
- **peer ip-prefix export**
- **peer route-policy export**

只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被发布。

表1-7 配置 MBGP 路由信息的发布策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置发布路由信息时对引入的路由进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct** | **isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	必选其一 缺省情况下，发布路由信息时不对路由进行过滤
对发布给 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的路由指定路由策略	**peer** { *group-name* | *peer-address* } **route-policy** *route-policy-name* **export**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 ACL 的过滤策略	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **export**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 AS 路径过滤列表的 MBGP 路由过滤策略	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **export**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 IP 前缀列表的路由过滤策略	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **export**	

1.5.7 配置 MBGP 路由信息的接收策略

通过配置 MBGP 路由信息的接收策略，可以对从 IPv4 MBGP 对等体/IPv4 MBGP 对等体组接收的路由进行过滤，只有满足指定条件的路由才能被接收并加到 MBGP 路由表中。

用户可以根据需求选择过滤策略，同时配置几种过滤策略时，按照如下顺序执行：

- **filter-policy import**
- **peer filter-policy import**
- **peer as-path-acl import**
- **peer ip-prefix import**
- **peer route-policy import**

只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被接收。

表1-8 配置 MBGP 路由信息的接收策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** as-number	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
对接收的路由信息进行过滤	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	必选其一 缺省情况下，不对接收的路由信息进行过滤
对来自 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的路由指定路由策略	**peer** { *group-name* | *ip-address* } **route-policy** *policy-name* **import**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 ACL 的过滤策略	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **import**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 AS 路径过滤列表的 MBGP 路由过滤策略	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **import**	
为 IPv4 MBGP 对等体/IPv4 MBGP 对等体组设置基于 IP 前缀列表的路由过滤策略	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **import**	
配置允许从 IPv4 MBGP 对等体/IPv4 MBGP 对等体组接收的最大路由数	**peer** { *group-name* | *ip-address* } **route-limit** limit [*percentage*]	可选 缺省情况下，允许从 IPv4 MBGP 对等体/IPv4 MBGP 对等体组接收的最大路由数无限制

注意

对等体组的成员可以与所在的组使用不同的入方向路由策略，即接收路由时，各对等体可以选择自己的策略。

1.5.8 配置 MBGP 路由衰减

通过配置 MBGP 衰减，可以抑制不稳定的路由信息，不将这类路由加入到 MBGP 路由表中，也不将这类路由向其他 IPv4 MBGP 对等体发布。

表1-9 配置 MBGP 路由衰减

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置 MBGP 路由衰减参数	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	必选 缺省情况下，没有配置 MBGP 路由衰减

1.6 配置 MBGP 的路由属性

MBGP 具有很多路由属性，利用这些属性可以改变 MBGP 的选路策略。

1.6.1 配置准备

在配置 MBGP 的路由属性之前，需完成以下任务：配置 MBGP 基本功能。

1.6.2 配置 MBGP 路由管理优先级

应用路由策略，可以为匹配过滤条件的特定路由配置优先级。对于那些没有匹配的路由，使用缺省优先级。

表1-10 配置 MBGP 路由管理优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置 MBGP 路由的管理优先级	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	可选 缺省情况下，MBGP EBGP 路由的管理优先级为 255，MBGP IBGP 路由的管理优先级为 255，本地产生的 MBGP 路由的管理优先级为 130

1.6.3 配置本地优先级缺省值

表1-11 配置本地优先级缺省值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置本地优先级的缺省值	**default local-preference** *value*	可选 缺省情况下，本地优先级的缺省值为 100

1.6.4 配置 MED 属性

在其它条件相同的情况下，MED 较小的路由被优选作为自治系统的外部路由。

表1-12 配置 MED 属性

操作		命令	说明
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	-
进入 MBGP 地址族视图		**ipv4-family multicast**	-
配置MED属性	配置系统 MED 的缺省值	**default med** *med-value*	可选 缺省情况下，MED 的缺省值为 0
	配置允许比较来自不同 AS 邻居的路由路径的 MED 属性值	**compare-different-as-med**	可选 缺省情况下，不允许比较来自不同 AS 邻居的路由路径的 MED 属性值
	配置根据路由来自的 AS 进行分组对 MED 排序优选	**bestroute compare-med**	可选 缺省情况下，不根据路由来自的 AS 进行分组对 MED 排序优选
	配置允许比较联盟对等体的路由按 MED 值进行优选	**bestroute med-confederation**	可选 缺省情况下，比较联盟对等体的路由时不考虑 MED 值

1.6.5 配置发布路由时将自身地址作为下一跳

在某些组网环境中，本地路由器向 MBGP IBGP 对等体/MBGP IBGP 对等体组发布路由时，为保证 MBGP IBGP 邻居能够找到正确的下一跳，可以配置 **peer next-hop-local** 命令将自身地址作为下一跳地址。如果配置了负载分担，则不论是否配置了 **peer next-hop-local** 命令，本地路由器向 MBGP IBGP 对等体/MBGP IBGP 对等体组发布路由时都先将下一跳地址改变为自身地址。

在第三方下一跳（即两个 MBGP 连接在同一网段的广播网）这种特殊的组网环境中，缺省情况下，向 MBGP EBGP 对等体/MBGP EBGP 对等体组发布路由时，不将自身地址作为下一跳；只有配置了 **peer next-hop-local** 命令，才将自身地址作为下一跳。

表1-13 配置发布路由时将自身地址作为下一跳

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置发布路由时将自身地址作为下一跳	**peer** { *group-name* | *ip-address* } **next-hop-local**	可选 缺省情况下，向 MBGP EBGP 对等体/对等体组发布路由时，将自身地址作为下一跳；向 MBGP IBGP 对等体/对等体组发布路由时，不将自身地址作为下一跳

1.6.6 配置 AS_PATH 属性

通常情况下，MBGP 会检查对等体发来的路由的 AS_PATH 属性，如果其中已存在本地 AS 号，则 MBGP 会忽略此路由，以免形成路由环路。

表1-14 配置 AS_PATH 属性

操作		命令	说明
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	-
进入 MBGP 地址族视图		**ipv4-family multicast**	-
配置 AS_PATH 属性	配置允许本地 AS 号重复出现的次数	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	可选 缺省情况下，不允许本地 AS 重复
	禁止路由器将 AS_PATH 当作选路算法中的一个因素	**bestroute as-path-neglect**	可选 缺省情况下，路由器可以将 AS_PATH 当作选路算法中的一个因素
	配置发送 MBGP 更新报文时 AS_PATH 属性中不携带私有 AS 号	**peer** { *group-name* | *ip-address* } **public-as-only**	可选 缺省情况下，发送 MBGP 更新报文时，允许携带私有自治系统号

1.7 调整和优化 MBGP 网络

1.7.1 配置准备

在调整和优化 MBGP 网络之前，需完成以下任务：配置 MBGP 基本功能。

1.7.2 配置 MBGP 软复位

MBGP 的选路策略改变后，为了使新的策略生效，必须复位 MBGP 连接，但这样会造成短暂的 MBGP 连接中断。

通过使能 Route-Refresh 功能，当策略改变后，系统可以在不中断 MBGP 连接的情况下，自动对 MBGP 路由表进行动态刷新。

如果对等体不支持 Route-Refresh 功能，则可以将从对等体接收的所有路由更新保存在本地，当选路策略发生改变后，在不中断连接的情况下重新刷新 MBGP 路由表，并应用新的策略。

1. 通过 Route-Refresh 实现 MBGP 软复位

在对等体支持并使能 Route-Refresh 功能的情况下，如果 MBGP 的路由策略发生了变化，本地路由器会向 MBGP 对等体发布 Router-Refresh 消息，收到此消息的对等体会将其路由信息重新发给本地路由器。这样，在不中断 MBGP 连接的情况下，就可以对 MBGP 路由表进行动态更新，并应用新的选路策略。

表1-15 通过 Route-Refresh 实现 MBGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
使能 BGP 路由刷新功能	**peer** { *group-name* | *ip-address* } **capability-advertise route-refresh**	可选 缺省情况下，路由刷新功能处于使能状态

2. 通过将所有路由更新保存在本地实现 MBGP 软复位

当对等体不支持 Route-Refresh 功能时，可通过配置 **peer keep-all-routes** 命令实现软复位功能。

用户也可以通过执行 **refresh bgp ipv4 multicast** 命令对保存在本地的所有路由重新过一遍策略。

表1-16 通过将所有路由更新保存在本地实现 MBGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
禁止 BGP 路由刷新和多协议扩展功能	**peer** { *group-name* | *ip-address* } **capability-advertise conventional**	可选 缺省情况下，BGP 路由刷新和多协议扩展功能处于使能状态
进入 MBGP 地址族视图	**ipv4-family multicast**	-
保存所有来自 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的原始路由信息，即使这些路由没有通过已配置的入口策略	**peer** { *group-name* | *ip-address* } **keep-all-routes**	必选 缺省情况下，不保存 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的原始路由信息
退回用户视图	**return**	-
手工对 MBGP 连接进行软复位	**refresh bgp ipv4 multicast** { **all** | *ip-address* | **group** *group-name* | **external** | **internal** } { **export** | **import** }	可选

1.7.3 配置 MBGP 负载分担路由条数

表1-17 配置 MBGP 负载分担路由条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
配置进行 MBGP 负载分担的最大路由条数	**balance** *number*	必选 没有进行 MBGP 负载分担

1.8 配置大型 MBGP 网络

1.8.1 配置准备

在配置大型 MBGP 网络之前，需完成以下任务：相邻节点的网络层互通。

1.8.2 配置 IPv4 MBGP 对等体组

在大型网络中，对等体的数目众多，配置和维护极为不便。使用对等体组可以降低管理的难度，还可以提高路由发布效率。可以在 MBGP 中使能 IPv4 单播对等体组，从而达到共同管理组内 IPv4 MBGP 对等体的目的。MBGP 中不能单独创建对等体组。

表1-18 配置 IPv4 MBGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	-
创建 BGP 对等体组	**group** *group-name* [**external** | **internal**]	必选 缺省情况下，没有创建 BGP 对等体组
将对等体加入已存在的对等体组	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	必选 缺省情况下，对等体不属于任何对等体组
进入 MBGP 地址族视图	**ipv4-family multicast**	-
使能 IPv4 单播对等体组	**peer** *group-name* **enable**	必选
将 IPv4 MBGP 对等体加入对等体组中	**peer** *ip-address* **group** *group-name*	必选 缺省情况下，IPv4 MBGP 对等体不属于任何对等体组

注意

- 配置 IPv4 MBGP 对等体组的前提是，在 BGP-IPv4 组播地址族下存在已经使能的 IPv4 单播对等体组。
- 将 IPv4 MBGP 对等体加入对等体组，需要先在 BGP 视图下将相应的对等体加入对等体组。

1.8.3 配置 MBGP 团体

对等体组可以使一组对等体共享相同的策略，而利用团体可以使多个 AS 中的一组 BGP 路由器共享相同的策略。团体是一个路由属性，在 BGP 对等体之间传播，它并不受到 AS 范围的限制。

BGP 路由器在将带有团体属性的路由发布给其它对等体之前，可以改变此路由原有的团体属性。除了使用公认的团体属性外，用户还可以使用团体属性列表自定义扩展团体属性，以便更为灵活地控制路由策略。

表1-19 配置 MBGP 团体

操作		命令	说明
进入系统视图		**system-view**	-
进入 BGP 视图		**bgp** *as-number*	-
进入 MBGP 地址族视图		**ipv4-family multicast**	-
配置向 MBGP 对等体/MBGP 对等体组发布团体属性	配置发布团体属性	**peer** { *group-name* | *ip-address* } **advertise-community**	必选 缺省情况下，不将团体属性和扩展团体属性发布给任何 IPv4 MBGP 对等体/IPv4 MBGP 对等体组
	配置发布扩展团体属性	**peer** { *group-name* | *ip-address* } **advertise-ext-community**	
对发布给 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的路由指定路由策略		**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，不指定 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的路由策略

注意

- 配置 MBGP 团体时，必须使用路由策略来定义具体的团体属性，然后在发布路由信息时应用此路由策略。
- 关于路由策略的配置，请参见“IP 路由分册”中的“路由策略配置”。

1.8.4 配置 MBGP 路由反射器

在 AS 内部，为保证 MBGP IBGP 对等体之间的连通性，需要在 MBGP IBGP 对等体之间建立全连接关系。当 MBGP IBGP 对等体数目很多时，在 MBGP IBGP 对等体之间建立全连接的开销很大。使用路由反射器，可以解决这个问题。

表1-20 配置 MBGP 路由反射器

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 MBGP 地址族视图	**ipv4-family multicast**	-
将本设备配置为路由反射器，并指定 IPv4 MBGP 对等体（组）作为路由反射器的客户	**peer** { *group-name* | *peer-address* } **reflect-client**	必选 缺省情况下，没有配置路由反射器及其客户
配置允许客户到客户的路由反射	**reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
配置路由反射器的集群 ID	**reflector cluster-id** *cluster-id*	可选 缺省情况下，每个路由反射器是使用自己的 Router ID 作为集群 ID

注意

- 通常情况下，路由反射器的客户之间不要求是全连接的，路由缺省通过反射器从一个客户反射到其它客户；如果客户之间是全连接的，可以禁止客户间的反射，以便减少开销。
- 通常，一个集群里只有一个路由反射器。此时是由反射器的路由器 ID 来识别该集群的。设置多个路由反射器可提高网络的稳定性。如果一个集群中配有多个路由反射器，请使用相关命令为所有的路由反射器配置同样的集群 ID，以避免路由循环。

1.9 MBGP 显示和维护

1.9.1 MBGP 配置显示

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MBGP 的运行情况，通过查看显示信息验证配置的效果。

表1-21 MBGP 配置显示

操作	命令
显示 IPv4 BGP 组播路由表中的路由信息	**display ip multicast routing-table** [**verbose**]
显示指定目的地址的组播路由信息	**display ip multicast routing-table** *ip-address* [*mask-length* | *mask*] [**longer-match**] [**verbose**]
显示 IPv4 MBGP 对等体组信息	**display bgp multicast group** [*group-name*]
显示 MBGP 发布的路由信息	**display bgp multicast network**
显示 AS 路径信息	**display bgp multicast paths** [*as-regular-expression*]
显示 IPv4 MBGP 对等体/IPv4 MBGP 对等体组的信息	**display bgp multicast peer** [*ip-address*] [**verbose**]
显示 MBGP 路由信息	**display bgp multicast routing-table** [*ip-address* [{ *mask* | *mask-length* } [**longer-prefixes**]]]
显示匹配指定 AS 路径过滤列表的 MBGP 路由信息	**display bgp multicast routing-table as-path-acl** *as-path-acl-number*
显示 CIDR 的 MBGP 路由信息	**display bgp multicast routing-table cidr**
显示指定 MBGP 团体的 MBGP 路由信息	**display bgp multicast routing-table community**[*aa:nn&<1-13>*] [**no-advertise** | **no-export** | **no-export-subconfed**] * [**whole-match**]
显示匹配指定 MBGP 团体列表的 MBGP 路由	**display bgp multicast routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>
显示 MBGP 衰减的 MBGP 路由信息	**display bgp multicast routing-table dampened**
显示 MBGP 衰减的配置参数	**display bgp multicast routing-table dampening parameter**
显示源 AS 不一致的路由	**display bgp multicast routing-table different-origin-as**
显示 MBGP 路由振荡统计信息	**display bgp multicast routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *ip-address* [{ *mask* | *mask-length* } [**longer-match**]]]
显示向指定的 IPv4 MBGP 对等体发送或者从 IPv4 MBGP 对等体收到的路由信息	**display bgp multicast routing-table peer** *ip-address* { **advertised-routes** | **received-routes** } [*network-address* [*mask* | *mask-length*] | **statistic**]
显示与指定的 AS 路径正则表达式相匹配的 IPv4 MBGP 路由信息	**display bgp multicast routing-table regular-expression** *as-regular-expression*
显示 MBGP 的路由统计信息	**display bgp multicast routing-table statistic**

1.9.2 复位 MBGP 连接

当 MBGP 路由策略或协议发生变化后，如果需要通过复位 MBGP 连接使新的配置生效，请在用户视图下进行下列配置。

表1-22 复位 MBGP 连接

操作	命令
复位指定的 MBGP 连接	**reset bgp ipv4 multicast** { **all** | *as-number* | *ip-address* | **group** *group-name* | **external** | **internal** }

1.9.3 清除 MBGP 信息

在用户视图下，执行 **reset** 命令可以清除 MBGP 相关统计信息。

表1-23 清除 MBGP 信息

操作	命令
清除路由的衰减信息并释放被抑制的路由	**reset bgp ipv4 multicast dampening** [*ip-address* [*mask* | *mask-length*]]
清除路由的振荡统计信息	**reset bgp ipv4 multicast flap-info** [**regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number* | *ip-address* [*mask* | *mask-length*]]

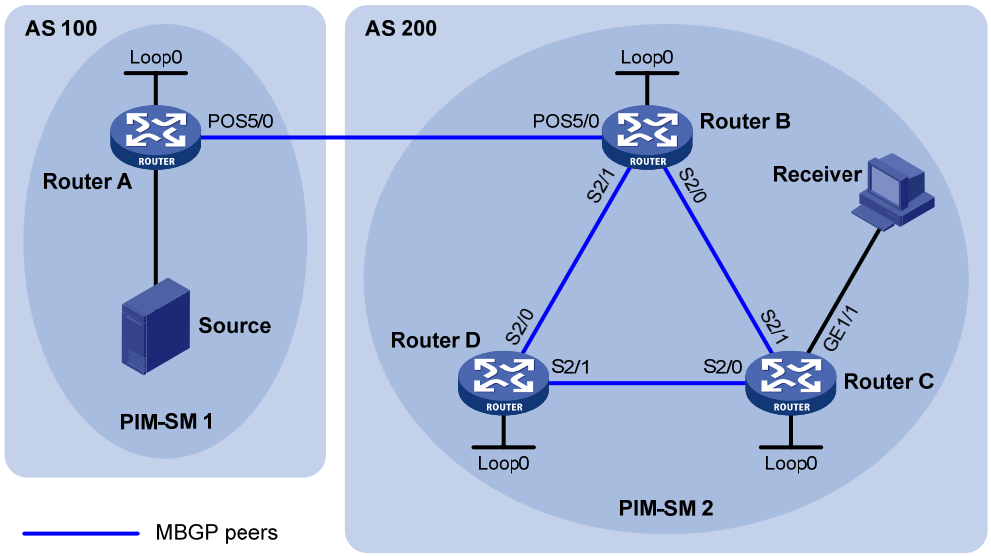
1.10 MBGP 典型配置举例

1. 组网需求

- 网络中存在两个自治系统：PIM-SM 1 属于 AS 100，PIM-SM 2 属于 AS 200。各 AS 内部采用 OSPF 进行互联，AS 之间采用 MBGP 交换组播路由信息；
- 组播源属于 AS 100 内的 PIM-SM 1，接收者则属于 AS 200 内的 PIM-SM 2；
- 将 Router A 和 Router B 各自的 Loopback0 接口分别配置为各自 PIM-SM 域的 C-BSR 和 C-RP；
- 在 Router A 与 Router B 之间通过 MBGP 建立 MSDP 对等体关系。

2. 组网图

图1-1 MBGP 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source	-	10.110.1.100/24	Router C	GE1/1	10.110.2.1/24
Router A	GE1/1	10.110.1.1/24		S2/0	192.168.4.1/24
	POS5/0	192.168.1.1/24		S2/1	192.168.2.2/24
	Loop0	1.1.1.1/32		Loop0	3.3.3.3/32
Router B	POS5/0	192.168.1.2/24	Router D	S2/0	192.168.3.2/24
	S2/0	192.168.2.1/24		S2/1	192.168.4.2/24
	S2/1	192.168.3.1/24		Loop0	4.4.4.4/32
	Loop0	2.2.2.2/32			

3. 配置步骤

(1) 配置各路由器接口的 IP 地址和单播路由协议

- 请按照 图 1-1配置各接口的IP地址和子网掩码，具体配置过程略。
- 配置 AS200 内的各路由器之间采用 OSPF 路由协议交换路由信息（AS 内各路由器创建的 OSPF 进程号为 1），确保各 AS 内部在网络层互通，能学到彼此 Loopback 接口的路由，具体配置过程略。

(2) 使能 IP 组播路由，使能 PIM-SM 和 IGMP，并配置 BSR 的服务边界

在 Router A 上使能 IP 组播路由，在各接口上使能 PIM-SM。

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitGigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] pim sm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface pos 5/0
[RouterA-Pos5/0] pim sm
[RouterA-Pos5/0] quit
```

Router B 和 Router D 上的配置与 Router A 相似，配置过程略。

在 Router C 上使能 IP 组播路由，在各接口上使能 PIM-SM，并在主机侧接口 GigabitEthernet1/1 上使能 IGMP。

```
<RouterC> system-view
[RouterC] multicast routing-enable
[RouterC] interface serial 2/0
[RouterC-Serial2/0] pim sm
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] pim sm
[RouterC-Serial2/1] quit
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] pim sm
[RouterC-GigabitEthernet1/1] igmp enable
[RouterC-GigabitEthernet1/1] quit
```

在 Router A 上配置 BSR 的服务边界。

```
[RouterA] interface pos 5/0
[RouterA-Pos5/0] pim bsr-boundary
[RouterA-Pos5/0] quit
```

在 Router B 上配置 BSR 的服务边界。

```
[RouterB] interface pos 5/0
[RouterB-Pos5/0] pim bsr-boundary
[RouterB-Pos5/0] quit
```

(3) 配置 Loopback0 接口和 C-BSR、C-RP 的位置

在 Router A 上配置 Loopback0 接口和 C-BSR、C-RP 的位置。

```
[RouterA] interface loopback 0
```

```
[RouterA-LoopBack0] ip address 1.1.1.1 32
[RouterA-LoopBack0] pim sm
[RouterA-LoopBack0] quit
[RouterA] pim
[RouterA-pim] c-bsr loopback 0
[RouterA-pim] c-rp loopback 0
[RouterA-pim] quit
```

在 Router B 上配置 Loopback0 接口和 C-BSR、C-RP 的位置。

```
[RouterB] interface loopback 0
[RouterB-LoopBack0] ip address 2.2.2.2 32
[RouterB-LoopBack0] pim sm
[RouterB-LoopBack0] quit
[RouterB] pim
[RouterB-pim] c-bsr loopback 0
[RouterB-pim] c-rp loopback 0
[RouterB-pim] quit
```

(4) 配置 BGP 协议，配置 MBGP 对等体

在 Router A 上配置 EBGP 邻接关系、配置 MBGP 对等体。

```
[RouterA] bgp 100
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] peer 192.168.1.2 as-number 200
[RouterA-bgp] import-route direct
[RouterA-bgp] ipv4-family multicast
[RouterA-bgp-af-mul] peer 192.168.1.2 enable
[RouterA-bgp-af-mul] import-route direct
[RouterA-bgp-af-mul] quit
[RouterA-bgp] quit
```

在 Router B 上配置 EBGP 邻接关系、配置 MBGP 对等体，并引入 OSPF 路由。

```
[RouterB] bgp 200
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] peer 192.168.1.1 as-number 100
[RouterB-bgp] import-route ospf 1
[RouterB-bgp] ipv4-family multicast
[RouterB-bgp-af-mul] peer 192.168.1.1 enable
[RouterB-bgp-af-mul] import-route ospf 1
[RouterB-bgp-af-mul] quit
[RouterB-bgp] quit
```

(5) 配置 MSDP 对等体

在 Router A 上配置 MSDP 对等体。

```
[RouterA] msdp
[RouterA-msdp] peer 192.168.1.2 connect-interface pos 5/0
[RouterA-msdp] quit
```

在 Router B 上配置 MSDP 对等体。

```
[RouterB] msdp
[RouterB-msdp] peer 192.168.1.1 connect-interface pos 5/0
```

```
[RouterB-msdp] quit
```

(6) 检验配置效果

通过使用 **display bgp multicast peer** 命令可以查看路由器之间 MBGP 对等体的关系。例如：

查看 Router B 上 MBGP 对等体关系的信息。

```
[RouterB] display bgp multicast peer

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.1 4   100       56       56     0       0 00:40:54 Established
```

通过使用 **display msdp brief** 命令可以查看路由器之间 MSDP 对等体建立情况。例如：

查看 Router B 上 MSDP 对等体建立情况的简要信息。

```
[RouterB] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0
  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.1.1       Up       00:07:17        100    1          0
```

目 录

1 IPv6 组播路由与转发配置 1-1
1.1 IPv6 组播路由与转发简介 1-1
1.1.1 RPF检查机制 1-1
1.2 IPv6 组播路由与转发配置任务简介 1-3
1.3 使能IPv6 组播路由 1-3
1.4 配置IPv6 组播路由与转发 1-4
1.4.1 配置准备 1-4
1.4.2 配置IPv6 组播路由策略 1-4
1.4.3 配置IPv6 组播转发范围 1-4
1.4.4 配置IPv6 组播转发表容量 1-5
1.5 IPv6 组播路由与转发显示和维护 1-5
1.6 常见配置错误举例 1-6
1.6.1 IPv6 组播数据异常终止 1-6

1 IPv6 组播路由与转发配置

1.1 IPv6 组播路由与转发简介

在 IPv6 组播实现中，组播路由和转发分为三种表：

- 每个 IPv6 组播路由协议都有一个协议自身的路由表，如 IPv6 PIM 路由表（IPv6 PIM Routing-Table）；
- 各 IPv6 组播路由协议的组播路由信息经过综合形成一个总的 IPv6 组播路由表（IPv6 Multicast Routing-Table）；
- IPv6 组播转发表（IPv6 Multicast Forwarding-Table）直接用于控制 IPv6 组播数据包的转发，是真正指导 IPv6 组播数据转发的转发表。

IPv6 组播路由表由一组（S，G）表项组成，其中（S，G）表示由源 S 向 IPv6 组播组 G 发送 IPv6 组播数据的路由信息。如果路由器支持多种 IPv6 组播路由协议，则其 IPv6 组播路由表中将包括由多种协议生成的组播路由。路由器根据组播路由和转发策略，从 IPv6 组播路由表中选出最优的组播路由，并下发到 IPv6 组播转发表中。

1.1.1 RPF 检查机制

IPv6 组播路由协议依赖于现有的 IPv6 单播路由信息或 IPv6 MBGP 路由来创建 IPv6 组播路由表项。IPv6 组播路由协议在创建 IPv6 组播路由表项时，运用了 RPF（Reverse Path Forwarding，逆向路径转发）检查机制，以确保 IPv6 组播数据能够沿正确的路径传输，同时还能避免由于各种原因而造成的环路。

1. RPF 检查过程

执行 RPF 检查的依据是 IPv6 单播路由或 IPv6 MBGP 路由：

- IPv6 单播路由表中汇集了到达各个目的网段的最短路径；
- IPv6 MBGP 路由表直接提供 IPv6 组播路由信息。

在执行 RPF 检查时，路由器同时查找 IPv6 单播路由表和 IPv6 MBGP 路由表，具体过程如下：

(1) 首先，分别从 IPv6 单播路由表和 IPv6 MBGP 路由表中各选出一条最优路由：

- 以“报文源”的 IPv6 地址为目的地址查找 IPv6 单播路由表，自动选取一条最优 IPv6 单播路由。对应表项中的出接口为 RPF 接口，下一跳为 RPF 邻居。路由器认为来自 RPF 邻居且由该 RPF 接口收到的 IPv6 组播报文所经历的路径是从源 S 到本地的最短路径。
- 以“报文源”的 IPv6 地址为目的地址查找 IPv6 MBGP 路由表，自动选取一条最优 IPv6 MBGP 路由。对应表项中的出接口为 RPF 接口，下一跳为 RPF 邻居。

(2) 然后，从这两条最优路由中选择一条作为 RPF 路由：

- 如果配置了按照最长匹配选择路由，则从这两条路由中选出最长匹配的那条路由；如果这两条路由的前缀长度一样，则选择其中优先级最高的那条路由；如果它们的优先级也相同，则按照 IPv6 MBGP 路由、IPv6 单播路由的顺序进行选择。

- 如果没有配置按照最长匹配选择路由，则从这两条路由中选出优先级最高的那条路由；如果它们的优先级相同，则按照 IPv6 MBGP 路由、IPv6 单播路由的顺序进行选择。



说明

根据 IPv6 组播报文传输的具体情况不同，“报文源”所代表的具体含义也不同：

- 如果当前报文沿从组播源到接收者或 RP（Rendezvous Point，汇集点）的 SPT（Shortest Path Tree，最短路径树）进行传输，则以组播源为“报文源”进行 RPF 检查；
- 如果当前报文沿从 RP 到接收者的 RPT（Rendezvous Point Tree，共享树）进行传输，则以 RP 为“报文源”进行 RPF 检查；
- 如果当前报文为 BSR（BootStrap Router，自举路由器）报文，沿从 BSR 到各路由器的路径进行传输，则以 BSR 为“报文源”进行 RPF 检查。

有关 SPT、RPT、RP 和 BSR 的详细介绍，请参见“IP 组播分册”中的“IPv6 PIM 配置”。

---

2. RPF 检查在 IPv6 组播转发中的应用

对每一个收到的 IPv6 组播数据报文都进行 RPF 检查会给路由器带来较大负担，而利用 IPv6 组播转发表可以解决这个问题。在建立 IPv6 组播路由和转发表时，会把 IPv6 组播数据报文（S，G）的 RPF 接口记录为（S，G）表项的入接口。当路由器收到 IPv6 组播数据报文（S，G）后，查找 IPv6 组播转发表：

(1) 如果 IPv6 组播转发表中不存在（S，G）表项，则对该报文执行 RPF 检查，将其 RPF 接口作为入接口，结合相关路由信息创建相应的表项，并下发到 IPv6 组播转发表中：

- 若该报文实际到达的接口正是其 RPF 接口，则 RPF 检查通过，向所有的出接口转发该报文；
- 若该报文实际到达的接口不是其 RPF 接口，则 RPF 检查失败，丢弃该报文。

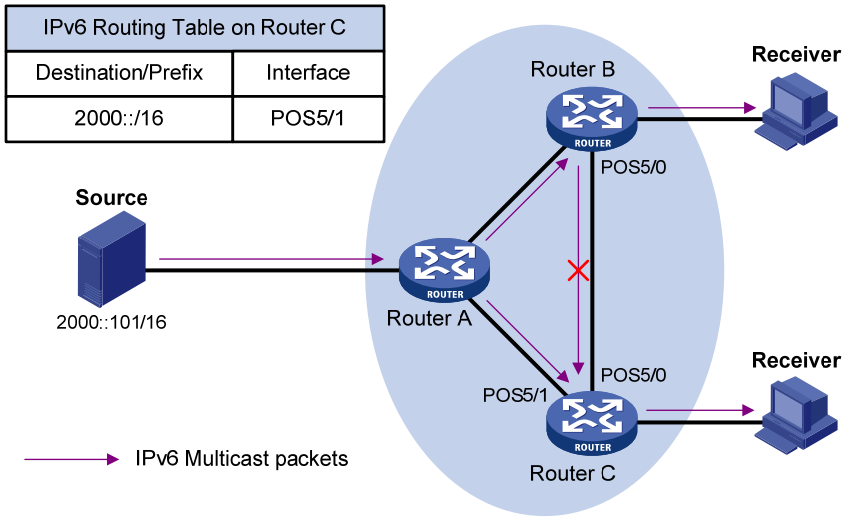
(2) 如果 IPv6 组播转发表中已存在（S，G）表项，且该报文实际到达的接口与入接口相匹配，则向所有的出接口转发该报文。

(3) 如果 IPv6 组播转发表中已存在（S，G）表项，但该报文实际到达的接口与入接口不匹配，则对此报文执行 RPF 检查：

- 若其 RPF 接口与入接口一致，则说明（S，G）表项正确，丢弃这个来自错误路径的报文；
- 若其 RPF 接口与入接口不符，则说明（S，G）表项已过时，于是把入接口更新为 RPF 接口。如果该报文实际到达的接口正是其 RPF 接口，则向所有的出接口转发该报文，否则将其丢弃。

如 图 1-1所示，假设网络中IPv6 单播路由畅通，未配置IPv6 MBGP。IPv6 组播报文（S，G）沿从组播源（Source）到接收者（Receiver）的SPT进行传输。假定Router C上的IPv6 组播转发表中已存在（S，G）表项，其记录的入接口为POS5/1。

图1-1 RPF 检查过程



- 如果该 IPv6 组播报文从接口 POS5/1 到达 Router C，与（S，G）表项的入接口相匹配，则向所有的出接口转发该报文。
- 如果该 IPv6 组播报文从接口 POS5/0 到达 Router C，与（S，G）表项的入接口不匹配，则对其执行 RPF 检查：通过查找 IPv6 单播路由表发现到达 Source 的出接口（即 RPF 接口）是 POS5/1，与（S，G）表项的入接口一致。这说明（S，G）表项是正确的，该报文来自错误的路径，RPF 检查失败，于是丢弃该报文。

1.2 IPv6 组播路由与转发配置任务简介

表1-1 IPv6 组播路由与转发配置任务简介

配置任务		说明	详细配置
使能 IPv6 组播路由		必选	1.3
配置 IPv6 组播路由与转发	配置 IPv6 组播路由策略	可选	1.4.2
	配置 IPv6 组播转发范围	可选	1.4.3
	配置 IPv6 组播转发表容量	可选	1.4.4

1.3 使能 IPv6 组播路由

在配置各项三层 IPv6 组播功能之前，必须首先使能 IPv6 组播路由。

表1-2 使能 IPv6 组播路由

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 组播路由	**multicast ipv6 routing-enable**	必选 缺省情况下，IPv6 组播路由处于关闭状态

1.4 配置 IPv6 组播路由与转发

1.4.1 配置准备

在配置 IPv6 组播路由与转发之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 配置 IPv6 PIM-DM（或 IPv6 PIM-SM）

在配置 IPv6 组播路由与转发之前，需准备以下数据：

- 单条 IPv6 组播转发表项的最大下行节点数目
- IPv6 组播转发表的最大表项数

1.4.2 配置 IPv6 组播路由策略

可以配置组播路由器按照最长匹配原则来选择RPF路由，有关RPF路由选择的详细介绍，请参见“1.1.1 1. RPF检查过程”一节。此外，通过配置根据组播源或组播源组进行IPv6 组播流量的负载分担，还可以优化存在多条IPv6 组播数据流时的网络流量。

表1-3 配置 IPv6 组播路由策略

操作	命令	说明
进入系统视图	**system-view**	-
配置按照最长匹配选择 RPF 路由	**multicast ipv6 longest-match**	可选 缺省情况下，选择优先级最高的路由作为 RPF 路由
配置对 IPv6 组播流量进行负载分担	**multicast ipv6 load-splitting {source | source-group }**	可选 缺省情况下，不对 IPv6 组播流量进行负载分担

1.4.3 配置 IPv6 组播转发范围

IPv6 组播信息在网络中的转发并不是漫无边际的，每个 IPv6 组播组对应的 IPv6 组播信息都必须在确定的范围内传递。目前定义 IPv6 组播转发范围的方式为确定充当 IPv6 组播转发边界的接口，以形成一个封闭的 IPv6 组播转发区域。

可以在所有支持 IPv6 组播转发的接口上配置针对某个 IPv6 组播组的转发边界。IPv6 组播转发边界为指定范围的 IPv6 组播组划定了边界条件，如果 IPv6 组播报文的目的地址与边界条件匹配，就停止转发。当在一个接口上配置了 IPv6 组播转发边界后，将不能从该接口转发 IPv6 组播报文（包括本机发出的 IPv6 组播报文），也不能从该接口接收 IPv6 组播报文。

表1-4 配置 IPv6 组播转发范围

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 IPv6 组播转发边界	**multicast ipv6 boundary** *ipv6-group-address prefix-length*	必选 缺省情况下，没有配置 IPv6 组播转发边界

1.4.4 配置 IPv6 组播转发表容量

路由器为每个收到的 IPv6 组播数据报文都维护相应的转发表项。但是，过量的 IPv6 组播转发表项可能会耗尽路由器内存，从而导致路由器性能下降。用户可以根据实际组网情况和业务性能要求对 IPv6 组播转发表中的表项数量进行限制。如果 IPv6 组播转发表最大表项数的配置值小于当前值，则超出数目的表项并不会立刻被删除，而必须由 IPv6 组播路由协议来删除，同时也无法添加新的 IPv6 组播转发表项。

路由器为每个下行节点复制一份 IPv6 组播数据报文并发送出去，每个下行节点就形成 IPv6 组播分发树的一条分支。用户可以根据实际组网情况和业务性能要求对组播转发表中单条表项的下行节点数目（即出接口数目）进行限制，以缓解路由器的复制压力。如果单条 IPv6 组播转发表项的最大下行节点数目的配置值小于当前值，则超出数目的下行节点并不会被立刻删除，而必须由 IPv6 组播路由协议来删除，同时新增的下行节点将无法添加到转发表该表项中。

表1-5 配置 IPv6 组播转发表容量

操作	命令	说明
进入系统视图	**system-view**	-
配置 IPv6 组播转发表的最大表项数	**multicast ipv6 forwarding-table route-limit** *limit*	可选 缺省情况下，IPv6 组播转发表的最大表项数为系统所允许的最大值，该值随设备的不同而有所差别，请以设备的实际情况为准
配置单条 IPv6 组播转发表项的最大下行节点数目	**multicast ipv6 forwarding-table downstream-limit** *limit*	可选 缺省情况下，单条 IPv6 组播转发表项的最大下行节点数目为系统所允许的最大值，该值随设备的不同而有所差别，请以设备的实际情况为准

1.5 IPv6 组播路由与转发显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 组播路由与转发的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv6 组播路由与转发的统计信息。

表1-6 IPv6 组播路由与转发显示和维护

操作	命令
查看 IPv6 组播边界信息	**display multicast ipv6 boundary** [*ipv6-group-address* [*prefix-length*]] [**interface** *interface-type interface-number*]
查看 IPv6 组播转发表信息（集中式设备）	**display multicast ipv6 forwarding-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics**] * [**port-info**]
查看 IPv6 组播转发表信息（分布式设备）	**display multicast ipv6 forwarding-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics** | **slot** *slot-number*] * [**port-info**]
查看 IPv6 组播路由表信息	**display multicast ipv6 routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } }] *

操作	命令
查看 IPv6 组播源的 RPF 信息	**display multicast ipv6 rpf-info** *ipv6-source-address* [*ipv6-group-address*]
清除 IPv6 组播转发表中的转发项	**reset multicast ipv6 forwarding-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }
清除 IPv6 组播路由表中的路由项	**reset multicast ipv6 routing-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }

注意

- 执行 **reset** 命令将清除 IPv6 组播路由表或 IPv6 组播转发表中的信息，可能导致 IPv6 组播信息无法正常传输；
- 清除 IPv6 组播路由表中的路由项后，IPv6 组播转发表中的相应表项也将随之删除；
- 清除 IPv6 组播转发表中的转发项后，IPv6 组播路由表中的相应表项也将随之删除。

1.6 常见配置错误举例

1.6.1 IPv6 组播数据异常终止

1. 故障现象

- 当某主机发送了加入 IPv6 组播组 G 的报文后，离该主机最近的路由器上却没有 IPv6 组播组 G 的组成员信息。中间路由器能成功接收 IPv6 组播数据，但数据不能到达末梢网络。
- 中间路由器的接口收到 IPv6 组播数据，但在 IPv6 PIM 路由表里没有相应的（S，G）表项。

2. 分析

- 命令 **multicast ipv6 boundary** 用来过滤接口上收到的 IPv6 组播报文。如果报文没有通过这个命令的 IPv6 ACL 匹配规则，IPv6 PIM 不能创建路由项。
- 此外，IPv6 PIM 中的命令 **source-policy** 用来过滤接收的 IPv6 组播报文。如果报文没有通过这个命令的 IPv6 ACL 匹配规则，IPv6 PIM 也不能创建路由项。

3. 处理过程

(1) 使用命令 **display current-configuration** 查看组播转发边界上配置的 IPv6 ACL 过滤规则。更改 **multicast ipv6 boundary** 命令的 IPv6 ACL 规则，使 IPv6 组播数据的源地址和 IPv6 组播组地址通过 IPv6 ACL 过滤。

(2) 检查组播过滤器配置。使用命令 **display current-configuration** 查看组播过滤器的配置，更改 **source-policy** 命令的 IPv6 ACL 规则，使 IPv6 组播数据的源地址和 IPv6 组播组地址通过 IPv6 ACL 过滤。

目 录

1 MLD配置 1-1
1.1 MLD简介 1-1
1.1.1 MLD的版本 1-1
1.1.2 MLDv1 原理简介 1-1
1.1.3 MLDv2 原理简介 1-3
1.1.4 MLD报文类型 1-4
1.1.5 MLD SSM Mapping 1-6
1.1.6 MLD Proxying 1-7
1.1.7 协议规范 1-8
1.2 MLD配置任务简介 1-9
1.3 配置MLD基本功能 1-9
1.3.1 配置准备 1-9
1.3.2 使能MLD 1-9
1.3.3 配置MLD版本 1-10
1.3.4 配置静态加入 1-10
1.3.5 配置IPv6 组播组过滤器 1-11
1.3.6 配置IPv6 组播组数量限制 1-11
1.4 调整MLD性能 1-12
1.4.1 配置准备 1-12
1.4.2 配置MLD报文选项 1-12
1.4.3 配置MLD查询和响应 1-13
1.4.4 配置MLD快速离开 1-15
1.5 配置MLD SSM Mapping 1-16
1.5.1 配置准备 1-16
1.5.2 使能MLD SSM Mapping 1-16
1.5.3 配置MLD SSM Mapping规则 1-17
1.6 配置MLD Proxying 1-17
1.6.1 配置准备 1-17
1.6.2 使能MLD Proxying 1-17
1.6.3 配置下行接口的IPv6 组播转发能力 1-18
1.7 MLD显示和维护 1-18
1.8 MLD典型配置举例 1-19
1.8.1 MLD基本功能配置举例 1-19
1.8.2 MLD SSM Mapping功能配置举例 1-21
1.8.3 MLD Proxying功能配置举例 1-24
1.9 常见配置错误举例 1-26
1.9.1 接收者侧路由器上无组成员信息 1-26
1.9.2 同一网段各路由器上组成员关系不一致 1-27

1 MLD 配置

1.1 MLD 简介

MLD 是 Multicast Listener Discovery Protocol（组播侦听者发现协议）的简称，它用于 IPv6 路由器在其直连网段上发现组播侦听者。组播侦听者（Multicast Listener）是那些希望接收组播数据的主机节点。

路由器通过 MLD 协议，可以了解自己的直连网段上是否有 IPv6 组播组的侦听者，并在数据库里做相应记录。同时，路由器还维护与这些 IPv6 组播地址相关的定时器信息。

MLD 路由器使用 IPv6 单播链路本地地址作为源地址发送 MLD 报文。MLD 使用 ICMPv6（Internet Control Message Protocol for IPv6，针对 IPv6 的互联网控制报文协议）报文类型。所有的 MLD 报文被限制在本地链路上，跳数为 1。

1.1.1 MLD 的版本

到目前为止，MLD 有两个版本：

- MLDv1（由 RFC 2710 定义），源自 IGMPv2
- MLDv2（由 RFC 3810 定义），源自 IGMPv3

所有版本的 MLD 协议都支持 ASM（Any-Source Multicast，任意信源组播）模型；MLDv2 可以直接应用于 SSM（Source-Specific Multicast，指定信源组播）模型，而 MLDv1 则需要在 MLD SSM Mapping 技术的支持下才能应用于 SSM 模型。


说明

有关 ASM 和 SSM 模型的介绍，请参见“IP 组播分册”中的“组播概述”。

1.1.2 MLDv1 原理简介

MLDv1 主要基于查询和响应机制完成对 IPv6 组播组成员的管理。

1. 查询器选举机制

当一个网段内有多台 IPv6 组播路由器时，由于它们都能从主机那里收到 MLD 成员关系报告报文（Multicast Listener Report Message），因此只需要其中一台路由器发送 MLD 查询报文（Query Message）就足够了。这就需要有一个查询器（Querier）的选举机制来确定由哪台路由器作为 MLD 查询器，其选举过程如下：

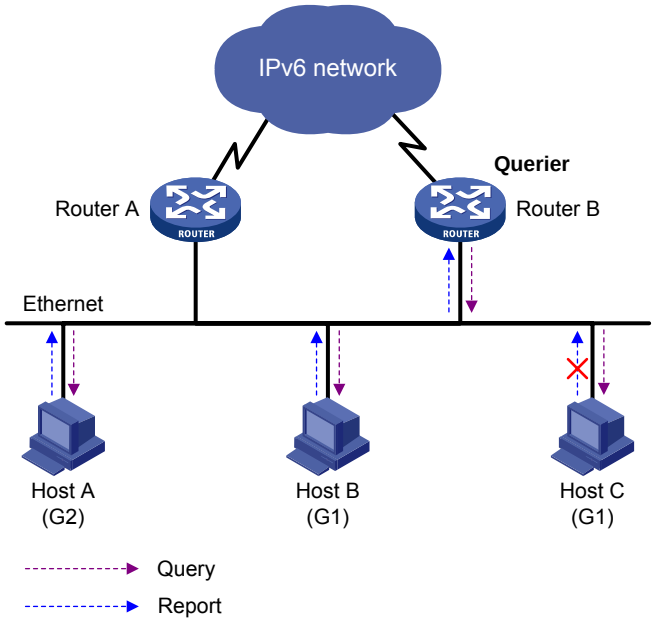
(1) 所有 MLD 路由器在初始时都认为自己是查询器，并向本地网段内的所有主机和路由器发送 MLD 普遍组查询（General Query）报文（目的地址为 FF02::1）；

(2) 本地网段中的其它 MLD 路由器在收到该报文后，将报文的源 IPv6 地址与自己的接口地址作比较。通过比较，IPv6 地址最小的路由器将成为查询器，其它路由器成为非查询器（Non-Querier）；

(3) 所有非查询器上都会启动一个定时器（即其它查询器存在时间定时器 Other Querier Present Timer）。在定时器超时前，如果收到了来自查询器的 MLD 查询报文，则重置该定时器；否则，就认为原查询器失效，并发起新的查询器选举过程。

2. 加入 IPv6 组播组机制

图1-1 MLD 查询响应示意图



如 图 1-1所示，假设Host B与Host C想要收到发往IPv6 组播组G1 的IPv6 组播数据，而Host A想要收到发往IPv6 组播组G2 的IPv6 组播数据，那么主机加入IPv6 组播组以及MLD查询器（Router B）维护IPv6 组播组成员关系的基本过程如下：

(1) 主机会主动向其要加入的 IPv6 组播组发送 MLD 成员关系报告报文以声明加入，而不必等待 MLD 查询器发来的 MLD 查询报文；

(2) MLD 查询器（Router B）周期性地以组播方式向本地网段内的所有主机和路由器发送普遍组查询报文（目的地址为 FF02::1）；

(3) 在收到该查询报文后，关注 G1 的 Host B 与 Host C 其中之一（这取决于谁的延迟定时器先超时）——譬如 Host B 会首先以组播方式向 G1 发送 MLD 成员关系报告报文，以宣告其属于 G1。由于本地网段中的所有主机都能收到 Host B 发往 G1 的报告报文，因此当 Host C 收到该报告报文后，将不再发送同样针对 G1 的报告报文，因为 MLD 路由器（Router A 和 Router B）已知道本地网段中有对 G1 感兴趣的主机了。这个机制称为主机上的 MLD 成员关系报告抑制机制，该机制有助于减少本地网段的信息流量；

(4) 与此同时，由于 Host A 关注的是 G2，所以它仍将以组播方式向 G2 发送报告报文，以宣告其属于 G2；

(5) 经过以上的查询和响应过程，MLD 路由器了解到本地网段中有 G1 和 G2 的成员，于是由 IPv6 组播路由协议（如 IPv6 PIM）生成（*，G1）和（*，G2）组播转发项作为 IPv6 组播数据的转发依据，其中的“*”代表任意 IPv6 组播源；

(6) 当由 IPv6 组播源发往 G1 或 G2 的 IPv6 组播数据经过组播路由到达 MLD 路由器时，由于 MLD 路由器上存在（*，G1）和（*，G2）组播转发项，于是将该 IPv6 组播数据转发到本地网段，接收者主机便能收到该 IPv6 组播数据了。

3. 离开 IPv6 组播组机制

当一个主机离开某 IPv6 组播组时：

(1) 该主机向本地网段内的所有 IPv6 组播路由器（目的地址为 FF02::2）发送离开组（Done）报文；

(2) 当查询器收到该报文后，向该主机所声明要离开的那个 IPv6 组播组发送特定组查询（Multicast-Address-Specific Query）报文（目的地址字段和组地址字段均填充为所要查询的 IPv6 组播组地址）；

(3) 如果该网段内还有该 IPv6 组播组的其它成员，则这些成员在收到特定组查询报文后，会在该报文中所设定的最大响应时间（Maximum Response Delay）内发送成员关系报告报文；

(4) 如果在最大响应时间内收到了该 IPv6 组播组其它成员发送的成员关系报告报文，查询器就会继续维护该 IPv6 组播组的成员关系；否则，查询器将认为该网段内已无该 IPv6 组播组的成员，于是不再维护这个 IPv6 组播组的成员关系。

1.1.3 MLDv2 原理简介

MLDv2 的原理与 MLDv1 基本相同，并新增了以下特性：

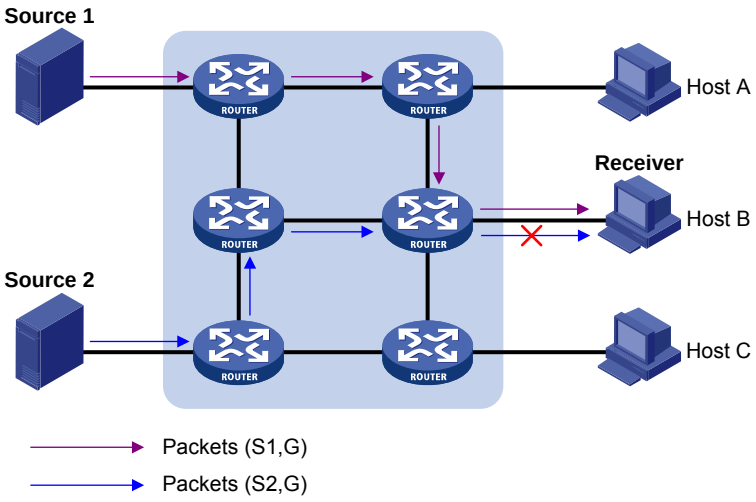
1. 对 IPv6 组播源的过滤

MLDv2 增加了针对 IPv6 组播源的过滤模式（INCLUDE/EXCLUDE），使主机在加入某 IPv6 组播组 G 的同时，能够明确要求接收或拒绝来自某特定 IPv6 组播源 S 的 IPv6 组播信息。当主机加入 IPv6 组播组时：

- 若要求只接收来自指定 IPv6 组播源如 S1、S2、……发来的 IPv6 组播信息，则其报告报文中可以标记为 INCLUDE Sources（S1，S2，……）；
- 若拒绝接收来自指定 IPv6 组播源如 S1、S2、……发来的 IPv6 组播信息，则其报告报文中可以标记为 EXCLUDE Sources（S1，S2，……）。

如 图 1-2所示，网络中存在Source 1（S1）和Source 2（S2）两个IPv6 组播源，均向IPv6 组播组G发送IPv6 组播报文。Host B仅对从Source 1 发往G的信息感兴趣，而对来自Source 2 的信息没有兴趣。

图1-2 指定源组的 IPv6 组播流路经



如果主机与路由器之间运行的是 MLDv1，Host B 加入 IPv6 组播组 G 时无法对 IPv6 组播源进行选择，因此无论 Host B 是否需要，来自 Source 1 和 Source 2 的 IPv6 组播信息都将传递给 Host B。

当主机与路由器之间运行了 MLDv2 之后，Host B 就可以要求只接收来自 Source 1、发往 G 的 IPv6 组播信息（S1，G），或要求拒绝来自 Source 2、发往 G 的 IPv6 组播信息（S2，G），这样就只有来自 Source 1 的 IPv6 组播信息才能传递给 Host B 了。

2. MLD 状态

运行 MLDv2 的组播路由器按每条直连链路上的组播地址（per multicast address per attached link）来保持 IPv6 组播组的状态。IPv6 组播组的状态包括：

- 过滤模式：保持对 INCLUDE 或 EXCLUDE 的状态跟踪。
- 源列表：保持对新增或删除 IPv6 组播源的跟踪。
- 定时器：表示 IPv6 组播地址超时后切换到 INCLUDE 模式的过滤定时器、关于源记录的源定时器等。

3. 接收者主机的状态侦听

运行 MLDv2 的组播路由器通过侦听接收者主机的状态，记录和维护网段上加入到源组的主机的信息。

1.1.4 MLD 报文类型

下面以 MLDv2 为例对 MLD 的报文类型进行介绍：

1. MLD 查询报文

MLD查询器通过发送MLD查询报文来了解相邻接口的组播侦听状态。MLD查询报文的格式如 图 1-3 所示，图中深蓝色部分为MLDv1 的报文格式，各字段的含义如 表 1-1所示。

图1-3 MLDv2 查询报文格式

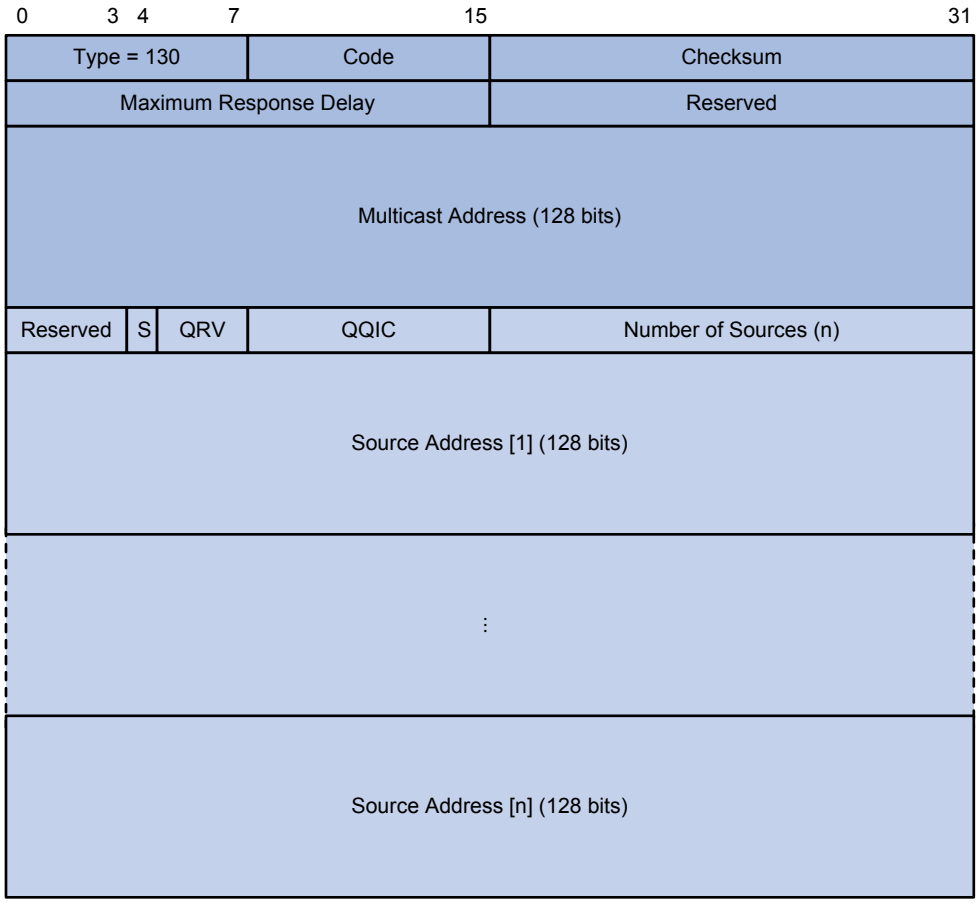


表1-1 MLDv2 查询报文各字段含义

字段	描述
Type = 130	报文类型，130 代表查询报文
Code	初始化为 0
Checksum	标准的 IPv6 校验和
Maximum Response Delay	主机发送报告报文前允许的最大响应时间
Reserved	保留字段，初始化为 0
Multicast Address	• 普遍组查询中，此字段设置为 0 • 特定组或特定源组查询中，此字段设置为待查询的 IPv6 组播组地址
S	标识位，表示路由器接收到查询报文后是否对定时器更新进行抑制
QRV	查询器的健壮性变量（Querier’s Robustness Variable）
QQIC	查询器发送普遍组查询报文的查询间隔（Querier’s Query Interval Code）
Number of Sources	• 普遍组查询或特定组查询中，此字段设置为 0 • 特定源组查询中，此字段表示查询报文中包含的源地址个数
Source Address(i)	特定源组查询中的 IPv6 组播源地址（i＝1, 2, …, n，其中 n 表示源地址的个数）

2. MLD 报告报文

主机通过发送MLD报告报文来汇报当前的组播侦听状态。MLD报告报文的格式如 图 1-4所示，各字段的含义如 表 1-2所示。

图1-4 MLDv2 报告报文格式

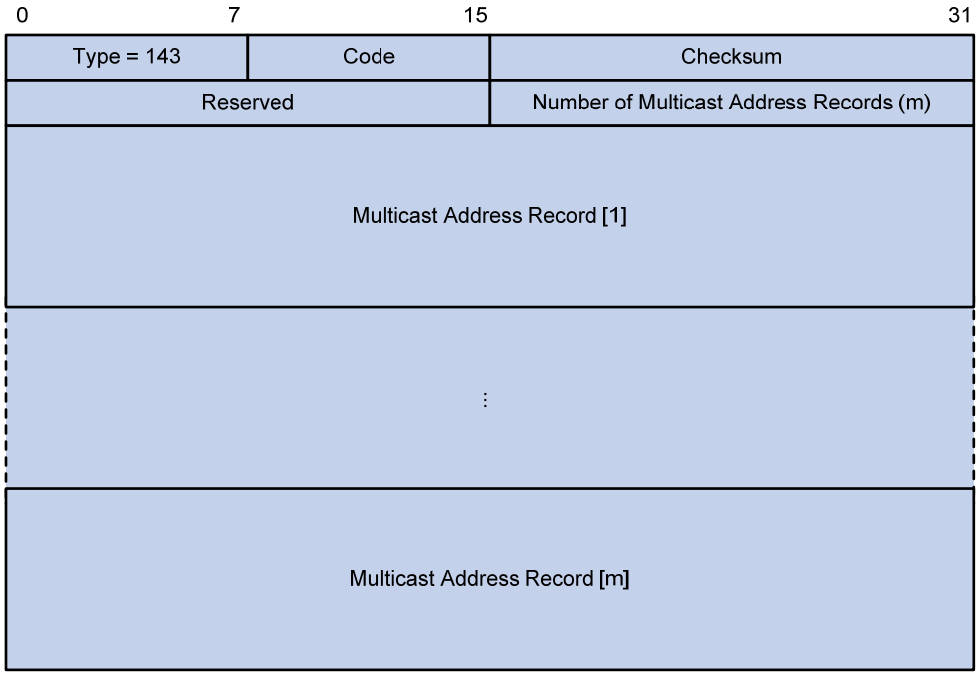


表1-2 MLDv2 报告报文各字段含义

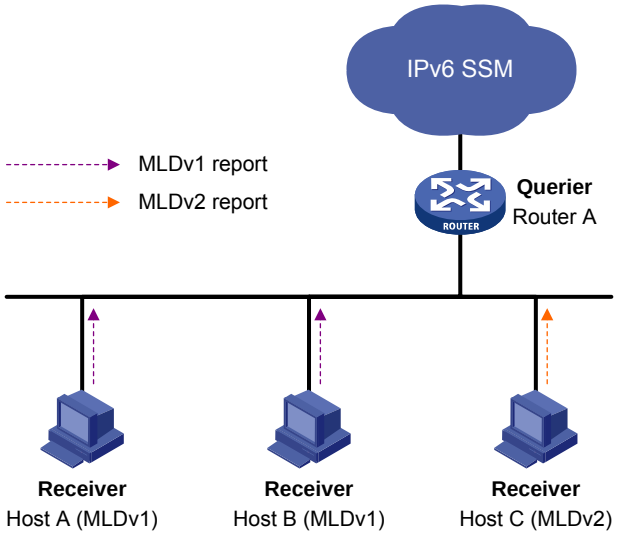
字段	描述
Type = 143	报文类型，143 代表报告报文
Reserved	保留字段，发送时设置为 0，接收时忽略此值
Checksum	标准的 IPv6 校验和
Number of Multicast Address Records	IPv6 组播地址记录的个数
Multicast Address Record(i)	组播地址记录，表示主机在接口上侦听到的每个 IPv6 组播地址信息，包括记录类型、IPv6 组播地址、IPv6 源地址等（i＝1, 2, …, m，其中 m 表示 IPv6 组播地址记录的个数）

1.1.5 MLD SSM Mapping

MLD SSM Mapping 通过在路由器上配置 SSM 静态映射规则，从而为运行 MLDv1 的接收者主机提供对 SSM 模型的支持。

SSM 模型要求在接收者主机所在的网段，路由器能够了解主机加入 IPv6 组播组时所指定的 IPv6 组播源。如果接收者主机上运行的是 MLDv2，则可以在 MLDv2 的报告报文中直接指定 IPv6 组播源的地址；如果某些接收者主机只能运行 MLDv1，则在 MLDv1 的报告报文中无法指定 IPv6 组播源的地址。这种情况下需要通过在路由器上配置 MLD SSM Mapping 功能，将 MLDv1 报告报文中所包含的（*，G）信息映射为（G，INCLUDE，（S1，S2...））信息。

图1-5 MLD SSM Mapping 组网图



在如 图 1-5所示的IPv6 SSM网络中，Host A、Host B和Host C上分别运行MLDv1 和MLDv2。在不允许将Host A和Host B升级为MLDv2 的情况下，若要为Host A和Host B也提供SSM组播服务，则需在Router A上配置MLD SSM Mapping功能。

配置完成后，当 Router A 收到来自主机的 MLDv1 报告报文时，首先检查该报文中所携带的 IPv6 组播组地址 G，然后根据检查结果的不同分别进行处理：

(1) 如果 G 不在 IPv6 SSM 组地址范围内，则提供 ASM 组播服务。

(2) 如果 G 在 IPv6 SSM 组地址范围内：

- 若 Router A 上没有 G 对应的 MLD SSM Mapping 规则，则无法提供 SSM 组播服务，丢弃该报文；
- 若 Router A 上有 G 对应的 MLD SSM Mapping 规则，则依据规则将报告报文中所包含的（*，G）信息映射为（G，INCLUDE，（S1，S2...））信息，可以提供 SSM 组播服务。

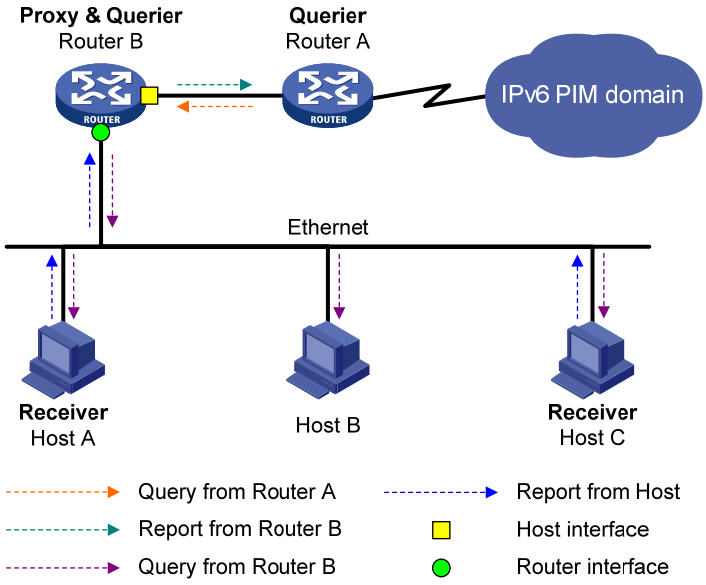
说明

- MLD SSM Mapping 不对 MLDv2 的报告报文进行处理。
- 有关 IPv6 SSM 组地址范围的介绍，请参见“IP 组播分册”中的“IPv6 PIM 配置”。

1.1.6 MLD Proxying

在一些简单的树型网络拓扑中，边缘设备上并不需要运行复杂的 IPv6 组播路由协议（如 IPv6 PIM），可以通过在这些设备上配置 MLD Proxying（MLD 代理）功能，使其代理下游主机来发送 MLD 报文及维护组成员关系，并基于该关系进行 IPv6 组播转发。在上游设备看来，配置了 MLD Proxying 功能的设备（称为 MLD 代理设备）不再是一个 IPv6 PIM 邻居，而只是一台主机。

图1-6 MLD Proxying 组网图



如 图 1-6所示，MLD Proxying模型中定义了以下两种接口类型：

- 上行接口：又称代理接口，指 MLD 代理设备上运行 MLD Proxying 功能的接口，即朝向组播分发树树根方向的接口。由于该接口执行 MLD 协议的主机行为，因此也称为主机接口（Host Interface）。
- 下行接口：指 MLD 代理设备上除上行接口外其它运行 MLD 协议的接口，即背向组播分发树树根方向的接口。由于该接口执行 MLD 协议的路由器行为，因此也称为路由器接口（Router Interface）。

MLD 代理设备上维护着一个组成员关系数据库（Membership Database），将所有下行接口维护的组成员关系记录都存到这个数据库中。组成员关系记录的结构如下：（Multicast-address，Filter-mode，Source-list），每条记录都是各下行接口上具有相同组地址的成员关系记录的合集。

上行接口正是依据这个数据库来执行主机行为——当收到查询报文时根据当前数据库状态响应报告报文，或者当数据库变化时主动发送报告或离开报文；而下行接口则执行路由器行为——参与查询器的选举、发送查询报文并根据报告报文维护组成员关系等。

1.1.7 协议规范

与 MLD 相关的协议规范有：

- RFC 2710：Multicast Listener Discovery (MLD) for IPv6
- RFC 3810：Multicast Listener Discovery Version 2 (MLDv2) for IPv6
- RFC 4605：Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")

1.2 MLD 配置任务简介

表1-3 MLD 配置任务简介

配置任务		说明	详细配置
配置 MLD 基本功能	使能 MLD	必选	1.3.2
	配置 MLD 版本	可选	1.3.3
	配置静态加入	可选	1.3.4
	配置 IPv6 组播组过滤器	可选	1.3.5
	配置 IPv6 组播组数量限制	可选	1.3.6
调整 MLD 性能	配置 MLD 报文选项	可选	1.4.2
	配置 MLD 查询和响应	可选	1.4.3
	配置 MLD 快速离开	可选	1.4.4
配置 MLD SSM Mapping	使能 MLD SSM Mapping	可选	1.5.2
	配置 MLD SSM Mapping 规则	可选	1.5.3
配置 MLD Proxying	使能 MLD Proxying	可选	1.6.2
	配置下行接口的 IPv6 组播转发能力	可选	1.6.3

说明

- MLD 视图下的配置具有全局性，接口视图下的配置仅对所在接口有效。
- 若没有在接口视图下进行配置，则该接口继承 MLD 视图下的全局配置；若两个视图都进行了配置，则优先采用该接口视图下所作的配置。

1.3 配置 MLD 基本功能

1.3.1 配置准备

在配置 MLD 基本功能之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 配置 IPv6 PIM-DM（或 IPv6 PIM-SM）

在配置 MLD 基本功能之前，需准备以下数据：

- MLD 的版本
- 以静态方式加入的 IPv6 组播组和 IPv6 组播源的地址
- IPv6 组播组过滤的 ACL 规则
- 允许接口加入的 IPv6 组播组最大数量

1.3.2 使能 MLD

在需要建立和维护 IPv6 组播组成员关系的接口上使能 MLD。

表1-4 使能 MLD

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 组播路由	**multicast ipv6 routing-enable**	必选 缺省情况下，IPv6 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 MLD	**mld enable**	必选 缺省情况下，没有使能 MLD

说明

有关 **multicast ipv6 routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“IPv6 组播路由与转发命令”。

1.3.3 配置 MLD 版本

由于不同版本 MLD 协议的报文结构与种类不同，因此需要为同一网段上的所有路由器配置相同版本的 MLD，否则 MLD 将不能正常运行。

1. 全局配置 MLD 版本

表1-5 全局配置 MLD 版本

操作	命令	说明
进入系统视图	**system-view**	-
进入 MLD 视图	**mld**	-
配置 MLD 的版本	**version** *version-number*	可选 缺省情况下，MLD 的版本为 MLDv1

2. 在接口上配置 MLD 版本

表1-6 在接口上配置 MLD 版本

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 MLD 的版本	**mld version** *version-number*	可选 缺省情况下，MLD 的版本为 MLDv1

1.3.4 配置静态加入

在配置了静态加入 IPv6 组播组或组播源组后，接口将作为该 IPv6 组播组的虚拟组成员存在，从而可以接收发往该组的 IPv6 组播数据，以测试 IPv6 组播数据的转发。

表1-7 配置静态加入

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置静态加入IPv6组播组或组播源组	**mld static-group** *ipv6-group-address* [**source** *ipv6-source-address*]	必选 缺省情况下，接口没有以静态方式加入任何IPv6组播组或组播源组

说明

- 在运行IPv6 PIM-SM的设备上配置静态加入时，如果待配接口上使能了IPv6 PIM-SM，则该接口必须为IPv6 PIM-SM的DR，否则该接口将不能加入IPv6组播组或组播源组；如果待配接口上未使能IPv6 PIM-SM但使能了MLD，则该接口必须为MLD查询器，否则该接口也不能加入IPv6组播组或组播源组。有关IPv6 PIM-SM和DR的介绍，请参见“IP组播分册”中的“IPv6 PIM配置”。
- 在配置了静态加入后，接口并不会对MLD查询器发出的查询报文进行响应；在配置或取消静态加入时，接口也不会主动发送MLD成员关系报告报文或MLD离开组报文。也就是说，该接口并没有真正成为该IPv6组播组的成员。

1.3.5 配置IPv6组播组过滤器

如果不希望接口所在网段上的主机加入某些IPv6组播组，可以在该接口上设置一个IPv6 ACL规则作为过滤器，对收到的报告报文进行过滤，以限制接口所服务的IPv6组播组范围。

表1-8 配置IPv6组播组过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置IPv6组播组过滤器	**mld group-policy** *acl6-number* [*version-number*]	必选 缺省情况下，没有配置IPv6组播组过滤器

1.3.6 配置IPv6组播组数量限制

通过配置允许接口加入的IPv6组播组最大数量，可以方便用户灵活控制接口所能加入的IPv6组播组个数。

表1-9 配置IPv6组播组数量限制

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置允许接口加入的IPv6组播组最大数量	**mld group-limit** *limit*	必选 缺省情况下，允许接口加入的IPv6组播组最大数量为1024

1.4 调整 MLD 性能

说明

对于本节中的配置任务来说:

- 在 MLD 视图下所作的配置对所有接口生效，在接口视图下所作的配置仅对当前接口生效;
- 如果在这两个视图下进行了相同功能或参数的配置，则不论配置先后，接口视图下的配置将被优先采用。

1.4.1 配置准备

在调整 MLD 性能之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 配置 MLD 基本功能

在调整 MLD 性能之前，需准备以下数据：

- MLD 查询器的启动查询间隔
- MLD 查询器的启动查询次数
- 发送 MLD 普遍组查询报文的时间间隔
- MLD 查询器的健壮系数
- MLD 普遍组查询的最大响应时间
- 发送 MLD 特定组查询报文的时间间隔
- MLD 其它查询器的存在时间

1.4.2 配置 MLD 报文选项

由于 MLD 特定组查询报文和特定源组查询报文的存在，而 IPv6 组播组是千变万化的，设备不可能加入所有的组，因此 MLD 需要借助 Router-Alert（路由器报警）选项来将送达本地、但没有加入的 IPv6 组播报文送往上层协议进行处理。有关 Router-Alert 选项的详细介绍，请参考 RFC 2113。

对于 IPv6 头中是否携带 Router-Alert 选项的 MLD 报文，设备会做出不同的处理：

- 出于兼容性考虑，缺省情况下设备不对 Router-Alert 选项进行检查，即处理所有收到的 MLD 报文。此时，MLD 报文中无论是否携带有 Router-Alert 选项，设备都会将其送给上层协议进行处理。
- 为了提高设备性能、减少不必要的开支，同时出于协议安全性的考虑，可以配置设备丢弃未携带 Router-Alert 选项的 MLD 报文，此时，当设备收到 MLD 报文时，会检查该报文的 Router-Alert 选项，如果没有携带该选项，就丢弃该报文。

1. 全局配置 MLD 报文选项

表1-10 全局配置 MLD 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入 MLD 视图	**mld**	-

操作	命令	说明
配置丢弃未携带 Router-Alert 选项的 MLD 报文	**require-router-alert**	可选 缺省情况下，设备不对 Router-Alert 选项进行检查
配置在发送的 MLD 报文中携带 Router-Alert 选项	**send-router-alert**	可选 缺省情况下，在发送的 MLD 报文中携带 Router-Alert 选项

2. 在接口上配置 MLD 报文选项

表1-11 在接口上配置 MLD 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置丢弃未携带 Router-Alert 选项的 MLD 报文	**mld require-router-alert**	可选 缺省情况下，设备不对 Router-Alert 选项进行检查
配置在发送的 MLD 报文中携带 Router-Alert 选项	**mld send-router-alert**	可选 缺省情况下，在发送的 MLD 报文中携带 Router-Alert 选项

1.4.3 配置 MLD 查询和响应

当 MLD 查询器启动时，会以“MLD 查询器启动查询间隔”为时间间隔发送“MLD 查询器启动查询次数”次 MLD 普遍组查询报文。

当 MLD 查询器启动后，会周期性地发送 MLD 普遍组查询报文，以判断网络上是否有 IPv6 组播组成员，发送间隔即为“发送 MLD 普遍组查询报文的时间间隔”。可以根据网络的实际情况来修改周期性发送 MLD 普遍组查询报文的时间间隔。

当 MLD 查询器收到 MLD 离开组报文后，也会发送“健壮系数”次 MLD 特定组查询报文，发送间隔即为“发送 MLD 特定组查询报文的时间间隔”。由此可见，MLD 查询器的健壮系数越大，MLD 查询器就越“健壮”，但是 IPv6 组播组超时所需的时间也就越长。

在收到 MLD 查询报文（包括普遍组查询和特定组查询）后，主机会为其所加入的每个 IPv6 组播组都启动一个延迟定时器，其值在 0 到最大响应时间（该时间值从 MLD 查询报文的最大响应时间字段获得）中随机选定，当定时器的值减为 0 时，主机就会向该定时器对应的 IPv6 组播组发送 MLD 成员关系报告报文。

合理配置 MLD 查询的最大响应时间，既可以使主机对 MLD 查询报文做出快速响应，又可以减少由于定时器同时超时，造成大量主机同时发送报告报文而引起的网络拥塞：

- 对于 MLD 普遍组查询报文来说，通过配置 MLD 普遍组查询的最大响应时间来填充其最大响应时间字段；
- 对于 MLD 特定组查询报文来说，所配置的发送 MLD 特定组查询报文的时间间隔将被填充到其最大响应时间字段。也就是说，MLD 特定组查询的最大响应时间从数值上与发送 MLD 特定组查询报文的时间间隔相同。

当同一网段上有多台组播路由器时，由 MLD 查询器负责发送 MLD 查询报文。如果非查询器在“其它查询器存在时间”超时前没有收到来自查询器的 MLD 查询报文，就会认为原有查询器失效，从而触发新的查询器选举过程；否则，非查询器将重置“其它查询器存在时间定时器”。

1. 全局配置 MLD 查询和响应

表1-12 全局配置 MLD 查询和响应

操作	命令	说明
进入系统视图	**system-view**	-
进入 MLD 视图	**mld**	-
配置 MLD 查询器的启动查询间隔	**startup-query-interval** *interval*	可选 关于MLD查询器启动查询间隔的缺省取值请参见 表 1-13后的说明
配置 MLD 查询器的启动查询次数	**startup-query-count** *value*	可选 关于MLD查询器启动查询次数的缺省取值请参见 表 1-13后的说明
配置发送 MLD 普遍组查询报文的时间间隔	**timer query** *interval*	可选 缺省情况下，发送 MLD 普遍组查询报文的时间间隔为 125 秒
配置 MLD 查询器的健壮系数	**robust-count** *robust-value*	可选 缺省情况下，MLD 查询器的健壮系数为 2
配置 MLD 普遍组查询的最大响应时间	**max-response-time** *interval*	可选 缺省情况下，MLD 普遍组查询的最大响应时间为 10 秒
配置发送 MLD 特定组查询报文的时间间隔	**last-listener-query-interval** *interval*	可选 缺省情况下，发送 MLD 特定组查询报文的时间间隔为 1 秒
配置 MLD 其它查询器的存在时间	**timer other-querier-present** *interval*	可选 关于MLD其它查询器存在时间的缺省取值请参见 表 1-13后的说明

2. 在接口上配置 MLD 查询和响应

表1-13 在接口上配置 MLD 查询和响应

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 MLD 查询器的启动查询间隔	**mld startup-query-interval** *interval*	可选 关于MLD查询器启动查询间隔的缺省取值请参见 表 1-13后的说明
配置 MLD 查询器的启动查询次数	**mld startup-query-count** *value*	可选 关于MLD查询器启动查询次数的缺省取值请参见 表 1-13后的说明

操作	命令	说明
配置发送 MLD 普遍组查询报文的时间间隔	**mld timer query** *interval*	可选 缺省情况下，发送 MLD 普遍组查询报文的时间间隔为 125 秒
配置 MLD 查询器的健壮系数	**mld robust-count** *robust-value*	可选 缺省情况下，MLD 查询器的健壮系数为 2
配置 MLD 普遍组查询的最大响应时间	**mld max-response-time** *interval*	可选 缺省情况下，MLD 普遍组查询的最大响应时间为 10 秒
配置发送 MLD 特定组查询报文的时间间隔	**mld last-listener-query-interval** *interval*	可选 缺省情况下，发送 MLD 特定组查询报文的时间间隔为 1 秒
配置 MLD 其它查询器的存在时间	**mld timer other-querier-present** *interval*	可选 关于MLD其它查询器存在时间的缺省取值请参见 表 1-13后的说明

说明

对于 MLD 查询器的启动查询间隔和次数，以及 MLD 其它查询器的存在时间这三个值，若进行了手工配置则取其各自的配置值，否则：

- 对于 MLD 查询器的启动查询间隔，其取值为发送 MLD 普遍组查询报文时间间隔的 1/4。缺省情况下，发送 MLD 普遍组查询报文的时间间隔为 125 秒，则此时 MLD 查询器的启动查询间隔 = 125 ÷ 4 = 31.25（秒）。
- 对于 MLD 查询器的启动查询次数，其取值等于 MLD 查询器的健壮系数。缺省情况下，MLD 查询器的健壮系数为 2，则此时 MLD 查询器的启动查询次数也为 2 次。
- 对于 MLD 其它查询器的存在时间，其取值 = 发送 MLD 普遍组查询报文的时间间隔 × MLD 查询器的健壮系数 + MLD 普遍组查询的最大响应时间 ÷ 2。缺省情况下，以上三个参数的取值分别为 125 秒、2 次和 10 秒，则此时 MLD 其它查询器的存在时间 = 125 × 2 + 10 ÷ 2 = 255（秒）。

注意

- 应确保 MLD 其它查询器的存在时间大于发送 MLD 普遍组查询报文的时间间隔，否则有可能导致网络内的 MLD 查询器反复变化。
- 应确保发送 MLD 普遍组查询报文的时间间隔大于 MLD 普遍组查询的最大响应时间，否则有可能造成对 IPv6 组播组成员的误删。

1.4.4 配置 MLD 快速离开

在某些应用（如 ADSL 拨号上网）中，MLD 查询器的一个端口唯一对应着一台接收者主机，当主机在多个 IPv6 组播组间频繁切换（如进行电视选台）时，为了快速响应主机的离开组报文，可以在 MLD 查询器上开启 MLD 快速离开功能。

在开启了 MLD 快速离开功能之后，当 MLD 查询器收到来自主机的离开组报文时，不再发送 MLD 特定组查询报文，而是直接向上游发送离开通告，这样一方面减小了响应延迟，另一方面也节省了网络带宽。

1. 全局配置 MLD 快速离开

表1-14 全局配置 MLD 快速离开

操作	命令	说明
进入系统视图	**system-view**	-
进入 MLD 视图	**mld**	-
配置 IPv6 组播组成员快速离开	**fast-leave** [**group-policy** *acl6-number*]	必选 缺省情况下，IPv6 组播组成员快速离开功能是关闭的

2. 在接口上配置 MLD 快速离开

表1-15 在接口上配置 MLD 快速离开

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置 IPv6 组播组成员快速离开	**mld fast-leave** [**group-policy** *acl6-number*]	必选 缺省情况下，IPv6 组播组成员快速离开功能是关闭的

1.5 配置 MLD SSM Mapping

在 IPv6 SSM 网络中，由于各种可能的限制，某些接收者主机只能运行 MLDv1。为了向这些仅支持 MLDv1 的接收者主机提供 SSM 服务，可以在路由器上配置 MLD SSM Mapping 功能。

1.5.1 配置准备

在配置 MLD SSM Mapping 之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 配置 MLD 基本功能

1.5.2 使能 MLD SSM Mapping

表1-16 使能 MLD SSM Mapping

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 MLD SSM Mapping 功能	**mld ssm-mapping enable**	必选 缺省情况下，接口上的 MLD SSM Mapping 功能处于关闭状态

说明

为保证本网段内运行任意版本 MLD 的接收者主机都能得到 SSM 服务，建议在处于该网段的接口上运行 MLDv2。

1.5.3 配置 MLD SSM Mapping 规则

通过多次配置可以实现同一 IPv6 组播组到多个 IPv6 组播源的映射。

表1-17 配置 MLD SSM Mapping 规则

操作	命令	说明
进入系统视图	**system-view**	-
进入 MLD 视图	**mld**	-
配置 MLD SSM Mapping 规则	**ssm-mapping** *ipv6-group-address prefix-length ipv6-source-address*	必选 缺省情况下，未配置 MLD SSM Mapping 规则

1.6 配置 MLD Proxying

1.6.1 配置准备

在配置 MLD Proxying 之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 使能 IPv6 组播路由

1.6.2 使能 MLD Proxying

在设备朝向组播分发树树根方向的接口上使能了 MLD 代理功能之后，该设备就成为了 MLD 代理设备。

表1-18 使能 MLD Proxying

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能 MLD 代理功能	**mld proxying enable**	必选 缺省情况下，接口上的 MLD 代理功能处于关闭状态

说明

- 一台设备上只允许有一个接口使能 MLD 代理功能。
- 在进行 MLD 的相关配置时，在已使能 MLD 代理功能的接口上不允许再使能 MLD，且只有 **mld require-router-alert**、**mld send-router-alert** 和 **mld version** 命令可在该接口上生效。
- 在已使能 MLD 代理功能的设备上不允许再使能其它 IPv6 组播路由协议（如不能在接口上使能 IPv6 PIM-DM 或 IPv6 PIM-SM，但在 IPv6 PIM 视图下配置的 **source-lifetime**、**source-policy** 和 **ssm-policy** 命令仍会生效），反之亦然。

1.6.3 配置下行接口的 IPv6 组播转发能力

通常，IPv6 组播数据只能被查询器转发，非查询器不具备 IPv6 组播转发能力，这样可以避免 IPv6 组播数据被重复转发。对于 MLD 代理设备也同样如此，只有当其下行接口是查询器时，IPv6 组播数据才能通过该接口转发给下游主机。

但在某些情况下，MLD 代理设备的下行接口不能在查询器竞选中获胜，此时应使能该下行接口在非查询器状态下的 IPv6 组播转发能力。

表1-19 配置下行接口的 IPv6 组播转发能力

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能下行接口在非查询器状态下的 IPv6 组播转发能力	**mld proxying forwarding**	必选 缺省情况下，当 MLD 代理设备的下行接口处于非查询器状态时，不转发 IPv6 组播数据

注意

在共享网段内存在多台 MLD 代理设备的情况下，当其中一台 MLD 代理设备的下行接口已竞选为查询器时，不应再在其它 MLD 代理设备的下行接口上进行此配置，否则该网段内将收到多份重复的 IPv6 组播数据流。

1.7 MLD 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 MLD 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 MLD 的统计信息。

表1-20 MLD 显示和维护

操作	命令
查看 MLD 组播组的信息	**display mld group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]
查看接口上 MLD 配置和运行信息	**display mld interface** [*interface-type interface-number*] [**verbose**]
查看 MLD 代理组的信息	**display mld proxying group** [*group-address*] [**verbose**]

操作	命令
查看 MLD 路由表的信息	**display mld routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **flags** { **act** | **suc** }] *
查看 MLD SSM Mapping 规则	**display mld ssm-mapping** *ipv6-group-address*
查看依据 MLD SSM Mapping 规则创建的 IPv6 组播组信息	**display mld ssm-mapping group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**verbose**]
清除 MLD 组播组的信息	**reset mld group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }
清除依据 MLD SSM Mapping 规则创建的 IPv6 组播组信息	**reset mld ssm-mapping group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }

说明

通过 **reset mld group** 命令不能够清除以静态方式加入的 MLD 组播组的信息。

注意

执行 **reset mld group** 命令可能导致接收者中断组播信息的接收。

1.8 MLD 典型配置举例

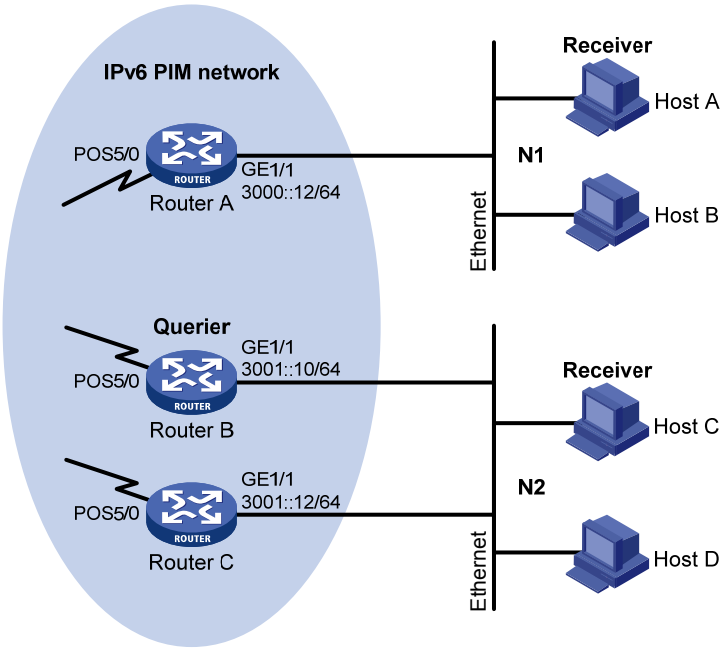
1.8.1 MLD 基本功能配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者组成末梢网络 N1 和 N2，Host A 与 Host C 分别为 N1 和 N2 中的组播信息接收者。
- IPv6 PIM 网络中的 Router A 连接 N1，Router B 与 Router C 共同连接 N2。
- Router A 通过 GigabitEthernet1/1 连接 N1，通过 POS5/0 连接 IPv6 PIM 网络中的其它设备。
- Router B 与 Router C 分别通过各自的 GigabitEthernet1/1 连接 N2，并分别通过各自的 POS5/0 连接 IPv6 PIM 网络中的其它设备。
- Router A 与 N1 之间运行 MLDv1；Router B、Router C 与 N2 之间也分别运行 MLDv1，且由于 Router B 的接口 IPv6 地址较小，因此在 N2 中通常由其来充当 MLD 查询器。

2. 组网图

图1-7 MLD 基本功能配置组网图



3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址和 IPv6 单播路由协议

使能各路由器的IPv6 转发功能，并按照 图 1-7配置各接口的IPv6 地址和前缀长度，具体配置过程略。

配置 IPv6 PIM 网络内的各路由器之间采用 OSPFv3 协议进行互连，确保 IPv6 PIM 网络内部在网络层互通，并且各路由器之间能够借助单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-DM 和 MLD

在 Router A 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-DM，并在主机侧接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mld enable
[RouterA-GigabitEthernet1/1] pim ipv6 dm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface pos 5/0
[RouterA-Pos5/0] pim ipv6 dm
[RouterA-Pos5/0] quit
```

在 Router B 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-DM，并在主机侧接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterB> system-view
[RouterB] multicast ipv6 routing-enable
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] mld enable
[RouterB-GigabitEthernet1/1] pim ipv6 dm
```

```
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface pos 5/0
[RouterB-Pos5/0] pim ipv6 dm
[RouterB-Pos5/0] quit
```

在 Router C 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-DM，并在主机侧接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterC> system-view
[RouterC] multicast ipv6 routing-enable
[RouterC] interface GigabitEthernet 1/1
[RouterC-GigabitEthernet1/1] mld enable
[RouterC-GigabitEthernet1/1] pim ipv6 dm
[RouterC-GigabitEthernet1/1] quit
[RouterC] interface pos 5/0
[RouterC-Pos5/0] pim ipv6 dm
[RouterC-Pos5/0] quit
```

(3) 检查配置效果

通过使用 **display mld interface** 命令可以查看各路由器接口上 MLD 的配置和运行情况。例如：

查看 Router B 在 GigabitEthernet1/1 上的 MLD 信息。

```
[RouterB] display mld interface GigabitEthernet 1/1
 GigaibitEthernet1/1(FE80::200:5EFF:FE66:5100):
   MLD is enabled
   Current MLD version is 1
   Value of query interval for MLD(in seconds): 125
   Value of other querier present interval for MLD(in seconds): 255
   Value of maximum query response time for MLD(in seconds): 10
   Querier for MLD: FE80::200:5EFF:FE66:5100 (this router)
  Total 1 MLD Group reported
```

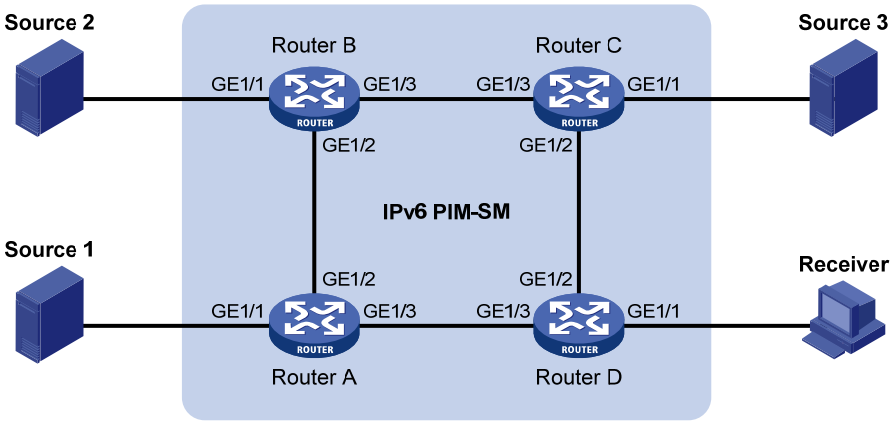
1.8.2 MLD SSM Mapping 功能配置举例

1. 组网需求

- IPv6 PIM-SM 网络中同时采用 ASM 和 SSM 方式提供 IPv6 组播服务，将 Router D 的接口 GigabitEthernet1/3 配置为 C-BSR 和 C-RP，IPv6 SSM 组播组的范围为 FF3E::/64。
- Router D 的接口 GigabitEthernet1/1 上运行 MLDv2，接收者主机上运行 MLDv1，且不能升级至 MLDv2，因此该主机在加入 IPv6 组播组时无法指定 IPv6 组播源。
- Source 1、Source 2 和 Source 3 都向 IPv6 SSM 组范围内的组播组发送 IPv6 组播数据，要求通过在 Router D 上配置 MLD SSM Mapping 功能，使接收者主机只能接收来自 Source 1 和 Source 3 的 IPv6 组播数据。

2. 组网图

图1-8 MLD SSM Mapping 功能配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source 1	-	1001::1/64	Source 3	-	3001::1/64
Source 2	-	2001::1/64	Receiver	-	4001::1/64
Router A	GE1/1	1001::2/64	Router C	GE1/1	3001::2/64
	GE1/2	1002::1/64		GE1/2	3002::1/64
	GE1/3	1003::1/64		GE1/3	2002::2/64
Router B	GE1/1	2001::2/64	Router D	GE1/1	4001::2/64
	GE1/2	1002::2/64		GE1/2	3002::2/64
	GE1/3	2002::1/64		GE1/3	1003::2/64

3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址和 IPv6 单播路由协议

使能各路由器的IPv6 转发功能，并按照 图 1-8配置各接口的IPv6 地址和前缀长度，具体配置过程略。

配置 IPv6 PIM-SM 域内的各路由器之间采用 OSPFv3 协议进行互连，确保 IPv6 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助 IPv6 单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-SM、MLD 和 MLD SSM Mapping 功能

在 Router D 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-SM，并在主机侧接口 GigabitEthernet1/1 上使能 MLD 和 MLD SSM Mapping 功能，配置 MLD 版本为 2。

```
<RouterD> system-view
[RouterD] multicast ipv6 routing-enable
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigabitEthernet1/1] mld enable
[RouterD-GigabitEthernet1/1] mld version 2
[RouterD-GigabitEthernet1/1] mld ssm-mapping enable
[RouterD-GigabitEthernet1/1] pim ipv6 sm
[RouterD-GigabitEthernet1/1] quit
[RouterD] interface GigabitEthernet 1/2
[RouterD-GigabitEthernet1/2] pim ipv6 sm
[RouterD-GigabitEthernet1/2] quit
[RouterD] interface gigabitethernet 1/3
[RouterD-GigabitEthernet1/3] pim ipv6 sm
[RouterD-GigabitEthernet1/3] quit
```

在 Router A 上使能 IPv6 组播路由，并在各接口上使能 IPv6 PIM-SM。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] pim ipv6 sm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface GigabitEthernet 1/2
[RouterA-GigabitEthernet1/2] pim ipv6 sm
[RouterA-GigabitEthernet1/2] quit
[RouterA] interface GigabitEthernet 1/3
[RouterA-GigabitEthernet1/3] pim ipv6 sm
[RouterA-GigabitEthernet1/3] quit
```

Router B 和 Router C 的配置与 Router A 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP

在 Router D 上配置 C-BSR 和 C-RP 的位置。

```
[RouterD] pim ipv6
[RouterD-pim6] c-bsr 1003::2
[RouterD-pim6] c-rp 1003::2
[RouterD-pim6] quit
```

(4) 配置 IPv6 SSM 组播组的地址范围

在 Router D 上配置 IPv6 SSM 组播组的地址范围为 FF3E::/64。

```
[RouterD] acl ipv6 number 2000
[RouterD-acl6-basic-2000] rule permit source ff3e:: 64
[RouterD-acl6-basic-2000] quit
[RouterD] pim ipv6
[RouterD-pim6] ssm-policy 2000
[RouterD-pim6] quit
```

Router A、Router B 和 Router C 的配置与 Router D 相似，配置过程略。

(5) 配置 MLD SSM Mapping 规则

在 Router D 上配置 MLD SSM Mapping 规则。

```
[RouterD] mld
[RouterD-mld] ssm-mapping ff3e:: 64 1001::1
[RouterD-mld] ssm-mapping ff3e:: 64 3001::1
[RouterD-mld] quit
```

(6) 检验配置效果

通过使用 **display mld ssm-mapping** 命令可以查看路由器上的 MLD SSM Mapping 规则。例如：

查看 Router D 上 IPv6 组播组 FF3E::101 的 MLD SSM Mapping 规则。

```
[RouterD] display mld ssm-mapping ff3e::101
Group: FF3E::101
Source list:
       1001::1
       3001::1
```

通过使用 **display mld ssm-mapping group** 命令可以查看路由器上依据 MLD SSM Mapping 规则创建的 IPv6 组播组信息。例如：

查看 Router D 上依据 MLD SSM Mapping 规则创建的 IPv6 组播组信息。

```
[RouterD] display mld ssm-mapping group
Total 1 MLD SSM-mapping Group(s).
Interface group report information
 GigabitEthernet1/1(4001::2):
  Total 1 MLD SSM-mapping Group reported
   Group Address: FF3E::101
    Last Reporter: 4001::1
    Uptime: 00:02:04
    Expires: off
```

通过使用 **display pim ipv6 routing-table** 命令可以查看路由器的 IPv6 PIM 路由表信息。例如：

查看 Router D 上 IPv6 PIM 路由表信息。

```
[RouterD] display pim ipv6 routing-table
Total 0 (*, G) entry; 2 (S, G) entry

 (1001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: GigabitEthernet1/3
         Upstream neighbor: 1003::1
         RPF prime neighbor: 1003::1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: GigabitEthernet1/1
                 Protocol: mld, UpTime: 00:13:25, Expires: -

 (3001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: GigabitEthernet1/2
         Upstream neighbor: 3002::1
         RPF prime neighbor: 3002::1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: GigabitEthernet1/1
                 Protocol: mld, UpTime: 00:13:25, Expires: -
```

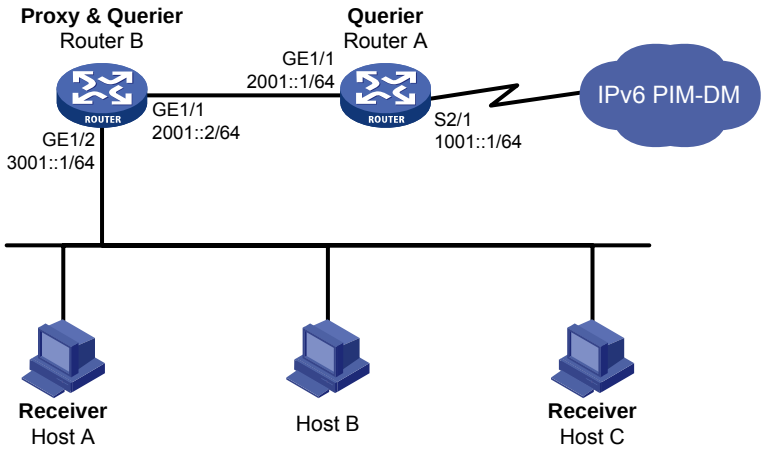
1.8.3 MLD Proxying 功能配置举例

1. 组网需求

- 核心网络中运行 IPv6 PIM-DM，末梢网络中的接收者 Host A 和 Host C 通过 IPv6 组播组 FF3E::101 点播视频节目。
- 要求通过在 Router B 上配置 MLD Proxying 功能，使其在不运行 IPv6 PIM-DM 的情况下实现组成员关系的维护和 IPv6 组播数据的正常转发。

2. 组网图

图1-9 MLD Proxying 功能配置组网图



3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址

使能各路由器的IPv6 转发功能，并按照 图 1-9配置各接口的IPv6 地址和前缀长度，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-DM、MLD 和 MLD Proxying

在 Router A 上使能 IPv6 组播路由，在接口 Serial2/1 上使能 IPv6 PIM-DM，并在接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface serial 2/1
[RouterA-Serial2/1] pim ipv6 dm
[RouterA-Serial2/1] quit
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mld enable
[RouterA-GigabitEthernet1/1] pim ipv6 dm
[RouterA-GigabitEthernet1/1] quit
```

在 Router B 上使能 IPv6 组播路由，在接口 GigabitEthernet1/1 上使能 MLD Proxying，并在接口 GigabitEthernet1/2 上使能 MLD。

```
<RouterB> system-view
[RouterB] multicast ipv6 routing-enable
[RouterB] interface GigabitEthernet 1/1
[RouterB-GigabitEthernet1/1] mld proxying enable
[RouterB-GigabitEthernet1/1] quit
[RouterB] interface GigabitEthernet 1/2
[RouterB-GigabitEthernet1/2] mld enable
[RouterB-GigabitEthernet1/2] quit
```

(3) 检验配置效果

通过使用 **display mld interface** 命令可以查看各路由器接口上 MLD 的配置和运行情况。例如：

查看 Router B 在 GigabitEthernet1/1 上 MLD 配置和运行的详细信息。

```
[RouterB] display mld interface GigabitEthernet 1/1 verbose
 GigabitEthernet1/1(2001::2):
   MLD proxy is enabled
   Current MLD version is 1
   Multicast routing on this interface: enabled
   Require-router-alert: disabled
```

通过使用 **display mld group** 命令可以查看 MLD 组播组的信息。例如：

查看 Router A 上的 MLD 组播组信息。

```
[RouterA] display mld group
Total 1 MLD Group(s).
Interface group report information
 GigabitEthernet1/1(2001::1):
  Total 1 MLD Groups reported
   Group Address    Last Reporter    Uptime      Expires
   ff3e::101        2001::2          00:02:04    00:01:15
```

由此可见，主机的 MLD 报告通过 Router B 的代理接口 GigabitEthernet1/1 发给 Router A。

1.9 常见配置错误举例

1.9.1 接收者侧路由器上无组成员信息

1. 故障现象

当某主机发送了加入 IPv6 组播组 G 的报文后，离该主机最近的路由器上却没有 IPv6 组播组 G 的组成员信息。

2. 分析

- 组网、接口连线的正确与否以及接口的协议层是否 up 将直接影响 IPv6 组播组成员信息的生成；
- 在路由器上必须使能 IPv6 组播路由，在连接主机的接口上必须使能 MLD；
- 如果路由器接口上运行的 MLD 版本比主机的低，那么路由器将无法识别主机发来的较高版本的 MLD 报告报文；
- 如果在接口上使用命令 **mld group-policy** 对加入 IPv6 组播组 G 进行了限制后，该接口将不再接收未通过过滤的那些要求加入 IPv6 组播组 G 的报文。

3. 处理过程

(1) 检查组网是否正确，接口间的连线是否正确，以及接口状态是否正常，是否配置了正确的 IPv6 地址。通过命令 **display mld interface** 查看接口信息。若无接口信息输出，说明接口状态异常，原因通常是接口上配置了 **shutdown** 命令，或者接口连线不正确，或者接口上没有配置正确的 IPv6 地址。

(2) 检查是否使能了 IPv6 组播路由。通过命令 **display current-configuration** 查看是否配置了命令 **multicast ipv6 routing-enable**。若缺少该配置，则需要在系统视图下执行命令 **multicast ipv6 routing-enable** 使能 IPv6 组播路由，同时也需要在相应接口上使能 MLD。

(3) 检查接口上运行的 MLD 版本。通过命令 **display mld interface** 来检查接口上运行的 MLD 版本是否低于主机所使用的版本。

(4) 检查接口上是否配置了 IPv6 ACL 规则来限制主机加入 IPv6 组播组 G。通过命令 **display current-configuration interface** 观察是否配置了 **mld group-policy** 命令。如果配置的 IPv6 ACL 规则对加入 IPv6 组播组 G 进行了限制，则需要修改该 IPv6 ACL 规则，允许接受 IPv6 组播组 G 的报告报文。

1.9.2 同一网段各路由器上组成员关系不一致

1. 故障现象

在同一网段的不同 MLD 路由器上，各自维护的组成员关系不一致。

2. 分析

- 运行 MLD 的路由器为每个接口维护多个参数，各参数之间相互影响，非常复杂。如果同一网段路由器的 MLD 接口参数配置不一致，必然导致组成员关系的混乱。
- 另外，MLD 目前有 2 个版本，版本不同的 MLD 路由器与主机之间虽然可以兼容，但是连接在同一网段的所有路由器必须运行相同版本的 MLD。如果同一网段路由器的 MLD 版本不一致，也将导致 MLD 组成员关系的混乱。

3. 处理过程

(1) 检查 MLD 配置。通过命令 **display current-configuration** 观察接口上 MLD 的配置信息。

(2) 在同一网段的所有路由器上执行命令 **display mld interface** 来检查 MLD 相关定时器的参数，确保配置一致。

(3) 通过命令 **display mld interface** 来检查各路由器上运行的 MLD 版本是否一致。

目 录

1 IPv6 PIM配置 1-1
1.1 IPv6 PIM简介 1-1
1.1.1 IPv6 PIM-DM简介 1-1
1.1.2 IPv6 PIM-DM工作机制 1-1
1.1.3 IPv6 PIM-SM简介 1-4
1.1.4 IPv6 PIM-SM工作机制 1-5
1.1.5 SSM模型在IPv6 PIM中的实现 1-10
1.1.6 协议规范 1-11
1.2 配置IPv6 PIM-DM 1-12
1.2.1 IPv6 PIM-DM配置任务简介 1-12
1.2.2 配置准备 1-12
1.2.3 使能IPv6 PIM-DM 1-12
1.2.4 使能状态刷新能力 1-13
1.2.5 配置状态刷新参数 1-13
1.2.6 配置IPv6 PIM-DM定时器 1-14
1.3 配置IPv6 PIM-SM 1-14
1.3.1 IPv6 PIM-SM配置任务简介 1-14
1.3.2 配置准备 1-15
1.3.3 使能IPv6 PIM-SM 1-15
1.3.4 配置RP 1-16
1.3.5 配置BSR 1-18
1.3.6 配置IPv6 组播源注册 1-21
1.3.7 配置SPT切换 1-22
1.4 配置IPv6 PIM-SSM 1-22
1.4.1 IPv6 PIM-SSM配置任务简介 1-22
1.4.2 配置准备 1-22
1.4.3 使能IPv6 PIM-SM 1-23
1.4.4 配置IPv6 SSM组播组范围 1-23
1.5 配置IPv6 PIM公共特性 1-24
1.5.1 IPv6 PIM公共特性配置任务简介 1-24
1.5.2 配置准备 1-24
1.5.3 配置IPv6 组播数据过滤器 1-25
1.5.4 配置Hello报文过滤器 1-25
1.5.5 配置Hello报文选项 1-26
1.5.6 配置IPv6 PIM公共定时器 1-27
1.5.7 配置加入/剪枝报文规格 1-29
1.6 IPv6 PIM显示和维护 1-29
1.7 IPv6 PIM典型配置举例 1-30
1.7.1 IPv6 PIM-DM典型配置举例 1-30
1.7.2 IPv6 PIM-SM典型配置举例 1-34
1.7.3 IPv6 PIM-SSM典型配置举例 1-39

1 IPv6 PIM 配置

1.1 IPv6 PIM 简介

IPv6 PIM 是 Protocol Independent Multicast for IPv6（IPv6 协议无关组播）的简称，表示可以利用静态路由或者任意 IPv6 单播路由协议（包括 RIPng、OSPFv3、IS-ISv6、BGP4+等）所生成的 IPv6 单播路由表为 IPv6 组播提供路由。IPv6 组播路由与所采用的 IPv6 单播路由协议无关，只要能够通过 IPv6 单播路由协议产生相应的 IPv6 组播路由表项即可。IPv6 PIM 借助 RPF（Reverse Path Forwarding，逆向路径转发）机制实现对 IPv6 组播报文的转发。当 IPv6 组播报文到达本地设备时，首先对其进行 RPF 检查：若 RPF 检查通过，则创建相应的 IPv6 组播路由表项，从而进行 IPv6 组播报文的转发；若 RPF 检查失败，则丢弃该报文。有关 RPF 的详细介绍，请参见“IP 组播分册”中的“IPv6 组播路由与转发配置”。

根据实现机制的不同，IPv6 PIM 分为以下两种模式：

- IPv6 PIM-DM（Protocol Independent Multicast-Dense Mode for IPv6，IPv6 协议无关组播—密集模式）
- IPv6 PIM-SM（Protocol Independent Multicast-Sparse Mode for IPv6，IPv6 协议无关组播—稀疏模式）


说明

为了描述的方便，本文中把由支持 IPv6 PIM 协议的组播路由器所组成的网络简称为“IPv6 PIM 域”。

1.1.1 IPv6 PIM-DM 简介

IPv6 PIM-DM 属于密集模式的 IPv6 组播路由协议，使用“推（Push）模式”传送 IPv6 组播数据，通常适用于 IPv6 组播组成员相对比较密集的小型网络。

IPv6 PIM-DM 的基本原理如下：

- IPv6 PIM-DM 假设网络中的每个子网都存在至少一个 IPv6 组播组成员，因此 IPv6 组播数据将被扩散（Flooding）到网络中的所有节点。然后，IPv6 PIM-DM 对没有 IPv6 组播数据转发的分支进行剪枝（Prune），只保留包含接收者的分支。这种“扩散—剪枝”现象周期性地发生，被剪枝的分支也可以周期性地恢复成转发状态。
- 当被剪枝分支的节点上出现了 IPv6 组播组的成员时，为了减少该节点恢复成转发状态所需的时间，IPv6 PIM-DM 使用嫁接（Graft）机制主动恢复其对 IPv6 组播数据的转发。

一般说来，密集模式下数据包的转发路径是有源树（Source Tree，即以 IPv6 组播源为“根”、IPv6 组播组成员为“枝叶”的一棵转发树）。由于有源树使用的是从 IPv6 组播源到接收者的最短路径，因此也称为最短路径树（Shortest Path Tree，SPT）。

1.1.2 IPv6 PIM-DM 工作机制

IPv6 PIM-DM 的工作机制可以概括如下：

- 邻居发现

- 构建 SPT
- 嫁接
- 断言

1. 邻居发现

在 IPv6 PIM 域中，路由器通过周期性地的向所有 IPv6 PIM 路由器以组播方式发送 IPv6 PIM Hello 报文（以下简称 Hello 报文），以发现 IPv6 PIM 邻居，维护各路由器之间的 IPv6 PIM 邻居关系，从而构建和维护 SPT。



说明

路由器每个运行了 IPv6 PIM 协议的接口都会周期性地发送 Hello 报文，从而了解与该接口相关的 IPv6 PIM 邻居信息。

---

2. 构建 SPT

构建 SPT 的过程也就是“扩散—剪枝”的过程：

(1) 在 IPv6 PIM-DM 域中，IPv6 组播源 S 向 IPv6 组播组 G 发送 IPv6 组播报文时，首先对 IPv6 组播报文进行扩散：路由器对该报文的 RPF 检查通过后，便创建一个（S，G）表项，并将该报文向网络中的所有下游节点转发。经过扩散，IPv6 PIM-DM 域内的每个路由器上都会创建（S，G）表项。

(2) 然后对那些下游没有接收者的节点进行剪枝：由没有接收者的下游节点向上游节点发剪枝报文（Prune Message），以通知上游节点将相应的接口从其组播转发表项（S，G）所对应的出接口列表中删除，并不再转发该 IPv6 组播组的报文至该节点。

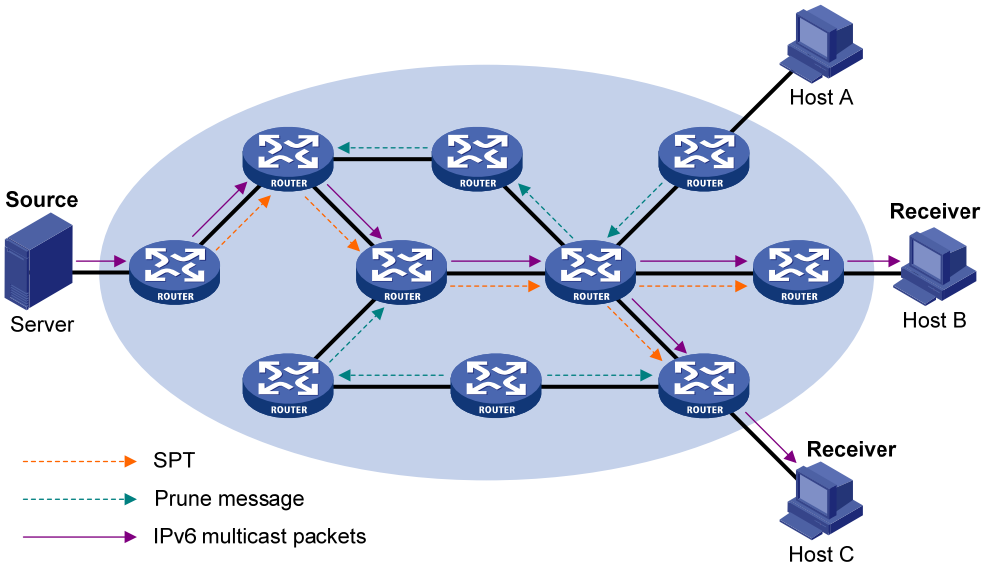


说明

- （S，G）表项包括 IPv6 组播源的地址 S、IPv6 组播组的地址 G、出接口列表和入接口等。
- 路由器上收到 IPv6 组播数据的接口称为“上游”，转发 IPv6 组播数据的接口称为“下游”。

剪枝过程最先由叶子路由器发起，如 图 1-1所示，没有接收者（Receiver）的路由器（如与Host A直连的路由器）主动发起剪枝，并一直持续到IPv6 PIM-DM域中只剩下必要的分支，这些分支共同构成了SPT。

图1-1 IPv6 PIM-DM 中构建 SPT 示意图



“扩散—剪枝”的过程是周期性发生的。各个被剪枝的节点提供超时机制，当剪枝超时后便重新开始这一过程。

说明

剪枝在 IPv6 PIM-SM 中有着相似的应用。

---

3. 嫁接

当被剪枝的节点上出现了 IPv6 组播组的成员时，为了减少该节点恢复成转发状态所需的时间，IPv6 PIM-DM 使用嫁接机制主动恢复其对 IPv6 组播数据的转发，过程如下：

(1) 需要恢复接收 IPv6 组播数据的节点向其上游节点发送嫁接报文（Graft Message）以申请重新加入到 SPT 中；

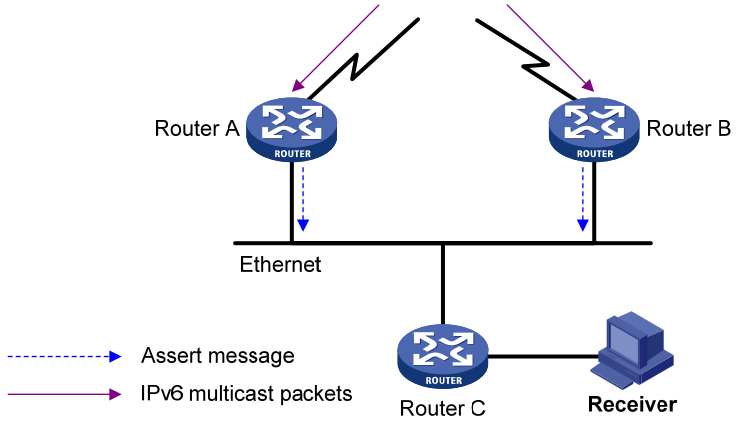
(2) 当上游节点收到该报文后恢复该下游节点的转发状态，并向其回应一个嫁接应答报文（Graft-Ack Message）以进行确认；

(3) 如果发送嫁接报文的下游节点没有收到来自其上游节点的嫁接应答报文，将重新发送嫁接报文直到被确认为止。

4. 断言

在一个网段内如果存在多台组播路由器，则相同的 IPv6 组播报文可能会被重复发送到该网段。为了避免出现这种情况，就需要通过断言（Assert）机制来选定唯一的 IPv6 组播数据转发者。

图1-2 Assert 机制示意图



如 图 1-2所示，当Router A和Router B从上游节点收到（S，G）的IPv6 组播报文后，都会向本地网段转发该报文，于是处于下游的节点Router C就会收到两份相同的IPv6 组播报文，Router A和Router B也会从各自的本地接口收到对方转发来的该IPv6 组播报文。此时，Router A和Router B会通过本地接口向所有IPv6 PIM路由器以组播方式发送断言报文（Assert Message），该报文中携带有以下信息：IPv6 组播源地址S、IPv6 组播地址G、到IPv6 组播源的IPv6 单播路由的优先级和度量值。通过一定的规则对这些参数进行比较后，Router A和Router B中的获胜者将成为（S，G）IPv6 组播报文在本网段的转发者，比较规则如下：

(1) 到 IPv6 组播源的 IPv6 单播路由的优先级较高者获胜；

(2) 如果到 IPv6 组播源的 IPv6 单播路由的优先级相等，那么到 IPv6 组播源的度量值较小者获胜；

(3) 如果到 IPv6 组播源的度量值也相等，则本地接口 IPv6 链路本地地址较大者获胜。

1.1.3 IPv6 PIM-SM 简介

IPv6 PIM-DM 使用以“扩散—剪枝”方式构建的 SPT 来传送 IPv6 组播数据。尽管 SPT 的路径最短，但是其建立的过程效率较低，并不适合大中型网络。

IPv6 PIM-SM 属于稀疏模式的 IPv6 组播路由协议，使用“拉（Pull）模式”传送 IPv6 组播数据，通常适用于 IPv6 组播组成员分布相对分散、范围较广的大中型网络。

IPv6 PIM-SM 的基本原理如下：

- IPv6 PIM-SM 假设所有主机都不需要接收 IPv6 组播数据，只向明确提出需要 IPv6 组播数据的主机转发。IPv6 PIM-SM 实现组播转发的核心任务就是构造并维护 RPT（Rendezvous Point Tree，共享树或汇集树），RPT 选择 IPv6 PIM 域中某台路由器作为公用的根节点 RP（Rendezvous Point，汇集点），IPv6 组播数据通过 RP 沿着 RPT 转发给接收者；
- 连接接收者的路由器向某 IPv6 组播组对应的 RP 发送加入报文，该报文被逐跳送达 RP，所经过的路径就形成了 RPT 的分支；
- IPv6 组播源如果要向某 IPv6 组播组发送 IPv6 组播数据，首先由 IPv6 组播源侧 DR（Designated Router，指定路由器）负责向 RP 进行注册，把注册报文（Register Message）通过单播方式发送给 RP，该报文到达 RP 后触发建立 SPT。之后 IPv6 组播源把 IPv6 组播数据沿着 SPT 发向 RP，当 IPv6 组播数据到达 RP 后，被复制并沿着 RPT 发送给接收者。

说明

复制仅发生在分发树的分支处，这个过程能够自动重复直到数据包最终到达接收者。

1.1.4 IPv6 PIM-SM 工作机制

IPv6 PIM-SM 的工作机制可以概括如下：

- 邻居发现
- DR 选举
- RP 发现
- 嵌入式 RP
- 构建 RPT
- IPv6 组播源注册
- SPT 切换
- 断言

1. 邻居发现

IPv6 PIM-SM使用与IPv6 PIM-DM类似的邻居发现机制，具体请参见“1.1.2 1. 邻居发现”一节。

2. DR 选举

借助 Hello 报文还可以为共享网络（如 Ethernet）选举 DR，DR 将作为该共享网络中 IPv6 组播数据的唯一转发者。

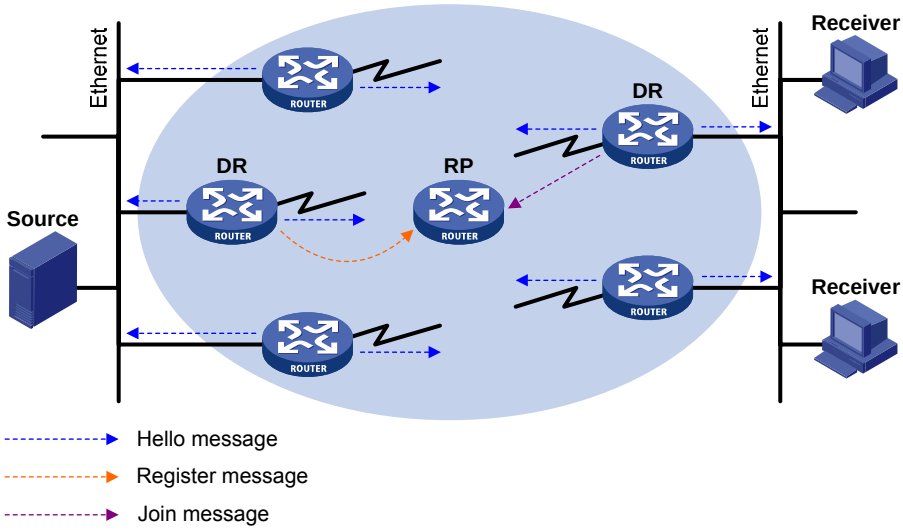
无论是与 IPv6 组播源相连的网络，还是与接收者相连的网络，都需要选举 DR。接收者侧的 DR 负责向 RP 发送加入报文；IPv6 组播源侧的 DR 负责向 RP 发送注册报文。

说明

- 各路由器之间通过比较 Hello 报文中所携带的优先级和 IPv6 链路本地地址，可以为多路由器网段选举 DR。
- 在充当接收者侧 DR 的设备上必须使能 MLD，否则连接在该 DR 上的接收者将不能通过该 DR 加入 IPv6 组播组。

有关 MLD 的介绍，请参见“IP 组播分册”中的“MLD 配置”。

图1-3 DR 选举示意图



如 图 1-3所示，DR的选举过程如下：

(1) 共享网络上的各路由器相互之间发送 Hello 报文（携带有竞选 DR 优先级的参数），拥有最高优先级的路由器将成为 DR；

(2) 如果优先级相同，或者网络中至少有一台路由器不支持在 Hello 报文中携带竞选 DR 优先级的参数，则根据各路由器的 IPv6 链路本地地址大小来竞选 DR，IPv6 链路本地地址最大的路由器将成为 DR。

当 DR 出现故障时，其余路由器在超时后仍没有收到来自 DR 的 Hello 报文，则会触发新的 DR 选举过程。

3. RP 发现

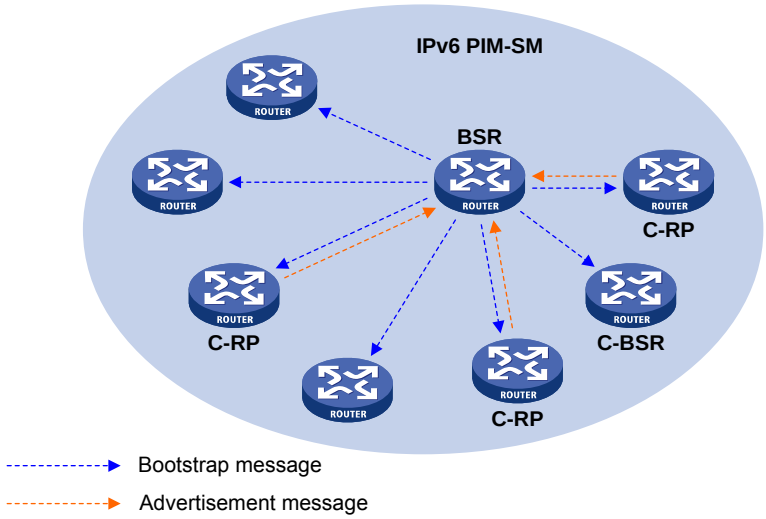
RP 是 IPv6 PIM-SM 域中的核心设备。在结构简单的小型网络中，IPv6 组播信息量少，整个网络仅依靠一个 RP 进行 IPv6 组播信息的转发即可，此时可以在 IPv6 PIM-SM 域中的各路由器上静态指定 RP 的位置；但是在更多的情况下，IPv6 PIM-SM 域的规模都很大，通过 RP 转发的 IPv6 组播信息量巨大。为了缓解 RP 的负担并优化 RPT 的拓扑结构，可以在 IPv6 PIM-SM 域中配置多个 C-RP（Candidate-RP，候选 RP），通过自举机制来动态选举 RP，使不同的 RP 服务于不同的组播组，此时需要配置 BSR（BootStrap Router，自举路由器）。BSR 是 IPv6 PIM-SM 域中的管理核心，一个 IPv6 PIM-SM 域内只能有一个 BSR，但可以配置多个 C-BSR（Candidate-BSR，候选 BSR）。这样，一旦 BSR 发生故障，其余 C-BSR 能够通过自动选举产生新的 BSR，从而确保业务免受中断。

说明

- 一个 RP 可以同时服务于多个 IPv6 组播组，但一个 IPv6 组播组只能唯一对应一个 RP。
- 一台设备可以同时充当 C-RP 和 C-BSR。

如 图 1-4所示，BSR负责收集网络中由C-RP发来的宣告报文（Advertisement Message），该报文中携带有C-RP的地址和优先级以及其服务的IPv6 组范围，BSR将这些信息汇总为RP-Set（RP集，即IPv6 组播组与RP的映射关系数据库），封装在自举报文（Bootstrap Message）中并发布到整个IPv6 PIM-SM域。

图1-4 RP 与 BSR 信息交互示意图



网络中的各路由器将依据 RP-Set 提供的信息，使用相同的规则从众多 C-RP 中为特定 IPv6 组播组选择其对应的 RP，具体规则如下：

(1) 首先比较 C-RP 的优先级，优先级较高者获胜。

(2) 若优先级相同，则使用哈希（Hash）函数计算哈希值，该值较大者获胜。

(3) 若优先级和哈希值都相同，则 C-RP 地址较大者获胜。

哈希函数的表达式为：Value (G, M, C_i) = (1103515245 * ((1103515245 * (G & M) + 12345) XOR C_i) + 12345) mod 2^{31}，其中各符号的含义如 表 1-1所示。

表1-1 哈希函数中各符号含义

符号	含义
Value	哈希值
G	将 IPv6 组播组地址的每 32 比特作为一节，各节之间通过异或运算而得的数值。假设 IPv6 组播组地址为 FF0E:C20:1A3:63::101，则 G = 0xFF0E0C20 XOR 0x01A30063 XOR 0x00000000 XOR 0x00000101
M	哈希掩码长度（Hash Mask Length）
C_i	将 C-RP IPv6 地址的每 32 比特作为一节，各节之间通过异或运算而得的数值。假设 C-RP 的 IPv6 地址为 3FFE:B00:C18:1::10，则 C_i = 0x3FFE0B00 XOR 0x0C180001 XOR 0x00000000 XOR 0x00000010
&	逻辑运算符，表示与运算
XOR	逻辑运算符，表示异或运算
mod	算术运算符，表示整除取余

4. 嵌入式 RP

通过嵌入式 RP（Embedded RP）机制可以从 IPv6 组播地址中解析出 RP 的地址，从而实现 IPv6 组播组到 RP 的映射，以取代静态配置的 RP 或由 BSR 机制动态计算出来的 RP，DR 不再需要预先知道 RP 的信息，只需对组播报文进行分析即可知道 RP 的地址。其工作原理如下：

- 接收者侧：

(1) 接收者主机发送 MLD 报告报文声明加入某 IPv6 组播组；

(2) 接收者侧的 DR 提取内嵌在 IPv6 组播地址中的 RP 地址，并向该 RP 发送加入报文（Join Message）。

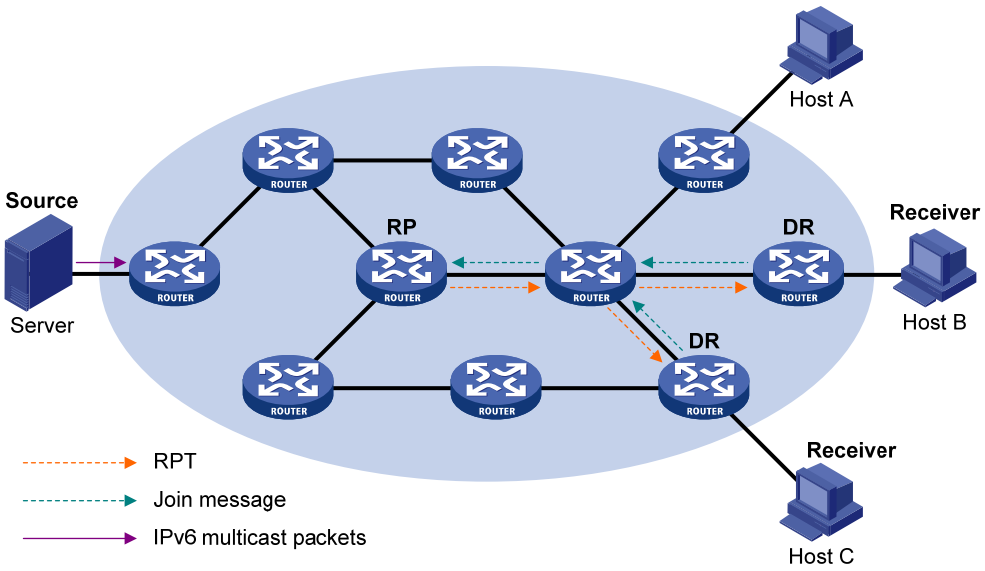
- IPv6 组播源侧：

(1) IPv6 组播源要向某 IPv6 组播组发送 IPv6 组播数据；

(2) IPv6 组播源侧的 DR 提取内嵌在 IPv6 组播地址中的 RP 地址，并向该 RP 发送注册报文。

5. 构建 RPT

图1-5 IPv6 PIM-SM 中构建 RPT 示意图



如 图 1-5所示，RPT的构建过程如下：

(1) 当接收者加入一个 IPv6 组播组 G 时，先通过 MLD 报文通知与其直连的 DR；

(2) DR 掌握了 IPv6 组播组 G 的接收者的信息后，向该组所对应的 RP 方向逐跳发送加入报文；

(3) 从 DR 到 RP 所经过的路由器就形成了 RPT 的分支，这些路由器都在其转发表中生成了（*，G）表项，这里的“*”表示来自任意 IPv6 组播源。RPT 以 RP 为根，以 DR 为叶子。

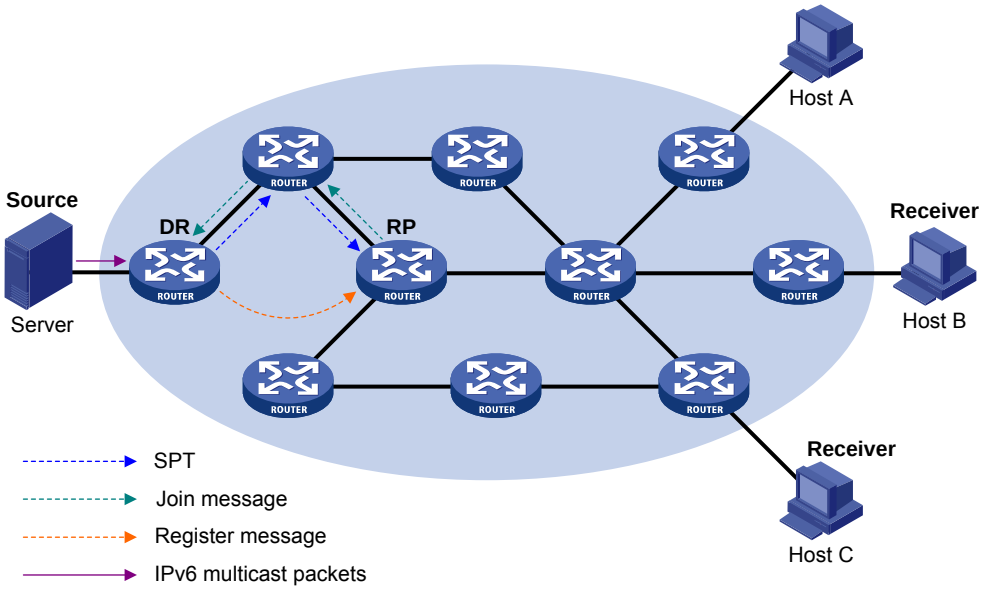
当发往 IPv6 组播组 G 的 IPv6 组播数据流经 RP 时，数据就会沿着已建立好的 RPT 到达 DR，进而到达接收者。

当某接收者对 IPv6 组播组 G 的信息不再感兴趣时，与其直连的 DR 会逆着 RPT 向该组的 RP 方向逐跳发送剪枝报文；上游节点收到该报文后在其出接口列表中删除与下游节点相连的接口，并检查自己是否拥有该 IPv6 组播组的接收者，如果没有则继续向其上游转发该剪枝报文。

6. IPv6 组播源注册

IPv6 组播源注册的目的是向 RP 通知 IPv6 组播源的存在。

图1-6 IPv6 组播源注册示意图



如 图 1-6所示，IPv6 组播源向RP注册的过程如下：

(1) 当 IPv6 组播源 S 向 IPv6 组播组 G 发送了一个 IPv6 组播报文时，与 IPv6 组播源直连的 DR 在收到该报文后，就将其封装成注册报文，并通过单播方式发送给相应的 RP；

(2) 当 RP 收到该报文后，一方面解封装注册报文并将封装在其中的 IPv6 组播报文沿着 RPT 转发给接收者，另一方面向 IPv6 组播源方向逐跳发送（S，G）加入报文。这样，从 RP 到 IPv6 组播源所经过的路由器就形成了 SPT 的分支，这些路由器都在其转发表中生成了（S，G）表项。SPT 以 IPv6 组播源侧的 DR 为根，以 RP 为叶子。

(3) IPv6 组播源发出的 IPv6 组播数据沿着已建立好的 SPT 到达 RP，然后由 RP 把 IPv6 组播数据沿着 RPT 向接收者进行转发。当 RP 收到沿着 SPT 转发来的 IPv6 组播数据后，通过单播方式向与 IPv6 组播源直连的 DR 发送注册停止报文（Register-Stop Message），IPv6 组播源注册过程结束。

说明

本节中假设允许 RP 发起 SPT 切换，否则 IPv6 组播源侧 DR 将一直用注册报文封装 IPv6 组播报文，注册过程不会结束，除非 RP 上（S，G）表项的出接口变为空。

---

7. SPT 切换

在 IPv6 PIM-SM 域中，一个 IPv6 组播组唯一对应一个 RP 和一棵 RPT。在 SPT 切换前，所有发往该组的 IPv6 组播报文都必须先由 IPv6 组播源侧 DR 封装在注册报文中发往 RP，由 RP 解封装后再沿 RPT 分发给接收者侧的 DR，RP 是所有 IPv6 组播数据必经的中转站。

(1) RP 发起的 SPT 切换

RP 周期性地检测 IPv6 组播数据（S，G）的转发速率，缺省情况下，收到组播报文后，立即向 IPv6 组播源方向发送（S，G）加入报文，在 IPv6 组播源侧 DR 与 RP 之间建立起 SPT 分支，后续的 IPv6 组播报文都直接沿该分支到达 RP。

说明

由RP发起的SPT切换的详细过程，请参见“1.1.4 6. IPv6 组播源注册”一节。

(2) 接收者侧 DR 发起的 SPT 切换

接收者侧 DR 周期性地检测 IPv6 组播数据（S，G）的转发速率，缺省情况下，收到组播报文后，立即发起 SPT 切换，过程如下：

- 首先，接收者侧 DR 向 IPv6 组播源方向逐跳发送（S，G）加入报文，并最终送达 IPv6 组播源侧 DR，沿途经过的所有路由器在其转发表中都生成了（S，G）表项，从而建立了 SPT 分支；
- 随后，当 IPv6 组播数据沿 SPT 到达 RPT 与 SPT 分叉的路由器时，该路由器开始丢弃沿 RPT 到达的 IPv6 组播数据，同时向 RP 逐跳发送含 RP 位的剪枝报文，RP 收到该报文后继续向 IPv6 组播源方向发送剪枝报文（假设此时只有这一个接收者），从而完成了 SPT 切换；
- 最终，IPv6 组播数据将沿 SPT 从 IPv6 组播源到达到接收者。

通过 SPT 切换，IPv6 PIM-SM 能够以比 IPv6 PIM-DM 更经济的方式建立 SPT。

8. 断言

IPv6 PIM-SM使用与IPv6 PIM-DM类似的断言机制，具体请参见“1.1.2 4. 断言”一节。

1.1.5 SSM 模型在 IPv6 PIM 中的实现

SSM（Source-Specific Multicast，指定信源组播）模型和 ASM（Any-Source Multicast，任意信源组播）模型是两个完全对等的模型。目前，ASM 模型包括 IPv6 PIM-DM 和 IPv6 PIM-SM 两种模式，SSM 模型能够借助 IPv6 PIM-SM 的部分技术来实现。

SSM 模型为指定源组播提供了解决方案，通过 MLDv2 来维护主机与路由器之间的关系。鉴于 IPv6 PIM-DM 模式以“扩散—剪枝”方式构建以 IPv6 组播源为根的 SPT，虽然 SPT 的路径最短，但是分发树的建立过程效率较低，不适合大中型网络。

在实际应用中，通常采用 IPv6 PIM-SM 模式的一部分技术来实现 SSM 模型。由于接收者已经通过其它渠道（如广告咨询等）知道了 IPv6 组播源的具体位置，因此在 SSM 模型中无需 RP，无需构建 RPT，也无需 IPv6 组播源注册过程来发现其它 IPv6 PIM 域内的 IPv6 组播源。

与 ASM 模型相比，SSM 模型仅需要 MLDv2 和 IPv6 PIM-SM 部分子集的支持，SSM 模型在 IPv6 PIM 域内的工作机制可以简单概括如下：

- 邻居发现
- DR 选举
- 构建 SPT

1. 邻居发现

IPv6 PIM-SSM使用与IPv6 PIM-SM完全相同的邻居发现机制，具体请参见“1.1.2 1. 邻居发现”一节。

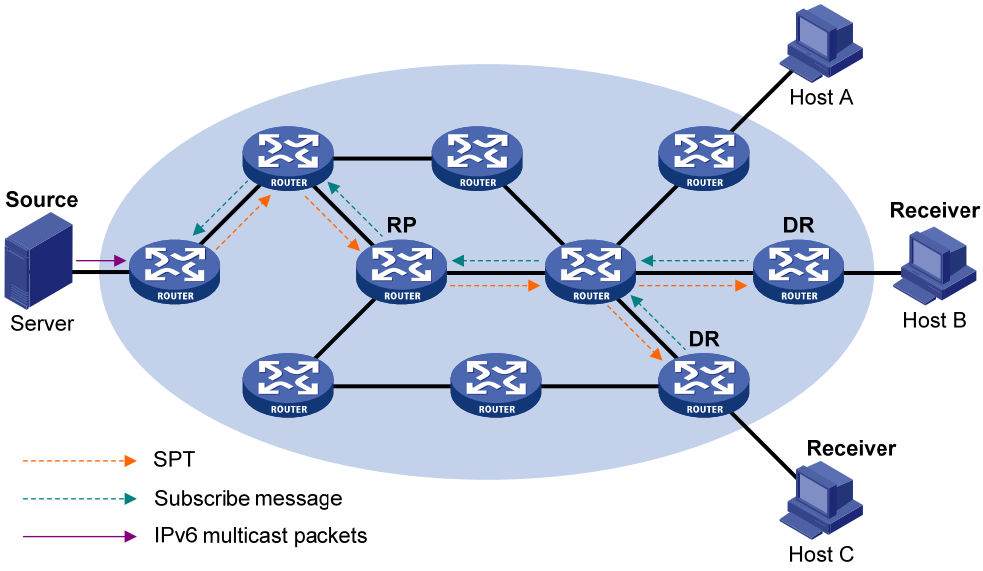
2. DR 选举

IPv6 PIM-SSM使用与IPv6 PIM-SM完全相同的DR选举机制，具体请参见 “1.1.4 2. DR选举”一节。

3. 构建 SPT

构建为 IPv6 PIM-SM 服务的 RPT，还是构建为 IPv6 PIM-SSM 服务的 SPT，关键在于接收者准备加入的 IPv6 组播组是否属于 IPv6 SSM 组地址范围（IANA 保留的 IPv6 SSM 组地址范围为 FF3x::/32，其中 x 表示任意合法的 scope）。

图1-7 IPv6 PIM-SSM 中构建 SPT 示意图



如 图 1-7所示，Host B和Host C为IPv6 组播信息的接收者（Receiver），由其借助MLDv2 的报告报文向DR报告自己对来自IPv6 组播源S、发往IPv6 组播组G的信息感兴趣。收到该报告报文的DR先判断该报文中的IPv6 组地址是否在IPv6 SSM组地址范围内：

- 如果在 IPv6 SSM 组地址范围内，则构建 IPv6 PIM-SSM，并向 IPv6 组播源 S 逐跳发送通道（Channel）的订阅报文（Subscribe Message）。沿途所有路由器上都创建（S，G）表项，从而在网络内构建了一棵以 IPv6 组播源 S 为根、以接收者为叶子的 SPT，该 SPT 就是 IPv6 PIM-SSM 中的传输通道；
- 如果不在 IPv6 SSM 组地址范围内，则仍旧按照 IPv6 PIM-SM 的流程进行后续处理，此时 DR 需要向 RP 发送（*，G）加入报文，同时需要进行 IPv6 组播源的注册。

说明

在 IPv6 PIM-SSM 中，借助 “通道” 的概念表示 IPv6 组播组，借助 “订阅报文” 的概念表示加入报文。

1.1.6 协议规范

与 IPv6 PIM 相关的协议规范有：

- RFC 4601：Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)

- RFC 3973：Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification (Revised)
- RFC 3956：Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address
- RFC 4607：Source-Specific Multicast for IP
- RFC 5059：Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- draft-ietf-ssm-overview-05：An Overview of Source-Specific Multicast (SSM)

1.2 配置 IPv6 PIM-DM

1.2.1 IPv6 PIM-DM 配置任务简介

表1-2 IPv6 PIM-DM 配置任务简介

配置任务	说明	详细配置
使能 IPv6 PIM-DM	必选	1.2.3
使能状态刷新能力	可选	1.2.4
配置状态刷新参数	可选	1.2.5
配置 IPv6 PIM-DM 定时器	可选	1.2.6
配置 IPv6 PIM 公共特性	可选	1.5

1.2.2 配置准备

在配置 IPv6 PIM-DM 之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通

在配置 IPv6 PIM-DM 之前，需准备以下数据：

- 发送状态刷新报文的时间间隔
- 等待接收新状态刷新报文的最小时间
- 状态刷新报文的 Hop Limit 值
- 嫁接报文的重传时间

1.2.3 使能 IPv6 PIM-DM

在接口上使能了 IPv6 PIM-DM 后，路由器会定期发送 Hello 报文以发现 IPv6 PIM 邻居，并对收到的来自 IPv6 PIM 邻居的报文进行处理。在部署 IPv6 PIM-DM 域时，建议在其所有非边界接口上均使能 IPv6 PIM-DM。

表1-3 使能 IPv6 PIM-DM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 组播路由	**multicast ipv6 routing-enable**	必选 缺省情况下，IPv6 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能 IPv6 PIM-DM	**pim ipv6 dm**	必选 缺省情况下，IPv6 PIM-DM 处于关闭状态

注意

- 同一台设备所有接口上所启用的 IPv6 PIM 模式必须相同。
- IPv6 PIM-DM 不能与处于 IPv6 SSM 组地址范围内的 IPv6 组播组同时使用。

说明

有关 **multicast ipv6 routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“IPv6 组播路由与转发命令”。

1.2.4 使能状态刷新能力

只有当一个共享网段内的所有 IPv6 PIM 路由器上都使能了状态刷新能力时，该共享网段才具备状态刷新能力。

请在 IPv6 PIM-DM 域内的所有路由器上进行如下配置。

表1-4 使能状态刷新能力

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
使能状态刷新能力	**pim ipv6 state-refresh-capable**	可选 缺省情况下，状态刷新能力处于使能状态

1.2.5 配置状态刷新参数

为了避免各路由器上被剪枝的接口因为超时而恢复转发，与 IPv6 组播源直连的路由器会周期性地发送（S，G）状态刷新报文，该报文沿着 IPv6 PIM-DM 域最初的扩散路径逐跳进行转发，从而刷新沿途所有路由器上的剪枝定时器的状态。

路由器可能在短时间内收到多个状态刷新报文，而其中有些报文可能是重复的。为了避免接收这些重复的报文，可以配置接收新状态刷新报文的等待时间：路由器将丢弃在该时间内收到的状态刷新报文；当该时间超时后，路由器将正常接收新的状态刷新报文，并更新自己的 IPv6 PIM-DM 状态，同时重置该等待时间。

在收到状态刷新报文时，路由器会将该报文的 Hop Limit 值减 1 后转发给其下游，直至该报文的 Hop Limit 值减为 0，当网络规模很小时，状态刷新报文将在网络中循环传递。因此，为了有效控制刷新报文的传递范围，需要根据网络规模大小配置合适的 Hop Limit 值。

请在 IPv6 PIM-DM 域内的所有路由器上进行如下配置。

表1-5 配置状态刷新参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置发送状态刷新报文的时间间隔	**state-refresh-interval** *interval*	可选 缺省情况下，发送状态刷新报文的时间间隔为 60 秒
配置接收新状态刷新报文的等待时间	**state-refresh-rate-limit** *interval*	可选 缺省情况下，接收新状态刷新报文的等待时间为 30 秒
配置状态刷新报文的 Hop Limit 值	**state-refresh-hoplimit** *hoplimit-value*	可选 缺省情况下，状态刷新报文的 Hop Limit 值为 255

1.2.6 配置 IPv6 PIM-DM 定时器

嫁接报文是 IPv6 PIM-DM 中唯一使用确认机制的报文。在 IPv6 PIM-DM 域中，下游路由器发出嫁接报文后，如果在指定时间内没有收到来自其上游路由器的嫁接应答报文，则会重发嫁接报文，直到被确认。

表1-6 配置 IPv6 PIM-DM 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置嫁接报文的重传时间	**pim ipv6 timer graft-retry** *interval*	可选 缺省情况下，嫁接报文的重传时间为 3 秒


说明

有关IPv6 PIM-DM其它定时器的相关配置，请参见“1.5.6 配置IPv6 PIM公共定时器”。

1.3 配置 IPv6 PIM-SM

1.3.1 IPv6 PIM-SM 配置任务简介

表1-7 IPv6 PIM-SM 配置任务简介

配置任务		说明	详细配置
使能 IPv6 PIM-SM		必选	1.3.3
配置 RP	配置静态 RP	可选	1.3.4 1.
	配置 C-RP	可选	1.3.4 2.
	使能嵌入式 RP	可选	1.3.4 3.
	全局配置 C-RP 定时器	可选	1.3.4 4.

配置任务		说明	详细配置
配置 BSR	配置 C-BSR	可选	1.3.5 1.
	配置 BSR 服务边界	可选	1.3.5 2.
	全局配置 C-BSR 参数	可选	1.3.5 3.
	配置 C-BSR 定时器	可选	1.3.5 4.
配置 IPv6 组播源注册		可选	1.3.6
配置 SPT 切换		可选	1.3.7
配置 IPv6 PIM 公共特性		可选	1.5

1.3.2 配置准备

在配置 IPv6 PIM-SM 之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通

在配置 IPv6 PIM-SM 之前，需准备以下数据：

- 静态 RP 的 IPv6 地址及表示其所服务 IPv6 组范围的 IPv6 ACL 规则
- C-RP 的优先级及表示其所服务 IPv6 组范围的 IPv6 ACL 规则
- 合法 C-RP 的地址范围及表示其所服务 IPv6 组范围的 IPv6 ACL 规则
- 发送宣告报文的时间间隔
- C-RP 超时时间
- C-BSR 的优先级
- 哈希掩码长度
- 表示合法 BSR 地址范围的 IPv6 ACL 规则
- 自举时间间隔
- 自举超时时间
- 表示注册报文过滤规则的 IPv6 ACL 规则
- 注册抑制时间
- 注册探测时间

1.3.3 使能 IPv6 PIM-SM

在接口上使能了 IPv6 PIM-SM 后，路由器会定期发送 Hello 报文以发现 IPv6 PIM 邻居，并对收到的来自 IPv6 PIM 邻居的报文进行处理。在部署 IPv6 PIM-SM 域时，建议在其所有非边界接口上均使能 IPv6 PIM-SM。

表1-8 使能 IPv6 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 组播路由	**multicast ipv6 routing-enable**	必选 缺省情况下，IPv6 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
使能 IPv6 PIM-SM	**pim ipv6 sm**	必选 缺省情况下，IPv6 PIM-SM 处于关闭状态

注意

同一台设备所有接口上所启用的 IPv6 PIM 模式必须相同。

说明

有关 **multicast ipv6 routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“IPv6 组播路由与转发命令”。

1.3.4 配置 RP

RP 可以通过手工方式静态配置，也可以通过 BSR 机制动态选举。对于大型 IPv6 PIM 网络，配置静态 RP 将会非常繁琐。所以，通常静态 RP 是作为动态选举 RP 机制的备份手段，以提高网络的健壮性，增强组播网络的运营管理能力。

1. 配置静态 RP

当网络内仅有一个动态 RP 时，通过手工配置静态 RP 可以避免因单一节点故障而引起的通信中断，同时也可以避免 C-RP 与 BSR 之间频繁的信息交互而占用带宽。

请在 PIM-SM 域内的所有路由器上进行如下配置。

表1-9 配置静态 RP

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置静态 RP	**static-rp** *ipv6-rp-address* [*acl6-number*] [**preferred**]	必选 缺省情况下，没有配置静态 RP

注意

为了让静态 RP 功能正常发挥作用，必须在所有路由器上指定相同的静态 RP 地址。

2. 配置 C-RP

在 PIM-SM 域中，可以把有意成为 RP 的路由器配置为 C-RP。BSR 通过接收来自 C-RP 的 C-RP 信息，或者接收来自其它路由器的自动 RP 宣告，收集 C-RP 信息并将其汇总为 RP-Set 信息，然后在全网内扩散。之后，网络内的其它路由器根据 RP-Set 信息计算出特定组播组范围所对应的 RP。建议在骨干网路由器上配置 C-RP。

为了防止 C-RP 欺骗，需要在 BSR 上配置合法的 C-RP 地址范围及其所服务的组播组范围。同时由于每个 C-BSR 都可能成为 BSR，因此必须在 IPv6 PIM-SM 域内的所有 C-BSR 上都配置相同的过滤策略。

表1-10 配置 C-RP

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置某接口为 C-RP	**c-rp** *ipv6-address* [**group-policy** *acl6-number* | **priority** *priority* | **holdtime** *hold-interval* | **advertisement-interval** *adv-interval*] *	必选 缺省情况下，没有配置 C-RP
合法 C-RP 的地址范围及其所服务的 IPv6 组播组范围	**crp-policy** *acl6-number*	可选 缺省情况下，C-RP 的地址范围及其所服务的 IPv6 组播组范围不受任何限制

说明

- 在配置 C-RP 时，应在 C-RP 与 IPv6 PIM-SM 域中的其它设备之间保留较大的通信带宽。
- 一个 RP 可以为多个 IPv6 组播组服务，也可以为所有 IPv6 组播组服务。每个 IPv6 组播组在任意时刻，只能由唯一的一个 RP 为其转发数据，而不能由多个 RP 转发数据。

3. 使能嵌入式 RP

嵌入式 RP 机制可以从 IPv6 组播地址中直接解析出 RP，从而取代静态配置的 RP 或由 BSR 机制动态计算出来的 RP，DR 不再需要预先知道 RP 的信息，只需对组播报文进行分析即可知道 RP 的地址。

表1-11 使能嵌入式 RP

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
使能嵌入式 RP 功能	**embedded-rp** [*acl6-number*]	可选 缺省情况下，默认嵌入式 RP 地址范围内的 IPv6 组播组均可以使用嵌入式 RP 功能

说明

默认的嵌入式 RP 地址范围为 FF7x::/12 和 FFFx::/12，其中 x 表示任意合法的 scope。有关 scope 字段的详细介绍，请参见“IP 组播分册”中的“组播概述”。

4. 全局配置 C-RP 定时器

为了使 BSR 能够在 IPv6 PIM-SM 域内分发 RP-Set 信息，C-RP 必须周期性地向 BSR 发送宣告报文，BSR 从该报文中学习 RP-Set 信息，并将该信息与自己的 IPv6 地址一起封装在自举报文中向域中的所有 IPv6 PIM 路由器进行宣告。

C-RP 在其宣告报文中封装一个保持时间，BSR 在收到该报文后，从中获得该时间值并启动 C-RP 超时定时器，如果超时后 BSR 仍没有收到来自 C-RP 后续的宣告报文，则认为目前网络中的 C-RP 失效或不可达。

请在已配置为 C-RP 的路由器上进行如下配置。

表1-12 全局配置 C-RP 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置发送宣告报文的时间间隔	**c-rp advertisement-interval** *interval*	可选 缺省情况下，发送宣告报文的时间间隔为 60 秒
配置 C-RP 超时时间	**c-rp holdtime** *interval*	可选 缺省情况下，C-RP 的超时时间为 150 秒



说明

有关IPv6 PIM-SM其它定时器的相关配置，请参见“1.5.6 配置IPv6 PIM公共定时器”。

1.3.5 配置 BSR

在一个 IPv6 PIM-SM 域中只能有一个 BSR，但需要配置至少一个 C-BSR。任意一台路由器都可以被配置为 C-BSR。在 C-BSR 之间通过自动选举产生 BSR，BSR 负责在 IPv6 PIM-SM 域中收集并发布 RP 信息。

1. 配置 C-BSR

C-BSR 应配置在骨干网的路由器上，在将路由器配置为 C-BSR 时，必须同时指定一个使能了 IPv6 PIM-SM 的接口的 IPv6 地址。C-BSR 间的自动选举机制简单描述如下：

- 最初，每个 C-BSR 都认为自己是本 IPv6 PIM-SM 域的 BSR，并使用接口的 IPv6 地址作为 BSR 地址，发送自举报文；
- 当某 C-BSR 收到其它 C-BSR 发来的自举报文时，首先比较自己与后者的优先级，优先级较高者获胜；在优先级相同的情况下，再比较自己与后者的 BSR 地址，拥有较大 IPv6 地址者获胜。如果后者获胜，则用后者的 BSR 地址替换自己的 BSR 地址，并不再认为自己是 BSR；否则，保留自己的 BSR 地址，并继续认为自己是 BSR。

通过在路由器上配置合法 BSR 的地址范围，可以对收到的 BSR 报文按照地址范围进行过滤，从而防止某些恶意主机非法伪装成 BSR，以避免合法的 BSR 被恶意取代。必须在 IPv6 PIM-SM 域内的所有路由器上进行相同的配置。通常针对以下两类情况实施预防措施：

(1) 某些恶意主机通过伪造 BSR 报文以欺骗路由器，试图更改 RP 映射关系。这种攻击通常发生在边缘路由器上，由于 BSR 处于网络内部，主机在网络外部，因此边缘路由器通过对收到的 BSR 报文进行邻居检查和 RPF 检查，丢弃不符合要求的报文，就可以避免外部网络用户对内部网络 BSR 的攻击；

(2) 网络中某台路由器被攻击者控制，或者有非法接入的路由器时，攻击者可以将这样的路由器配置为 C-BSR，并使其在竞争中获胜，从而控制网络中 RP 信息的发布权。由于在被配置为 C-BSR 后，路由器会自动向整个网络扩散 BSR 报文，而 BSR 报文是 Hop Limit 值为 1 的 IPv6 组播报文，所以只要其邻居路由器不接收该 BSR 报文，就不会影响整个网络。因此，通过在整个网络的所有路由器上都配置合法 BSR 的地址范围，从而丢弃合法范围之外的 BSR 报文，就可以防止此类攻击。

以上两种预防策略可以部分地保护网络中 BSR 的安全。但是如果某台合法的 BSR 路由器被攻击者控制，还是可能导致问题。

表1-13 配置 C-BSR

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置某接口为 C-BSR	**c-bsr** *ipv6-address* [*hash-length* [*priority*]]	必选 缺省情况下，没有配置 C-BSR
配置合法的 BSR 地址范围	**bsr-policy** *acl6-number*	可选 缺省情况下，BSR 的地址范围不受任何限制


说明

由于 BSR 与 IPv6 PIM-SM 域中的其它设备需要交换大量信息，因此应在 C-BSR 与 IPv6 PIM-SM 域中的其它设备之间保留较大的通信带宽。

2. 配置 BSR 服务边界

BSR 作为 IPv6 PIM-SM 域中的管理核心，负责将收集到的 RP-Set 信息以自举报文的形式发向 IPv6 PIM-SM 域中的所有路由器。

BSR 的服务边界，即 IPv6 PIM-SM 域的边界。BSR 是针对特定的服务范围而言的，众多的 BSR 服务边界接口将网络划分成不同的 IPv6 PIM-SM 域，自举报文无法通过 IPv6 PIM-SM 域的边界，BSR 服务边界之外的路由器也不能参与本 IPv6 PIM-SM 域内的组播转发。

请在有待成为 BSR 服务边界的路由器上进行如下配置。

表1-14 配置 BSR 服务边界

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-

操作	命令	说明
配置 BSR 的服务边界	**pim ipv6 bsr-boundary**	必选 缺省情况下，没有配置 BSR 的服务边界

3. 全局配置 C-BSR 参数

在一个 IPv6 PIM-SM 域中，从众多 C-BSR 中选举出唯一的 BSR。IPv6 PIM-SM 域内的 C-RP 向 BSR 发送宣告报文，由 BSR 汇总为 RP-Set，并向本 IPv6 PIM-SM 域内的所有路由器进行宣告。所有路由器都使用统一的哈希算法，得到特定 IPv6 组播组所对应 RP 的地址。

请在已配置为 C-BSR 的路由器上进行如下配置。

表1-15 全局配置 C-BSR 参数

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置哈希掩码长度	**c-bsr hash-length** *hash-length*	可选 缺省情况下，哈希掩码长度为 126
配置 C-BSR 的优先级	**c-bsr priority** *priority*	可选 缺省情况下，C-BSR 的优先级为 0

4. 配置 C-BSR 定时器

当某 C-BSR 竞选成为 BSR 后，它会通过自举报文向其所服务的区域以组播方式发送自己的 IPv6 地址和 RP-Set 信息。BSR 以自举时间间隔周期性地向网络发送自举报文，收到该报文的 C-BSR 会在自举超时时间内将其保持，此时 BSR 选举过程暂停；当自举超时时间超时后，C-BSR 之间会触发新一轮的 BSR 选举过程。

请在已配置为 C-BSR 的路由器上进行如下配置。

表1-16 配置 C-BSR 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置自举时间间隔	**c-bsr interval** *interval*	可选 缺省取值请参见 表 1-16后的相关说明
配置自举超时时间	**c-bsr holdtime** *interval*	可选 缺省取值请参见 表 1-16后的相关说明

说明

关于自举时间间隔的取值：

- 如果没有进行手工配置，则其取值由如下公式决定：自举时间间隔 =（自举超时时间 − 10）÷ 2。缺省情况下，自举超时时间为 130 秒，则自举时间间隔的缺省值 =（130 − 10）÷ 2 = 60（秒）。
- 如果进行了手工配置，则取配置值。

关于自举超时时间的取值：

- 如果没有进行手工配置，则其取值由如下公式决定：自举超时时间 = 自举时间间隔 × 2 + 10。缺省情况下，自举时间间隔为 60 秒，则自举超时时间的缺省值 = 60 × 2 + 10 = 130（秒）。
- 如果进行了手工配置，则取配置值。

注意

配置时，应保证自举时间间隔小于自举超时时间。

1.3.6 配置 IPv6 组播源注册

在 IPv6 PIM-SM 域内，IPv6 组播源侧 DR 向 RP 发送注册报文，而这些注册报文拥有不同的 IPv6 组播源或 IPv6 组播地址。为了让 RP 服务于特定的 IPv6 组播组，可以对注册报文进行过滤。如果某个（S，G）表项被过滤规则拒绝，或者过滤规则中没有定义对它的操作，RP 都会向 DR 发送注册停止报文，以停止该 IPv6 组播数据的注册过程。

出于对注册报文在传递过程中完整性的考虑，可以配置根据整个报文来计算校验和。但为了减少往注册报文中封装数据报文的工作量并考虑到互通性，一般情况下不建议配置根据注册报文的全部内容来计算校验和的方式。

当接收者不再通过 RP 接收发往某 IPv6 组播组的数据（即 RP 不再服务于该 IPv6 组播组），或 RP 正式开始接收来自 IPv6 组播源的 IPv6 组播数据时，RP 将向 IPv6 组播源侧 DR 发送注册停止报文，DR 收到该报文后将停止发送封装有 IPv6 组播数据的注册报文并启动注册停止定时器（Register-Stop Timer）。当注册停止定时器超时后，DR 会向 RP 发送一个空注册报文（Null-Register Message，即不封装 IPv6 组播数据的注册报文）：如果 DR 在注册探测时间（Register_Probe_Time）内收到了来自 RP 的注册停止报文，DR 将刷新其注册停止定时器；否则，DR 将重新开始发送封装有 IPv6 组播数据的注册报文。

注册停止定时器的超时时间是一个随机值，由其它两个时间值决定：注册抑制时间（Register_Suppression_Time）和注册探测时间。其具体取值范围如下：（0.5×注册抑止时间，1.5×注册抑止时间）－注册探测时间。

请在所有已配置为 C-RP 的路由器上配置注册报文的过滤规则和根据注册报文的全部内容来计算校验和；请在所有可能成为组播源侧 DR 的路由器上配置注册抑制时间和注册探测时间。

表1-17 配置 IPv6 组播源注册

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置注册报文的过滤规则	**register-policy** *acl6-number*	可选 缺省情况下，没有配置注册报文的过滤规则

操作	命令	说明
配置根据注册报文的全部内容来计算校验和	**register-whole-checksum**	可选 缺省情况下，仅根据注册报文头来计算校验和
配置注册抑制时间	**register-suppression-timeout** *interval*	可选 缺省情况下，注册抑制时间为 60 秒
配置注册探测时间	**probe-interval** *interval*	可选 缺省情况下，注册探测时间为 5 秒

1.3.7 配置 SPT 切换

请在所有可能成为接收者侧 DR 和所有已配置为 C-RP 的路由器上进行如下配置。

表1-18 配置 SPT 切换

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置发起 SPT 切换的条件	**spt-switch-threshold infinity** [**group-policy** *acl6-number* [**order** *order-value*]]	可选 缺省情况下，设备收到第一个 IPv6 组播数据包后便立即向 SPT 切换

1.4 配置 IPv6 PIM-SSM

说明

IPv6 PIM-SSM 模型需要 MLDv2 的支持，因此应确保连接有接收者的 IPv6 PIM 路由器上使能了 MLDv2。

1.4.1 IPv6 PIM-SSM 配置任务简介

表1-19 IPv6 PIM-SSM 配置任务简介

配置任务	说明	详细配置
使能 IPv6 PIM-SM	必选	1.4.3
配置 IPv6 SSM 组播组范围	可选	1.4.4
配置 IPv6 PIM 公共特性	可选	1.5

1.4.2 配置准备

在配置 IPv6 PIM-SSM 之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通

在配置 IPv6 PIM-SSM 之前，需准备以下数据：

- IPv6 SSM 组播组范围

1.4.3 使能 IPv6 PIM-SM

由于 SSM 模型是通过 IPv6 PIM-SM 的部分子集功能实现的，因此系统在使能了 IPv6 PIM-SM 的同时也具备了 IPv6 PIM-SSM 能力。在部署 IPv6 PIM-SM 域时，建议在其所有非边界接口上均使能 IPv6 PIM-SM。

表1-20 使能 IPv6 PIM-SM

操作	命令	说明
进入系统视图	**system-view**	-
使能 IPv6 组播路由	**multicast ipv6 routing-enable**	必选 缺省情况下，IPv6 组播路由处于关闭状态
进入接口视图	**interface** *interface-type interface-number*	-
使能 IPv6 PIM-SM	**pim ipv6 sm**	必选 缺省情况下，IPv6 PIM-SM 处于关闭状态

注意

同一台设备所有接口上所启用的 IPv6 PIM 模式必须相同。

说明

有关 **multicast ipv6 routing-enable** 命令的详细介绍，请参见“IP 组播分册”中的“IPv6 组播路由与转发命令”。

1.4.4 配置 IPv6 SSM 组播组范围

在把来自 IPv6 组播源的信息传递给接收者的过程中，是采用 IPv6 PIM-SSM 模型还是 IPv6 PIM-SM 模型，这取决于接收者订阅通道（S，G）中的 IPv6 组播组是否在 IPv6 SSM 组播组范围之内，所有使能了 IPv6 PIM-SM 的接口将会认为属于该范围内的 IPv6 组播组采用了 IPv6 SSM 模型。

请在 IPv6 PIM-SM 域内的所有路由器上进行如下配置。

表1-21 配置 IPv6 SSM 组播组范围

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置 IPv6 SSM 组播组的范围	**ssm-policy** *acl6-number*	可选 缺省情况下，IPv6 SSM 组播组的范围为 FF3x::/32，其中 x 表示任意合法的 scope

注意

- 应确保域内所有路由器上配置的 IPv6 SSM 组播组地址范围都一致，否则组播信息将无法通过 IPv6 SSM 模型进行传输。
- 如果某 IPv6 组播组属于 IPv6 SSM 组播组范围，但该组成员使用 MLDv1 发送加入报文，则设备不会触发（*，G）加入报文。

1.5 配置 IPv6 PIM 公共特性

说明

对于本节中那些既可在 IPv6 PIM 视图、又可在接口视图下进行配置的功能或参数来说：

- 在 IPv6 PIM 视图下所作的配置对所有接口生效，在接口视图下所作的配置仅对当前接口生效；
- 如果在这两个视图下对某功能或参数均进行了配置，则不论配置先后，接口视图下的配置将被优先采用。

1.5.1 IPv6 PIM 公共特性配置任务简介

表1-22 IPv6 PIM 公共特性配置任务简介

配置任务	说明	详细配置
配置 IPv6 组播数据过滤器	可选	1.5.3
配置 Hello 报文过滤器	可选	1.5.4
配置 Hello 报文选项	可选	1.5.5
配置 IPv6 PIM 公共定时器	可选	1.5.6
配置加入/剪枝报文规格	可选	1.5.7

1.5.2 配置准备

在配置 IPv6 PIM 公共特性之前，需完成以下任务：

- 配置任一 IPv6 单播路由协议，实现域内网络层互通
- 配置 IPv6 PIM-DM（或 IPv6 PIM-SM、IPv6 PIM-SSM）

在配置 IPv6 PIM 公共特性之前，需准备以下数据：

- 表示 IPv6 组播数据过滤规则的 IPv6 ACL 规则
- 表示合法 Hello 报文的源地址范围的 IPv6 ACL 规则
- 竞选 DR 的优先级（全局值/接口值）
- 保持 IPv6 PIM 邻居可达状态的时间（全局值/接口值）
- 剪枝延迟时间（全局值/接口值）
- 剪枝否决时间（全局值/接口值）
- 发送 Hello 报文的时间间隔（全局值/接口值）
- 发送 Hello 报文的最大延迟时间（接口值）

- 保持断言状态的时间（全局值/接口值）
- 发送加入/剪枝报文的时间间隔（全局值/接口值）
- 保持加入/剪枝状态的时间（全局值/接口值）
- IPv6 组播源生存时间
- 加入/剪枝报文的最大长度
- 加入/剪枝报文中（S，G）表项的最大数量

1.5.3 配置 IPv6 组播数据过滤器

无论在 IPv6 PIM-DM 还是 IPv6 PIM-SM 域内，各路由器都可以对流经自己的 IPv6 组播数据进行检查，通过比较是否符合过滤规则，从而决定是否继续转发 IPv6 组播数据。也就是说 IPv6 PIM 域内的路由器能够成为 IPv6 组播数据的过滤器。过滤器的存在一方面有助于实现信息流量控制，另一方面可以在安全性方面限定下游接收者能够获得的信息。

表1-23 配置 IPv6 组播数据过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置 IPv6 组播数据过滤器	**source-policy** *acl6-number*	必选 缺省情况下，没有配置 IPv6 组播数据过滤器

说明

- 通常，过滤器的位置离 IPv6 组播源距离越近，过滤影响越明显。
- 过滤器不仅过滤独立的 IPv6 组播数据，还过滤封装在注册报文中的 IPv6 组播数据。

1.5.4 配置 Hello 报文过滤器

随着 IPv6 PIM 协议的推广和应用，对其安全性的要求也越来越高。建立正确的 IPv6 PIM 邻居是 IPv6 PIM 协议安全应用的前提。如果在接口上指定了合法 Hello 报文的源地址范围，便能够保证 IPv6 PIM 邻居的正确建立，从而有效防止各种 IPv6 PIM 协议报文攻击，提高设备对 IPv6 PIM 协议报文处理的安全性。

表1-24 配置 Hello 报文过滤器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置合法 Hello 报文的源地址范围	**pim ipv6 neighbor-policy** *acl6-number*	必选 缺省情况下，Hello 报文的源地址范围不受任何限制

说明

当 Hello 报文过滤器的配置生效后，对于之前已建立的 IPv6 PIM 邻居，若由于其 Hello 报文被过滤而导致无法收到后续的 Hello 报文，将会在老化超时后被自动删除

1.5.5 配置 Hello 报文选项

无论在 IPv6 PIM-DM 域还是在 IPv6 PIM-SM 域内，各路由器之间发送的 Hello 报文都包含很多可供配置的选项，对各选项的介绍如下：

- DR_Priority（仅用于 IPv6 PIM-SM）：表示竞选 DR 的优先级，优先级高的设备被选举为 DR。可以在与 IPv6 组播源或接收者直连的共享网络中的所有路由器上都配置此参数。
- Holdtime：表示保持 IPv6 PIM 邻居可达状态的时间，若超时后仍没有收到 Hello 报文，则认为 IPv6 PIM 邻居失效或不可达。
- LAN_Prune_Delay：表示在共享网络上传递剪枝报文的延迟时间，该选项由三部分组成：Lan-delay（剪枝延迟时间）、Override-interval（剪枝否决时间）和禁止加入报文抑制能力。当共享网段内各 IPv6 PIM 路由器的剪枝延迟或否决时间值不同时，取其中最大的值；当要禁止加入报文抑制能力时，须在共享网段内的所有 IPv6 PIM 路由器上都禁止该能力。

剪枝延迟时间将使上游路由器延迟处理收到的剪枝报文。如果配置的剪枝延迟时间过小，可能导致下游路由器还未否决剪枝动作，上游路由器就已经停止转发 IPv6 组播数据，因此配置时需要慎重。

剪枝否决时间表示允许下游路由器否决剪枝动作的时间。路由器在收到下游路由器发来的剪枝报文后并不立即执行剪枝动作，而是仍然将当前的转发状态保持剪枝延迟时间＋剪枝否决时间。如果下游路由器需要继续接收 IPv6 组播数据，则必须在剪枝否决时间内否决剪枝动作；否则上游路由器会在剪枝延迟时间＋剪枝否决时间超时后执行剪枝动作。

IPv6 PIM 路由器发送 Hello 报文时，会生成一个随机数 Generation ID 并携带在该报文中。一台 IPv6 PIM 路由器的 Generation ID 一般不会变化，除非该路由器的状态发生了改变（如接口刚使能了 IPv6 PIM 或设备进行了重启），此时，当其开始或重新开始发送 Hello 报文时，会生成一个新的 Generation ID。这样，如果 IPv6 PIM 路由器发现其上游邻居发来的 Hello 报文中 Generation ID 发生了改变，就会认为该邻居的状态已经丢失或其上游邻居已经改变，从而触发发送加入报文以进行状态刷新。

在禁止加入报文抑制能力（即使能邻居跟踪）时，应在共享网段的所有 IPv6 PIM 路由器上都禁止该能力，否则上游路由器无法跟踪每台下游路由器的加入报文。

1. 全局配置 Hello 报文选项

表1-25 全局配置 Hello 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置竞选 DR 的优先级	**hello-option dr-priority** *priority*	可选 缺省情况下，竞选 DR 的优先级为 1
配置保持 IPv6 PIM 邻居可达状态的时间	**hello-option holdtime** *interval*	可选 缺省情况下，保持 IPv6 PIM 邻居可达状态的时间为 105 秒

操作	命令	说明
配置剪枝延迟时间	**hello-option lan-delay** *interval*	可选 缺省情况下，剪枝延迟时间为 500 毫秒
配置剪枝否决时间	**hello-option override-interval** *interval*	可选 缺省情况下，剪枝否决时间为 2500 毫秒
禁止加入报文抑制能力	**hello-option neighbor-tracking**	必选 缺省情况下，加入报文抑制能力处于使能状态

2. 在接口上配置 Hello 报文选项

表1-26 在接口上配置 Hello 报文选项

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置竞选 DR 的优先级	**pim ipv6 hello-option dr-priority** *priority*	可选 缺省情况下，竞选 DR 的优先级为 1
配置保持 IPv6 PIM 邻居可达状态的时间	**pim ipv6 hello-option holdtime** *interval*	可选 缺省情况下，保持 IPv6 PIM 邻居可达状态的时间为 105 秒
配置剪枝延迟时间	**pim ipv6 hello-option lan-delay** *interval*	可选 缺省情况下，剪枝延迟时间为 500 毫秒
配置剪枝否决时间	**pim ipv6 hello-option override-interval** *interval*	可选 缺省情况下，剪枝否决时间为 2500 毫秒
禁止加入报文抑制能力	**pim ipv6 hello-option neighbor-tracking**	必选 缺省情况下，加入报文抑制能力处于使能状态
配置不接受无 Generation ID 的 Hello 报文	**pim ipv6 require-genid**	必选 缺省情况下，接受无 Generation ID 的 Hello 报文

1.5.6 配置 IPv6 PIM 公共定时器

IPv6 PIM 路由器通过周期性地发送 Hello 报文，以发现 IPv6 PIM 邻居，并维护各路由器之间的 IPv6 PIM 邻居关系。

为了避免多个 IPv6 PIM 路由器同时发送 Hello 报文而导致冲突，当 IPv6 PIM 路由器在收到 Hello 报文时，将延迟一段时间后再发送 Hello 报文，该时间值为小于“触发 Hello 报文的最大延迟时间”的一个随机值。

IPv6 PIM 路由器通过周期性地向其上游路由器发送加入/剪枝报文以更新状态，在该报文中携带有保持时间，上游路由器为被剪枝的下游接口设置保持加入/剪枝状态定时器。

在断言中落选的路由器将会剪掉其下游的转发接口，并把这种断言状态保持一段时间。超时后，落选的路由器会重新恢复转发 IPv6 组播数据。

当路由器没有收到来自 IPv6 组播源 S 的后续 IPv6 组播数据时，不会立即删除（S，G）表项，而是将其维持一段时间后再删除，这段时间就称为 IPv6 组播源的生存时间。

1. 全局配置 IPv6 PIM 公共定时器

表1-27 全局配置 IPv6 PIM 公共定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置发送 Hello 报文的时间间隔	**timer hello** *interval*	可选 缺省情况下，发送 Hello 报文的时间间隔为 30 秒
配置发送加入/剪枝报文的时间间隔	**timer join-prune** *interval*	可选 缺省情况下，发送加入/剪枝报文的时间间隔为 60 秒
配置保持加入/剪枝状态的时间	**holdtime join-prune** *interval*	可选 缺省情况下，保持加入/剪枝状态的时间为 210 秒
配置保持断言状态的时间	**holdtime assert** *interval*	可选 缺省情况下，保持断言状态的时间为 180 秒
配置 IPv6 组播源生存时间	**source-lifetime** *interval*	可选 缺省情况下，IPv6 组播源的生存时间为 210 秒

2. 在接口上配置 IPv6 PIM 公共定时器

表1-28 在接口上配置 IPv6 PIM 公共定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入接口视图	**interface** *interface-type interface-number*	-
配置发送 Hello 报文的时间间隔	**pim ipv6 timer hello** *interval*	可选 缺省情况下，发送 Hello 报文的时间间隔为 30 秒
配置触发 Hello 报文的最大延迟时间	**pim ipv6 triggered-hello-delay** *interval*	可选 缺省情况下，触发 Hello 报文的最大延迟时间为 5 秒
配置发送加入/剪枝报文的时间间隔	**pim ipv6 timer join-prune** *interval*	可选 缺省情况下，发送加入/剪枝报文的时间间隔为 60 秒
配置保持加入/剪枝状态的时间	**pim ipv6 holdtime join-prune** *interval*	可选 缺省情况下，保持加入/剪枝状态的时间为 210 秒
配置保持断言状态的时间	**pim ipv6 holdtime assert** *interval*	可选 缺省情况下，保持断言状态的时间为 180 秒

说明

如果网络没有特殊要求，建议采用缺省值。

1.5.7 配置加入/剪枝报文规格

如果加入/剪枝报文的尺寸较大，则丢失一个报文将导致较多信息的遗失；如果加入/剪枝报文的尺寸较小，则单个报文的丢失所产生的影响也将降低。

通过控制加入/剪枝报文中（S，G）表项的数目，可以有效减少单位时间内发送的（S，G）表项数量。

表1-29 配置加入/剪枝报文规格

操作	命令	说明
进入系统视图	**system-view**	-
进入 IPv6 PIM 视图	**pim ipv6**	-
配置加入/剪枝报文的最大长度	**jp-pkt-size** *packet-size*	可选 缺省情况下，加入/剪枝报文的最大长度为 8100 字节
配置加入/剪枝报文中（S，G）表项的最大数量	**jp-queue-size** *queue-size*	可选 缺省情况下，加入/剪枝报文中（S，G）表项的最大数量为 1020 个

1.6 IPv6 PIM 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 PIM 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下执行 **reset** 命令可以清除 IPv6 PIM 统计信息。

表1-30 IPv6 PIM 显示和维护

操作	命令
查看 IPv6 PIM-SM 域中的 BSR 信息，以及本地配置并生效的 C-RP 信息	**display pim ipv6 bsr-info**
查看 IPv6 PIM 所使用的 IPv6 单播路由信息	**display pim ipv6 claimed-route** [*ipv6-source-address*]
查看 IPv6 PIM 控制报文的数量	**display pim ipv6 control-message counters** [**message-type** { **probe** | **register** | **register-stop** } | [**interface** *interface-type interface-number* | **message-type** { **assert** | **bsr** | **crp** | **graft** | **graft-ack** | **hello** | **join-prune** | **state-refresh** }] *]
查看尚未被确认的嫁接信息	**display pim ipv6 grafts**
查看接口上的 IPv6 PIM 信息	**display pim ipv6 interface** [*interface-type interface-number*] [**verbose**]
查看待发送的加入/剪枝报文信息	**display pim ipv6 join-prune mode** { **sm** [**flags** *flag-value*] | **ssm** } [**interface** *interface-type interface-number* | **neighbor** *ipv6-neighbor-address*] * [**verbose**]
查看 IPv6 PIM 邻居信息	**display pim ipv6 neighbor** [**interface** *interface-type interface-number* | *ipv6-neighbor-address* | **verbose**] *

操作	命令
查看 IPv6 PIM 路由表的内容	**display pim ipv6 routing-table** [*ipv6-group-address* [*prefix-length*] | *ipv6-source-address* [*prefix-length*] | **incoming-interface** [*interface-type interface-number* | **register**] | **outgoing-interface** { **include** | **exclude** | **match** } { *interface-type interface-number* | **register** } | **mode** *mode-type* | **flags** *flag-value* | **fsm**] *
查看 RP 的信息	**display pim ipv6 rp-info** [*ipv6-group-address*]
重置 IPv6 PIM 控制报文计数器	**reset pim ipv6 control-message counters** [**interface** *interface-type interface-number*]

1.7 IPv6 PIM 典型配置举例

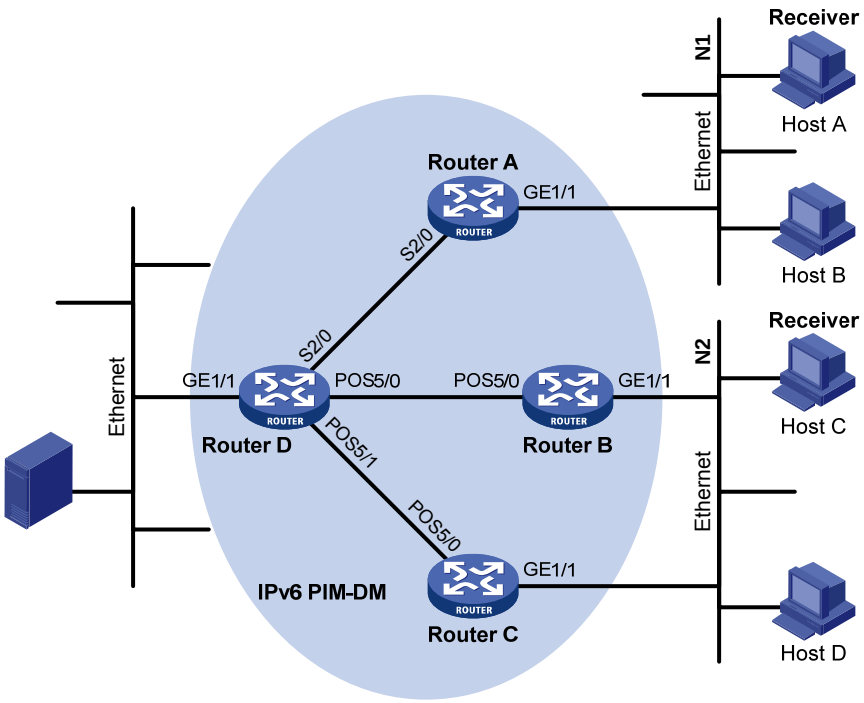
1.7.1 IPv6 PIM-DM 典型配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 IPv6 PIM 域采用 DM 方式。
- Host A 和 Host C 为两个末梢网络中的 IPv6 组播信息接收者；Router D 通过 GigabitEthernet1/1 接口与 IPv6 组播源（Source）所在的网络连接；Router A 通过 GigabitEthernet1/1 接口连接末梢网络 N1，通过 Serial2/0 接口连接 Router D；Router B 和 Router C 分别通过各自的 GigabitEthernet1/1 接口连接末梢网络 N2，并分别通过各自的 POS5/0 接口连接 Router D。
- Router A 与末梢网络 N1 之间运行 MLDv1；Router B 和 Router C 与末梢网络 N2 之间也运行 MLDv1。

2. 组网图

图1-8 IPv6 PIM-DM 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE1/1	1001::1/64	Router D	GE1/1	4001::1/64
	S2/0	1002::1/64		S2/0	1002::2/64
Router B	GE1/1	2001::1/64		POS5/0	2002::2/64
	POS5/0	2002::1/64		POS5/1	3001::2/64
Router C	GE1/1	2001::2/64			
	POS5/0	3001::1/64			

3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址和 IPv6 单播路由协议

使能各路由器的IPv6 转发功能，并按照 图 1-8配置各接口的IPv6 地址和前缀长度，具体配置过程略。

配置 IPv6 PIM-DM 域内的各路由器之间采用 OSPFv3 协议进行互连，确保 IPv6 PIM-DM 域内部在网络层互通，并且各路由器之间能够借助 IPv6 单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-DM 和 MLD

在 Router A 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-DM，并在其连接末梢网络的接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mld enable
[RouterA-GigabitEthernet1/1] pim ipv6 dm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface serial 2/0
[RouterA-Serial2/0] pim ipv6 dm
[RouterA-Serial2/0] quit
```

Router B 和 Router C 的配置与 Router A 相似，配置过程略。

在 Router D 上使能 IPv6 组播路由，并在各接口上使能 IPv6 PIM-DM。

```
<RouterD> system-view
[RouterD] multicast ipv6 routing-enable
[RouterD] interface GigabitEthernet 1/1
[RouterD-GigaitEthernet1/1] pim ipv6 dm
[RouterD-GigabitEthernet1/1] quit
[RouterD] interface serial 2/0
[RouterD-Serial2/0] pim ipv6 dm
[RouterD-Serial2/0] quit
[RouterD] interface pos 5/0
[RouterD-Pos5/0] pim ipv6 dm
[RouterD-Pos5/0] quit
[RouterD] interface pos 5/1
[RouterD-Pos5/1] pim ipv6 dm
[RouterD-Pos5/1] quit
```

(3) 检验配置效果

通过使用 **display pim ipv6 interface** 命令可以查看路由器接口上 IPv6 PIM 的配置和运行情况。例如：

查看 Router D 上 IPv6 PIM 的配置信息。

```
[RouterD] display pim ipv6 interface
 Interface           NbrCnt HelloInt   DR-Pri    DR-Address
 GE1/1               0      30         1         4001::1
                                                  (local)
 Ser2/0              0      30         1         1002::2
                                                  (local)
 Pos5/0              1      30         1         2002::2
                                                  (local)
 Pos5/1              1      30         1         3001::2
                                                  (local)
```

通过使用 **display pim ipv6 neighbor** 命令可以查看路由器之间的 IPv6 PIM 邻居关系。例如：

查看 Router D 上 IPv6 PIM 的邻居关系信息。

```
[RouterD] display pim ipv6 neighbor
 Total Number of Neighbors = 3

 Neighbor        Interface           Uptime   Expires  Dr-Priority
 1002::1         Ser2/0              00:04:00 00:01:29 1
 2002::1         Pos5/0              00:04:16 00:01:29 3
 3001::1         Pos5/1              00:03:54 00:01:17 5
```

假如 Host A 需要接收 IPv6 组播组 G（FF0E::101）的信息，当 IPv6 组播源 S（4001::100/64）向 IPv6 组播组 G 发送 IPv6 组播数据时，通过扩散生成 SPT，SPT 路径中各路由器（Router A 和 Router D）上都存在（S，G）表项，Host A 向 Router A 发送 MLD 报告以加入 IPv6 组播组 G，在 Router A 上生成（*，G）表项。通过使用 **display pim ipv6 routing-table** 命令可以查看路由器的 IPv6 PIM 路由表信息。例如：

查看 Router A 上的 IPv6 PIM 路由表信息。

```
[RouterA] display pim ipv6 routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::101)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:01:24
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet1/1
             Protocol: mld, UpTime: 00:01:20, Expires: never

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:01:20
     Upstream interface: Serial2/0
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet1/1
             Protocol: pim-dm, UpTime: 00:01:20, Expires: never
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 IPv6 PIM 路由表信息。

```
[RouterD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: LOC ACT
     UpTime: 00:02:19
     Upstream interface: GigabitEthernet1/1
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 3
         1: Serial2/0
             Protocol: pim-dm, UpTime: 00:02:19, Expires: never
         2: Pos5/0
             Protocol: pim-dm, UpTime: 00:02:19, Expires: never
         3: Pos5/1
             Protocol: pim-dm, UpTime: 00:02:19, Expires: never
```

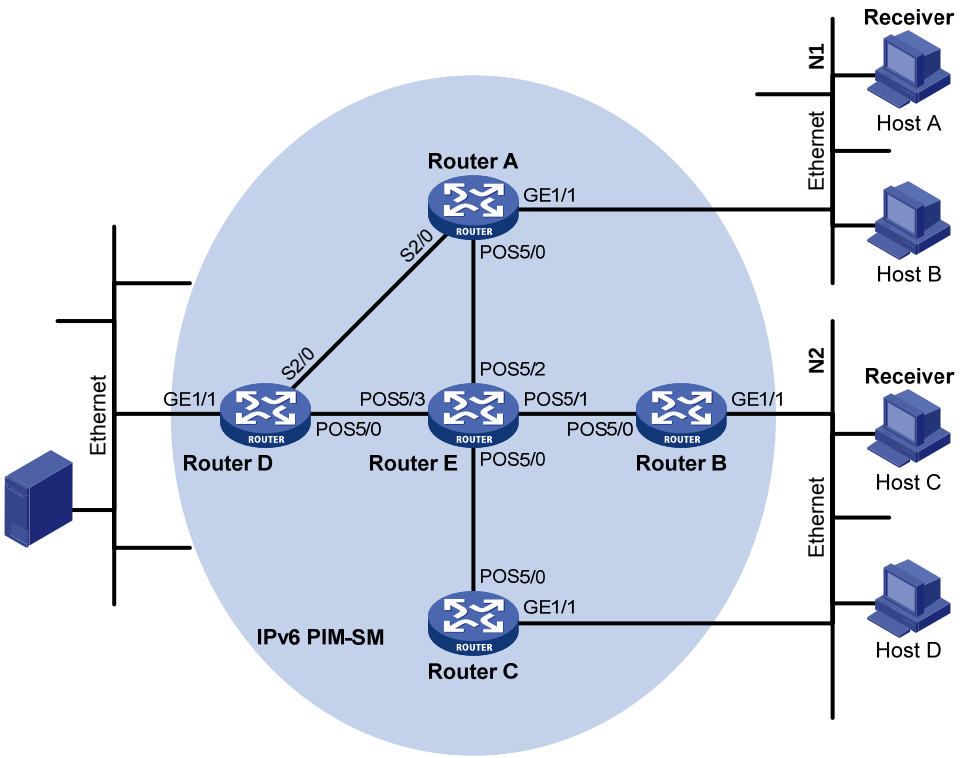
1.7.2 IPv6 PIM-SM 典型配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 IPv6 PIM 域采用 SM 方式。
- Host A 和 Host C 为两个末梢网络中的 IPv6 组播信息接收者；Router D 通过 GigabitEthernet1/1 接口与 IPv6 组播源（Source）所在网络连接；Router A 通过 GigabitEthernet1/1 接口连接末梢网络 N1，通过 Serial2/0 接口和 POS5/0 接口分别连接 Router D 和 Router E；Router B 和 Router C 分别通过各自的 GigabitEthernet1/1 接口连接末梢网络 N2，并分别通过各自的 POS5/0 接口连接 Router E；
- 将 Router D 的 POS5/0 接口和 Router E 的 POS5/2 接口都配置为 C-BSR 和 C-RP，其中 Router E 上 C-BSR 的优先级较高；C-RP 所服务的 IPv6 组播组范围为 FF0E::101/64，通过改变哈希掩码长度使此范围内的 IPv6 组地址间隔映射到这两个 C-RP 上。
- Router A 与末梢网络 N1 之间运行 MLDv1；Router B 和 Router C 与末梢网络 N2 之间也运行 MLDv1。

2. 组网图

图1-9 IPv6 PIM-SM 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE1/1	1001::1/64	Router D	GE1/1	4001::1/64
	S2/0	1002::1/64		S2/0	1002::2/64
	POS5/0	1003::1/64		POS5/0	4002::1/64
Router B	GE1/1	2001::1/64	Router E	POS5/0	3001::2/64
	POS5/0	2002::1/64		POS5/1	2002::2/64
Router C	GE1/1	2001::2/64		POS5/2	1003::2/64
	POS5/0	3001::1/64		POS5/3	4002::2/64

3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址和 IPv6 单播路由协议

使能各路由器的IPv6 转发功能，并按照 图 1-9配置各接口的IPv6 地址和前缀长度，具体配置过程略。

配置 IPv6 PIM-SM 域内的各路由器之间采用 OSPFv3 协议进行互连，确保 IPv6 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助 IPv6 单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-SM 和 MLD

在 Router A 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-SM，并在其连接末梢网络的接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mld enable
[RouterA-GigabitEthernet1/1] pim ipv6 sm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface Serial 2/0
[RouterA-Serial2/0] pim ipv6 sm
[RouterA-Serial2/0] quit
[RouterA] interface Pos 5/0
[RouterA-Pos5/0] pim ipv6 sm
[RouterA-Pos5/0] quit
```

Router B 和 Router C 的配置与 Router A 相似，Router D 和 Router E 除了不需要在相应接口上使能 MLD 外，其它的配置也与 Router A 相似，配置过程略。

(3) 配置 C-BSR 和 C-RP

在 Router D 上配置 RP 通告的服务范围，以及 C-BSR 和 C-RP 的位置，并指定哈希掩码长度为 128，C-BSR 的优先级为 10。

```
<RouterD> system-view
[RouterD] acl ipv6 number 2005
[RouterD-acl6-basic-2005] rule permit source ff0e::101 64
[RouterD-acl6-basic-2005] quit
[RouterD] pim ipv6
[RouterD-pim6] c-bsr 4002::1 128 10
[RouterD-pim6] c-rp 4002::1 group-policy 2005
[RouterD-pim6] quit
```

在 Router E 上配置 RP 通告的服务范围，以及 C-BSR 和 C-RP 的位置，并指定哈希掩码长度为 128，C-BSR 的优先级为 20。

```
<RouterE> system-view
[RouterE] acl ipv6 number 2005
[RouterE-acl6-basic-2005] rule permit source ff0e::101 64
[RouterE-acl6-basic-2005] quit
[RouterE] pim ipv6
[RouterE-pim6] c-bsr 1003::2 128 20
[RouterE-pim6] c-rp 1003::2 group-policy 2005
[RouterE-pim6] quit
```

(4) 检验配置效果

通过使用 **display pim ipv6 interface** 命令可以查看路由器接口上 IPv6 PIM 的配置和运行情况。例如：

查看 Router A 上 IPv6 PIM 的配置信息。

```
[RouterA] display pim ipv6 interface
 Interface            NbrCnt HelloInt   DR-Pri    DR-Address
 GE1/1                0      30         1         1001::1
                                                   (local)
 Ser2/0               1      30         1         1002::2
 Pos5/0               1      30         1         1003::2
```

通过使用 **display pim ipv6 bsr-info** 命令可以查看路由器上 BSR 选举的信息，以及本地配置并生效的 C-RP 信息。例如：

查看 Router A 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterA] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Accept Preferred
     Uptime: 00:04:22
     Expires: 00:01:46
```

查看 Router D 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterD] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Elected
     Uptime: 00:05:26
     Expires: 00:01:45
 Candidate BSR Address: 4002::1
     Priority: 10
     Hash mask length: 128
     State: Candidate

 Candidate RP: 4002::1(Pos5/0)
     Priority: 0
     HoldTime: 130
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:48
```

查看 Router E 上的 BSR 信息，以及本地配置并生效的 C-RP 信息。

```
[RouterE] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Elected
     Uptime: 00:01:10
```

```
    Next BSR message scheduled at: 00:01:48
 Candidate BSR Address: 1003::2
    Priority: 20
    Hash mask length: 128
    State: Elected

 Candidate RP: 1003::2(Pos5/2)
    Priority: 0
    HoldTime: 130
    Advertisement Interval: 60
    Next advertisement scheduled at: 00:00:48
```

通过使用 **display pim ipv6 rp-info** 命令可以查看路由器上获取的 RP 信息。例如：

查看 Router A 上的 RP 信息。

```
[RouterA] display pim ipv6 rp-info
 PIM-SM BSR RP information:
 prefix/prefix length: FF0E::101/64
    RP: 4002::1
    Priority: 0
    HoldTime: 130
    Uptime: 00:05:19
    Expires: 00:02:11

    RP: 1003::2
    Priority: 0
    HoldTime: 130
    Uptime: 00:05:19
    Expires: 00:02:11
```

假如 Host A 需要接收 IPv6 组播组 G（FF0E::100）的信息，由于根据哈希算法得出 G 对应的 RP 为 Router E，因此 Router A 和 Router E 之间会生成 RPT。当 IPv6 组播源 S（4001::100/64）向 RP 发起注册后，Router D 和 Router E 之间会生成 SPT。当 Router A 收到 IPv6 组播数据后立即执行从 RPT 到 SPT 的切换。RPT 路径中的路由器（Router A 和 Router E）上存在（*，G）表项，而 SPT 路径中的路由器（Router A 和 Router D）上存在（S，G）表项，通过使用 **display pim ipv6 routing-table** 命令可以查看路由器的 IPv6 PIM 路由表信息。例如：

查看 Router A 上的 IPv6 PIM 路由表信息。

```
[RouterA] display pim ipv6 routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::100)
    RP: 1003::2
    Protocol: pim-sm, Flag: WC
    UpTime: 00:03:45
    Upstream interface: Pos5/0
        Upstream neighbor: 1003::2
        RPF prime neighbor: 1003::2
    Downstream interface(s) information:
    Total number of downstreams: 1
```

```
        1: GigabitEthernet1/1
            Protocol: mld, UpTime: 00:02:15, Expires: 00:03:06

 (4001::100, FF0E::100)
     RP: 1003::2
     Protocol: pim-sm, Flag: SPT ACT
     UpTime: 00:02:15
     Upstream interface: Serial2/0
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet1/1
             Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:06
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 IPv6 PIM 路由表信息。

```
[RouterD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF0E::100)
     RP: 1003::2
     Protocol: pim-sm, Flag: SPT LOC ACT
     UpTime: 00:14:44
     Upstream interface: GigabitEthernet1/1
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Pos5/0
             Protocol: mld, UpTime: 00:14:44, Expires: 00:02:26
```

查看 Router E 上的 IPv6 PIM 路由表信息。

```
[RouterE] display pim ipv6 routing-table
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, FF0E::100)
     RP: 1003::2 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:16:56
     Upstream interface: Register
         Upstream neighbor: 4002::1
         RPF prime neighbor: 4002::1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Pos5/2
             Protocol: pim-sm, UpTime: 00:16:56, Expires: 00:02:34
```

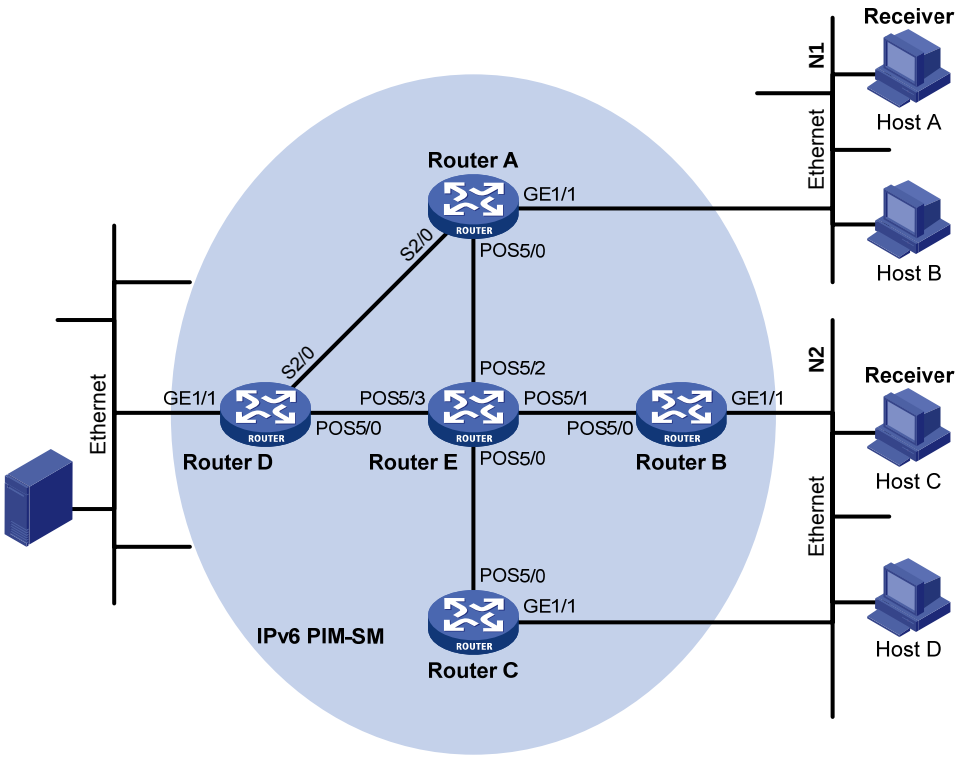
1.7.3 IPv6 PIM-SSM 典型配置举例

1. 组网需求

- 接收者通过组播方式接收视频点播信息，不同组织的接收者群体组成末梢网络，每个末梢网络中都存在至少一个接收者，整个 IPv6 PIM 域采用 SSM 方式。
- Host A 和 Host C 为两个末梢网络中的 IPv6 组播信息接收者；Router D 通过 GigabitEthernet1/1 接口与 IPv6 组播源（Source）所在网络连接；Router A 通过 GigabitEthernet1/1 接口连接末梢网络 N1，通过 Serial2/0 接口和 POS5/0 接口分别连接 Router D 和 Router E；Router B 和 Router C 分别通过各自的 GigabitEthernet1/1 接口连接末梢网络 N2，并分别通过各自的 POS5/0 接口连接 Router E；Router E 连接 Router A、Router B、Router C 和 Router D。
- SSM 组播组的范围是 FF3E::/64。
- Router A 与末梢网络 N1 之间运行 MLDv2；Router B 和 Router C 与末梢网络 N2 之间也运行 MLDv2。

2. 组网图

图1-10 IPv6 PIM-SSM 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Router A	GE1/1	1001::1/64	Router D	GE1/1	4001::1/64
	S2/0	1002::1/64		S2/0	1002::2/64
	POS5/0	1003::1/64		POS5/0	4002::1/64
Router B	GE1/1	2001::1/64	Router E	POS5/0	3001::2/64
	POS5/0	2002::1/64		POS5/1	2002::2/64
Router C	GE1/1	2001::2/64		POS5/2	1003::2/64
	POS5/0	3001::1/64		POS5/3	4002::2/64

3. 配置步骤

(1) 使能 IPv6 转发功能，并配置 IPv6 地址和 IPv6 单播路由协议

使能各路由器的IPv6 转发功能，并按照 图 1-10配置各接口的IPv6 地址和前缀长度，具体配置过程略。

配置 IPv6 PIM-SM 域内的各路由器之间采用 OSPFv3 协议进行互连，确保 IPv6 PIM-SM 域内部在网络层互通，并且各路由器之间能够借助 IPv6 单播路由协议实现动态路由更新，具体配置过程略。

(2) 使能 IPv6 组播路由，并使能 IPv6 PIM-SM 和 MLD

在 Router A 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-SM，并在其连接末梢网络的接口 GigabitEthernet1/1 上使能 MLD，且配置其版本为 2。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/1
[RouterA-GigabitEthernet1/1] mld enable
[RouterA-GigabitEthernet1/1] mld version 2
[RouterA-GigabitEthernet1/1] pim ipv6 sm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface Serial 2/0
[RouterA-Serial2/0] pim ipv6 sm
[RouterA-Serial2/0] quit
[RouterA] interface Pos 5/0
[RouterA-Pos5/0] pim ipv6 sm
[RouterA-Pos5/0] quit
```

Router B 和 Router C 的配置与 Router A 相似，Router D 和 Router E 除了不需要在相应接口上使能 MLD 外，其它的配置也与 Router A 相似，配置过程略。

(3) 配置 IPv6 SSM 组播组的地址范围

在 Router A 上配置 IPv6 SSM 组播组的地址范围为 FF3E::/64。

```
[RouterA] acl ipv6 number 2000
[RouterA-acl6-basic-2000] rule permit source ff3e:: 64
[RouterA-acl6-basic-2000] quit
[RouterA] pim ipv6
[RouterA-pim6] ssm-policy 2000
[RouterA-pim6] quit
```

Router B、Router C、Router D 和 Router E 的配置与 Router A 相似，配置过程略。

(4) 检验配置效果

通过使用 **display pim ipv6 interface** 命令可以查看路由器接口上 IPv6 PIM 的配置和运行情况。例如：

查看 Router A 上 IPv6 PIM 的配置信息。

```
[RouterA] display pim ipv6 interface
 Interface             NbrCnt HelloInt   DR-Pri   DR-Address
 GE1/1                 0      30         1        1001::1
                                                   (local)
 Ser2/0                1      30         1        1002::2
```

```
 Pos5/0                1      30         1        1003::2
```

假如 Host A 需要接收指定 IPv6 组播源 S（4001::100/64）发往 IPv6 组播组 G（FF3E::101）的信息，Router A 会向 IPv6 组播源方向构造 SPT，SPT 路径中的路由器（Router A 和 Router D）上生成（S，G）表项，而 SPT 路径之外的路由器（Router E）上没有 IPv6 组播路由项，通过使用 **display pim ipv6 routing-table** 命令可以查看路由器的 IPv6 PIM 路由表信息。例如：

查看 Router A 上的 IPv6 PIM 路由表信息。

```
[RouterA] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:00:11
     Upstream interface: Serial2/0
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: GigabitEthernet1/1
             Protocol: mld, UpTime: 00:00:11, Expires: 00:03:25
```

Router B 和 Router C 上的显示信息与 Router A 类似。

查看 Router D 上的 IPv6 PIM 路由表信息。

```
[RouterD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF3E::101)
     Protocol: pim-ssm, Flag: LOC
     UpTime: 00:08:02
     Upstream interface: GigabitEthernet1/1
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Serial2/0
             Protocol: pim-ssm, UpTime: 00:08:02, Expires: 00:03:25
```

目 录

1 IPv6 MBGP配置 1-1
1.1 IPv6 MBGP简介 1-1
1.2 IPv6 MBGP配置任务简介 1-1
1.3 配置IPv6 MBGP的基本功能 1-2
1.3.1 配置准备 1-2
1.3.2 配置IPv6 MBGP对等体 1-2
1.3.3 配置路由首选值 1-2
1.4 控制路由信息的发布与接收 1-3
1.4.1 配置准备 1-3
1.4.2 配置IPv6 MBGP发布本地IPv6 路由 1-3
1.4.3 配置IPv6 MBGP引入其他路由 1-3
1.4.4 配置IPv6 MBGP路由聚合 1-4
1.4.5 配置向IPv6 MBGP对等体/对等体组发送缺省路由 1-4
1.4.6 配置IPv6 MBGP路由信息的发布策略 1-5
1.4.7 配置IPv6 MBGP路由信息的接收策略 1-5
1.4.8 配置IPv6 MBGP路由衰减 1-6
1.5 配置IPv6 MBGP的路由属性 1-6
1.5.1 配置准备 1-7
1.5.2 配置IPv6 MBGP路由管理优先级 1-7
1.5.3 配置本地优先级缺省值 1-7
1.5.4 配置MED属性 1-7
1.5.5 配置发布路由时将自身地址作为下一跳 1-8
1.5.6 配置AS_PATH属性 1-8
1.6 调整和优化IPv6 MBGP网络 1-9
1.6.1 配置准备 1-9
1.6.2 配置IPv6 MBGP软复位 1-9
1.6.3 配置最大等价路由的条数 1-10
1.7 配置大型IPv6 MBGP网络 1-10
1.7.1 配置准备 1-10
1.7.2 配置IPv6 MBGP对等体组 1-10
1.7.3 配置IPv6 MBGP团体 1-11
1.7.4 配置IPv6 MBGP路由反射器 1-12
1.8 IPv6 MBGP显示和维护 1-12
1.8.1 IPv6 MBGP显示 1-12
1.8.2 复位IPv6 MBGP连接 1-13
1.8.3 清除IPv6 MBGP信息 1-13
1.9 IPv6 MBGP典型配置举例（路由应用） 1-14

1 IPv6 MBGP 配置

说明

本章只列出了 IPv6 MBGP 专有的配置与操作，其他相关内容请参见“IP 路由分册”中的“IPv6 BGP 配置”。

1.1 IPv6 MBGP 简介

为了提供对多种网络层协议的支持，IETF（Internet Engineering Task Force，因特网工程任务组）对 BGP-4 进行了扩展，形成 MP-BGP（Multi-Protocol BGP，多协议 BGP），使 BGP 能够为多种路由应用提供路由信息。

由于 IPv6 组播拓扑和 IPv6 单播拓扑有可能不同，需要通过 MP-BGP 扩展使得 BGP 能够将 IPv6 单播 NLRI（Network Layer Reachability Information，网络层可达性信息）和 IPv6 组播 NLRI 分开运载，其中组播 NLRI 专用于执行 RPF（Reverse Path Forwarding，逆向路径转发）功能。这使得在 IPv6 单播路由表和组播路由表中，对于相同的前缀有不同的路径选择，维护了一致的单播转发并使域间组播正常工作。

目前的 MP-BGP 标准是 RFC 2858（Multiprotocol Extensions for BGP-4，BGP-4 的多协议扩展）。MP-BGP 在 IPv6 组播上的应用简称为 IPv6 MBGP（IPv6 Multicast BGP，IPv6 组播 BGP）。

说明

- 本章主要介绍 MP-BGP 应用于 IPv6 组播的配置事项，即 IPv6 MBGP 配置事项，有关 BGP 的详细内容以及配置事项，请参见“IP 路由分册”中的“BGP 配置”。
- 有关 RPF 检查的详细内容以及配置事项，请参见“IP 组播分册”中的“组播路由与转发配置”。

1.2 IPv6 MBGP 配置任务简介

表1-1 IPv6 MBGP 配置任务简介

配置任务		说明	详细配置
配置 IPv6 MBGP 基本功能	配置 IPv6 MBGP 对等体	必选	1.3.2
	配置路由首选值	可选	1.3.3
控制路由信息的发布与接收	配置 IPv6 MBGP 发布本地 IPv6 路由	可选	1.4.2
	配置 IPv6 MBGP 引入其他路由	可选	1.4.3
	配置 IPv6 MBGP 路由聚合	可选	1.4.4
	配置向 MBGP 对等体/对等体组发送缺省路由	可选	1.4.5
	配置 IPv6 MBGP 路由信息的发布策略	可选	1.4.6
	配置 IPv6 MBGP 路由信息的接收策略	可选	1.4.7
	配置 IPv6 MBGP 路由衰减	可选	1.4.8

配置任务		说明	详细配置
配置 IPv6 MBGP 的路由属性	配置 IPv6 MBGP 路由管理的优先级、缺省 LOCAL_PREF 及 NEXT_HOP 属性	可选	1.5.2
	配置 MED 属性	可选	1.5.3
	配置 AS_PATH 属性	可选	1.5.5
调整和优化 IPv6 MBGP 网络	配置 IPv6 MBGP 软复位	可选	1.6.2
	配置最大等价路由的条数	可选	1.6.3
组建大型 IPv6 MBGP 网络	配置 IPv6 MBGP 对等体组	可选	1.7.2
	配置 IPv6 MBGP 团体	可选	1.7.3
	配置 IPv6 MBGP 路由反射器	可选	1.7.4

1.3 配置 IPv6 MBGP 的基本功能

1.3.1 配置准备

由于 IPv6 MBGP 是 MP-BGP 的一种应用，因此在配置 IPv6 MBGP 之前，需完成以下任务：

- 使能 IPv6 功能
- 配置接口的网络层地址
- 配置 BGP 基本功能

1.3.2 配置 IPv6 MBGP 对等体

表1-2 配置 IPv6 MBGP 对等体

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	必选 缺省情况下，系统没有运行 BGP
进入 IPv6 地址族视图	**ipv6-family**	-
配置对等体的 AS 号	**peer** *ipv6-address* **as-number** *as-number*	必选 缺省情况下，对等体无 AS 号
创建并进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	必选
激活指定对等体	**peer** *ipv6-address* **enable**	必选 缺省情况下，对等体在 BGP IPv6 组播地址族视图下处于非激活状态

1.3.3 配置路由首选值

表1-3 配置路由首选值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-

操作	命令	说明
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
为从 IPv6 MBGP 对等体/对等体组接收的路由分配首选值	**peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value*	可选 缺省情况下，从 IPv6 MBGP 对等体/对等体组接收的路由的首选值为 0

注意

- 如果同时通过路由策略和 **peer preferred-value** 命令为从 IPv6 MBGP 对等体/对等体组接收的 IPv6 MBGP 路由设置首选值，通过路由策略过滤的 IPv6 MBGP 路由的首选值将取路由策略中设置的首选值而不是 **peer preferred-value** 命令设置的首选值；
- 只有当路由策略里设置的首选值为 0 时，IPv6 MBGP 路由的首选值才会取 **peer preferred-value** 命令设置的值；
- 没有通过路由策略过滤的 IPv6 组播路由的首选值仍取 **peer preferred-value** 命令设置的首选值。
- 通过路由策略配置 BGP IPv6 组播路由信息首选值的相关配置可参考命令 **peer route-policy** 和“IP 路由分册/路由策略命令”中的 **apply preferred-value**。

1.4 控制路由信息的发布与接收

控制 IPv6 MBGP 的路由信息的发布和接收，包括对路由信息进行过滤、应用路由策略和路由衰减。

1.4.1 配置准备

在控制路由信息的发布与接收之前，需完成以下任务：配置 IPv6 MBGP 基本功能。

1.4.2 配置 IPv6 MBGP 发布本地 IPv6 路由

表1-4 配置 IPv6 MBGP 发布本地 IPv6 路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
将路由通告到 IPv6 MBGP 路由表中	**network** *ipv6-address prefix-length* [**route-policy** *route-policy-name* | **short-cut**]	必选 缺省情况下，IPv6 MBGP 不通告任何路由

1.4.3 配置 IPv6 MBGP 引入其他路由

表1-5 配置 IPv6 MBGP 引入其他路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-

操作	命令	说明
允许将缺省路由引入到 IPv6 MBGP 路由表中	**default-route imported**	可选 缺省情况下，IPv6 MBGP 不允许将缺省路由引入到 IPv6 MBGP 路由表中
配置引入其它协议路由信息并通告	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	必选 缺省情况下，IPv6 MBGP 不引入且不通告其它协议的路由



说明

通过 **import-route** 命令引入 IGP 路由时，不能引入 IGP 的缺省路由，必须通过配置 **default-route imported** 命令将缺省路由引入到 IPv6 MBGP 路由表中。

1.4.4 配置 IPv6 MBGP 路由聚合

在中型或大型 BGP 网络中，在向对等体发布 IPv6 组播路由信息时，可以配置路由聚合，减小对等体路由表中的路由数量。BGP 仅支持 IPv6 组播路由的手动聚合。

表1-6 配置路由聚合

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置手动路由聚合	**aggregate** *ipv6-address prefix-length* [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	必选 缺省情况下，不进行路由聚合

1.4.5 配置向 IPv6 MBGP 对等体/对等体组发送缺省路由

表1-7 配置向 IPv6 MBGP 对等体/对等体组发送缺省路由

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置向 IPv6 MBGP 对等体/对等体组发送缺省路由	**peer** { *ipv6-group-name* | *ipv6-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	必选 缺省情况下，不向 IPv6 MBGP 对等体/对等体组发布缺省路由

说明

执行 **peer default-route-advertise** 命令后，不论本地路由表中是否存在缺省路由，都将向指定 IPv6 MBGP 对等体/对等体组发布一条下一跳地址为本地地址的缺省路由。

1.4.6 配置 IPv6 MBGP 路由信息的发布策略

表1-8 配置 IPv6 MBGP 路由信息的发布策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
对发布的路由进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol process-id*]	任选其一 缺省情况下，不对引入的路由在发布时进行过滤 可以根据需求选择过滤策略，同时配置几种过滤策略时，按照如下顺序执行： • **filter-policy export** • **peer filter-policy export** • **peer as-path-acl export** • **peer ipv6-prefix export** • **peer route-policy export** 只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被发布
配置基于 IPv6 ACL 的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **export**	
配置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **export**	
配置基于 IPv6 地址前缀列表的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **export**	
配置出方向路由策略	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	

说明

- IPv6 MBGP 对等体组的成员必须与所在的组使用相同的出方向路由更新策略，即对外发布路由时，一个对等体组遵循的策略是相同的。
- IPv6 MBGP 对引入的路由信息进行过滤后，会将符合条件的路由信息发布给 IPv6 MBGP 对等体。

1.4.7 配置 IPv6 MBGP 路由信息的接收策略

对 IPv6 MBGP 接收的路由进行过滤，只有满足某些条件的路由才能被 IPv6 MBGP 接收，并加到路由表中。

表1-9 配置 IPv6 MBGP 路由信息的接收策略

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-

操作	命令	说明
对接收的路由信息进行过滤	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	任选其一 缺省情况下，不对接收的路由信息进行过滤 可以根据需求选择过滤策略，同时配置几种过滤策略时，按照如下顺序执行： • **filter-policy import** • **peer filter-policy import** • **peer as-path-acl import** • **peer ip-prefix import** **peer route-policy import** 只有通过前面的过滤策略，才能继续执行后面的过滤策略；只有通过所有配置的过滤策略后，路由信息才能被接收
对接收的路由信息应用路由策略	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **import**	
配置基于 IPv6 ACL 的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **import**	
配置基于 AS 路径过滤列表的 IPv6 BGP 路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **import**	
配置基于 IPv6 地址前缀列表的路由过滤策略	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **import**	
配置允许从对等体/对等体组接收的最大 IPv6 地址前缀数	**peer** { *ipv6-group-name* | *ipv6-address* } **route-limit** *limit* [*percentage*]	可选 缺省情况下，允许从组播对等体/对等体组接收的最大路由数无限制

说明

IPv6 MBGP 对等体组的成员可以与所在的组使用不同的入方向路由策略，即接收路由时，各对等体可以选择自己的策略。

1.4.8 配置 IPv6 MBGP 路由衰减

表1-10 配置 IPv6 MBGP 路由衰减

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置 IPv6 MBGP 路由衰减	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	可选 缺省情况下，没有配置 IPv6 MBGP 路由衰减

1.5 配置 IPv6 MBGP 的路由属性

本节主要介绍使用各类路由属性来改变 IPv6 MBGP 的选路策略。包括如下属性：

- IPv6 MBGP 路由管理优先级
- 缺省 LOCAL_PREF 属性值
- MED 属性
- NEXT_HOP 属性
- AS_PATH 属性

1.5.1 配置准备

在配置 IPv6 MBGP 的路由属性之前，需完成以下任务：配置 IPv6 MBGP 基本功能。

1.5.2 配置 IPv6 MBGP 路由管理优先级

表1-11 配置 IPv6 MBGP 路由管理优先级

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置 IPv6 MBGP 路由管理优先级	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	可选 缺省情况下，EBGP 路由的管理优先级为 255，IBGP 路由的管理优先级为 255，本地产生的 IPv6 BGP 路由的管理优先级为 130

1.5.3 配置本地优先级缺省值

表1-12 配置本地优先级缺省值

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置本地优先级缺省值	**default local-preference** *value*	可选 缺省情况下，本地优先级缺省值为 100

1.5.4 配置 MED 属性

表1-13 配置 MED 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置系统的缺省 MED 值	**default med** *med-value*	可选 缺省情况下，系统缺省的 MED 值为 0
配置允许比较来自不同 AS 邻居的路由的 MED 属性值	**compare-different-as-med**	可选 缺省情况下，不允许比较来自不同 AS 邻居的路由的 MED 属性值
配置根据路由来自的 AS 进行分组对 MED 排序优选	**bestroute compare-med**	可选 缺省情况下，不根据路由来自的 AS 进行分组对 MED 排序优选

操作	命令	说明
配置允许比较联盟对等体的路由按 MED 值进行优选	**bestroute med-confederation**	可选 缺省情况下，比较联盟对等体的路由时不考虑 MED 值

1.5.5 配置发布路由时将自身地址作为下一跳

在某些组网环境中，为保证 IPv6 组播 IBGP 邻居能够找到正确的下一跳，可以配置在向 IPv6 组播 IBGP 对等体/对等体组发布路由时，改变下一跳地址为自身地址。如果配置了 IPv6 MBGP 负载分担，则不论是否配置了 **peer next-hop-local** 命令，本地路由器向 IPv6 组播 IBGP 对等体/对等体组发布路由时都先将下一跳地址改变为自身地址。

在第三方下一跳（即两个 IPv6 MBGP 连接在同一网段的广播网）这种特殊的组网环境中，缺省情况下，向 IPv6 组播 EBGP 对等体/对等体组发布路由时，不将自身地址作为下一跳。

表1-14 配置发布路由时将自身地址做为下一跳

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置发布路由时将自身地址作为下一跳	**peer** { *ipv6-group-name* | *ipv6-address* } **next-hop-local**	可选 缺省情况下，向 EBGP 对等体/对等体组发布路由时，将自身地址作为下一跳；向 IBGP 对等体/对等体发布路由时，不将自身地址作为下一跳

1.5.6 配置 AS_PATH 属性

表1-15 配置 AS_PATH 属性

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置允许本地 AS 号在所接收的路由的 AS_PATH 属性中出现，并可同时配置允许重复的次数	**peer** { *ipv6-group-name* | *ipv6-address* } **allow-as-loop** [*number*]	可选 缺省情况下，不允许本地 AS 号重复
禁止路由器将 AS_PATH 当作选路算法中的一个因素	**bestroute as-path-neglect**	可选 缺省情况下，路由器可以将 AS_PATH 当作选路算法中的一个因素
配置发送 IPv6 MBGP 更新报文时不携带私有 AS 编号	**peer** { *ipv6-group-name* | *ipv6-address* } **public-as-only**	可选 缺省情况下，发送 IPv6 MBGP 更新报文时携带私有自治系统号

1.6 调整和优化 IPv6 MBGP 网络

1.6.1 配置准备

在调整和优化 IPv6 MBGP 网络之前，需完成以下任务：

- 使能 IPv6 能力
- 配置 IPv6 MBGP 基本功能

1.6.2 配置 IPv6 MBGP 软复位

IPv6 MBGP 的选路策略改变后，为了使新的策略生效，必须复位 IPv6 MBGP 连接，但这样会造成短暂的 IPv6 MBGP 连接中断。

通过使能 Route-refresh 功能，当策略改变后，系统可以在不中断 IPv6 MBGP 连接的情况下，自动对 IPv6 MBGP 路由表进行动态刷新。

如果对等体不支持 Route-Refresh 功能，则可以将从对等体接收的所有路由更新保存在本地，当选路策略发生改变后，在不中断连接的情况下重新刷新 IPv6 MBGP 路由表，并应用新的策略。

1. 通过 Route-Refresh 实现 IPv6 MBGP 软复位

在对等体支持并使能 Route-Refresh 功能的情况下，如果 IPv6 MBGP 的路由策略发生了变化，本地路由器会向 IPv6 MBGP 对等体发布 Route-Refresh 消息，收到此消息的对等体会将其路由信息重新发给本地路由器。这样，在不中断 IPv6 MBGP 连接的情况下，就可以对 IPv6 MBGP 路由表进行动态更新，并应用新的选路策略。

表1-16 通过 Route-Refresh 实现 IPv6 MBGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 IPv6 地址族视图	**ipv6-family**	-
使能 IPv6 BGP 路由刷新功能	**peer** { *ipv6-group-name* | *ipv6-address* } **capability-advertise route-refresh**	可选 缺省情况下，路由刷新功能处于使能状态

2. 通过将所有路由更新保存在本地实现 IPv6 MBGP 软复位

当对等体不支持 Route-Refresh 功能时，可通过配置 **peer keep-all-routes** 命令实现软复位功能。

用户也可以通过执行 **refresh bgp ipv6 multicast** 命令对保存在本地的所有路由重新过一遍策略。

表1-17 通过将所有路由更新保存在本地实现 IPv6 MBGP 软复位

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 IPv6 地址族视图	**ipv6-family**	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-

操作	命令	说明
保存所有来自对等体/对等体组的原始路由信息，即使这些路由没有通过已配置的入口策略	**peer** { *ipv6-group-name* | *ipv6-address* } **keep-all-routes**	必选 缺省情况下，不保存对等体/对等体组的原始路由信息
手工对 IPv6 MBGP 连接进行软复位	**refresh bgp ipv6 multicast** { **all** | *ipv6-address* | **group** *ipv6-group-name* | **external** | **internal** } { **export** | **import** }	可选

1.6.3 配置最大等价路由的条数

表1-18 配置最大等价路由的条数

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置进行 IPv6 MBGP 负载分担的路由条数	**balance** *number*	必选 缺省情况下，不进行负载分担

1.7 配置大型 IPv6 MBGP 网络

1.7.1 配置准备

在配置 IPv6 MBGP 对等体组之前，需完成以下任务：配置 IPv6 MBGP 基本功能。

1.7.2 配置 IPv6 MBGP 对等体组

为方便管理，减少重复配置操作，管理员将一些享有相同更新策略的 IPv6 BGP 对等体划分到一个逻辑组织，这些组织就称为对等体组。只须对一个对等体组配置策略，该策略对组内所有成员都有效。

表1-19 配置 IPv6 MBGP 对等体组

操作	命令	说明
进入系统视图	**system-view**	-
启动 BGP，进入 BGP 视图	**bgp** *as-number*	-
进入 IPv6 地址族视图	**ipv6-family**	-
创建 BGP 对等体组	**group** *ipv6-group-name* [**external** | **internal**]	必选
将对等体加入已存在的对等体组	**peer** *ipv6-address* **group** *ipv6-group-name* [**as-number** *as-number*]	必选 缺省情况下，对等体不属于任何对等体组
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
使能 IPv6 单播对等体组	**peer** *ipv6-group-name* **enable**	必选

操作	命令	说明
在已使能的对等体组中加入使能的IPv6 组播对等体	**peer** *ipv6-address* **group** *ipv6-group-name*	必选 缺省情况下，对等体不属于任何对等体组

注意

配置 IPv6 MBGP 对等体组：

- 需要先在 IPv6 地址族视图下将对等体加入对等体组；
- 然后在 BGP IPv6 组播地址族视图下使能 IPv6 MBGP 对等体组，并把 IPv6 MBGP 对等体加入 IPv6 MBGP 对等体组。

1.7.3 配置 IPv6 MBGP 团体

对等体组可以使一组对等体共享相同的策略，而利用团体可以使多个 AS 中的一组 IPv6 MBGP 路由器共享相同的策略。团体是一个路由属性，在 IPv6 MBGP 对等体之间传播，它并不受到 AS 范围的限制。

BGP 路由器在将带有团体属性的路由发布给其它对等体之前，可以改变此路由原有的团体属性。除了使用公认的团体属性外，用户还可以使用团体属性列表自定义扩展团体属性，以便更为灵活地控制路由策略。

表1-20 配置 IPv6 MBGP 团体

操作		命令	说明
进入系统视图		**system-view**	-
启动 BGP，进入 BGP 视图		**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图		**ipv6-family multicast**	-
配置向 IPv6 MBGP 对等体/IPv6 MBGP 对等体组发布团体属性	配置发布团体属性	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-community**	必选 缺省情况下，不将团体属性发布给任何对等体/对等体组
	配置发布扩展团体属性	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-ext-community**	必选 缺省情况下，不将扩展团体属性发布给任何对等体/对等体组
对发布给 MBGP 对等体/MBGP 对等体组的路由指定路由策略		**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	必选 缺省情况下，不指定对等体/对等体组的路由策略

说明

- 配置 IPv6 MBGP 团体时，必须使用路由策略来定义具体的团体属性，然后在发布路由信息时应用此路由策略。
- 关于路由策略的配置，请参见“IP 路由分册”中的“路由策略配置”。

1.7.4 配置 IPv6 MBGP 路由反射器

为保证 IBGP 对等体之间的连通性，需要在 IPv6 组播 IBGP 对等体之间建立全连接关系。当 IPv6 组播 IBGP 对等体数目很多时，建立全连接网的开销很大。路由反射器可以解决这个问题。

表1-21 配置 IPv6 MBGP 路由反射器

操作	命令	说明
进入系统视图	**system-view**	-
进入 BGP 视图	**bgp** *as-number*	-
进入 BGP IPv6 组播地址族视图	**ipv6-family multicast**	-
配置将本机作为路由反射器，并将 IPv6 MBGP 对等体/对等体组作为路由反射器的客户	**peer** { *ipv6-group-name* | *ipv6-address* } **reflect-client**	必选 缺省情况下，没有配置路由反射器及其客户
使能客户机之间的路由反射	**reflect between-clients**	可选 缺省情况下，允许客户到客户的路由反射
配置路由反射器的集群 ID	**reflector cluster-id** *cluster-id*	可选 缺省情况下，每个路由反射器是使用自己的 Router ID 作为集群 ID

说明

- 通常情况下，路由反射器的客户之间不要求是全连接的，路由通过反射器从一个客户反射到其它客户；如果客户之间是全连接的，可以禁止客户间的反射，以便减少开销。
- 当一个集群里有多个路由反射器时，需要给所有位于同一个集群内的路由反射器配置相同的 Cluster_ID，以避免路由循环。

1.8 IPv6 MBGP 显示和维护

1.8.1 IPv6 MBGP 显示

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 IPv6 MBGP 的运行情况，通过查看显示信息验证配置的效果。

表1-22 IPv6 MBGP 显示

操作	命令
显示对等体组信息	**display bgp ipv6 multicasst group** [*ipv6-group-name*]
显示 IPv6 MBGP 发布的 IPv6 路由信息	**display bgp ipv6 multicast network**
显示 AS 路径信息	**display bgp ipv6 multicast paths** [*as-regular-expression*]
显示 IPv6 MBGP 对等体/对等体组信息	**display bgp ipv6 multicast peer** [[*ipv6-address*] **verbose**]
显示 IPv6 MBGP 路由信息	**display bgp ipv6 multicast routing-table** [*ipv6-address prefix-length*]
显示匹配指定 AS 路径过滤列表的路由	**display bgp ipv6 multicast routing-table as-path-acl** *as-path-acl-number*

操作	命令
显示指定 IPv6 MBGP 团体的路由信息	**display bgp ipv6 multicast routing-table community** [*aa:nn*<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**]* [**whole-match**]
显示匹配指定 IPv6 MBGP 团体列表的路由信息	**display bgp ipv6 multicast routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>
显示 IPv6 MBGP 衰减的路由	**display bgp ipv6 multicast routing-table dampened**
显示 IPv6 MBGP 路由衰减参数	**display bgp ipv6 multicast routing-table dampening parameter**
显示来自不同自治系统的 IPv6 MBGP 路由	**display bgp ipv6 multicast routing-table different-origin-as**
显示 IPv6 MBGP 路由振荡统计信息	**display bgp ipv6 multicast routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *network-address* [*prefix-length* [**longer-match**]]]
显示向指定的 IPv6 MBGP 对等体发送或者从指定的 IPv6 BGP 对等体收到的路由信息	**display bgp ipv6 multicast routing-table peer** *ipv6-address* { **advertised-routes** | **received-routes** } [*network-address prefix-length* | **statistic**]
显示匹配 AS 正则表达式的 IPv6 组播路由信息	**display bgp ipv6 multicast routing-table regular-expression** *as-regular-expression*
显示 IPv6 MBGP 的路由统计信息	**display bgp ipv6 multicast routing-table statistic**
显示 IPv6 MBGP 路由表中的路由信息	**display ipv6 multicast routing-table** [**verbose**]
显示指定目的地址的组播路由信息	**display ipv6 multicast routing-table** *ipv6-address prefix-length* [**longer-match**] [**verbose**]

1.8.2 复位 IPv6 MBGP 连接

当 MBGP 路由策略或协议发生变化后，如果需要通过复位 IPv6 MBGP 连接使新的配置生效，请在用户视图下进行下列配置。

表1-23 复位 IPv6 MBGP 连接

操作	命令
复位指定的 IPv6 MBGP 连接	**reset bgp ipv6 multicast** { *as-number* | *ipv6-address* | **all** | **group** *ipv6-group-name* | **external** | **internal** }

1.8.3 清除 IPv6 MBGP 信息

在用户视图下执行 **reset** 命令可清除 IPv6 MBGP 路由的统计信息。

表1-24 清除 IPv6 MBGP 信息

操作	命令
清除 IPv6 MBGP 路由的衰减信息并释放被抑制的路由	**reset bgp ipv6 multicast dampening** [*ipv6-address prefix-length*]
清除 IPv6 MBGP 路由的振荡统计信息	**reset bgp ipv6 multicast flap-info** [*ipv6-address/prefix-length* | **regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number*]

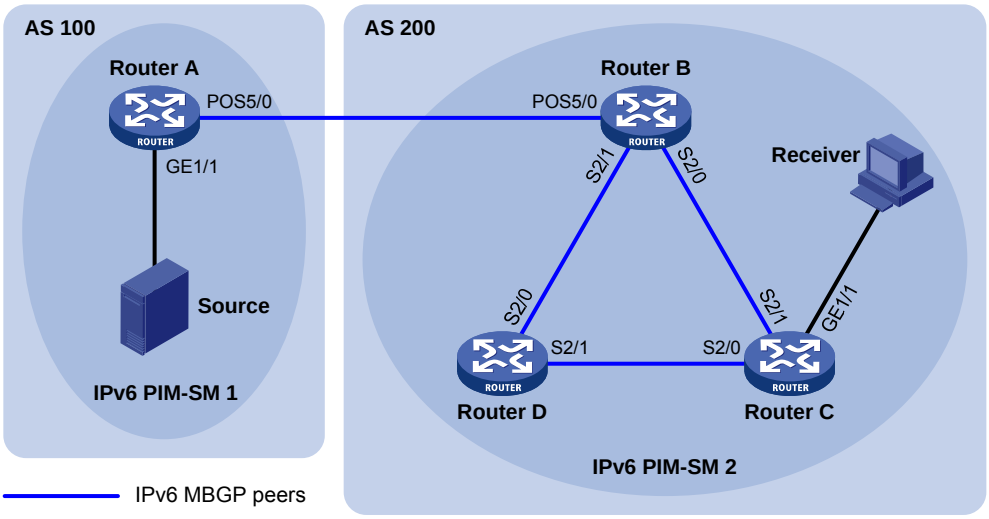
1.9 IPv6 MBGP 典型配置举例（路由应用）

1. 组网需求

- 网络中存在两个自治系统：IPv6 PIM-SM 1 属于 AS 100，IPv6 PIM-SM 2 属于 AS 200。各 AS 内部采用 OSPFv3 进行互联，AS 之间采用 IPv6 MBGP 交换 IPv6 组播路由信息；
- IPv6 组播源属于 AS 100 内的 IPv6 PIM-SM 1，接收者则属于 AS 200 内的 IPv6 PIM-SM 2；
- 将 Router A 和 Router B 各自的 POS5/0 接口分别配置为各自 IPv6 PIM-SM 域的 C-BSR 和 C-RP。

2. 组网图

图1-1 IPv6 MBGP 典型配置组网图



设备	接口	IP 地址	设备	接口	IP 地址
Source	-	1002::100/64	Router C	GE1/1	3002::1/64
Router A	GE1/1	1002::1/64		S2/0	3001::1/64
	POS5/0	1001::1/64		S2/1	2001::2/64
Router B	POS5/0	1001::2/64	Router D	S2/0	2002::2/64
	S2/0	2001::1/64		S2/1	3001::2/64
	S2/1	2002::1/64			

3. 配置步骤

(1) 配置各路由器接口的 IPv6 地址和 IPv6 单播路由协议

- 请按照 图 1-1配置各接口的IPv6 地址和前缀长度，具体配置过程略。
- 配置 AS200 内的各路由器之间采用 OSPFv3 路由协议交换路由信息（AS 内各路由器创建的 OSPF 进程号为 1），确保各 AS 内部在网络层互通，具体配置过程略。

(2) 使能 IPv6 组播路由，使能 IPv6 PIM-SM 和 MLD，并配置 BSR 的服务边界

在 Router A 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-SM。

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface gigabitethernet 1/1
[RouterA-GigabitEthernet1/1] pim ipv6 sm
[RouterA-GigabitEthernet1/1] quit
[RouterA] interface pos 5/0
```

```
[RouterA-Pos5/0] pim ipv6 sm
[RouterA-Pos5/0] quit
```

Router B 和 Router D 上的配置与 Router A 相似，配置过程略。

在 Router C 上使能 IPv6 组播路由，在各接口上使能 IPv6 PIM-SM，并在主机侧接口 GigabitEthernet1/1 上使能 MLD。

```
<RouterC> system-view
[RouterC] multicast ipv6 routing-enable
[RouterC] interface serial 2/0
[RouterC-Serial2/0] pim ipv6 sm
[RouterC-Serial2/0] quit
[RouterC] interface serial 2/1
[RouterC-Serial2/1] pim ipv6 sm
[RouterC-Serial2/1] quit
[RouterC] interface gigabitethernet 1/1
[RouterC-GigabitEthernet1/1] pim ipv6 sm
[RouterC-GigabitEthernet1/1] mld enable
[RouterC-GigabitEthernet1/1] quit
```

在 Router A 上配置 BSR 的服务边界。

```
[RouterA] interface pos 5/0
[RouterA-Pos5/0] pim ipv6 bsr-boundary
[RouterA-Pos5/0] quit
```

在 Router B 上配置 BSR 的服务边界。

```
[RouterB] interface pos 5/0
[RouterB-Pos5/0] pim ipv6 bsr-boundary
[RouterB-Pos5/0] quit
```

(3) 配置 C-BSR 和 C-RP 的位置

在 Router A 上配置 C-BSR 和 C-RP 的位置。

```
[RouterA] pim ipv6
[RouterA-pim6] c-bsr 1001::1
[RouterA-pim6] c-rp 1001::1
[RouterA-pim6] quit
```

在 Router B 上配置 C-BSR 和 C-RP 的位置。

```
[RouterB] pim ipv6
[RouterB-pim6] c-bsr 1001::2
[RouterB-pim6] c-rp 1001::2
[RouterB-pim6] quit
```

(4) 配置 BGP 协议，并配置 IPv6 MBGP 对等体

在 Router A 上配置 EBGP 邻接关系、并配置 IPv6 MBGP 对等体。

```
[RouterA] ipv6
[RouterA] bgp 100
[RouterA-bgp] router-id 1.1.1.1
[RouterA-bgp] ipv6-family
[RouterA-bgp-af-ipv6] peer 1001::2 as-number 200
[RouterA-bgp-af-ipv6] import-route direct
```

```
[RouterA-bgp-af-ipv6] quit
[RouterA-bgp] ipv6-family multicast
[RouterA-bgp-af-ipv6-mul] peer 1001::2 enable
[RouterA-bgp-af-ipv6-mul] import-route direct
[RouterA-bgp-af-ipv6-mul] quit
[RouterA-bgp] quit
```

在 Router B 上配置 EBGP 邻接关系、配置 IPv6 MBGP 对等体，并引入 OSPFv3 路由。

```
[RouterB] ipv6
[RouterB] bgp 200
[RouterB-bgp] router-id 2.2.2.2
[RouterB-bgp] ipv6-family
[RouterB-bgp-af-ipv6] peer 1001::1 as-number 100
[RouterB-bgp-af-ipv6] import-route ospfv3 1
[RouterB-bgp-af-ipv6] quit
[RouterB-bgp] ipv6-family multicast
[RouterB-bgp-af-ipv6-mul] peer 1001::1 enable
[RouterB-bgp-af-ipv6-mul] import-route ospfv3 1
[RouterB-bgp-af-ipv6-mul] quit
[RouterB-bgp] quit
```

(5) 检验配置效果

通过使用 **display bgp ipv6 multicast peer** 命令可以查看路由器之间 IPv6 MBGP 对等体的关系。例如：

查看 Router B 上 IPv6 MBGP 对等体关系的信息。

```
[RouterB] display bgp ipv6 multicast peer

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  1001::1     4   100       56       56     0       0 00:40:54 Established
```

目 录

1 OAP单板配置 1-1
1.1 OAP单板简介 1-1
1.2 登录OAP单板的操作系统 1-1
1.2.1 通过OAP的Console口登录 1-1
1.2.2 通过OAP的管理以太网口以SSH方式登录 1-1
1.2.3 通过OAP单板内部以太网口以SSH方式登录 1-2
1.2.4 从设备侧重定向到OAP单板 1-2
1.3 复位OAP单板系统 1-2

1 OAP 单板配置

1.1 OAP 单板简介

OAA（Open Application Architecture，开放应用架构）是一个开放的软硬件体系。H3C 的 OAA 技术以 H3C 设备为基础，提供了一套完整的软、硬件标准接口。第三方合作厂商可以根据自己的优势生产具有特殊功能的产品，只要这些产品遵循 OAA 标准接口，就可以与 H3C 的设备互相兼容，使单一网络产品的功能得到扩充，为客户创造更大的价值。

OAP（Open Application Platform，开放应用平台）是基于 OAA 架构的物理平台。它可以是一台独立的网络设备，也可以是一块插卡，作为设备扩展部件。我们把这种插卡形式的 OAP 称为 OAP 单板。OAP 单板上运行独立的操作系统，客户可根据需要在该操作系统下加载安全、语音等业务软件，为客户提供多样化的服务。同时，OAP 单板插入设备的扩展插槽，通过内部业务接口与设备进行数据交互、状态交互以及控制交互。

1.2 登录 OAP 单板的操作系统

说明

OAP 单板操作系统的登录方式与 OAP 单板的硬件结构相关，具体支持以下哪些登录方式请以实际情况为准。

1.2.1 通过 OAP 的 Console 口登录

通过 OAP 单板上的 Console 口可以直接登录单板的操作系统。以一台 PC 作为终端为例：

(1) 将 PC 串口通过配置电缆与 OAP 的 Console 口相连；

(2) 打开 PC，并在 PC 上运行终端仿真程序（如超级终端）。选择连接方式为 COM，并设置终端参数，要求与设备 Console 用户界面下的相关参数配置一致，在设备使用参数缺省值时，终端参数为：波特率为 9600，数据位为 8，奇偶校验为无，停止位为 1，流量控制为无。

配置完成后，即可通过 PC 上的终端仿真程序登录 OAP 单板的操作系统。

1.2.2 通过 OAP 的管理以太网口以 SSH 方式登录

通过 OAP 单板上的管理以太网口可以直接登录单板的操作系统。OAP 单板默认开启 SSH 服务器功能，用户可以使用 SSH 客户端设备（如 PC），通过单板的管理以太网口，以 SSH 方式登录单板的操作系统。下面以一台交换机作为 SSH 客户端为例说明登录过程：

(1) 将网线的一端接到交换机的以太网口上，另一端连到 OAP 的管理以太网口。

(2) 确保交换机和管理以太口之间路由可达。

(3) 在交换机上使用 **ssh2** 命令，输入 OAP 管理以太网口的 IP 地址作为 SSH 服务器地址，连接成功后即可登录单板的操作系统。

1.2.3 通过 OAP 单板内部以太网口以 SSH 方式登录

OAP 单板安装到设备的扩展插槽后，通过两个内部的业务口与设备进行信息交互。一个是串口，一个是快速以太网口，这种登录方式用到的就是快速以太网口。

OAP 单板默认开启 SSH 服务器功能，用户可以使用 SSH 客户端设备（如 PC），通过设备，以 SSH 方式登录单板的操作系统。下面以一台 PC 作为 SSH 客户端为例说明登录过程：

(1) 用网线将 PC 和设备上的以太网口连接起来。

(2) 确保 PC 与内部业务口中的快速以太网口之间路由可达。

(3) 在 PC 上使用 SSH 客户端功能输入 OAP 单板的快速以太网口的 IP 地址作为 SSH 服务器地址，连接成功后即可登录单板的操作系统。

1.2.4 从设备侧重定向到 OAP 单板

从设备侧通过以下操作可以重定向连接到单板的操作系统，显示界面将从设备的命令行操作界面切换到 OAP 单板操作系统的操作界面，从而可以对 OAP 单板上的系统及应用软件进行管理。切换以后，可以通过快捷键<Ctrl+k>返回到设备的命令行操作界面。

表1-1 从设备侧重定向到 OAP 单板

操作	命令	说明
从设备侧重定向到 OAP 单板	**oap connect slot** *slot-number*	必选 该操作在用户视图下执行

1.3 复位 OAP 单板系统

在操作系统出现故障或其他异常情况下，可以通过下面的命令复位 OAP 单板系统，使得 OAP 单板再次上电启动。该操作相当于使用 OAP 单板上的复位按钮进行硬件复位 OAP 单板。

OAP 单板有独立的 CPU 系统，复位 OAP 单板系统，设备侧仍然可以对 OAP 单板进行识别和控制。

表1-2 复位 OAP 单板系统

操作	命令	说明
复位 OAP 单板系统	**oap reboot slot** *slot-number*	必选 该操作在用户视图下执行



注意

复位操作可能会造成数据丢失以及业务中止，因此执行此操作请前先保存 OAP 单板操作系统的业务数据，并对 OAP 单板操作系统执行关机操作，以免业务中止及硬盘数据丢失等情况发生。

目 录

1 ACFP配置 1-1
1.1 ACFP简介 1-1
1.1.1 ACFP体系结构 1-1
1.1.2 ACFP联动 1-1
1.1.3 ACFP管理 1-2
1.1.4 ACFP信息概述 1-2
1.1.5 ACFP使用说明 1-5
1.2 ACFP配置任务简介 1-6
1.3 配置ACFP server 1-6
1.4 配置ACFP client 1-6
1.5 开启ACFP Trap功能 1-6
1.6 ACFP显示和维护 1-7
1.7 ACFP典型配置举例 1-7
1.7.1 组网需求 1-7
1.7.2 组网图 1-8
1.7.3 配置步骤 1-8

1 ACFP 配置

1.1 ACFP 简介

数据通信的基础网络主要由路由器和交换机组成，由这些设备完成数据报文的转发。随着数据网络的逐步发展，数据网络上运行的业务越来越多，但是传统的路由器、交换机并不适合处理很多新兴业务，因此产生了一些专门处理某些业务的产品，如防火墙、IDS（Intrusion Detection System，入侵检测系统）、IPS（Intrusion Prevention System，入侵抵御系统）等安全产品及语音、无线等产品。

为了更好地支撑新兴业务，传统网络设备（这里指路由器、交换机）生产厂家纷纷推出多种专用业务板（卡），或者提供一套软硬件接口，允许其他厂家提供板（卡）或者设备的硬件或软件，插入或连接到传统网络设备上，协作处理这些业务，从而能发挥各厂家在各自领域的优势，更有效地支撑新兴业务，同时减少用户投资。OAA（Open Application Architecture，开放应用架构）就是基于该思想而提出的开放业务架构，它能够让众多不同厂商生产的设备和软件集成在一起，象一台设备那样工作，为客户提供一体化的解决方案。

ACFP（Application Control Forwarding Protocol，应用控制转发协议）是基于 OAA 架构设计的应用控制转发协议。例如，联动的 IPS/IDS 卡或者 IPS/IDS 设备作为 ACFP 客户端，上面运行其他厂家的软件，支撑 IPS/IDS 业务。路由器或交换机收到 IP 报文后，通过匹配 ACFP 的联动策略规则，将报文镜像或重定向给 ACFP 客户端，ACFP 客户端上的软件对报文做监控、检测等业务处理，然后根据监控、检测的结果，再通过联动 MIB（Management Information Base，管理信息库）反馈给路由器或交换机，指示路由器或交换机做出相应处理（如过滤某些报文）。

1.1.1 ACFP 体系结构

图1-1 ACFP 体系结构示意图



如 图 1-1所示，ACFP体系可以分成三部分：

- 路由交换部件：是路由器和交换机的主体部分，这部分有着完整的路由器或交换机的功能，也是用户管理控制的核心，此部分称为 ACFP server；
- 独立业务部件：是可以开放给第三方合作开发的主体，主要用来提供各种独特的业务服务功能，此部分称为 ACFP client；
- 接口连接部件：是路由交换部件和独立业务部件的接口连接体，通过这个部件将两个不同厂商的设备连接在一起，以形成一个整体。

1.1.2 ACFP 联动

ACFP 联动就是指独立业务部件可以向路由交换部件发指令，改变路由交换部件的功能。联动功能主要是通过 SNMP 协议实现的：独立业务部件仿照网管系统的功能，向路由交换部件发送各种

SNMP 命令；而路由交换部件上支持 SNMP Agent 功能，可以执行收到的这些命令。在这个过程中，联系双方的关键就是联动的 MIB。

1.1.3 ACFP 管理

ACFP 联动提供了一套机制，使得 ACFP client 能够控制 ACFP server 上的流量，包括：

- 将 ACFP server 上的流量镜像、重定向到 ACFP client；
- 允许、拒绝 ACFP server 上的流量通过；
- 对 ACFP server 上的流量进行限速；
- 在报文中携带上下文 ID，通过上下文 ID 使得 ACFP server 和 ACFP client 能互通报文上下文环境。具体过程如下：ACFP server 中维护一个上下文表，通过上下文 ID 查询上下文表，每个上下文 ID 对应一个 ACFP 联动策略（联动策略的内容包括报文入接口、报文出接口、联动规则等信息）。当 ACFP server 收到的报文由于匹配某个联动规则而被重定向或镜像到 ACFP client 时，报文中会携带该联动规则所在联动策略对应的上下文 ID；当被重定向的报文从 ACFP client 返回时，报文中也会携带该上下文 ID，通过上下文 ID，ACFP server 就知道该报文是重定向返回的报文，然后进行正常的转发处理。

为便于 ACFP client 更好地控制流量，联动内容中设置联动策略与联动规则两级组织，基于策略管理匹配规则的流量，达到一种弹性管理的目的。

为有效支撑 Client/Server 这种联动模式，细粒度、弹性地设置各种规则，联动内容分成四块组织：ACFP server 信息、ACFP client 信息、ACFP 联动策略、ACFP 联动规则。这四块内容存储在 ACFP server 中。

由于一个 ACFP server 可以支持多个 ACFP client，因此，ACFP client 信息、ACFP 联动策略、ACFP 联动规则都是以表的形式组织的。

ACFP server 信息由 ACFP server 自己生成，ACFP client 信息、ACFP 联动策略、ACFP 联动规则都是由 ACFP client 生成并通过联动 MIB 或联动协议发送到 ACFP server。


说明

H3C SR6600 路由器不支持对 ACFP server 上的流量进行限速。

1.1.4 ACFP 信息概述

1. ACFP server 信息

ACFP server 信息包含的内容及其含义如下：

- 所能支持的工作模式：分为主机、穿透、镜像和重定向四种。ACFP server 可以同时支持其中的多种工作模式。只有当 ACFP server 所支持的工作模式包含 ACFP client 的工作模式时，这两个主体才能进行联动。
- 联动策略的最长有效期：说明了 ACFP server 的联动策略所能存活的最长时间。
- 联动策略存储的持久性：说明了 ACFP server 是否具备永久保存联动策略的能力，主要指 ACFP server 重新启动后还能否保有原来的联动策略。

- 当前所支持的上下文 ID 的类型：不同的 ACFP server 中，上下文 ID 在报文中所处的位置可能不同。上下文 ID 的类型分为 5 种：no-context（不携带上下文 ID）、HG-context（使用 HG 前导码作为上下文 ID）、HGPlus-context（使用加强版 HG 前导码作为上下文 ID）、FlowID-context（使用 FlowID 前导码作为上下文 ID）和 VLANID-context（使用 VLAN ID 作为上下文 ID）。

上述这些信息表明了一个 ACFP server 的联动能力，各 ACFP client 可通过联动协议、联动 MIB 的途径来获取这些信息。


说明

H3C SR6600 路由器只支持 LANID-context（使用 VLAN ID 作为上下文 ID）一种 ID 类型。

2. ACFP client 信息

ACFP client 信息包含的内容及其含义如下：

- ACFP client ID：ACFP client 的标识，可以通过联动协议由 ACFP server 分配，也可以由网络管理员指定，目的都是要确保各 ACFP client 在 ACFP server 中 client ID 的唯一性。
- 描述：ACFP client 的描述信息。
- 硬件版本：ACFP client 的硬件类别及版本号等信息。
- 操作系统版本：ACFP client 的系统名称及版本号等信息。
- 应用软件版本：ACFP client 的应用软件类别名称及其版本号等信息。
- IP 地址：ACFP client 的 IP 地址。
- 支持的工作模式：ACFP client 当前所能支持的工作模式，指主机、穿透、镜像、重定向这些模式的组合。

3. ACFP 联动策略

ACFP 联动策略指 ACFP client 发送给 ACFP server 所要实施的联动策略，策略信息包含的内容及其含义如下：

- ACFP client ID：ACFP client 的标识。
- 策略号：策略的标识。
- 策略入接口：报文进入 ACFP server 的接口。
- 策略出接口：报文被正常转发的出接口。
- 策略目的接口：ACFP server 连接该 ACFP client 的接口。
- 报文上下文 ID：会在镜像或重定向报文给 ACFP client 时用到；当发送的策略指定了连接 ACFP client 的接口时，由 ACFP server 为该策略分配一个全局的序号，即报文上下文 ID，每个上下文 ID 对应一个 ACFP 联动策略。
- 策略管理状态：表示该策略是否允许生效。
- 策略有效期：表示该策略有效的期限，借此来控制策略下的所有规则有效期限。
- 策略开始时间：表示该策略生效的起始时间，单位为每天的时、分、秒，借此来控制策略下的所有规则。

- 策略结束时间：表示该策略生效的结束时间，单位为每天的时、分、秒，借此来控制策略下的所有规则。
- 策略目的接口 down 时，该策略下所有规则处理动作：对于转发优先设备，在目的接口 down 后希望重定向和镜像的报文继续转发，此时选择 delete 动作；对于安全优先设备，在目的接口 down 后希望重定向和镜像的报文直接丢弃，此时选择 reserve 动作。
- 策略优先级：表示该策略的优先级，用数字 1～8 表示，数字越大，优先级越高。

4. ACFP 联动规则

ACFP 联动规则指 ACFP client 发送给 ACFP server 所要实施的联动规则，联动规则可以分为 3 类：

- 监控规则：即将哪些报文递给 ACFP client 做监控分析及业务处理。该规则对应的动作类型目前有重定向、镜像。
- 过滤规则：即明确哪些报文被丢弃、哪些报文允许通过。该规则对应的动作类型有丢弃、通过。
- 限制规则：即明确哪些报文将被限速。该规则对应的动作类型为限速。

规则信息包含的内容及其含义如下：

- ACFP client ID：ACFP client 的标识；
- 策略号：策略的标识；
- 规则号：规则的标识；
- 规则操作状态：表示规则是否应用成功；
- 动作类型：包括镜像、重定向、丢弃、通过和限速 5 种动作；
- 是否匹配所有报文：表示该规则是否要匹配所有的报文，如果是的话，则不需要进行下面的匹配；
- 源 MAC 地址；
- 目的 MAC 地址；
- 起始 VLAN ID；
- 结束 VLAN ID；
- IP 中的协议号；
- 源 IP 地址；
- 源 IP 地址反掩码；
- 源端口号操作符：类型为等于、不等于、大于、小于、之间，只有类型为之间时，下面的结束源端口号才有意义，标识所匹配的报文的源端口应该大于起始源端口号而小于结束源端口号；
- 起始源端口号；
- 结束源端口号；
- 目的 IP 地址；
- 目的 IP 地址反掩码；
- 目的端口号操作符：类型为等于、不等于、大于、小于、之间，只有类型为之间时，下面的结束目的端口号才有意义，标识所匹配的报文的目的端口应该大于起始目的端口号而小于结

束目的端口号；

- 起始目的端口号；
- 结束目的端口号；
- 报文的协议类型：包括 GRE、ICMP、IGMP、OSPF、TCP、UDP、IP 等；
- IP 优先级：报文优先级，用数字表示，取值范围为 0～7；
- IP ToS：IP 报文的服务类型；
- IP DSCP：IP 报文的差分服务编码点；
- TCP 标志：表示关心 TCP 六个标志位（URG、ACK、PSH、RST、SYN 和 FIN）中的某些位；
- IP 分片：是否是 IP 分片报文；
- 限制速率。

联动规则隶属于联动策略，通过联动策略可以管理策略下的规则。

1.1.5 ACFP 使用说明

- 对于 VLANID-context 上下文 ID 类型的设备，ACFP 功能使能后，某些 VLAN ID 不应再被其它模块使用，否则可能导致部分数据报文转发错误。不应使用的 VLAN ID 的范围与设备的型号有关，请以设备的实际情况为准。此限制由用户配置保证。
- ACFP 策略在 GRE 隧道环境中应用时只能配置在 Tunnel 口上。
- ACFP 不支持策略路由业务和 Netstream 业务。
- ACFP 重定向的报文处理和部分 QoS 处理（FR-DE 匹配、ATM-CLP 匹配、入接口匹配、QoS local-id 及本地优先级等）互斥，重定向到 ACFP client 后返回的报文不进行上述 QoS 处理。
- ACFP 重定向或镜像的报文在目的接口仅支持二层 QoS 处理（包括队列、WRED 等），不支持其它业务处理（包括非二层的 QoS 处理及非 QoS 业务的处理）。
- ACFP 不支持将同一个流镜像或重定向到多个 ACFP client。
- ACFP 不能处理本地发出的报文。
- 对于多核和软转发设备，ACFP 不支持下列报文的处理：广播报文、组播报文、MPLS 报文、发往本地的报文、IPv6 报文。
- 对于多核和软转发设备，ACFP 镜像和重定向功能只处理小于等于 1500 字节（指的是三层报文的长度，不包括链路层报文头）的 IP 报文，大于 1500 的报文将被丢弃。
- 对于软转发设备，当报文在入接口和出接口下匹配的规则的内容超出五元组信息时，该报文在普通转发流程中不会建立快转 Cache，不走快转流程。

1.2 ACFP 配置任务简介

表1-1 ACFP 配置任务简介

配置任务	说明	详细配置
配置 ACFP server	必选	1.3
配置 ACFP client	必选	1.4
开启 ACFP Trap 功能	可选	1.5

1.3 配置 ACFP server

表1-2 配置 ACFP server

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACFP server 功能	**acfp server enable**	必选 缺省情况下，ACFP server 功能处于关闭状态

1.4 配置 ACFP client

用户需在 ACFP client 上通过 MIB 配置 ACFP 联动策略和 ACFP 联动规则，具体配置与 ACFP client 上所使用的业务软件有关。

1.5 开启 ACFP Trap 功能

为确保 ACFP 功能正常运行，必须允许设备发送 ACFP 模块的 Trap 报文。

开启ACFP模块的Trap功能后，该模块会生成Trap报文，用于报告该模块的重要事件。ACFP Trap 报文对应的级别如 表 1-3所示。

表1-3 ACFP Trap 报文对应的级别

Trap 报文	级别
上下文 ID 类型改变	notifications
ACFP client 注册	notifications
ACFP client 注销	notifications
ACSEI 协议检测到 ACFP client 没有响应	warnings
ACFP server 不支持 ACFP client 的工作模式	errors
ACFP 联动策略的有效期改变	notifications
ACFP 联动规则创建	informational
ACFP 联动规则删除	informational

Trap 报文	级别
ACFP 联动规则发生错误	errors
ACFP 联动策略的有效期超时	notifications

生成的 Trap 报文将被发送到设备的信息中心，通过设置信息中心的参数，最终决定 Trap 报文的输出规则（即是否允许输出以及输出方向）。（有关信息中心参数的配置请参见“系统分册”中的“信息中心配置”。）

表1-4 开启 ACFP Trap 功能

操作	命令	说明
进入系统视图	**system-view**	-
开启 ACFP 模块的 Trap 功能	**snmp-agent trap enable acfp** [**client** | **policy** | **rule** | **server**]	可选 缺省情况下，ACFP 模块的 Trap 功能处于开启状态


说明

有关 **snmp-agent trap enable** 命令的详细介绍，请参见“系统分册”中的“SNMP 命令”。

1.6 ACFP 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 ACFP 的运行情况，通过查看显示信息验证配置的效果。

表1-5 ACFP 显示和维护

操作	命令
查看 ACFP server 信息	**display acfp server-info**
查看 ACFP client 信息	**display acfp client-info** [*client-id*]
查看 ACFP 策略信息	**display acfp policy-info** [**client** *client-id* [*policy-index*] | **dest-interface** *interface-type interface-number* | **in-interface** *interface-type interface-number* | **out-interface** *interface-type interface-number*] [**active** | **inactive**]
查看 ACFP 规则信息	**display acfp rule-info** { **in-interface** [*interface-type interface-number*] | **out-interface** [*interface-type interface-number*] | **policy** [*client-id policy-index*] }
查看 ACFP 规则缓存信息	**display acfp rule-cache** [**in-interface** *interface-type interface-number* | **out-interface** *interface-type interface-number*] *
查看 ACFP Trap 的配置情况	**display snmp-agent trap-list**

1.7 ACFP 典型配置举例

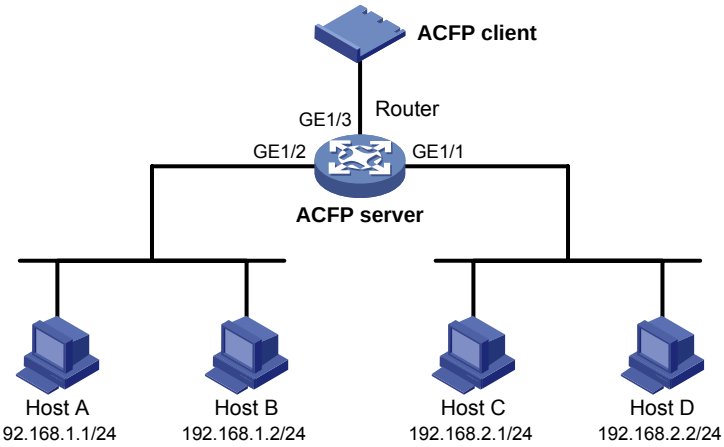
1.7.1 组网需求

- 某公司企业网通过 Router 实现各部门的互连，该 Router 为 ACFP server；

- ACFP client 与 Router 相连并控制 Router 上的流量，分析入接口为 GigabitEthernet1/2 的流量。ACFP client 经过分析后，允许入接口为 GigabitEthernet1/2、源 IP 为 192.168.1.1/24 的所有报文通过，并拒绝入接口为 GigabitEthernet1/2、源 IP 为 192.168.1.2/24 的所有报文通过。

1.7.2 组网图

图1-2 ACFP 典型配置组网图



1.7.3 配置步骤

1. 配置 Router

使能 ACFP server 功能。

```
<Router> system-view
[Router] acfp server enable
```

2. 通过 MIB 配置 ACFP client 的联动策略和监控规则

配置 ACFP client。

通过 MIB Browser 配置 ACFP client，并将信息发送到 Router。其中，client 索引为 1，支持主机、穿透、重定向和镜像四种工作模式（通过设置 h3cAcfpClientMode 节点实现），其它参数均为默认值。

配置 ACFP 策略。

通过 MIB Browser 配置 ACFP 策略，并将信息发送到 Router。其中，client 索引为 1，策略索引为 1，策略入接口为 GigabitEthernet1/2（通过设置 h3cAcfpPolicyInIfIndex 节点实现），策略目的接口为 GigabitEthernet1/3（通过设置 h3cAcfpPolicyDestIfIndex 节点实现），其它参数均为默认值。

配置 ACFP 规则。

通过 MIB Browser 配置 ACFP 规则，并将信息发送到 Router。其中，client 索引为 1，策略索引为 1，规则索引为 1，动作类型重定向（通过设置 h3cAcfpRuleAction 节点实现），其它参数均为默认值。

3. 通过 MIB 配置 ACFP client 的联动策略和过滤规则

配置 ACFP 策略。

通过 MIB Browser 配置 ACFP 策略，并将信息发送到 Router。其中，client 索引为 1，策略索引为 2，策略入接口为 GigabitEthernet1/2（通过设置 h3cAcfpPolicyInIfIndex 节点实现），其它参数均为默认值。

配置 ACFP 规则。

通过 MIB Browser 配置 ACFP 规则，并将信息发送到 Router。其中，client 索引为 1，策略索引为 2，规则索引为 1，动作类型为允许通过（通过设置 h3cAcfpRuleAction 节点实现），匹配源 IP 地址为 192.168.1.1 的报文（通过设置 h3cAcfpRuleSrcIP 节点实现），源 IP 地址的反掩码为 255.255.255.0（通过设置 h3cAcfpRuleSrcIPMask 节点实现），其它参数均为默认值。

通过 MIB Browser 配置 ACFP 规则，并将信息发送到 Router。其中，client 索引为 1，策略索引为 2，规则索引为 2，动作类型为拒绝通过（通过设置 h3cAcfpRuleAction 节点实现），匹配源 IP 地址为 192.168.1.2 的报文（通过设置 h3cAcfpRuleSrcIP 节点实现），源 IP 地址的反掩码为 255.255.255.0（通过设置 h3cAcfpRuleSrcIPMask 节点实现），其它参数均为默认值。

4. 验证配置效果

使用 **ping** 命令验证 Host A 与 Host C、以及 Host B 与 Host C 的连通性。测试结果表明：Host A 与 Host C 可以互通，Host B 与 Host C 则无法互通。

目 录

1 ACSEI配置 1-1
1.1 ACSEI简介 1-1
1.1.1 ACSEI的功能 1-1
1.1.2 ACSEI定时器 1-1
1.1.3 ACSEI的启动和运行过程 1-2
1.2 配置ACSEI server 1-2
1.2.1 使能ACSEI server 1-2
1.2.2 配置时钟同步定时器 1-2
1.2.3 配置监控定时器 1-3
1.2.4 关闭ACSEI client 1-3
1.2.5 重新启动ACSEI client 1-3
1.2.6 ACSEI server显示和维护 1-3

1 ACSEI 配置

1.1 ACSEI 简介

ACSEI 是一种私有协议，它为 ACFP server 和 client 提供了一种交互信息的方法，可以为 ACFP（Application Control Forwarding Protocol，应用控制转发协议）联动提供很好的支撑，保障 ACFP client 与 ACFP server 之间有效交互信息、协作运行某种业务。

作为 ACFP 的支撑协议，ACSEI 也包括 server 和 client 两种实体。

- ACSEI server 集成在设备软件系统（Comware）中，是设备支持的一项功能。
- ACSEI client 有两种实现方式：一种是集成在设备的系统软件（Comware）中，作为设备支持的一项功能；一种是集成在 OAP 单板的软件系统当中，是 OAP 单板支持的一项功能。这两种实现方式所需的硬件条件不同，配置方式也不一样，本文将分别进行介绍。



说明

- ACFP 是基于 OAA（Open Application Architecture，开放应用架构）架构设计的应用控制转发协议，联动的 IDS（Intrusion Detection System，入侵检测系统）卡或者 IDS 设备作为 ACFP 客户端，上面运行其他厂家的软件，支撑 IPS（Intrusion Prevention System，入侵抵御系统）/IDS 业务。有关 ACFP 的详细介绍请参见“OAA 分册”中的“ACFP 配置”。
- OAP（Open Application Platform，开放应用平台）是针对新兴业务提供的一个开放式应用平台。OAP 单板上运行操作系统，根据客户需要可加载安全、语音等业务软件，为客户提供多样化的服务。有关 OAP 单板的详细介绍请参见“OAA 分册”中的“OAP 单板配置”。

1.1.1 ACSEI 的功能

ACSEI 协议主要功能如下：

- ACSEI client 向 ACSEI server 注册、注销；
- ACSEI server 给 ACSEI client 分配 ID，用于保证各 ACSEI client 的唯一性与清晰性；
- ACSEI client 与 ACSEI server 之间的互相监控、互相感知；
- ACSEI server 与 ACSEI client 之间的信息交互（包括时钟同步等）；
- 通过 ACSEI server 对 ACSEI client 实施简单的控制，例如，关闭 ACSEI client、重启 ACSEI client。

ACSEI server 与 ACSEI client 为一对多的关系，一个 ACSEI server 允许注册 10 个 ACSEI client。

1.1.2 ACSEI 定时器

ACSEI server 用到两个定时器，分别是时钟同步定时器和监控定时器。

- 时钟同步定时器用来定时触发 ACSEI server 向 ACSEI client 发送时钟同步信息通告报文，用户可以通过命令行设置定时器的值。
- 监控定时器用来定时触发 ACSEI server 向 ACSEI client 发送监控请求报文，用户可以通过命令行设置定时器的值。

ACSEI client 启动两个定时器，分别为注册定时器和监控定时器。

- 注册定时器用来定时触发 ACSEI client 以组播方式（组播 MAC 地址为 010F-E200-0021）发送注册请求报文，用户不能设置定时器的值。
- 监控定时器用来定时触发 ACSEI client 向 ACSEI server 发送监控请求报文，用户不能设置定时器的值。

1.1.3 ACSEI 的启动和运行过程

ACSEI 启动和运行的整个过程可描述如下：

(1) 运行 ACSEI client 可执行程序，使能 ACSEI client 功能。

(2) 启动设备，使能 ACSEI server 功能。

(3) ACSEI client 以组播方式发送注册请求。

(4) ACSEI server 收到合法的注册请求后，与 ACSEI client 协商参数，协商通过后建立连接。

(5) 连接建立后相互监控连接的情况。

(6) 当 ACSEI server 检测到 ACSEI client 连接中断，ACFP server 会将对应 ACFP client 的配置、策略等删除。

1.2 配置 ACSEI server

1.2.1 使能 ACSEI server

表1-1 使能 ACSEI server

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACSEI server 功能	**acsei server enable**	必选 缺省情况下，ACSEI server 功能处于关闭状态

1.2.2 配置时钟同步定时器

表1-2 配置时钟同步定时器

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACSEI server 功能	**acsei server enable**	必选
进入 ACSEI 视图	**acsei server**	-
配置 ACSEI server 到 ACSEI client 的时钟同步定时器的值	**acsei timer clock-sync** *minutes*	可选 缺省情况下，ACSEI server 到 ACSEI client 的时钟同步定时器的值为 5 分钟

1.2.3 配置监控定时器

表1-3 配置监控定时器

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACSEI server 功能	**acsei server enable**	必选
进入 ACSEI 视图	**acsei server**	-
配置 ACSEI server 对 ACSEI client 的监控定时器的值	**acsei timer monitor** *seconds*	可选 缺省情况下，ACSEI server 对 ACSEI client 的监控定时器的值为 5 秒

1.2.4 关闭 ACSEI client

表1-4 关闭 ACSEI client

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACSEI server 功能	**acsei server enable**	必选
进入 ACSEI 视图	**acsei server**	-
关闭指定的 ACSEI client	**acsei client close** *client-id*	必选

1.2.5 重新启动 ACSEI client

表1-5 重新启动 ACSEI client

操作	命令	说明
进入系统视图	**system-view**	-
使能 ACSEI server 功能	**acsei server enable**	必选
进入 ACSEI 视图	**acsei server**	-
重新启动 ACSEI client	**acsei client reboot** *client-id*	必选

1.2.6 ACSEI server 显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 ACSEI server 系统的运行情况，通过查看显示信息验证配置的效果。

表1-6 ACSEI server 显示和维护

配置	命令
显示 ACSEI client 摘要信息	**display acsei client summary** [*client-id*]
显示 ACSEI client 信息	**display acsei client info** [*client-id*]