Redbooks
ibm.com/redbooks

IBM SAN42B-R Extension Switch and IBM b-type Gen 6 Extension Blade in Distance Replication Configurations (Disk and Tape)

Li Cao
José Cortés
Mark Detrick
Steve Guendert
Michael Mettler
Lukasz Razmuk
Megan Gilge

Storage

IBM

Redpaper

International Technical Support Organization

IBM SAN42B-R Extension Switch and IBM b-type Gen 6 Extension Blade in Distance Replication Configurations (Disk and Tape)

February 2017

REDP-5404-00

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (February 2017)

This edition applies to Fabric OS 8.01.

This document was created or updated on February 9, 2017.

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
DS6000™
DS8000®
FICON®
GDPS®
Geographically Dispersed Parallel Sysplex™
Global Technology Services®
IBM®
IBM z Systems®
Parallel Sysplex®
Redbooks®
Redpaper™
Redbooks (logo)  ®
Storwize®
System Storage®
System z®
Tivoli®
z Systems®
z/OS®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Preface

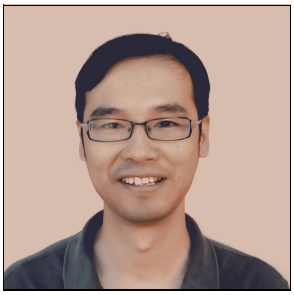
This IBM® Redpaper™ publication helps network and storage administrators understand how to implement the IBM SAN42B-R Extension Switch and the IBM b-type Gen 6 Extension Blade for distance replication. It provides an overview of the IBM System Storage® SAN42B-R extension switch hardware and software features, describes the extension architecture, shows example implementations, and explains how to troubleshoot your extension products.

IBM b-type extension products provide long-distance replication of your data for business continuity by using disaster recovery (BC/DR).

This paper provides an overview of extension, detailed information about IBM b-type extension technologies and products, preferred topologies, example implementations with FCIP and TS7760/7700 Grid IP Extension, monitoring, and troubleshooting.

Authors

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.



Li Cao is a Senior Storage Specialist from IBM China. He worked in IBM System Lab Services as a storage technical specialist since 2009. He is an IBM Certified IT Specialist (Level 2). Before joining IBM, he worked as a solution consultant in Hewlett-Packard China for three years. His areas of expertise include storage virtualization, storage area networks, data migration, and remote copy services. He holds a Master of Science degree in Computer Science from the University of Birmingham and a Bachelor degree from Fudan University.

José Cortés is a Test Specialist Engineer for TS7700 product at IBM Mexico. He is also the Focal Point for SAN and z Systems for the TS7700 Test Teams. Jose has worked with diverse IBM Storage products including TS3500, TS4500, TS3200, and TS3100. He has installed, implemented, and supported many of the SAN switches for TS7700 testing phases. He has knowledge in input/output definition file (IODF) configuration using IBM z/OS®. Jose holds a degree in Computer System Engineering from Tepic Institute of Technology in Mexico.

Mark Detrick is a Director, Principal Architect at Brocade based in Portland, Oregon, and has 15 years of experience in the IBM Fibre Connection (IBM FICON®), Fibre Channel (FC), and Extension fields. He holds a degree in Electrical Engineering from University of the Pacific and an MBA from Pepperdine University. Mark is a CCIE 6336 in Routing and Switching. His areas of expertise include large-scale SAN, extension, data center networks, IP networks, optical, and WAN networks. He has written extensively on various aspects of Fabric and Extension technologies, and IP networking. He has also contributed to other IBM Redbooks® publications.

Steve Guendert is a Principal Director with Brocade Communications where he is the CTO for mainframe connectivity products and strategy, as well as being responsible for the overall business. He has nearly twenty years of IBM z Systems® and mainframe storage experience with Brocade, McDATA, CNT, and IBM. He has a PhD in Management Information Systems (MIS), an MBA from Auburn University, and a B.S. in Economics from Northwestern University.

Michael Mettler is an IT Specialist at IBM Global Technology Services®, Zurich, Switzerland. As a member of the System Sales Implementation Service team, he designs and implements storage solutions that include various IBM products and technologies. Michael's areas of expertise include SAN, SVC, IBM Storwize® products, IBM Flash system, and IBM Tivoli® Storage Manager. He joined IBM in 2006 as an IT apprentice and will complete his part-time bachelor study in Business Information Technology at Zurich University for Applied Sciences in summer 2017.

Lukasz Razmuk is a Certified Infrastructure IT Architect at IBM Global Technology Services in Warsaw, Poland. He has 12 years of IBM experience in designing advanced and complex IT solutions based on IBM AIX®, Linux, IBM pSeries, virtualization, high availability, storage area network, Storage for Open Systems, and cloud. Lukasz is a technical leader for internal and external professional teams, and is a trusted advisor for IBM clients. He has expertise in aligning a company's IT strategy according to market trends to support its business needs, and acts as a Technical Account Advocate for Polish clients. Lukasz holds a Master of Science degree in Information Technology from the Polish-Japanese Institute of Information Technology. He is TOGAF 9 Certified and holds many technical certifications.

Megan Gilge is a Project Leader at the IBM International Technical Support Organization. Before joining the ITSO, she was an Information Developer for the IBM Semiconductor Solutions and IBM i products.

Thanks to the following people for their contributions to this project:

Jon Tate
IBM ITSO

Silviano Gaona
Brian Steffler
Brocade

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction

This chapter discusses the basics of data replication and storage networking extension. It provides a background on these technologies to serve as a foundation for the more detailed technical discussions in the subsequent chapters.

The first section provides an introduction to business continuity, data replication, and storage networking extension. This section includes a discussion of the need for data replication and extension technology for providing connectivity between data centers for a business continuity architecture. This section also provides some background on the IBM TS7700 and how it fits into IBM z Systems® clients' business continuity strategy.

The second section describes Fibre Channel over IP (FCIP) and IP Extension (IPEX) technologies, and the different types of IBM Storage and replication solutions that the technologies support.

It provides the following information:

- The need for data replication and storage networking extension
- Types of storage networking extension (FCIP and IPEX)

1.1 The need for data replication and storage networking extension

Data availability and business continuity offer a vital competitive edge that is crucial to an organization's success. Business continuity is the ability to adapt and respond to risks in order to maintain continuous business operations. There are three primary aspects or components of a business continuity strategy:

- Continuous operations: The capability to keep things running 24 hours a day, 365 days per year.
- High Availability (HA): The capability to provide access to applications regardless of local failures.
- Disaster Recovery: The capability to recover a data center at a different site if a disaster destroys or renders the primary site inoperable. Disaster recovery is only one (but a crucial) component of an overall business continuity strategy for z Systems clients.

One new term associated with business continuity that has gained recognition in the past five years is *IT resilience*, or *resilient architectures*. A resilient IT architecture gives an organization the ability to respond and adapt to a wide variety of external and internal demands, disruptions, disturbances, and threats while continuing business operations without any significant impact. Although related to planning for disaster recovery, planning for a resilient IT architecture is much broader in scope. A resilient IT architecture requires organizations to go beyond planning for recovery from an unplanned outage. In fact, a resilient IT architecture enables organizations to avoid outages entirely, which helps ensure business continuance.

1.1.1 Challenges in designing resilient z Systems architectures

Few events generate a stronger adverse business impact than an IT outage, even if it lasts for only a few minutes. The negative publicity that such events often generate in today's "always connected" news media world can be extremely detrimental. Clients, partners, and the market are driving the demand for continuously available, resilient IT architectures. The potential revenue loss from an outage and the damage to a company's reputation can be costly. In addition, the possible government sanctions in an increasingly complex regulatory environment can create even more challenges.

IBM z Systems clients across a diverse spectrum of industries face these growing challenges when planning a resilient IT architecture. Executive focus on resilient IT architecture solutions is increasing because it is no longer enough to merely plan for recovery from a disaster.

Enterprises that run z Systems need a greater level of availability to deal with a range of contingencies. These contingencies can be as simple as inadvertent power loss or common configuration errors, or as complex as major natural disasters like 2012's Hurricane Sandy, or human-induced disasters such as a terrorist attack. A business that plans for a resilient IT architecture must implement both traditional disaster recovery planning and additional planning for continuous availability.

The other major challenge is the cost tradeoff: What is the cost versus the potential loss? In an ideal world, every company can afford to implement a true continuous availability solution. However, clients must face fiscal realities. The closer you get to a continuous availability solution, the more complicated and the more costly it becomes. Companies must weigh the value of a potential loss caused by an outage compared with the cost to avoid the outage. You must decide which solution you need and can afford instead of merely deciding what you

want. At the same time, you still must make certain that your solution meets existing government regulatory requirements.

1.1.2 The need for data replication

In 1992, the SHARE user group in the United States, together with IBM, defined a set of Business Continuity tier levels to address the need to properly describe and quantify various different methodologies for successful mission-critical computer system Business Continuity implementations. These tiers are known as the “7 tiers of Business Continuity”. The tier concept is still used, and it is useful for describing today’s Business Continuity capabilities.

The seven tiers solutions offer a simple methodology for defining your current service level, the current risk, and the target service level and target environment. Each tier builds on the foundation of the previous tier. One common theme among the various tiers, other than tier 0 (no business continuity plan), is that there is some type of data replication technology involved. This replication technology ranges from simple backups to physical tape that are transported to a second site (the Pickup Truck Access Method or PTAM) all the way to the IBM Geographically Dispersed Parallel Sysplex™ (IBM GDPS®) solution. The replication technology used depends on a client’s Recovery Point Objective (RPO) and Recovery Time Objective (RTO).

Today, it is routine for z Systems clients to have multiple data centers that are interconnected for the purposes of continuous availability, disaster recovery, or both. This is in response to the ever-increasing dependence of businesses, governments, and society on IT services. It is also required to meet government regulations related to the availability of those services. Additionally, changes in connection technology mean that it is now possible to do things that previously were not possible or not feasible. In any event, having multiple data center sites necessitates a combination of local replication along with remote data replication (RDR).

RDR: RDR is the process of creating replicas of information assets at remote sites/locations. RDR helps z Systems clients mitigate the risks associated with regionally driven outages resulting from natural or human caused disasters that render a data center inoperable. During such a disaster, clients can move the production workload to a remote site to ensure business continuity. There are two types of RDR: Asynchronous and synchronous. Which type is used depends on several factors, including the RPO/RTO requirements, and the distance between sites. Generally speaking, synchronous replication is used primarily for stringent (near zero) RPO and RTO requirements at sites that are in relative close proximity, whereas asynchronous replication is used for less stringent RPO/RTO requirements and for replication between sites that can be separated by hundreds or thousands of kilometers.

1.1.3 The need for storage networking extension

Remote date replication requires some form of network connectivity between sites for data transmission. For z Systems clients, this RDR typically involves both disk (DASD) and virtual and physical tape replication. For parallel channels, the term “channel extender” referred to the networking connectivity between sites. This concept is more commonly referred to simply as “extension” because the reference has to do more with storage replication network connectivity than the physical z Systems channels themselves being “extended” over a long distance.

Over the ensuing two plus decades, a variety of protocols have been used for mainframe storage connectivity. During the 1990s and early 2000s, Enterprise Systems Connection (ESCON) was the predominant technology used for both DASD, and physical and virtual tape

connectivity to mainframes. ESCON was also used for RDR in conjunction with ESCON channel extension devices such as the CNT USDX that were connected into a WAN that spanned the distance between sites.

Depending on the type of replication performed (asynchronous or synchronous) and the storage device (DASD array, VTS, or VTL), the ESCON channel extension devices also had emulation software that would help replication performance over longer distances. At the time of its introduction in 1990, ESCON supported a maximum distance of 20 kilometers (km).

While ESCON used the relatively new Fibre Channel technology, it employed circuit switching rather than the packet switching that is typical of today's Fibre Channel implementations. This implementation provided significantly improved physical connectivity, but retained the same one-at-a-time Channel Command Word (CCW) challenge-response/permission seeking logical connectivity that was employed by parallel channels. As a result, the performance of ESCON is significantly reduced at extended distances by the multiple roundtrip delays required to fetch the channel programs, a performance deficiency commonly referred to as *ESCON droop*.

By the late 1990s, the shortcomings of ESCON were driving a new technical solution. IBM's Fibre Connection (FICON) evolved in the late 1990s to address the technical limitations of ESCON in bandwidth, channel/device addressing, and distance. Unlike parallel and ESCON channels, FICON channels rely on packet switching rather than circuit switching. FICON channels also rely on logical connectivity based on the notion of assumed completion rather than ESCON's permission-seeking schema. These two changes allow a much higher percentage of the available bandwidth to be employed for data transmission and for significantly better performance over extended distances.

Numerous tests comparing FICON to ESCON were done by IBM, other SAN equipment manufacturers, and independent consultants. The common theme among these tests is that they support the proposition that FICON is a much better performing protocol over a wide range of topologies when compared with traditional ESCON configurations. This performance advantage is very notable as the distance between the channel and control unit increases.

The 2000s ushered in FICON as the defacto storage networking protocol of choice for z Systems connectivity to storage subsystems. One of the most important advantages FICON had over ESCON was support for longer distances between sites, and significantly better performance over these long distances. This advantage allowed z Systems clients to start locating their data centers/sites further apart geographically.

Storage networking extension continued to evolve as well. FICON was part of the Fibre Channel family of protocols, and as such could use the new Fibre Channel over TCP/IP (FCIP) protocol for RDR. FCIP supported similar emulation software that the older ESCON channel extensions devices supported. This feature proved to be important because with FCIP emulation technology, IBM GDPS clients could have greater separation between sites for their eXtended Remote Copy (XRC) implementations, and for their IBM 3494 VTS peer to peer implementations.

In the mid to late 2000s, IP replication was introduced for some vendor disk arrays and for the newer virtual tape systems. Some IP replication marketing material referred to it as a way to "eliminate the expense of extension technology." Clients could simply plug the new IP/Ethernet adapter ports on these storage devices directly into the WAN.

More often than not, however, the performance of this IP replication was not that good. Without the emulation and protocol enhancement technology present in the extension hardware, replication performance at long distances would not meet SLA requirements. Brocade started working on a new IP Extension technology known as IPEX.

There are alternatives to using storage networking extension technology. The previously mentioned native IP replication is one. Another is DWDM. A third alternative is simply running fiber in the ground between sites, and using native Fibre Channel (cascaded FICON). However, for distances greater than 50 km between sites, these alternatives either do not offer the performance that FCIP or IP Extension has, or are not as cost effective.

Studies show that the most expensive component of a business continuity architecture is the cost of the bandwidth between sites. Anything done to more efficiently use that bandwidth provides cost savings for z Systems clients. Extension technology is unique when compared to the alternative for long distance RDR because it uses the bandwidth between sites in a much more efficient manner.

1.1.4 IBM TS7760/7700, business continuity, and grid basics

A key component in a resilient IT architecture for IBM z Systems clients is the IBM TS7700/7760 with its multi-cluster grid configuration. The TS7760/7700 grid configuration is a series of clusters connected by a network to form a high-availability, resilient virtual tape storage architecture. Logical volume attributes and data are replicated through Internet Protocol (IP) across these clusters, which are joined by the grid network.

However, to the host, the grid configuration looks like a single storage subsystem. This configuration ensures high availability, and ensures that production work continues even if an individual cluster becomes unavailable. Each grid configuration is optimized to help eliminate downtime from planned and unplanned outages, upgrades, and maintenance. Therefore, the TS7760/7700 Grid enhances the resilience of an organization's IT architecture.

IBM TS7760/7700 grid basics

The technology of a prior generation produced the IBM 3494 Virtual Tape Server (VTS), which had a feature called peer-to-peer VTS. Peer-to-peer VTS was a multisite-capable business continuity and disaster recovery solution, and was to tape what peer-to-peer Remote Copy (PPRC) was to Direct Access Storage Devices (DASDs). Peer-to-peer, VTS-to-VTS data transmission originally was performed by using ESCON, then FICON, and finally Transmission Control Protocol/Internet Protocol (TCP/IP). The features described in this publication apply to both the TS7760 and the TS7700.

The TS7760/7700 Grid configuration introduces new flexibility for designing business continuity solutions. The base architecture and design of the TS7700 integrates peer-to-peer communication capability, eliminating the virtual tape controllers and remote channel extension hardware for the prior generation Peer-to-peer VTS. This change provided the potential for significant simplification in the infrastructure needed for a business continuity solution and for simplified management.

Hosts attach directly to the TS7700s. Instead of FICON or ESCON, the connections between the TS7700 clusters use standard TCP/IP. Similar to the peer-to-peer VTS of the previous generation, with the new TS7700 Grid configuration, data is replicated between the clusters, based on the customer's established policies. Any data can be accessed through any of the TS7700 clusters, regardless of which system the data is on, if the grid contains at least one available copy.

As a business continuity solution for high availability and disaster recovery, multiple TS7700 clusters are connected using standard Ethernet connections. Local and geographically separated connections are supported, and provide a great amount of flexibility to address client needs. This IP network for data replication between TS7700 clusters is more commonly known as a TS7700 Grid. A TS7700 Grid refers to two to six physically separate TS7700 clusters that are connected to each other with a customer-supplied IP network. The TCP/IP

infrastructure that connects a TS7700 Grid is the grid network. The grid configuration forms a high-availability disaster recovery solution and provides metro and remote logical volume replication.

The clusters in a TS7700 Grid can be, but do not need to be, geographically dispersed. In a multiple-cluster grid configuration, two TS7700 clusters are often located within 100 kilometers (km) of each other. The remaining clusters can be more than 1,000 km away. This solution provides a highly available and redundant regional solution. It also provides a remote disaster recovery solution outside of the region.

With the TS7700 Grid, data is replicated and stored in a remote location and supports truly continuous uptime. The IBM TS7700 includes multiple modes of synchronous and asynchronous replication. Replication modes are assigned to data volumes by using the IBM Data Facility Storage Management Subsystem (DFSMS) policy. This policy provides flexibility in implementing business continuity solutions so that organizations can simplify their storage environments and optimize storage utilization. This functionality is similar to IBM Metro Mirror and Global Mirror with advanced copy services support for IBM z Systems customers.

The TS7700 Grid is a robust business continuity and IT resilience solution. By using the TS7700 Grid, organizations can move beyond the inadequacies of on-site backup (disk-to-disk or disk-to-tape) that cannot protect against regional (nonlocal) natural or human-induced disasters. By using the TS7700 Grid, data is created and accessed remotely through the grid network. Many TS7700 Grid configurations rely on this remote access to further increase the importance of the TCP/IP fabric.

With increased storage flexibility, an organization can adapt quickly and dynamically to changing business environments. Switching production to a peer TS7700 is accomplished in a few seconds with minimal operator skills. With a TS7700 Grid solution, z Systems clients eliminate planned and unplanned downtime. This approach can potentially save thousands of dollars in lost time and business, and can address today's stringent government and institutional data protection regulations.

The network infrastructure that supports a TS7700 Grid solution faces challenges and requirements of its own. First, the network components must individually provide reliability, high availability, and resiliency: The overall solution is only as good as its individual parts. A TS7700 Grid network requires nonstop predictable performance with components that have "five-9s" availability. A TS7700 Grid network must be designed with highly efficient components that minimize operating costs. These components must also be highly scalable, to support business and data growth and application needs, and to help accelerate the deployment of new technologies. The IBM SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade in the IBM SAN 512B-6 or IBM SAN 256B-6 provide an IP Extension capability ideal for the TS7700 grid IP replication.

1.2 Types of storage networking extension (FCIP and IPEX)

Section 1.1.4, "IBM TS7760/7700, business continuity, and grid basics" on page 50 provides a brief, historical summary of mainframe extension technology. This section provides a more detailed background on FCIP and IPEX. Later chapters discuss the technical details of both extension technologies.

1.2.1 Fibre Channel over TCP/IP (FCIP)

Over the last decade, extension networks for storage have become commonplace and continue to grow in size and importance. Growth is not limited to new deployments, but also involves the expansion of existing deployments. Requirements for data protection never ease because the economies of many countries depend on successful and continued business operations, and thus have passed laws mandating data protection.

Modern-day dependence on RDR means there is little tolerance for lapses that leave data vulnerable to loss. In mainframe environments, reliable and resilient networks (to the point of no frame loss and in-order frame delivery) is necessary for error-free operation, high performance, and operational ease. This improves availability, reduces risk, reduces operating expenses, and, most of all, reduces risk of data loss.

Because of the higher costs of long-distance dark fiber connectivity compared with other communications services, use of the more common and more affordable IP network services is an attractive option for Fibre Channel (FC) extension between geographically separated data centers.

FCIP is a technology for interconnecting Fibre Channel-based storage networks over extended distances through IP networks. FCIP enables a user to use their existing IP WAN infrastructure to connect Fibre Channel SANs. FCIP is a means of encapsulating Fibre Channel frames within TCP/IP and sending these IP packets over an IP-based network specifically for linking Fibre Channel SANs over these WANs.

FCIP implements tunneling techniques to carry the Fibre Channel traffic over the IP network. The tunneling is transparent, which means that it is invisible to the Upper Level Protocols (ULPs), such as FICON and FCP, that might be in use. The result is that both FICON and FCP I/O traffic are sent through these FCIP tunnels over an IP-based network.

FCIP supports applications such as remote data replication (RDR), centralized SAN backup, and data migration over very long distances that are impractical or costly using native Fibre Channel connections. FCIP tunnels, built on a physical connection between two extension switches or blades, allow Fibre Channel I/O to pass through the IP WAN.

The TCP connections ensure in-order delivery of FC frames and lossless transmission. The Fibre Channel fabric and all Fibre Channel targets and initiators are unaware of the presence of the IP WAN. Figure 1-1 shows the relationship of FC and TCP/IP layers and the general concept of FCIP tunneling.

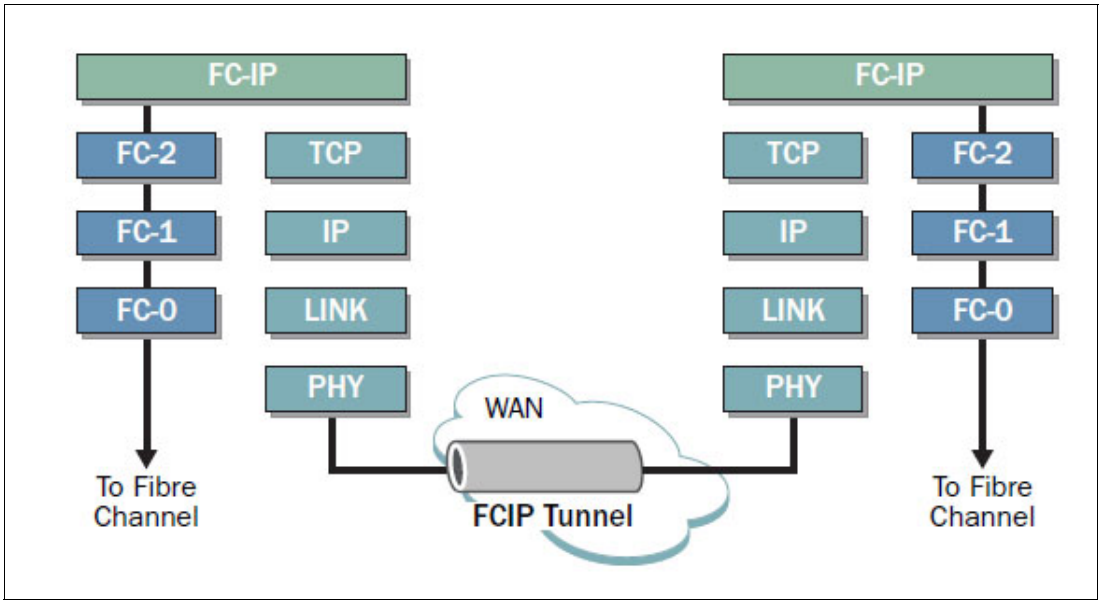


Figure 1-1 FCIP tunnel concept and TCP/IP layers

IP addresses and TCP connections are used only at the FCIP tunneling devices at each endpoint of the IP WAN “cloud.” Each IP network connection on either side of the WAN cloud is identified by an IP address and a TCP/IP connection between the two FCIP devices. An FCIP data engine encapsulates the entire Fibre Channel frame in TCP/IP and sends it across the IP network (WAN). At the receiving end, the IP and TCP headers are removed and a native Fibre Channel frame is delivered to the destination Fibre Channel node.

The existence of the FCIP devices and IP WAN cloud is transparent to the Fibre Channel switches, and the Fibre Channel content buried in the IP datagram is transparent to the IP network. TCP is required for transit across the IP tunnel to enforce in-order delivery of data and congestion control. The two FCIP devices in Figure 1-1 use the TCP connection to form a virtual inter-switch link (ISL), called a VE_Port, between them. They can pass Fibre Channel Class F traffic and data over this virtual ISL.

As a simple real-world analogy, this FCIP process is similar to writing a letter in one language and putting it into an envelope and sending it through the postal system to some other destination in the world. At the receiving end, the envelope is opened, and the contents of the letter are read. Along the route, those handling the letter do not have to understand the contents of the envelope, but only where its intended destination is and where it came from.

FCIP takes Fibre Channel frames regardless of what the frame is for (FICON, FCP, and so on) and places these into IP frames (envelopes) for transmission to the receiving destination. At the receiving destination, the envelopes are opened and the contents are placed back on the Fibre Channel network to continue their trip.

FCIP is standards based, and is discussed in detail in the IETF RFC 3821 document.

RFC 3821

The RFC 3821 specification is a 75-page document that describes mechanisms that allow the interconnection of islands of Fibre Channel storage area networks (SANs) to form a unified SAN in a single fabric by using connectivity between islands over IP. The chief motivation behind defining these interconnection mechanisms is a need to connect physically remote FC sites, allowing remote disk and DASD access, tape backup, and live/real-time mirroring of storage between remote sites.

The Fibre Channel standards have chosen nominal distances between switch elements that are less than the distances available in an IP network. Because Fibre Channel and IP networking technologies are compatible, it is logical to turn to IP networking for extending the allowable distances between Fibre Channel switch elements.

The critical, fundamental assumption made in RFC 3821 is that the Fibre Channel traffic is carried over the IP network in such a manner that the Fibre Channel fabric and all of the Fibre Channel devices in the fabric are unaware of the presence of the IP network. This means that the FC datagrams are delivered in time to comply with the existing Fibre Channel specifications. The Fibre Channel traffic can span local area networks (LANs), metropolitan area networks (MANs), and wide area networks (WANs), as long as this fundamental assumption is adhered to.

The role of TCP in FCIP Extension networks

The Transmission Control Protocol (TCP) is a standard protocol described by IETF RFC 793. Unlike UDP, which is connectionless, TCP is a connection-oriented transport protocol that guarantees reliable in-order delivery of a stream of bytes between the endpoints of a connection. TCP connections ensure in-order delivery of FC frames, error recovery, and lossless transmission in FCIP extension networks.

TCP achieves this by assigning each byte of data a unique sequence number, maintaining timers, acknowledging received data through use of Acknowledgements (ACKs), and retransmission of data if necessary. After a connection is established between the endpoints, data can be transferred. The data stream that passes across the connection is considered a single sequence of eight-bit bytes, each of which is given a sequence number.

TCP does not assume reliability from the lower-level protocols (such as IP), so TCP must guarantee this reliability itself.

FICON protocol emulation and FCIP

FICON device emulation and read/write tape pipelining technologies were first available on the Brocade USD-X and Brocade Edge M3000 extension products (formerly McDATA UltraNet Storage Director eXtended and UltraNet Edge Storage Router). These technologies provide for virtually unlimited distance extension of FICON tape and a popular mainframe disk mirroring solution from IBM, called Extended Remote Copy (XRC).

The Brocade USD-X and M3000 platforms with FICON emulation and pipelining capabilities set the industry standard used for FICON distance extension, and became the solution of choice for thousands of mainframe enterprises around the world. IBM and Brocade have used these technologies to expand the FICON extension capabilities of the IBM SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade for the IBM SAN512B-6 and SAN256B-6 FICON director platforms, setting yet another industry benchmark for extended FICON performance.

FICON emulation supports FICON traffic over IP WANs using FCIP as the underlying protocol. FICON emulation can be extended to support performance enhancements for these specific applications:

- IBM z/OS Global Mirror (formerly eXtended Remote Copy, or XRC)
- FICON Tape Emulation (tape read and write pipelining)

FCIP extension use cases in z Systems environments

FCIP extension provides an ideal RDR solution for the aforementioned FICON replication over long distance solutions such as IBM z/OS Global Mirror, and IBM tape read/write pipelining. Later chapters discuss the specific technology and features of FCIP in detail. For this section, note that the SAN42B-R and IBM Gen 6 Extension blade are also ideal for FCIP extension of asynchronous DASD RDR solutions such as IBM Global Mirror. In addition, the high availability features make them ideal for FCIP extension of synchronous DASD RDR solutions such as IBM Metro Mirror.

1.2.2 Internet protocol extension (IPEX)

Over the past several years, storage applications or replication solutions were developed that perform their replication by using IP instead of traditional mechanisms such as ESCON, FICON, or Fibre Channel over IP (FCIP).

Whereas a wide variety of IP replication solutions are available in the market, including a growing number of disk arrays and tape devices with native IP replication ports, these solutions are not optimized for replication over long distance. In a local or metropolitan environment, these replication solutions might deliver the needed performance and throughput to meet service levels and recovery objectives. However, when replicating over longer distance, these solutions have several inherent challenges that make it almost impossible to meet growing service level and recovery expectations. Such challenges include widespread replication throughput issues, network availability problems, and data security exposure.

Another challenge for array native IP replication solutions is that IP WAN connections are notorious for being problematic. In addition to network latency, WAN connections experience frequent disruptions and events that have enormous implications for replication traffic. Issues such as dropped packets, jitter, degraded or complete loss of network connectivity, and competing demands for bandwidth from the user community all negatively affect replication applications, making it difficult to achieve availability and recovery objectives.

These types of replication solutions were not designed to handle network interruptions. Keep in mind that each time a WAN link goes down, data in transit on the failed WAN link is lost, the replication application can time out and stop (including the main input/output (I/O) if performing synchronous replication), and the application goes into its restart/resync recovery mode. With each restart/resync, replication falls further and further behind. With the large (and growing) amount of data that needs to be replicated and the very high replication speeds, even a small unplanned outage can take days to recover from and can result in unrecoverable data.

To make matters worse, IP WAN connections typically involve multiple hops and often involve multiple service providers, making it complex and time-consuming to troubleshoot IP WAN problems. Native array IP replication solutions provide troubleshooting tools for issues related to the storage device itself, but they offer no visibility into the storage network or the state of the IP WAN connections. All that is known is that the replication time is exceeding target levels.

There is no proactive warning or insight that allows administrators to quickly identify and resolve potential issues. As a result, network issues require reactive management that affects operations and leads to downtime. Even ownership of storage network issues can be a challenge, often resulting in the assignment of blame within the IT organization and increased time to resolution.

Data security is another significant challenge for native IP replication solutions. Some arrays encrypt data to provide protection for data-at-rest but often do not provide encryption for data in-flight. This means that after the data leaves the confines of the secure data center, critical data is unprotected, making it vulnerable to security breaches, data theft, and "man-in-the-middle" attacks.

With the growing threats of hacking, snooping, and other high-profile cybercrimes, protecting data in-flight across the IP WAN is essential to meeting data security objectives. However, security cannot be achieved at throughput's expense. Some IP replication solutions do provide encryption of data in-flight as an option, but the performance penalty is unacceptable, often reducing throughput by 30 to 70 percent.

Unfortunately, there simply have not been IP switching platforms that had features similar to what FCIP devices offered. There is a benefit to using a technology designed for data storage traffic. The performance of this IP-based replication has not been on par with what was achievable with FCIP. The IBM SAN42B-R and the IBM b-type Gen 6 Extension Blade extension blade for the IBM SAN512B-6 and SAN256B-6 are the latest products from IBM that offer IPEX technology, and bring enterprise class extension benefits previously only available for FCIP extension, to IP storage replication.

The SAN42B-R and the IBM b-type Gen 6 Extension Blade incorporates several advanced technologies that are essential to ensuring maximum throughput and bandwidth utilization over distance, including these items:

- An aggressive Transmission Control Protocol (TCP) stack that is optimized for storage, called WAN-Optimized TCP (WO-TCP), a capability that is not available with solutions that use standard TCP stacks alone
- An advanced line-rate data compression architecture with three compression algorithms that are the most aggressive in the industry
- An encapsulation methodology that enables the industry's most efficient data transport across the WAN

These unique technologies combine to deliver industry-leading IP based replication performance over distance. They are discussed in more detail in later chapters.

IPEX use cases with IBM storage systems

The primary use case for the IPEX technology is with the IBM TS7700 family of products, specifically for the TS7760 or TS7700 Grid solution. IBM and Brocade jointly tested IPEX technology with the TS7760/7700 Grid. Significant performance improvements of IP replication using IPEX technology compared to native IP replication were discovered and documented. These performance improvements are greater as the distance increases between TS7700 clusters.

The performance of the grid replication using the IPEX technology also was significantly better than native IP replication when errors were introduced onto the WAN.

2

The IBM System Storage SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade

This chapter provides a high-level overview of the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade (FC 3892, 3893) software and hardware features.

It provides the following information:

- IBM SAN42B-R and IBM b-type Gen 6 Extension Blade product overview
- Hardware naming convention: IBM and Brocade
- Extension hardware architecture overview
- IBM SAN42B-R and IBM b-type Gen 6 Extension Blade software features
- IBM SAN42B-R hardware features
- Earlier b-type extension products
- Interoperability between IBM extension switches
- IBM Fabric Vision
- Terminology

2.1 IBM SAN42B-R and IBM b-type Gen 6 Extension Blade product overview

The IBM SAN42B-R Extension switch and the IBM b-type Gen 6 Extension Blade are enterprise-class extension switches designed to build high performance, highly available solutions for data replication and backup over two protocols:

- Fibre Channel over IP (FCIP)
- IP Extension (IPEX)

These switches integrate easily into any existing IP network and provide data replication over geography distributed data centers. They use cost-effective IP WANs to replicate open system logical unit number (LUN) or mainframe volume over distances that would otherwise be impossible, impractical, or too expensive with standard Fibre Channel connections that require dark fiber connections between data centers. IBM Extension switches support remote data replication (RDR), centralized backup, and data migration over very long distances.

2.2 Hardware naming convention: IBM and Brocade

Table 2-1 lists the b-type family products and their equivalent Brocade names. This publication refers to these switches using their IBM names.

Only feature codes are given for the extension blades because the machine type and model are associated with the IBM Storage Networking SAN512B-6 (8961-F08) and SAN256B-6 (8961-F04) directors in which they are installed.

Table 2-1 IBM b-type family product and Brocade equivalent names

IBM name	IBM machine type and model or feature code (blades only)	Brocade name
IBM SAN42B-R	2498-R42	7840
IBM b-type Gen 6 Extension Blade	Feature Code #3892	SX6
IBM SAN06B-R[a]	2498-R06	7800
8 Gbps Extension Blade (FC 3890)[b]	Feature Code #3891	Brocade FX8-24 Extension Blade

a. This switch cannot be configured for extension with the IBM SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade.
b. This blade cannot be configured for extension with the IBM SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade.

2.3 Extension hardware architecture overview

Each IBM SAN42B-R switch or extension blade contains two data processor (DP) complexes. DP complexes are synonymous with engines. Each DP complex contains a data processor attached to traditional switching ASICs, and consists of special purpose hardware for FCIP functions and multi-core network processors.

IBM SAN06B-R: The previous extension switch model, IBM SAN06B-R, contains one DP and cannot be configured for extension with the IBM SAN42B-R extension switch, the IBM b-type Gen 6 Extension Blade, and the 8 Gbps Extension blade (FC #3891).

The extension switches can be configured in either 10VE mode or 20VE mode. The 10VE mode accommodates nearly all environments, and is the default configuration. Changing the mode is disruptive because it requires you to restart the switch.

The IBM SAN42B-R can be connected to Fibre Channel through 16 Gbps FC ports (Gen5) and the IBM b-type Gen 6 Extension Blade can be connected to Fibre Channel through 32 Gbps FC ports. The Fibre Channel ports are not bound to a specific Data Processor. However, when you create a FC trunk, you need to consider FC port grouping. For more information, see 2.5.1, “IBM SAN42B-R Extension Switch” on page 32 and 2.5.2, “IBM b-type Gen 6 Extension Blade” on page 33.

You can configure the following switch modes:

- FCIP mode. In this mode, only FCIP traffic is sent over the extension tunnels. FCIP mode allows you to choose between 10VE mode and 20VE_Port modes:
  - In 10VE mode, a VE_Port can use all Fibre Channel bandwidth available, to a maximum of 20 Gbps per Data Processor.
  - In 20VE mode, a single VE_Port can use half the Fibre Channel bandwidth available, to a maximum of 10 Gbps per Data Processor. This option allows use of more VE_Ports, but at a lower maximum bandwidth.
- Hybrid mode. In this mode, FCIP traffic and IP traffic (IPEX) can be sent over the extension tunnels. In this mode, only 10VE_Port mode is available and the switch must be in 10VE mode before you can enable hybrid mode. Configuring the switch for hybrid mode is disruptive.

For detailed hardware configuration information, see 2.5, “IBM SAN42B-R hardware features” on page 32.

Figure 2-1 and Figure 2-2 on page 17 show components and connections for each DP complex in the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade. All connections that are shown in the illustrations are full-duplex and internal in the switch.

Figure 2-1 shows components and connections when the switch is enabled in 10VE mode.

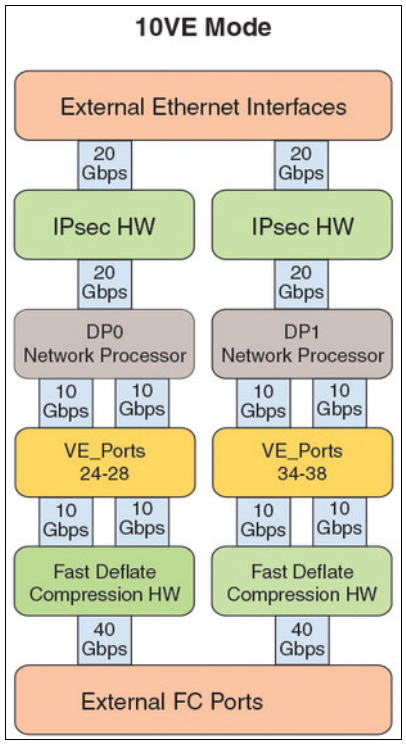


Figure 2-1 Data Processor components and VE_Port distribution in 10VE mode

Figure 2-2 illustrates components and connections for each DP complex when the switch is enabled in 20VE mode.

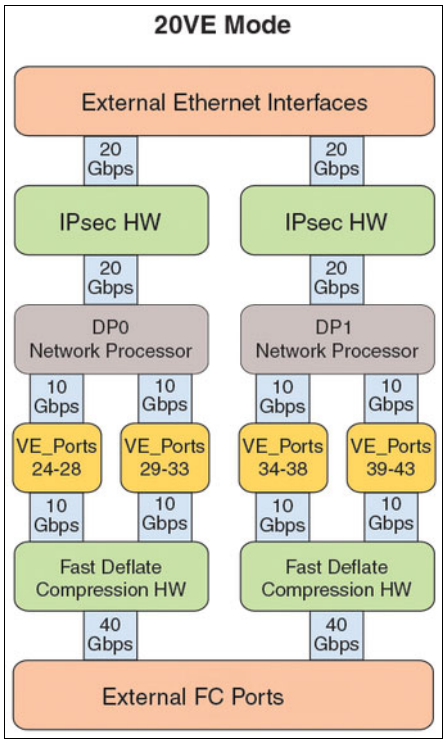


Figure 2-2 Data Processor components and VE_Port distribution in 20VE mode

2.4 IBM SAN42B-R and IBM b-type Gen 6 Extension Blade software features

This section describes extension and related software features.

2.4.1 Extension Trunking

This feature combines multiple WAN connections (FCIP circuits) into a single, logical, high-bandwidth, and highly available trunk. Extension Trunking provides high, aggregated bandwidth and shields end devices from IP network disruptions, making network path failures transparent to replication traffic.

Extension Trunking is a feature that is included in the base switch license. It includes the following key features:

- Allows you to aggregate the bandwidth of multiple 1/10 GbE or 40 GbE interfaces, which allows you to use WAN circuits from multiple service providers with different physical routes to ensure maximum availability.
- Supports aggregation of multiple WAN connections (up to eight circuits per trunk) with different latency or throughput characteristics.
- Enables configuration of redundant paths with an automatic failover and failback mechanism over WAN that can protect against transmission loss due to WAN failure.

- Provides lossless link loss (LLL), which ensures that all data lost in flight is retransmitted. This configuration prevents SCSI timeouts for open systems and interface control checks for mainframes. When there is a network path failure, extension trunking retransmits the lost packets and maintains data integrity without disruption.
- Provides granular load balancing on a weighted round-robin basis per batch.

Figure 2-3 shows an example of two tunnels that are trunking six circuits (physical connections) in Tunnel 1 (VE24) and two circuits in Tunnel 2 (VE34). Each tunnel operates on different Data Processors on an IBM SAN42B-R switch (DP0, DP1), and each circuit is assigned a unique IP address. In this case, each IP interface is assigned to a different Ethernet interface. The circuit flows from local IP interface to remote IP interface through the assigned Ethernet interfaces.

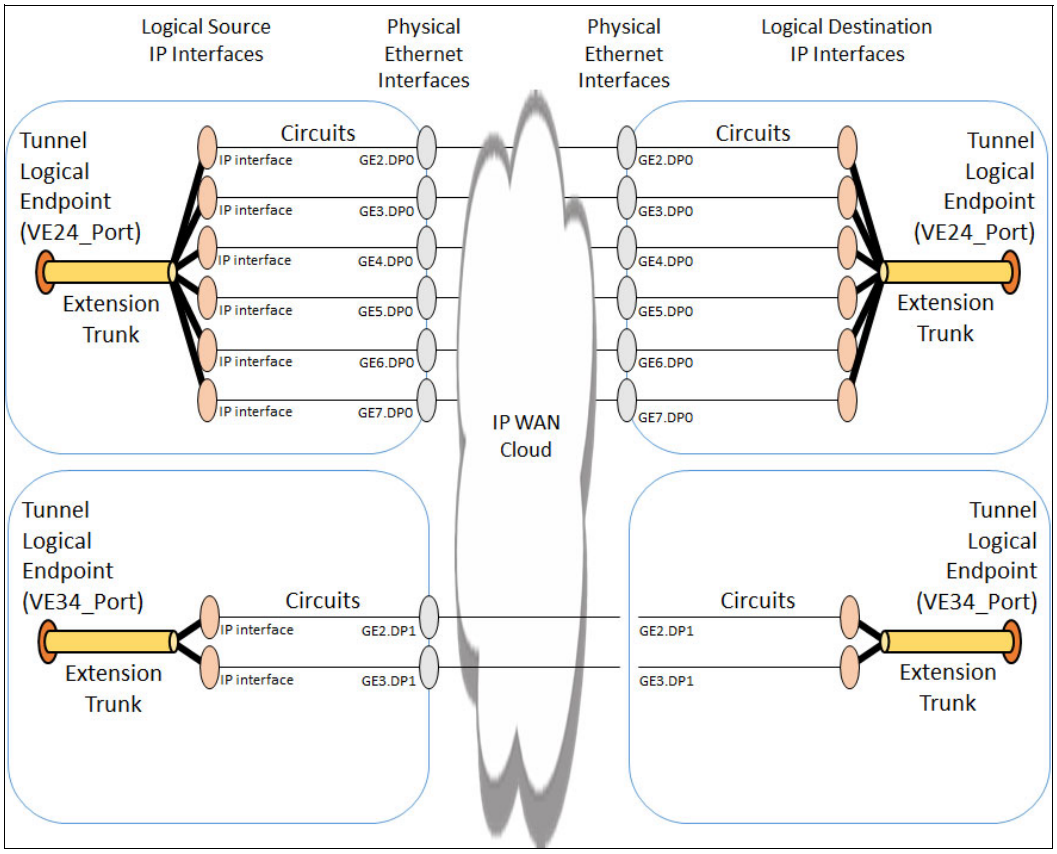


Figure 2-3 Tunnels, circuits, IP addresses, and Ethernet interfaces[1].

[1] The diagram comes from *Brocade Extension Architectures, Design and Best Practices Guide v2.0*, Mark Detrick.

Notes:

- The preferred method is to use one VE_Port to connect a fabric between two sites.
- As a preferred method, all tunnel and circuit settings should be identical on both sides of the tunnel.
- The difference between any two circuits on the same tunnel should not exceed the 4:1 ratio. This configuration is not required, but it is strongly preferred. For example, if one circuit is configured to 100 Mbps, any other circuit in that same tunnel should not exceed 400 Mbps.
- The maximum committed rate of a single circuit is 10 Gbps, whether configured on a 10 GbE or 40 GbE port.
- VEX_Ports are not supported on this platform.

2.4.2 Protocol optimization

Protocol extension offers an exceptional technology implementation for load balancing, failover, in-order delivery, and protocol optimization. It supports FastWrite disk I/O protocol optimization, OSTP, and FICON acceleration, and requires no additional hardware or licenses. Protocol acceleration requires a deterministic path both outbound and inbound. All FC frames from an exchange must pass through the same two VE_Ports in both directions. When more than one Data Processor is used, there is more than one path on which FC frames can be exchanged. In this situation, the remote side might choose to return these FC frames through a different path.

When you use protocol acceleration, it is necessary to confine traffic to a specific deterministic path bidirectionally. This configuration can be created in a few ways:

- Use only a single physical path, including a single VE_Port pair.
- Use Virtual Fabrics (VFs) with Logical Switches (LSs) that contain a single VE_Port pair.
- Configure Traffic Isolation Zones (TIZs).

All of these methods prevent the use of any type of load balancing and failover between VE_Ports.

Optimized traffic must be bidirectionally isolated to a specific path because protocol acceleration uses a state machine to perform the optimization. Protocol acceleration needs to know, in the correct order, what happened during a particular exchange. This configuration facilitates proper processing of the various sequences that make up the exchange until the exchange is finished, after which the state machine is discarded. These state machines are created and removed with every exchange that passes over the tunnel.

IBM SAN42B-R and IBM b-type Gen 6 Extension Blade have the capacity for tens of thousands of simultaneous state machines, or the equivalent number of flows, because they use 128 GB memory. The ingress FC frames are verified to be from a data flow that can be optimized. If a data flow cannot be optimized, it is merely passed across the tunnel without optimization. Therefore, if an end device does not support protocol optimization, it can share the same tunnel with other devices that support protocol optimization.

When you use IBM products, such as the IBM DS8000® and DS6000™ series, to prevent disruptions to your replication activities, do not enable protocol optimization on ISLs that are used for mirroring solutions such as Metro Mirror, Global Mirror, and Global Copy. These enterprise products use similar technology for remote replication and therefore do not benefit from write acceleration in the SAN. IBM SAN Volume Controller also does not use OSTP and FastWrite optimization.

2.4.3 Virtual Fabrics

Virtual Fabrics on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade play an important role in providing ways to achieve deterministic paths for protocol optimization, Virtual Fabrics is the preferred alternative over Traffic Isolation (TI) Zones to establish the deterministic paths that are necessary for protocol optimization. Protocol optimization requires that an exchange and all of its sequences and frames pass through the same VE_Port for both outbound and inbound. This configuration means that only a single VE_Port should exist within a virtual fabric logical switch.

By putting a single VE_Port in a logical switch, there is only one physical path between the two logical switches that are connected by using FCIP. A single physical path provides a deterministic path. When many devices or ports are connected for transmission across FCIP, as would be the case with tape for example, it is difficult to configure and maintain traffic isolation zones, whereas it is operationally simpler and more stable to use virtual fabric logical switch.

Configuring more than one VE_Port with one manually configured to have a higher Fabric Shortest Path First (FSPF) cost compared to the default is referred to as a *lay in wait* VE_Port. This configuration is not supported for protocol optimization such as FCIP-FW, OSTP, or FICON Emulation.

A “lay in wait” VE_Port can be used without protocol optimization and with RDR applications that can tolerate topology change with some frame loss. A small number of FC frames might be lost when you use “lay in wait” VE_Ports. If there are multiple VE_Ports within an logical switch, routing across those VE_Ports is performed according to the APT policy.

Virtual Fabrics are significant in mixed mainframe and open system environments. Mainframe and open system environments are configured differently and only virtual fabrics can provide autonomous logical switches accommodating the different configurations.

Here is a list of configuration differences between FICON and Open System environments that require virtual fabric logical switch when mixing environments on the same switch:

- The APT policy setting for FICON might not be the same as Open Systems.
- In-order delivery (IOD) can be used in FICON environments. IOD is disabled in Open Systems environments.
- Security ACLs are required in cascaded FICON environments. Security ACLs are not used in Open Systems environments.
- FICON Management Server (FMS) with Control Unit Port (CUP) can be enabled on FICON LSs.
- IBM Network Advisor management of the logical switch in FICON mode versus Open Systems mode.

Using a VE_Port in a selected logical switch does not preclude sharing an Ethernet interface with other VE_Ports from other logical switches. This is referred to as Ethernet Interface Sharing. For more information, see 2.4.4, “Ethernet interface sharing” on page 21. Figure 2-4 shows configuration of a circuit from two logical switches’ VE_Ports through a default switch to the WAN.

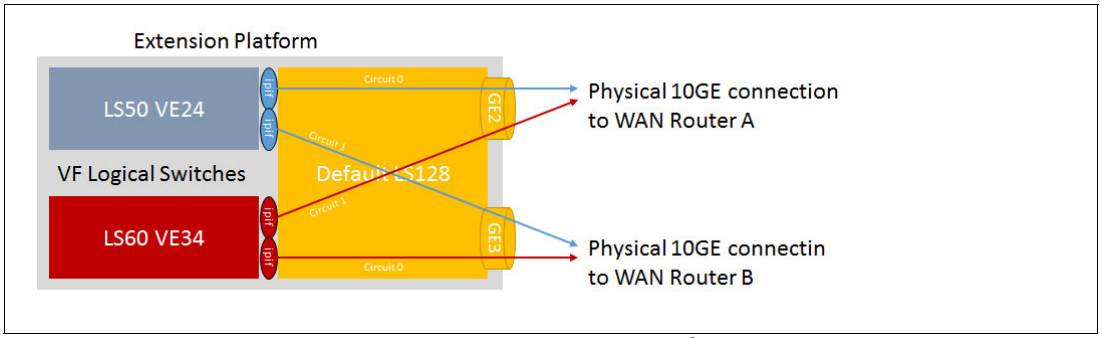


Figure 2-4 Detailed configuration of Logical Switches VE_Ports[2]

2.4.4 Ethernet interface sharing

An FCIP Trunk uses multiple Ethernet interfaces by assigning the circuits that belong to that trunk to different Ethernet interfaces. IP interfaces (ipif) are configured with IP addresses, subnet masks, and an Ethernet interface, which assigns the ipif to the interface. When the FCIP circuit is configured, the source IP address must be one that was used to configure an ipif, which in turn assigns the FCIP circuit to that Ethernet interface.

It is possible to assign multiple IP addresses and circuits to the same Ethernet interface by assigning multiple ipif to that same interface, each with its own unique IP address.

A circuit cannot be shared across more than one Ethernet interface. An IP address, an ipif, and a circuit can belong only to one Ethernet interface. Thus, if you want more than one Ethernet interface, you must use multiple circuits. If you attempt to configure the same IP address on more than one ipif, an error occurs.

It is possible to share an Ethernet interface with multiple circuits that belong to different virtual fabric logical switches. The Ethernet interface must be owned by the default switch (context 128). The ipif and iproute must also be configured within the default switch. The VE_Port is assigned to the logical switch that you want to extend with FCIP and is configured within that logical switch. The FCIP tunnel is also configured within that logical switch using the IP addresses of the ipif that are in the default switch. This technique permits efficient use of the Ethernet interfaces.

2.4.5 Circuit failover/failback and failover groups

Circuit failover/failback and failover groups are based on metrics that are assigned for each circuit. Each circuit has assigned metric 0 (primary, default value) or 1 (secondary). A circuit with metric 0 is used until a failure is encountered. The traffic is then automatically moved to a standby circuit with metric 1. Trunking first tries to retransmit any pending send traffic over another lowest metric circuit. Standby circuits are used only when all lower metric (0) circuits that are available within the tunnel fail.

[2] The diagram comes from *Brocade Extension Architectures, Design and Best Practices Guide v2.0*, Mark Detrick.

When the primary circuit is back, transfer is automatically moved back to it. Because of the mechanism and LLL, no data is lost and data remains in order during failover. All operations are transparent for upper layers, even if some frames are dropped.

Figure 2-5 shows a scenario where two circuits within a tunnel are configured as primary circuits with metric 0. This configuration means that normally both are used until a failure occurs. When one circuit fails, pending data is retransmitted in order over the remaining circuit.

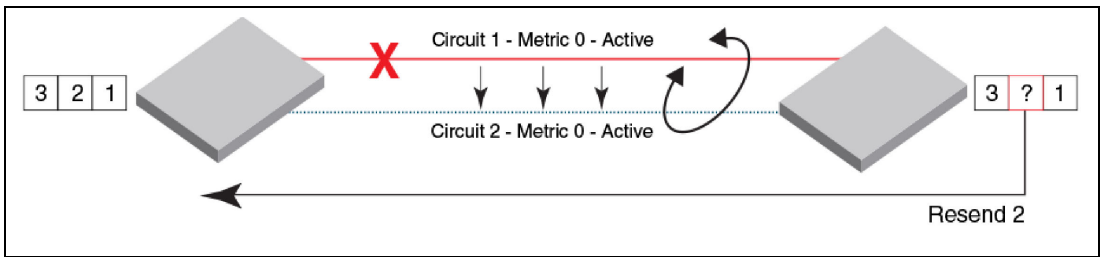


Figure 2-5 Link loss and retransmission over peer lowest metric circuit

Figure 2-6 shows a scenario where Circuit 1 within a tunnel is configured as the primary circuit with metric 0 and Circuit 2 as the standby circuit with metric 1. In this case, Circuit 2 is not used at all when Circuit 1 is up and running. Only when all primary circuits with a lower metric (0) are not available is data retransmitted over standby circuits with a higher metric (1).

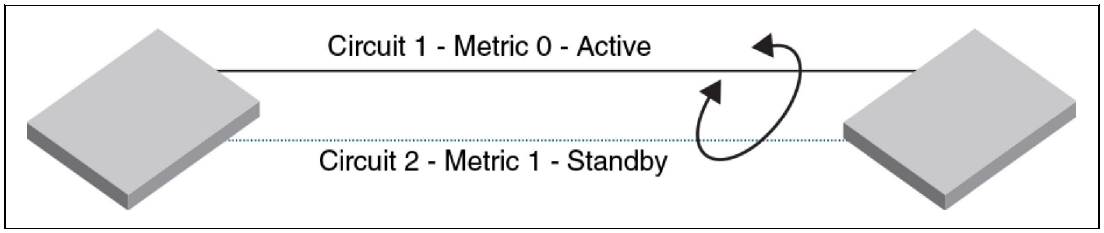


Figure 2-6 Active-Standby configuration with failover to a higher metric circuit

Because of failover group configuration, you can separate and dedicate circuits within a trunk. It is possible to create a maximum of four failover groups within a tunnel (VE_port), because the maximum number of circuits within a trunk is limited to eight. Failover groups allow you to define a set of metric 0 and metric 1 circuits that are part of a failover group.

When all metric 0 circuits in the group fail, metric 1 circuits take over operation, even if there are metric 0 circuits still active in other failover groups. With this configuration, you can better control which metric 1 circuits are activated if a metric 0 circuit fails. You could, for example, dedicate specific circuits for some critical systems with a dedicated HA solution.

Figure 2-7 shows an example where metrics and failover groups permit configuration of circuits over paths that should not be used unless the normal production path has gone down, for example, a backup path.

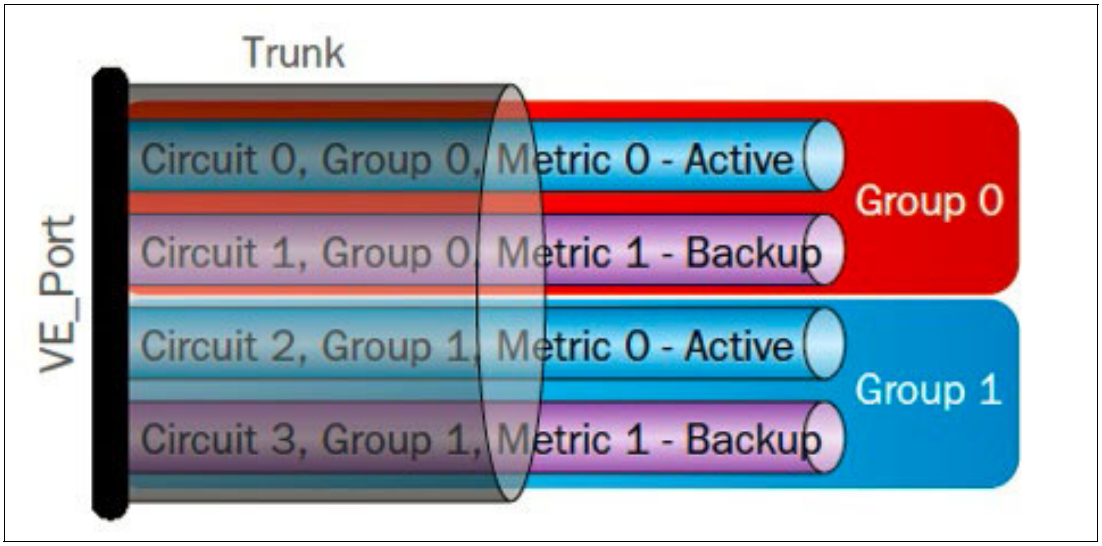


Figure 2-7 Extension Trunking: Circuit metrics and failover groups

Table 2-2 shows circuit details with failure group behavior.

Table 2-2 Tunnel with two failover groups with two circuits

Circuit	Failover group ID	Metric	When used
Circuit 0	FG 0	0	Active by default
Circuit 1	FG 0	1	Active only when Circuit 0 fails
Circuit 2	FG 1	0	Active by default
Circuit 3	FG 1	1	Active only when Circuit 2 fails

2.4.6 Circuit spillover

Circuit spillover is a new feature that is introduced in FOS 8.0.1 that allows you to define circuits that are used only when there is congestion or a period of high bandwidth utilization. This is a load-balancing method that also operates as a failover and failback scenario.

All standby circuits (metric 1) can be used automatically during high utilization or congestion periods in addition to primary (metric 0) circuits to provide HA with ad-hoc temporary additional bandwidth. Spillover is configured on a tunnel basis, and failover groups are not used when the tunnel is configured for spillover, and any failover groups that are defined are ignored.

When a tunnel is configured for spillover, traffic uses the metric 0 circuits until a high bandwidth utilization level is reached on those circuits. When bandwidth in metric 0 circuits is exceeded, the metric 1 circuits are also used in parallel to circuits with metric 0.

In general, the WAN utilization of 96 percent or greater must be sustained for approximately 15 to 20 seconds before spillover is activated. In contrast, when the utilization drops, the metric 1 spillover circuits are no longer used and only circuits with metric 0 are utilized. The metric 1 circuits are activated only if all the data cannot be sent over the metric 0 circuit and when the WAN utilization drops standby circuits are no longer utilized.

The utilization numbers might not be an exact value because they depend on the timing of how the data is collected. For example, a circuit might show 94 percent whereas the tunnel might show 93 percent high utilization. The circuit utilization is measured independently across all resource groups and the tunnel utilization is measured as a single resource group.

For example, a spillover tunnel with two circuits, one with metric 0 (primary) and the second with metric 1 (standby), is configured. Normally only circuit with metric 0 is used and a circuit with metric 1 acts as a standby circuit. When bandwidth utilization reaches approximately 96 percent for 15-20 seconds, both circuits (metric 0 and 1) are used until bandwidth decreases. With the same spillover configuration, HA is ensured, which means that in case of primary circuit (metric 0) failure, then the standby circuit (metric 1) is used automatically.

2.4.7 IPsec

The IPsec feature encrypts and decrypts data that is sent over WAN links with a standard 256-bit AES algorithm. IPsec is done on the hardware layer and introduces approximately 5 µs latency, so it can also be used for synchronous replications and does not introduce performance degradation.

The Extension Trunking feature is fully compatible with IPsec, and a mix of secure and non-secure tunnels can be configured on the same Ethernet port. IPsec supports network-level data integrity, data confidentiality, data origin authentication, and replay protection, so it protects data that is sent through the WAN from virtually every type of attack. It is an important feature for data replication between geographically distributed data centers when WAN security cannot be easily assured.

2.4.8 Adaptive Rate Limiting

The Adaptive Rate Limiting (ARL) feature is intended to maximize WAN utilization, especially when the WAN links are shared with other traffic. ARL uses dynamic bandwidth sharing between minimum (floor) and maximum (ceiling) rate limits to achieve maximum available performance based on conditions in the IP network and WAN. IBM SAN42B-R and IBM b-type Gen 6 Extension Blade sit in the middle of the aggregated IP storage flows and therefore they can monitor all data headed to the WAN and can manage it all properly.

ARL is performed on each circuit to change the rate at which the tunnel transmits data through the IP network. Circuit rate limiting is automatically adjusted based on available bandwidth. Absence of congestion events causes it to rise up towards the ceiling. A congestion event causes the rate limit to adjust down towards the floor.

ARL adjustments are not rapid, preventing additional congestion events from occurring. Over time, depending on network conditions, available bandwidth changes. If the available bandwidth has increased, ARL makes periodic attempts to adjust it upward. If adjusting it upward causes detected congestion, the rate remains stable.

When a circuit goes online or offline, ARL automatically adjusts the available IP bandwidth. ARL maintains utilization WAN bandwidth during periods of maintenance, firmware updates, or failures. ARL never attempts to exceed the maximum configured value and reserves at least the minimum configured value.

Figure 2-8 shows an ARL example. The blue circuit uses all available bandwidth (ceiling). When the yellow circuit starts to use bandwidth, ARL automatically reduces the available bandwidth for the blue one and established an equal balance for both flows. Next, the yellow circuit encounters a technical problem that causes the blue flow bandwidth to gradually increase again until the ceiling is reached.

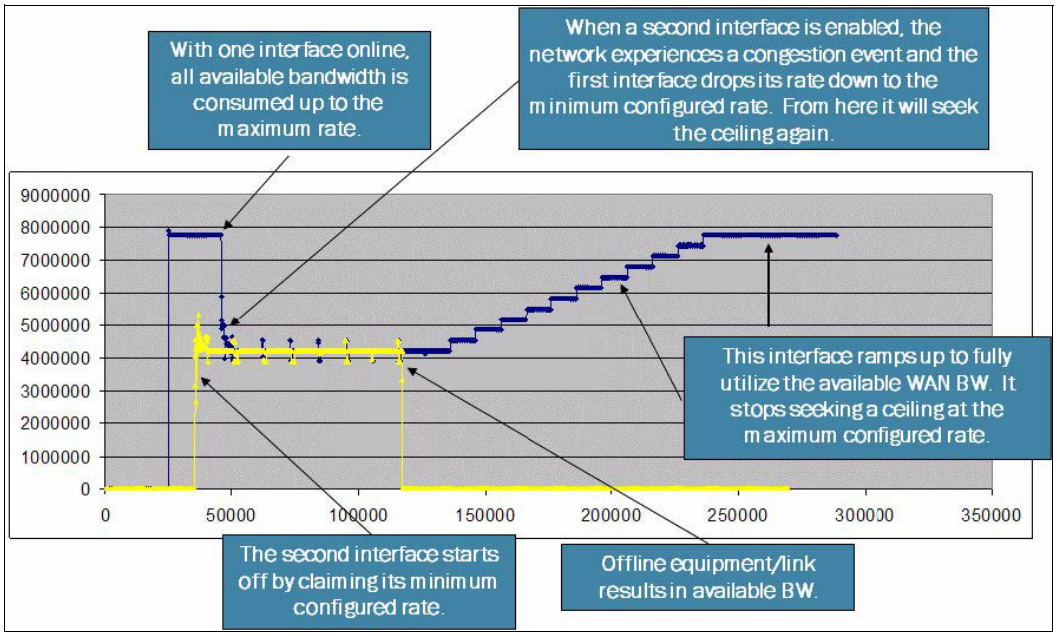


Figure 2-8 Adaptive Rate Limiting example for two circuits[3]

Note: The ratio between the minimum committed rate and the maximum committed rate for a circuit cannot exceed a 5:1 ratio.

2.4.9 Compression

Compression operates at both the hardware and software layer. This implementation is efficient and has almost no effect on other features, such as flow acceleration. When compression is used, the solution can be scaled up to 80 Gbps throughput (40 Gbps per Data Processor with compression ratio 4:1). When compression is used, hosts do not need to compress data, and it is possible to compress data that cannot be compressed by the host or storage array.

No CPU cycles are used for compression on the host side. Compression is turned off by default on extension switches, so you need to be aware of whether your data has already been compressed on host on the storage side, or you want to compress data on the extension switches. If your data has already been compressed, you probably will not be able to compress data more on the extension switches.

[3] This example is taken from *Brocade Extension Architectures, Design and Best Practices Guide v2.0*, Mark Detrick.

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade compress FC frames before they are encapsulated into FCIP packets. Compression is configured for each tunnel separately. A different compression option can be selected for each tunnel. There are four available compression options:

- None. No compression.
- Fast deflate. Pure hardware-based compression. This option provides the highest throughput per Data Processor before compression, but the least amount of compression. This option is suggested for synchronous replication. This is the only algorithm that is not available for IPEX.
- Deflate. Processor-based compression. It provides a lower speed than fast-deflate, but a faster speed than aggressive deflate. Deflate compression provides more compression than fast deflate, but is typically not as much compression as aggressive deflate. Data is precompressed per Data Processor.
- Aggressive deflate. Processor-based compression. Initiates the Data Processor in deflate mode with preference on compression. This mode is the slowest before compression, but typically provides the highest level of compression. Data is precompressed per Data Processor.

Auto compression is not supported on IBM SAN42B-R and IBM b-type Gen 6 Extension Blade.

Table 2-3 shows guidelines and compression algorithm comparison.

Table 2-3 Algorithms with tunnel bandwidth, supported protocols, and typical compression ratio

Algorithm	Total tunnel bandwidth links on each DP	Protocol support	Typical compression ratio
Fast deflate	More than 4 Gbps	FC	2:1
Deflate	2 Gbps - 4 Gbps	FC/IP	3:1
Aggressive deflate	2 Gbps or less	FC/IP	4:1

IBM makes no promises, guarantees, or any indication that a specific level of compression is possible for customer-specific data. Some data is highly compressible and some data cannot be compressed. The amount of application throughput varies depending on data compressibility and the selected compression mode.

2.4.10 Quality of service

Quality of service (QoS) improves and optimizes performance of critical systems when a contention of available WAN is observed. At this time, the critical systems' traffic is prioritized before less critical systems' traffic. Traffic flows performance is improved between initiators and targets within a tunnel. Frames with higher priority are sent first. The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade have PTQ (PerPriority TCP QoS) in which there are separate autonomous WO-TCP sessions for each QoS priority.

Each circuit handles one of the following priority traffic types:

- F class. F class is the highest priority, but is not available for IPEX. It uses a strict queue, which means that class-F frames (if any) are sent first without delay.
- QoS high. The default priority value is 50 percent of the available bandwidth.
- QoS medium. The default value is 30 percent of the available bandwidth.
- QoS low. The default value is 20 percent of the available bandwidth.

There is QoS in FC/FICON fabrics and across FC ISLs through virtual tunnels. There are different virtual tunnels for H/M/L/F-Class, each with its own set of Buffer-to-Buffer Credits (BBC) and flow control. There are five virtual tunnels for high, four virtual tunnels for medium, and two virtual tunnels for low.

After devices are assigned to these virtual tunnels, they use these virtual tunnels throughout the fabric. If data ingresses to an IBM extension switch through an ISL on a particular virtual tunnel, the data is automatically assigned to the associated TCP session for that priority. Devices that are directly connected to extension switches are also assigned to the proper priority TCP session based on the zone name prefix. The TCP marking is done at the IP layer by using Layer 3 Differentiated Services Code Point (DSCP) or at the Ethernet layer within the 802.1Q tag header using 802.1P. There are two options for TCP/IP network-based QoS:

- DSCP
- VLAN tagging and Layer 2 Class of Service (L2CoS)

If both FCIP and IPEX are being used simultaneously, the percentage of bandwidth that is applied to each during contention is configurable. The default group distribution is 50% FC and 50% IP.

Figure 2-9 shows the internal architecture of TCP connections that handle PP-TCP-QoS for one circuit.

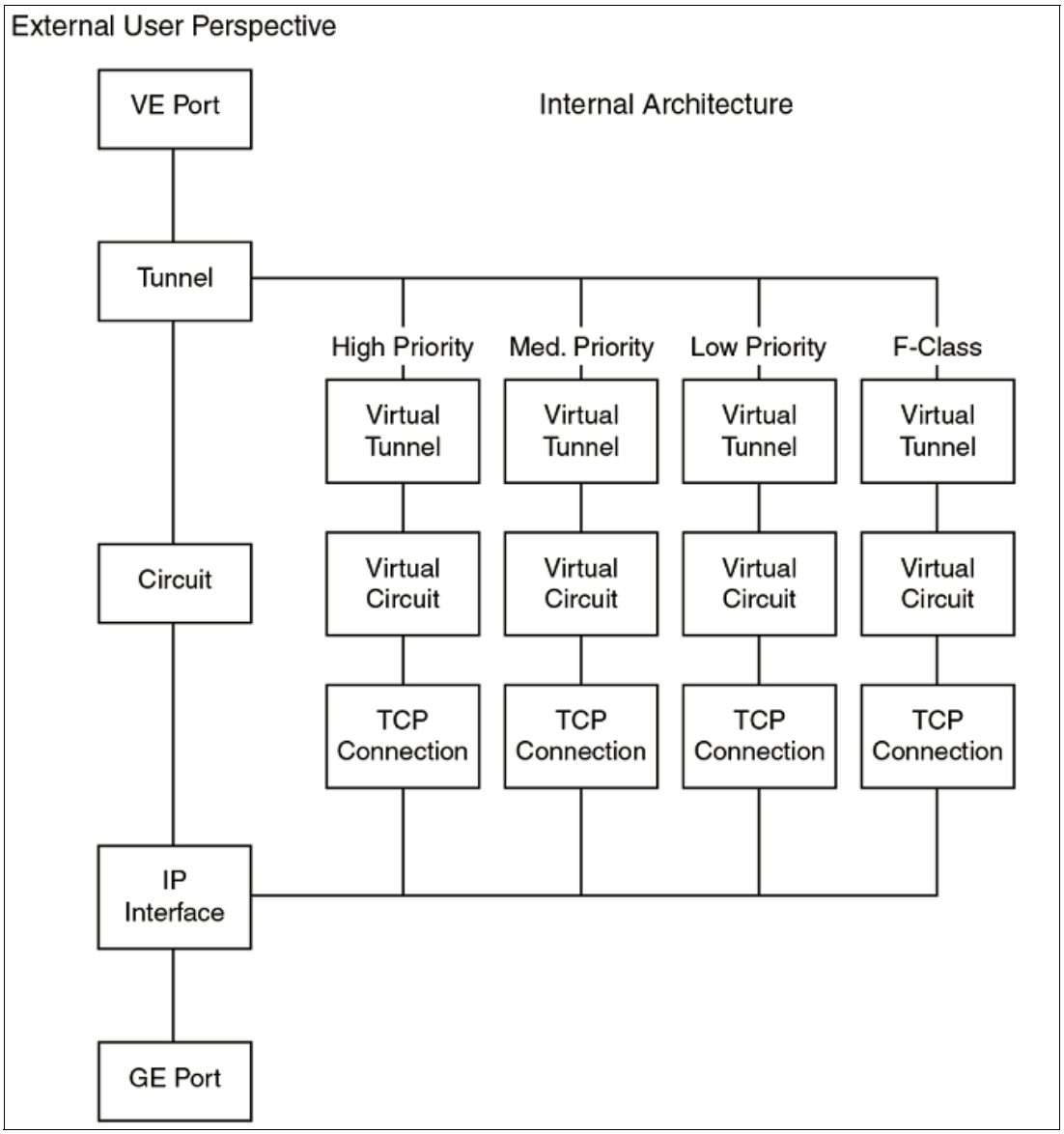


Figure 2-9 TCP connections for handling QoS

2.4.11 WAN Optimized TCP

Wan Optimized TCP (WO-TCP) is an aggressive TCP stack that optimizes TCP window size and flow control, which accelerates TCP transport for high-throughput storage applications by forming streams of bytes from storage I/O. IP Extension on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade introduces a new TCP stack named WO-TCP. It terminates IP storage TCP flows locally and transports the data across the WAN using WO-TCP.

The primary benefit here is the local ACK. By limiting ACKs only to the local data center, TCP that originates from an end IP storage device has to be capable of merely high-speed transport within the data center. Most native IP storage TCP stacks are capable only of high speeds over short distances.

Beyond the limits of the data center, “droop” becomes a significant factor. Droop refers to the inability of TCP to maintain line rate across distance. Droop worsens progressively as distance increases.

A stream of bytes is formed that is transported by WO-TCP. Sixteen data frames form a stream called a *batch*. Each batch has a single FCIP header, which reduces headers by 16:1. The batch is then compressed. By compressing the entire batch, it is possible to gain higher compression ratios.

The stream fills TCP segments to their maximum segment size. The maximum segment size is the IP maximum transmission unit (MTU) minus the IP and TCP headers (IP plus TCP headers is about 40 bytes). The result is full-size IP datagrams and minimal overhead, regardless of the compression.

TCP uses send and receive windows as its flow control mechanism. If an IP storage device on the receiving side needs to reduce an incoming flow, it closes the window. It is harmful to all flows that are using TCP if there is only one window for all data that is being transported. Having only one receive window means that all flows are slowed or halted. If only one end device asserts flow control, this can be a major problem for the whole environment.

The remedy is independent flow control for every stream. However, it is not practical to create a separate TCP connection for each flow because this configuration consumes excessive resources. Instead, autonomous streams are created using virtual TCP windows for each stream. The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade accommodate 512 streams per Data Processor. This configuration allows a total of 1024 streams for two DPs. Because a virtual TCP window is used for each stream, if a flow needs to slow down or stop, no other flows are affected and they continue to run at their full rate.

TCP provides these benefits:

- FCIP Batching is used to improve overall efficiency, maintain full utilization of links, and reduce *protocol overhead*. Simply put, a batch of FC frames is formed, after which the batch is compressed and processed as a single unit. This single unit is a “compressed byte stream,” and in this byte stream it is no longer relevant where frames begin and end. Frame boundaries become arbitrary for the purpose of transport across the tunnel. A batch consists of up to 16 FC frames (FCIP) or 4 FICON frames.

  All the frames must be from the same FCP exchange’s DATA_OUT sequence. SCSI commands, transfer readies, and responses are not batched. They are expedited by immediate transmission or protocol acceleration. If the last frame of the sequence arrives, the end-of-sequence bit is set, and the batch is known to be complete and processed immediately, even if fewer than 16 frames have been received.

  Figure 2-10 shows an example where two open systems batches are created. One is full with 16 frames and the other has two FC frames because of the set end-of-sequence bit.

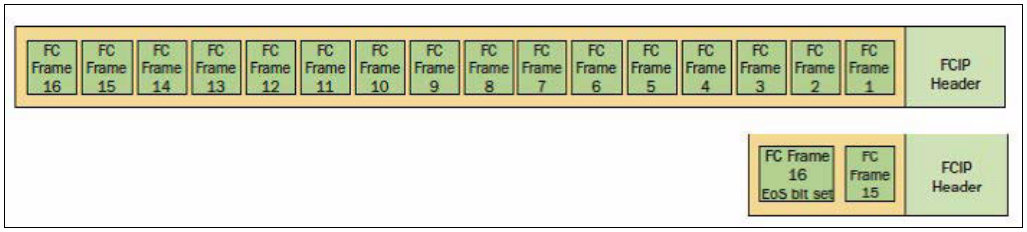


Figure 2-10 FCIP batch formation: Creating two batches.

- IP extension batching is a bit different from FCIP batching. IP extension batches are created on a stream basis. 1024 streams are supported on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade, which means 512 streams are supported per DP. A batch fills until it gets to 32 KB, after which no more IP datagrams are added.

  Figure 2-11 shows an example where the 16th IP datagram either meets or exceeds the 32-KB quantity. Therefore, it is the last IP datagram to be added to the batch. If data is arriving at 10 Gbps, it takes about 25 microseconds (µs) to fill a batch. If no more IP datagrams arrive for that stream, there is an up to 2 µs backstop timer that triggers the processing of that batch. Also, if a TCP frame is received with a PUSH Flag set, it triggers the processing of that batch. After an IP extension batch is formed, the processing of that batch is identical to that of FCIP batches.

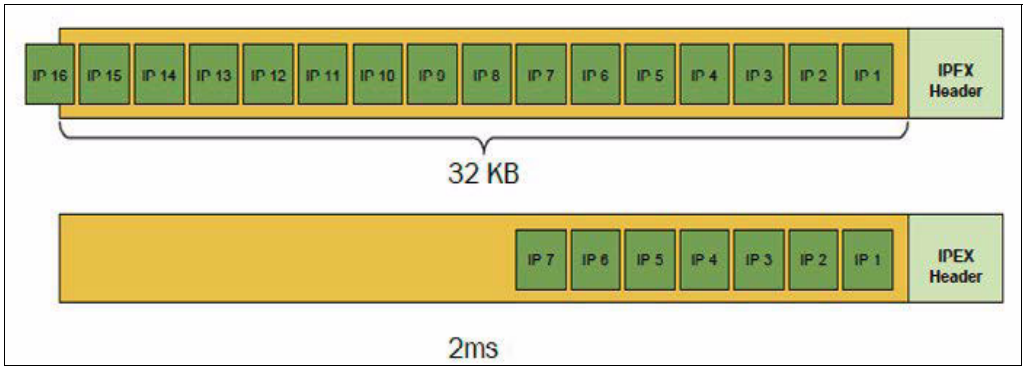


Figure 2-11 IP extension batch formation

2.4.12 Extension Hot Code Load

Extension Hot Code (HCL) allows non-disruptive firmware updates on Data Processors in the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade. An HCL tunnel provides high-availability (HA) support for FC traffic over the extension tunnels. HCL requires that you configure tunnels with four IP addresses per circuit, which includes both endpoints. The four addresses are the local and remote IP addresses.

The HA IP addresses are used when firmware is upgraded then traffic is failed over to the second Data Processor, and active FC traffic on Fibre Channel ports and VE_Ports are not disrupted. When the extension switch is operating in hybrid mode, tunnel groups are not supported by IP Extension, which means that IP traffic is disrupted during a firmware download.

The extension switch has two Data Processor complexes, referred to as DP0 and DP1. An Extension HCL firmware update occurs on one DP complex at a time. When a firmware update is initiated, the process always starts on DP0. Before DP0 is updated to the new firmware, traffic fails over to DP1 to maintain communication between the local and remote switch.

Figure 2-12 shows an example with a Main Tunnel that is used for normal operations when no firmware upgrade is in progress. This tunnel is normally created as an extension tunnel from a VE_Port. Another two tunnels (Local Backup Tunnel and Remote Backup Tunnel) are created automatically when the Main Tunnel is configured as HCL capable, and additional two HA-IP addresses are assigned.

The LBT is created upon specifying the local HA IP address for the circuit, and the RBT is created upon specifying the remote HA IP address for the circuit. All three tunnel groups (MT, LBT, and RBT) are associated with the same VE_Port. When an extension tunnel is configured to be HCL capable, the LBT and RBT tunnel groups are always present.

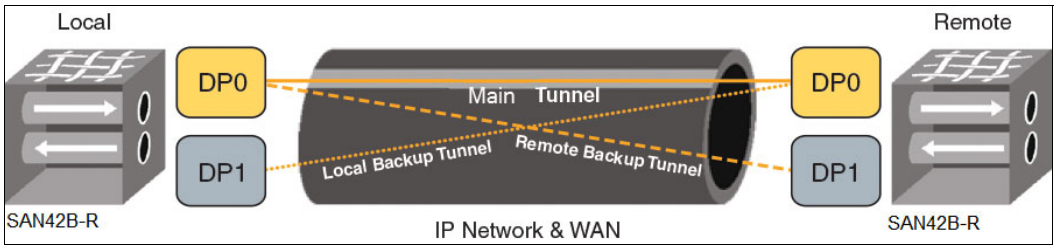


Figure 2-12 Extension HCL tunnels

2.4.13 Licensing

Table 2-4 shows the licensing options for IBM SAN42B-R and IBM b-type Gen 6 Extension Blade.

Table 2-4 License options

Product configuration	FC Ports	Ethernet Ports	WAN rate limiting
Base configuration	24x 16-Gbps (16x 16 Gbps for IBM b-type Gen 6 Extension Blade)	16x 1/10-GbE	5 Gbps
Medium configuration (Base + WAN Rate Upgrade 1)	24x 16-Gbps (16x 16 Gbps for IBM b-type Gen 6 Extension Blade)	16x 1/10-GbE	10 Gbps
Max configuration (Base + WAN Rate Upgrade 1 and WAN Rate Upgrade 2)	24x 16-Gbps (16x 16 Gbps for IBM b-type Gen 6 Extension Blade)	16x 1/10-GbE + 2x 40-GbE	Unlimited

All ports and interfaces on the switch are active except for the 40 GE interfaces. The 40 GE interfaces are enabled as part of WAN Rate Upgrade 2.

Note: The WAN rate limit is based on bandwidth, not on number of ports. The currently supported max WAN rate for Upgrade 1 and 2 with no compression is 40 Gbps.

The following features are available with the purchase of a specific license key for the IBM extension switches:

- Integrated Routing (IR)
- Advanced Acceleration for FICON
- FICON CUP
- WAN Rate Upgrade 1
- WAN Rate Upgrade 2

2.5 IBM SAN42B-R hardware features

This section describes basic IBM extension switch hardware features that are related to the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade.

2.5.1 IBM SAN42B-R Extension Switch

The IBM SAN42B-R is a 2U chassis extension switch that provides 2x Data Processors, 24x 16 Gbps FC ports (FC0-FC23) numbered 0 - 23 on the switch, 2x 40 GbE ports numbered 0 - 1, and 16x 1/10 GbE ports (ge2-ge17). Up to 20 VE_Ports are supported for tunnel configurations, with the default configuration being 10 VE_Ports. Figure 2-13 shows hardware ports and LED configuration.

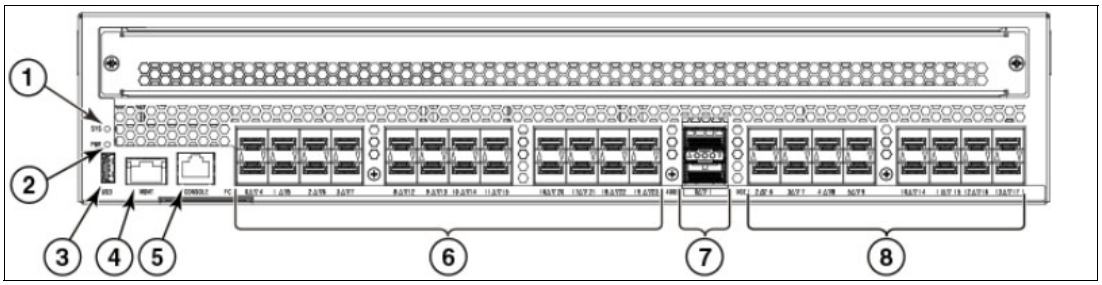


Figure 2-13 IBM SAN42B-R hardware ports and LEDs

The following components are shown in Figure 2-13:

1. System (SYS) status LED.
2. Power (PWR) LED.
3. USB port.
4. Ethernet management (mgmt) port.
5. Serial Console management port.
6. FC ports 0-23: 2, 4, 8, or 16 Gbps ports compatible with short wavelength (SWL), long wavelength (LWL), and extended long wavelength (ELWL) SFP+. FC ports can autonegotiate speeds with connecting ports.
7. 40 GbE QSFP ports 0 - 1 operate at 40 Gbps fixed speed.
8. 1 or 10 GbE SFP+ ports 2 - 17 operate at 10 Gbps or 1 Gbps fixed speeds with appropriate 10 Gbps or 1 Gbps transceivers installed.

When configuring a Fibre Channel trunk (ISL trunking) on the switch, be aware that there are three eight-port Fibre Channel port groups for configuring trunk groups or trunks on the blade:

- Port group 0: Ports 0-7
- Port group 1: Ports 8-15
- Port group 2: Ports 16-23

All ports in a trunk group must belong to the same port group. For example, to form an 8-port trunk, select all eight ports from port group 0 or port group 1. You cannot use ports from each port group for the trunk. You can use from 1 - 8 ports in a port group to form a trunk.

2.5.2 IBM b-type Gen 6 Extension Blade

The IBM b-type Gen 6 Extension Blade uses the Gen6 Condor-4 ASIC, supports the same extension features as IBM SAN42B-R. It provides 2x Data Processors, 16x 32 Gbps FC ports (FC0-FC15) numbered 0 - 15 on the switch, 2x 40 GbE ports numbered 0 - 1, and 16x 1/10 GbE ports (ge2-ge17) numbered 2 - 17. The IBM b-type Gen 6 Extension Blade can be installed in an empty Director's slot, only in slots 3 - 4 and 7 - 8, so the maximum configuration can contain four extension blades in the IBM Storage Networking SAN512B-6 (8961-F08) and SAN256B-6 (8961-F04) directors.

Up to 20 VE_Ports are supported for tunnel configurations, the default configuration being 10 VE_Ports. Figure 2-14 shows hardware port numbering.

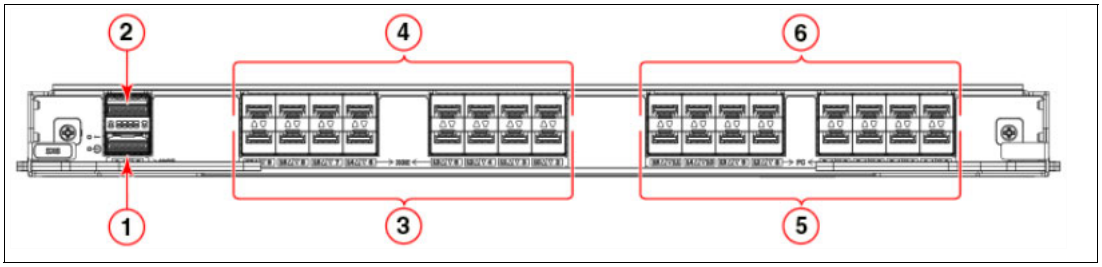


Figure 2-14 IBM b-type Gen 6 Extension Blade port numbering

The following components are shown in Figure 2-14:

1. 40 GbE QSFP port 0 operates at 40 Gbps fixed speed.
2. 40 GbE QSFP port 1 operates at 40 Gbps fixed speed.
3. 10/1 GbE SFP+ ports 2-9 (right to left) operate at 10 Gbps or 1 Gbps fixed speeds with appropriate 10 Gbps or 1 Gbps transceivers installed.
4. 10/1 GbE SFP+ ports 10 - 17 (right to left) operate at 10 Gbps or 1 Gbps fixed speeds with appropriate 10 Gbps or 1 Gbps transceivers installed.
5. FC ports 0-7 (right to left), support 32 Gbps transceivers operating at 8, 16, or 32 Gbps, 16 Gbps transceivers operating at 4, 8, or 16 Gbps, and 10 Gbps transceivers operating at fixed 10 Gbps.
6. FC ports 8-15 (right to left), support 32 Gbps transceivers operating at 8, 16, or 32 Gbps, 16 Gbps transceivers operating at 4, 8, or 16 Gbps, and 10 Gbps transceivers operating at fixed 10 Gbps.

When configuring a Fibre Channel trunk (ISL trunking) on the switch, be aware that there are two eight-ports Fibre Channel port groups for configuring trunk groups or trunks on the blade:

- Port group 0: ports 0 - 7
- Port group 1: ports 8 - 15

All ports in a trunk group must belong to the same port group. For example, to form an 8-port trunk, select all eight ports from port group 0 or port group 1. You cannot use ports from each port group for the trunk. You can use 1 - 8 ports in a port group to form a trunk.

2.5.3 VE_Port assignment

You can have a maximum of 20 VE_Ports on the switch. In the default 10VE mode, only 10 VE_Ports are enabled. In 20VE mode, all 20 VE_Ports are enabled. When the switch operates in hybrid mode, 20VE mode is not allowed. The 10VE mode accommodates nearly all environments, but the mode change is disruptive because it requires rebooting the switch. Detailed VE_Port assignment to a specific Data Processor is described here:

- In 10VE mode, 10 of the 20 VE_Ports are disabled. In 10VE mode, a VE_Port can use all Fibre Channel bandwidth available to the DP complex where it resides, up to a maximum of 20 Gbps. These are VE_Ports 29 - 33 and 39 - 43. Five VE_Ports are enabled on each DP complex as follows:
  - VE_Ports 24 - 28 are controlled by DP0.
  - VE_Ports 34 - 38 are controlled by DP1.
- In 20VE mode, there are four VE_Port groups and all 20 VE_Ports are enabled. In 20VE mode, a single VE_Port on a DP complex can use half the Fibre Channel bandwidth available to its DP complex, a maximum of 10 Gbps. Each port group can share 10 Gbps. When the switch operates in hybrid mode with IPEX, 20VE mode is not allowed.
  - VE_Ports 24 - 28 and VE_Ports 29 - 33 are controlled by DP0.
  - VE_Ports 34 - 38 and VE_Ports 39 - 43 are controlled by DP1.

2.6 Earlier b-type extension products

The IBM SAN06B-R and 8 Gbps Extension Blade (FC 3890) provide support for extension. However, they cannot be combined with the Storage SAN42B-R extension switch and IBM b-type Gen 6 Extension Blade in an extension tunnel.

2.6.1 IBM SAN06B-R (FC 7732)

The IBM SAN06B-R is an entry-level, rack mounted extension switch that is designed mainly for small and medium solutions that requires FCIP replication. The switch provides up to 16x 8 Gbps Fibre Channel ports and 6x 1 GbE ports, and aggregate bandwidth of up to 128 Gbps for non-blocking Fibre Channel switching and up to 6 Gbps for FCIP.

Dual redundant, hot-swappable power supplies with integrated fans maximize availability and minimize outages. In comparison to the new model IBM SAN42B-R, IBM SAN06B-R is still supported, but does not offer new features, such as 10 GbE and 40 GbE ports, Hot Code Load, and IPEX support. It uses only one Data Processor.

For more information about the IBM SAN06NB-R, see the following links:

- *IBM System Storage SAN06B-R Extension Switch*, TIPS1126

  http://www.redbooks.ibm.com/abstracts/tips1126.html
- *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331

  http://www.redbooks.ibm.com/abstracts/sg248331.html

- Brocade 7840 Extension Switch Technical Specifications

  http://bit.ly/2cTiPEc

2.6.2 IBM 8 Gbps Extension Blade (FC 3890)

The 8 Gbps Extension Blade (FC 3890) accelerates and optimizes replication, backup, and migration over any distance. The 12x 8 Gbps Fibre Channel ports, 10x 1 GbE ports, and up to 2x optional 10 GbE ports provide Fibre Channel and FCIP bandwidth, port density, and throughput for maximum application performance over IP wide area network (WAN) links.

In comparison to the new model IBM b-type Gen 6 Extension Blade port numbering, 8 Gbps Extension Blade (FC 3890) is still supported, but does not offer new features such as 40 GbE ports, Hot Code Load, and IPEX support. The extension blades can be installed in the IBM SAN512B-6 and IBM SAN256B-6 Gen6 Directors.

For more information about the 8 Gbps Extension Blade (FC 3890), see the following links:

- *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331

  http://www.redbooks.ibm.com/abstracts/sg248331.html

- *IBM Storage Networking SAN512B-6 and SAN256B-6 Directors*, REDP-5398

  http://www.redbooks.ibm.com/abstracts/redp5398.html

2.7 Interoperability between IBM extension switches

The IBM SAN42B-R, IBM b-type Gen 6 Extension Blade, IBM SAN06B-R, and 8 Gbps Extension Blade (FC 3890) all use the same IBM Fabric OS that supports the entire IBM storage networking product family. This technique helps ensure seamless interoperability with advanced features such as Fabric Vision technology, Integrated Routing, and Extension Trunking.

Note: Although IBM SAN42B-R, IBM b-type Gen 6 Extension Blade, and 8 Gbps Extension Blade (FC 3890) are compatible with current and previous generation Fibre Channel switches, the IBM SAN42B-R switch and IBM b-type Gen 6 Extension Blade do not support extension connections to the IBM SAN06B-R 8 Gbps Extension Blade (FC 3890).

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade can only connect to other IBM SAN42B-R switches and IBM b-type Gen 6 Extension Blades. You cannot connect extension tunnels created on an IBM SAN42B-R or IBM b-type Gen 6 Extension Blade to interfaces on any previous generation models other than these two switches.

Table 2-5 shows an interoperability matrix for extension connections between IBM extension switches.

Table 2-5 Extension connections interoperability matrix

Platform connection	IBM SAN42B-R	IBM b-type Gen 6 Extension Blade	IBM SAN06B-R	8 Gbps Extension Blade (FC 3890)
IBM SAN42B-R	✓	✓	-	-
IBM b-type Gen 6 Extension Blade	✓	✓	-	-
IBM SAN06B-R	-	-	✓	✓
8 Gbps Extension Blade (FC 3890)	-	-	✓	✓

Table 2-6 shows a comparison of platform capabilities.

Table 2-6 Extension capabilities by platform

Capabilities	IBM SAN42B-R	IBM b-type Gen 6 Extension Blade	IBM SAN06B-R	8 Gbps Extension Blade (FC 3890)
Extension Trunking	Yes	Yes	Yes	Yes
Adaptive Rate Limiting	Yes	Yes	Yes	Yes
10 GbE ports	Yes	Yes	No	Yes
40 GbE ports	Yes	Yes	No	No
FC ports	Yes (2, 4, 8, 16 Gbps)	Yes (4, 8, 16, 32 Gbps)	Yes (1, 2, 4, 8 Gbps)	Yes (1, 2, 4, 8 Gbps)
Compression	Yes Deflate, Aggressive Deflate, Fast Deflate	Yes Deflate, Aggressive Deflate, Fast Deflate	Yes LZ and Deflate	Yes LZ and Deflate
Protocol acceleration ► IBM Fastwrite ► Open Systems Tape Pipelining ► OSTP read ► OSTP write	Yes	Yes	Yes	Yes

Capabilities	IBM SAN42B-R	IBM b-type Gen 6 Extension Blade	IBM SAN06B-R	8 Gbps Extension Blade (FC 3890)
QoS ► Marking DSCP ► Marking 802.1P - VLAN tagging ► Enforcement 802.1P - VLAN tagging	Yes	Yes	Yes	Yes
FICON extension ► FICON emulation ► IBM z/OS Global Mirror acceleration ► Tape read acceleration ► Tape write acceleration ► Teradata emulation	Yes	Yes	Yes	Yes
IPsec ► AES-256-GCM ► SHA-512 HMAC ► IKEv2	Yes	Yes	Yes	Yes
VEX_Ports	No	No	Yes	Yes
Support for third-party WAN optimization hardware	No	No	Yes[a]	Yes[a]
IPv6 addresses for extension tunnels	Yes	Yes	Yes[b]	Yes[b]
Support for jumbo frames	Yes IP MTU of 9216 is maximum	Yes IP MTU of 9216 is maximum	No IP MTU of 1500 is maximum	No IP MTU of 1500 is maximum
Path Maximum Transmission Unit (PMTU) Discovery	Yes Maximum discoverable size is 9100 bytes.	Yes Maximum discoverable size is 9100 bytes.	No	No
Hot Code Load (Extension HCL)	Yes	Yes	No	No
IP Extension	Yes	Yes	No	No

a. Not supported in Fabric OS v7.0 and later.
b. IPv6 addressing is not supported in conjunction with IPsec.

Table 2-7 shows an IP Extension capabilities summary.

Table 2-7 IP Extension capabilities by platform

Capabilities	SAN42B-R	IBM b-type Gen 6 Extension Blade
Hybrid mode (FCIP and IP Extension)	Yes	Yes
Link access group (static LAG)	Yes	Yes
Switch virtual interface (SVI) IP interface (ipif)	IP traffic through an IBM extension tunnel.	IP traffic through an IBM extension tunnel.
IP compression	Deflate and aggressive deflate IP compression options are supported, but not fast deflate.	Deflate and aggressive deflate IP compression options are supported, but not fast deflate.
Traffic control list (TCL)	Yes	Yes
LAN side jumbo frames	Yes	Yes
Policy based routing (PBR)	Yes	Yes

2.8 IBM Fabric Vision

Fabric Vision technology is an advanced hardware and software architecture. It combines capabilities from Fabric Operating System (FOS), b-type devices, and IBM Network Advisor. It helps administrators address problems before operations are affected, and helps organizations meet their service level agreements (SLAs). IT organizations with large, complex, or highly virtualized data center environments often require advanced tools to help them more effectively monitor and manage their storage infrastructure.

Fabric Vision technology includes several breakthrough diagnostic, monitoring, and management capabilities that dramatically simplify day-to-day SAN administration and provide unprecedented visibility across the storage network. It automatically detects problems in the environment, takes predefined actions, and delivers integrated advanced monitoring and diagnostic tools that help increase your fabric resiliency, reduce downtime, and optimize performance. It is possible to fully monitor your IBM extension switches and blades with the IBM Network Advisor GUI.

Figure 2-15 shows an example of how easy it is to monitor GE_Ports utilization between two switches (sites) with a dashboard. For more information and examples, see Chapter 7, “Troubleshooting and monitoring” on page 213.

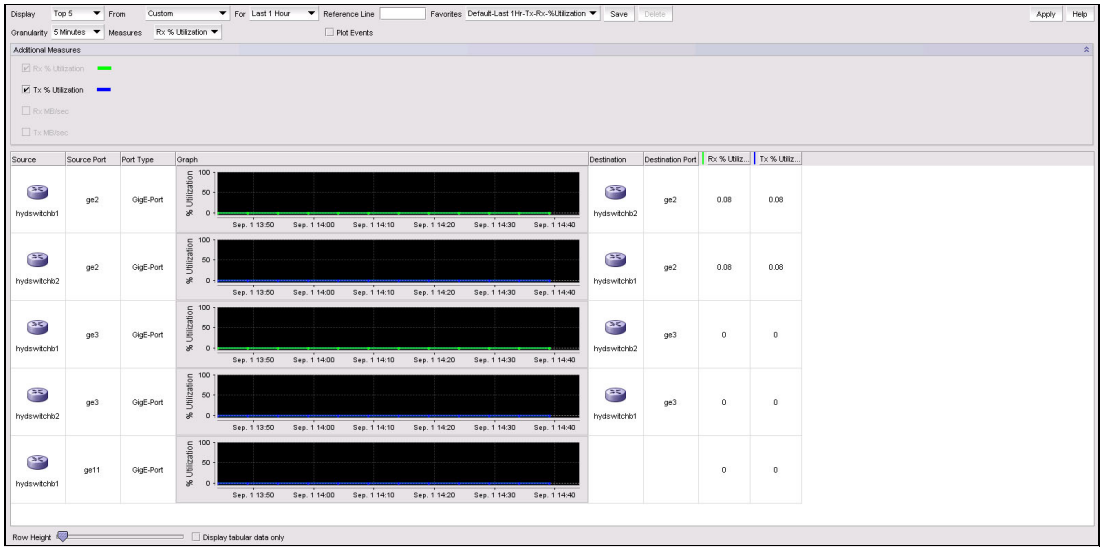


Figure 2-15 IBM Network Advisor dashboard with extension feature monitoring

Fabric Vision technology includes the following features for storage extension management:

- Monitoring and Alerting Policy Suite (MAPS) provides a prebuilt, easy-to-use solution policy-based threshold monitoring and alerting tool that proactively monitors storage extension network health based on a comprehensive set of metrics at tunnel, circuit, and QoS (tunnel and circuit) layers.

  MAPS simplifies fabric-wide threshold configuration, monitoring, and alerting. Administrators can configure multiple fabrics at one time using predefined or customized rules and policies for specific ports or switch elements through IBM Network Advisor. MAPS offers the following benefits:

  – Pre-defined monitoring groups and pre-validated monitoring policies that users can take advantage of. Pre-defined monitoring groups include switch ports attached to servers, switch ports attached to storage, E_Ports, short-wavelength SFPs, long-wavelength SFPs, and more. Pre-defined monitoring policies include aggressive, moderate, and conservative policies based on monitoring thresholds and actions.

  – Flexibility to create custom monitoring groups, such as switch ports attached to high-priority applications and another group of switch ports attached to low-priority applications, and monitor each group according to its own unique rules.

  – Flexible monitoring rules to monitor a given counter for different threshold values and take different actions when each threshold value is crossed. For example, users can monitor a CRC error counter at a switch port and generate a RASlog when the error rate reaches two per minute, send an email notification when the error rate is at five per minute, and fence a port when the error rate exceeds ten per minute.

  – Ability to monitor both sudden failures and gradually deteriorating conditions in the switch. For example, MAPS can detect and alert users if a CRC error counter suddenly increases to five per minute, or gradually increases to five per day.

- Support for multiple monitoring categories, enabling monitoring of the overall switch status, switch ports, SFPs, port blades, core blades, switch power supplies, fans, temperature sensors, security policy violations, fabric reconfigurations, CPU and memory utilization, traffic performance, FCIP circuit health, and so on.
- Support for multiple alerting mechanisms (RAS logs, SNMP traps, email notifications) and actions such as port fencing when errors exceed the specified threshold.

The new features introduced in IBM SAN42B-R and IBM b-type Gen 6 Extension Blade are described in Table 2-9 on page 42. For more information and detailed procedures for configuring new features related to the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade, see Chapter 7, “Troubleshooting and monitoring” on page 213.

- Fabric Performance Impact (FPI) monitoring uses predefined MAPS policies to automatically detect and alert administrators to different latency severity levels, and to identify slow drain devices that could affect the network. It uses advanced monitoring capabilities and intuitive MAPS dashboard reporting to indicate various latency severity levels.

  FPI monitoring is easy to configure and provides automatic mitigation or recovery from the effects of slow drain device. It replaces the previous FOS Bottleneck Detection feature.

- Dashboards provide integrated dashboards that display overall SAN and IP extension health, enabling administrators to get instant visibility into any hot spots at a switch level and quickly pinpoint issues occurring on a switch or in a fabric. These dashboard views are available:
  - Overall status of the switch health and the status of each monitoring category, including any out-of-range conditions and the rules that were triggered.
  - Historical information about the switch status for up to the last seven days that automatically provides raw counter information for various error counters. This integrated dashboard view also provides a single collection point for all dashboard data from a fabric for a specific application flow.

- Flow Vision helps administrators to identify, monitor, and analyze specific application flows to simplify troubleshooting, maximize performance, avoid congestion, and optimize resources. Flow Vision includes the following features:
  - Flow Learning enables administrators to non-disruptively, automatically learn (discover) all flows that go to or come from a specific host port or a storage port, or traverse ISLs, IFLs, or FCIP tunnels to monitor fabric-wide application performance. In addition, administrators can discover top and bottom bandwidth-consuming devices and manage capacity planning. This information helps administrators to discover what flows are active on a port without having to explicitly identify all the devices.
  - Flow Monitoring provides comprehensive visibility into flows within the fabric. Flows can be monitored from a specific host to multiple LUNs, from multiple hosts to a specific LUN, or across a specific ISL, IFL, or FCIP tunnel. Moreover administrators can perform LUN-level monitoring of specific frame types to identify resource contention or congestion that is impacting application performance.

    Examples of the frame types that can be monitored include SCSI Aborts, SCSI Read, SCSI Write, SCSI Reserve, and all rejected frames.
  - Flow Mirroring provides the ability to non-disruptively create real-time copies of specific application and data flows or frame types that can be captured for in-depth analysis. Flow Mirror is used for deeper analysis of flows of interest or specific frame types, such as analysis of SCSI Reservation frames, ABTS frames, or flows going to a bottlenecked device. Flow Mirroring allows you to analyze a live system without disturbing existing connections.

– Flow Generator provides a built-in traffic generator for pretesting and validating storage extension infrastructure, including route verification, QoS zone setup, extension trunking configuration, WAN access, IPsec policy setting, and integrity of optics, cables, and ports for robustness before the new infrastructure is deployed.

The new features introduced in the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade are described in Table 2-9 on page 42. For detailed procedures about configuring new features related to IBM SAN42B-R and IBM b-type Gen 6 Extension Blade, see Chapter 7, “Troubleshooting and monitoring” on page 213.

- Configuration and Operational Monitoring Policy Automation Services Suite (COMPASS) simplifies deployment, ensures consistency, and increases operational efficiencies of larger environments with automated switch and fabric configuration services. Administrators can configure a template or adopt an existing configuration as a template, and seamlessly deploy the configuration across the fabric. In addition, they can ensure that settings do not drift over time with COMPASS configuration and policy violation monitoring within IBM Network Advisor dashboards.
- Forward Error Correction (FEC) provides a data transmission error control method by including redundant data (error-correcting code) to ensure error-free transmission on a specified port or port range. When FEC is enabled, it can correct one burst of up to 11-bit errors in every 2112-bit transmission, whether the error is in a frame or a primitive.

  FEC is enabled by default. It is supported on E_Ports, and on the N_Ports and F_Ports of an access gateway by using the RDY, Normal (R_RDY), or Virtual Channel (VC_RDY) flow control modes.

  FEC operates on 16 Gbps and 32 Gbps ports only, and is mandatory in Gen 6 links to support 32 Gbps performance. It is enabled automatically when negotiation with a switch detects FEC capability and persists after the driver reloads and the system reboots. It functions with features such as QoS, trunking, and BB_Credit recovery. It is not supported on ports with DWDM devices.
- Credit Loss Recovery helps overcome performance degradation and congestion due to buffer credit loss, so it can significantly enhance application availability. It automatically detects and recovers buffer credit loss at the Virtual Channel (VC) level.
- ClearLink Diagnostic Port (D_Port) mode enables you to convert a Fibre Channel port into a diagnostic port for testing link traffic and running electrical loopback and optical loopback tests. The test results can be useful in diagnosing various physical layer issues without the need for special optical testers.

  In addition, it enables you to make tests to measure latency and distance across the switch links. It can be done when there are suspected physical layer issues or prior to implementation. With ClearLink Diagnostics, only the ports attached to the link being tested need to go offline, allowing the rest of the ports to continue to operate online.

Table 2-8 summarizes Fabric Vision features and support by IBM extension products.

Table 2-8 Fabric Vision technology support by feature and product

Feature	IBM SAN42B-R and IBM b-type Gen 6 Extension Blade	IBM SAN06B-R and 8 Gbps Extension Blade (FC 3890)
MAPS	✓	✓
Fabric Performance Impact (FPI) Monitoring	✓[a]	✓
Dashboards	✓	✓

Feature	IBM SAN42B-R and IBM b-type Gen 6 Extension Blade	IBM SAN06B-R and 8 Gbps Extension Blade (FC 3890)
Flow Learning	✓	-
Flow Monitoring	✓[a]	✓ See your product documentation for more information.
Flow Generator	✓[a]	-
COMPASS	✓	✓
Forward Error Correction	✓[a]	-
Credit Loss Recover	✓[a]	-
IBM ClearLink Diagnostics (D_Port)	✓[a]	-

a. Not available on IP Extension

Table 2-9 describes the Fabric Vision features supported by IBM extension products.

Table 2-9 Detailed Fabric Vision technology features supported by IBM extension products.

	IBM SAN42B-R and IBM b-type Gen 6 Extension Blade	IBM SAN06B-R and 8 Gbps Extension Blade (FC 3890)
Flow Generator	Generate, pass (including passing through VE_Port), and receive Flow Generator traffic	Pass (including passing through VE_Port) and receive, but does not generate Flow Generator traffic
MAPS	► **Per tunnel/VE:** Throughput, state change (VE fencing is supported for state change) ► **Per circuit:** RTT, jitter, throughput, packet loss, state change (circuit fencing is supported for state change) ► **Per QoS (at tunnel level):** Throughput, packet loss, RTT, jitter	► **Per circuit:** Throughput, packet loss, and state change (circuit fencing is supported for state change)
Flow Monitor	Report IOPS and throughput per (SID, DID, LUN, SCSIRead/Write) flow monitored on F_Port, E_Port; LUN-level supported on F_Port only	Report IOPS and throughput per (SID, DID, LUN, SCSIRead/Write) flow monitored on F_Port, E_Port; LUN-level supported on F_Port only
IBM Network Advisor FCIP SAN Extension Widget	Yes	Yes

2.9 Terminology

This section explains key terms that are used in this publication:

IP interfaces Are configured with IP addresses, subnet masks, and an Ethernet interface, which assigns the ipif to the interface. When the FCIP circuit is configured, the source IP address has to be one that was used to configure an ipif, which in turn assigns the FCIP circuit to that Ethernet interface.

It is possible to assign multiple IP addresses and circuits to the same Ethernet interface by assigning multiple ipif to that same interface, each with its own unique IP address. You must configure an IP interface (ipif) for each circuit that you intend to configure on a Ethernet port.

IP routing Is based on the destination IP address presented by an extension circuit. If the destination address is not on the same subnet as the Ethernet port IP address, you must configure an IP route to that destination with an IP gateway on the same subnet as the local Ethernet port IP address. You can define up 128 routes per GbE port on the IBM SAN42B-R switch, whereas you can define 120 routes per Data Processor.

Circuit Is a logical connection created between two IP addresses. It defines source and destination IP addresses on either end of a tunnel. When created, a committed rate can be configured. Each circuit requires a unique IP address pair. You can configure a maximum of eight circuits for a trunk (VE_Port). There is no limit on the number of circuits that you can configure on an Ethernet port.

Tunnels Are a collection of one or more circuits that create one logical connection between two devices. Each tunnel presents a VE_Port to the fabric. For each tunnel, you can configure a single circuit or a trunk consisting of multiple circuits and increase the bandwidth available to a tunnel.

A single-circuit tunnel is referred to as a *tunnel*, and a tunnel with multiple circuits is referred to as a *trunk* because multiple circuits are being trunked together. With the development of circuits, a tunnel is no longer bound to a single physical interface or a single connection to a peer Ethernet switch.

E/EX and VE/VEX Ports

E_Ports are Expansion Ports that attach to other E_Ports and create ISLs. EX_Ports are similar to E_Ports, and are used to connect to an FC routed port. Router ports are the demarcation point of fabric services for a fabric. Fabric services do not extend beyond an EX port. Virtual E_Port (VE_Port) is the port that enables communication across an extension tunnel. It is a virtual port because it is extension tunnel facing.

A tunnel is represented in a fabric as VE_Port. It is exactly like E_Port, but the underlying transport is IP and not Fibre Channel. VE_Ports do not use FC flow control mechanisms (BB credits). Rather, they use TCP flow control mechanisms. The VE_Port emulates an E_Port on either end of the tunnel and operates like E_Ports for all fabric services and FOS operations. VEX_Ports are not supported on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade.

Table 2-10 shows a comparison of Expansion and Virtual Extended ports.

Table 2-10 E/EX and VE/VEX ports

E port type	no FCE	FCR
Native FC	E_Port	EX_Port
Extended over tunnel	VE_Port	VEX_Port (not supported on IBM SAN42B-R and IBM b-type Gen 6 Extension Blades)

Extension Hot Code Load (eHCL)

Allows non disruptive firmware updates on Data Processors in the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade. An HCL tunnel provides HA support for FC traffic over the extension tunnels.

Maximum Transmission Unit (MTU) size

Is the largest-size IP datagram that an IP network can support end-to-end. If you are unsure what your Path MTU (PMTU) is, the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade can automatically determine the path MTU by using the PMTU feature.

Path Maximum Transmission Unit (PMTU)

Is the process of sending Internet Control Message Protocol (ICMP) datagrams of various known sizes across an IP network to determine the supported maximum datagram size.

Based on the largest ICMP Echo Reply datagram received, the PMTU discovery process sets the IP MTU for that circuit's IP interface (ipif). Each circuit initiates the PMTU discovery process before coming online.

The smallest supported MTU size is 1280 bytes. The largest supported IP MTU size is 9216 bytes. However, PMTU discovery cannot discover an MTU greater than 9100 bytes. If the IP network's MTU is known, set it manually.

VLAN (IEEE 802.1Q) Tags Ethernet frames on a circuit and all traffic over that circuit uses the specified VLAN. It is a method to assign data flows to specific WAN connections, especially when many circuits share the same physical Ethernet link.

WAN tool Allows you to verify a circuit for network performance, such as throughput, congestion, loss percentage, and out of order delivery. It generates some traffic over a pair of IP addresses to test the network link. It is useful mainly when you want to perform health check for a new link before configuring it as a circuit in a tunnel.

3

Extension architectures

Extension provides optimized Fibre Channel over IP (FCIP) and IP Extension (IPEX) storage communications over IP networks between data centers.

This chapter provides the following information:

- Overview
- The FC side
- The WAN side
- The LAN side

3.1 Overview

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade can be thought of as having three sides of connectivity:

- Fibre Channel (FC)/Fibre Channel connection (FICON)
- LAN
- WAN

These three sides join together internal to the IBM SAN42B-R/IBM b-type Gen 6 Extension Blade. The following sections describe these sides of connectivity:

- The FC side
- The WAN side
- The LAN side

3.2 The FC side

The FC side is a full-fledged FC switch. FC communication between F_Ports, E_Ports, and EX_Ports is identical to any other B type FC switch of the same generation and Fabric OS version. Internal to the platform are back-end ports that connect to two Data Processors (DP). These backend ports are referred to as VE_Ports. VEX_Ports are no longer supported on the SAN42B-R and the IBM b-type Gen 6 Extension Blade.

FC fabric design is beyond the scope of this document. Ample information is available on the Brocade website. For example, see the *SAN Design and Best Practices* white paper:

https://www.brocade.com/content/dam/common/documents/content-types/whitepaper/brocade-san-design-best-practices-wp.pdf

A replication network can be an integral part of any production SAN. However, Replication fabrics are usually stand-alone fabrics for many reasons:

- Simplicity: Functional segregation simplifies both the production and replication networks, which in turn can improve overall network availability.
- Operations: Separating the replication and production fabrics enables considerably easier operations.
- Serviceability: Both the production and replication networks are easier to troubleshoot, diagnose, and service if the complexities of both remain separate.
- Firmware Updates: Frequently, the requirements for a replication fabric firmware update do not coincide with the production fabric. By having separate fabrics, it becomes easier to apply a replication fabric firmware update without having to make any changes to the production fabric.

There are various ways to provide FC connectivity for replication:

- Isolated Replication Network:
  - Use the 24 available FC ports as F_Ports on the stand-alone SAN42B-R Extension switches.
  - Use the IBM b-type Gen 6 Extension Blade inside a SAN256B-6 or SAN512B-6 Director chassis. Form a VF LS and include only the needed FC ports. There are no ISLs.

- Integrated Replication Network:
  - Connect the SAN42B-R to a production fabric with ISLs.
  - Do not isolate the IBM b-type Gen 6 Extension Blade inside a SAN256B-6 or SAN512B-6 Director.

Keep in mind that integrating a distance technology like extension into a production network without confinement causes the entire fabric and all associated underlying fabric services to stretch across the WAN. This is usually an undesirable architecture because it can introduce instability into fabrics. There are ways to prevent or limit fabric services from extending across a WAN. Fibre Channel Routing (FCR) is an option.

FCR limits fabric services to within an edge fabric. FCR is not supported in FICON environments. Virtual Fabric Logical Switches (VF LS) is another way to limit fabric services to just certain ports using logical switches that make up a logical fabric. The Logical Fabric would have the minimum number of attached devices such that just those devices are exposed to the WAN, minimizing the number of devices applying pressure on the fabric during an event.

The FC switch within the IBM SAN42B-R or IBM b-type Gen 6 Extension Blade connects to two internal data processors, referred to as DP0 and DP1. The extension connection to the remote side is essentially an ISL between the FC switches at both ends. If multiple VE_Ports (For example, “multiple VE_Ports” can be one VE_Port on two different DPs, or two VE_Ports on the same DP) are used to connect an internal FC switch, at the FC switch level Fabric OS (FOS) routes traffic across those different paths using the configured settings.

The default setting is Exchange Based Routing (EBR) in which exchanges round-robin through the various VE_Ports going to the same remote domain. Internally, this process occurs before any processing for Extension Trunking. Therefore, Extension Trunking cannot be used to prevent out-of-order delivery. Exchange sequences and associated data are delivered in-order.

However, it is possible that exchanges themselves can be delivered out of order. If exchange 1 is sent to DP0 and exchange 2 is sent to DP1, the time to complete work on each DP can vary, potentially causing one exchange to pass the other. In addition, the trunk, circuits, IP network, and carrier for each DP might be completely different. If the storage application is not tolerant of out-of-order exchanges, then the use of multiple VE_Port connections between local and remote domains must be questioned.

3.3 The WAN side

Extension primarily focuses on WAN side transport and optimization. Application data feeds into the WAN side through the FC/FICON (goes into FCIP) side, the LAN (goes into IPEX) side, or both. The data processing unit for the WAN side is referred to as a DP. The SAN42B-R and IBM b-type Gen 6 Extension Blade have two DPs: DP0 and DP1. All FC/FICON ports and Ethernet interfaces can see both DPs equally, and there is no preference to use specific ports or interfaces to access a specific DP.

3.3.1 WAN side architectures

The WAN side is where extension tunnels and circuits originate and terminate. The following functions occur on the WAN side: Extension trunks (tunnels), circuits, WAN Optimized TCP/IP, QoS, IPsec, ARL, egress scheduling, WAN side monitoring, high-efficiency encapsulation, flow-control, PMTU, OSTP, FastWrite, and FICON Acceleration.

The WAN side is common to both FCIP and IPEX. In fact, it is preferred that FCIP and IPEX run through the same tunnel (VE_Port) to take advantage of common bandwidth management and flow-control from the same egress scheduler. The FC/FICON and LAN sides feed into the WAN side, forming FCIP and IPEX respectively, across a common tunnel.

From an IP network point of view, the extension WAN side can be thought of as a network interface card (NIC) in a server. When configuring IP network Ethernet connectivity for the SAN42B-R or IBM b-type Gen 6 Extension Blade, the same configuration is used to connect a host's NIC. The SAN42B-R and IBM b-type Gen 6 Extension Blade do not switch Ethernet, route IP, or participate in Spanning Tree Protocol (STP) or any other IP routing protocols. It is an Ethernet interface that originates and terminates point to point traffic flows with other SAN42B-R/IBM b-type Gen 6 Extension Blade platforms across an IP infrastructure.

3.3.2 Extension Trunking

Over the last decade, extension networks for storage have become common and continue to grow in size and importance. Growth is not limited to new deployments. The expansion of existing deployments is common as well. Requirements for data protection will never ease, as the economies of many countries depend on the successful and continued business operations of their enterprises and thus have passed laws mandating data protection. Modern-day dependence on Remote Data Replication (RDR) means there is little tolerance for lapses that leave data vulnerable.

In mainframe, open systems, and IP storage environments, reliable and resilient networks, to the point of no frame loss and in-order frame delivery, are necessary for error-free operation, high performance, and operational ease. This technique improves availability, reduces operating expenses, and most of all reduces risk of data loss.

Extension Trunking is one of the advanced extension features of the IBM Extension platforms that provides the following benefits:

- Single logical tunnel that consists of one or more individual circuits
- Efficient use of Virtual E_Ports, known as VE_Ports
- Aggregation of circuit bandwidth
- Failover/failback
- Failover groups and metrics
- Use of disparate characteristic WAN paths
- Lossless Link Loss (LLL)
- In-Order Delivery (IOD)
- Non-disruptive link loss
- Deterministic path for protocol acceleration

3.3.3 Circuits in an extension trunk

Extension Trunking, in essence, provides a single logical tunnel that consists of multiple circuits. A single-circuit is referred to as a tunnel. A tunnel with multiple circuits is referred to as a trunk, simply because multiple circuits are being trunked together. A tunnel or trunk belongs to a single VE_Port and is a single inter-switch link (ISL), and treated as such in architecture designs. These extension ISLs can carry FC and FICON using FCIP, and IP storage using IPEX. Circuits are individual connections within a trunk, each with its own unique pair of source and destination IP addresses.

Because a tunnel/trunk is an ISL, each end requires a VE_Port. With Extension Trunking, each circuit is not its own tunnel. If each circuit were its own tunnel, a VE_Port would be required for each circuit, which is not the case. Because a trunk is logically one tunnel, only one VE_Port is used regardless of the number of circuits making up the overall tunnel. Extension trunks are considered a preferred practice.

Figure 3-1 shows an example of two trunks. Trunk1 has four circuits and Trunk2 has two circuits. Both ends of each circuit have been assigned a unique IP address and physical Ethernet interface. In this case, each circuit has been assigned to a different Ethernet interface, which is not required but is a common practice.

The circuit to Ethernet interface assignment is flexible and depends on the environment's needs. For instance, multiple circuits can be assigned to a single Ethernet interface for the purpose of conserving ports on the data center LAN switch. There are no subnet restrictions. The same or different subnets can be used on each interface. A circuit originates from a local ipif and terminates on a remote ipif, passing through the assigned Ethernet interface.

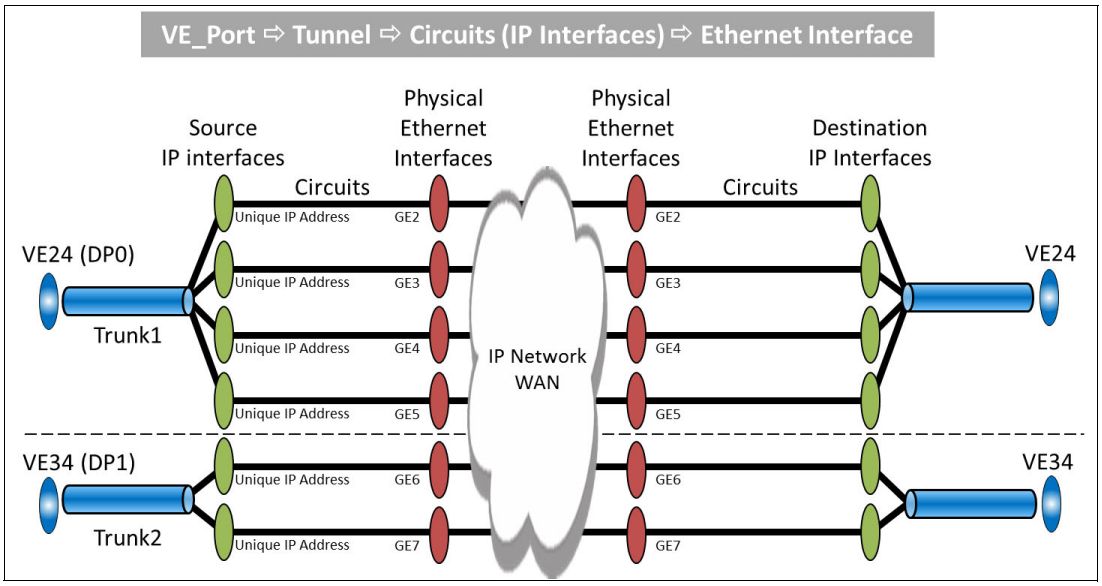


Figure 3-1 Trunks and their IP/Ethernet interfaces and circuits

3.3.4 Adaptive Rate Limiting

Adaptive Rate Limiting (ARL) is a dynamic rate limiting technology that is often used to readjust bandwidth rates to compensate for various failures or changes in the IP network. ARL is commonly used. For example, if a port, optic, or cable fails, ARL adjusts the remaining circuits upward to compensate for the offline circuits. If there are four circuits, each taking on 25% of the load, and one circuit fails because of a cable break, the remaining three circuits readjust to 33% each. If two circuits fail, the remaining two circuits readjust to 50% each.

ARL is also used for WAN connections that are shared with other traffic. This is done by designating a minimum amount of bandwidth always available to extension. Another amount of bandwidth is always dedicated to other non-extension flows. The gap in between is dynamically adaptable to either extension if the bandwidth is available, or released when it is not used by extension or demanded by non-extension applications. Ultimately, this configuration maximizes bandwidth usage and the available bandwidth for extension.

ARL is configured with minimum and maximum rate values. The rate limit is never less than the minimum value and never pushed to more than the maximum value. Between the minimum and maximum, the rate is adaptive based on IP network conditions at any instance. If the minimum and maximum values are set to be equal, this is called Committed Information Rate (CIR) and the circuit only pushes the CIR amount.

Circuits do not have to take the same or similar paths. The Round Trip Time (RTT) does not have to be the same. All circuits take on the RTT characteristic of the longest RTT circuit. For example, there are two circuits:

- Circuit0 RTT = 10 ms
- Circuit1 RTT = 25 ms

Both circuits effectively have a RTT of 25 ms, which is how the trunk scheduling algorithm behaves facilitating faster data transport to ULP. The delta between the longest and shortest RTT is not limitless, and must fall within supported guidelines. For more information, see the Fabric OS release notes.

Each circuit does not have to be configured identically to be added to a trunk. However, there are limitations to the maximum differences between circuits. The delta between the minimum and maximum rates for ARL is 5:1. For example, if the minimum rate is set at 1 Gbps, the maximum rate cannot exceed 5 Gbps, and this ratio is enforced in the configuration.

The rate difference between the maximum rates between any two circuits that belong to the same trunk is 4:1. For example, the maximum rate of circuit0 should not be more than 4x the maximum rate of circuit1. This configuration is preferred practice for optimal scheduler performance.

If there are multiple circuits assigned to an Ethernet interface, the aggregate of the minimum values cannot exceed the bandwidth of the interface. This limitation prevents oversubscribing the interface when guaranteed minimum rates have been configured.

It is possible, however, to configure the aggregate of the maximum bandwidth values beyond the capacity of the physical Ethernet interface. This configuration permits a circuit to use interface bandwidth greater than its minimum when another circuit is not using its minimum amount of bandwidth at the time.

For example, two circuits through a 10GE interface each have their maximum bandwidth set to full line-rate (10 Gbps). This is twice the rate that the physical interface can handle. If both circuits are demanding bandwidth, they equalize at 5 Gbps each. If one uses bandwidth mostly during the day and the other mostly at night, then each gets more available bandwidth during peak times.

If the SAN42B-R is a base unit (5 Gbps) or only has Upgrade1 (10 Gbps), the aggregate of the maximum values cannot exceed the license limits. There is no maximum limit when Upgrade1 + Upgrade2 has been activated. Upgrade2 removes maximum configuration limits.

3.3.5 VE_Ports of an extension trunk

Within the IBM SAN42B-R/IBM b-type Gen 6 Extension Blade architecture, there is a FC switching Application Specific Integrated Circuit (ASIC). Unlike E_Ports, VE_Ports are not part of this ASIC. Between the ASIC and the DP, VE_Ports are a logical representation of actual FC E_Ports.

Flows from multiple backend E_Ports feed into a DP permitting VE_Port data rates well above 16 and 32 Gbps, which is necessary for feeding the high data rates of the Fast-Deflate compression engine (40 Gbps) used for high-speed extension trunks (20 Gbps).

On the WAN side, the IBM SAN42B-R/IBM b-type Gen 6 Extension Blade has 10GE (10 Gigabit Ethernet) and 40GE interfaces. Think of VE_Ports as the transition point from the FC world into the TCP/IP world internal to the extension platforms.

Because a single VE_Port represents the ISL endpoint of multiple trunked circuits, this configuration affords the SAN42B-R/IBM b-type Gen 6 Extension Blade some benefits. First, fewer VE_Ports are needed, thus making the remaining virtual ports available for other trunks to different locations. Typically, only one VE_Port is needed per remote location. Second, bandwidth aggregation is achieved by merely adding circuits to the trunk.

Each IBM Extension platform has a maximum supported VE_Port bandwidth. The bandwidth of a VE_Port is the aggregate of the ARL minimum values from all trunk member circuits with the same metric (metric 0 or 1). Do not confuse this maximum with the maximum extension bandwidth that the platform can support. Assuming that the proper licenses are applied, these are the maximum VE_Port bandwidths:

- IBM SAN42B-R: 20 Gbps per DP
- IBM b-type Gen 6 Extension Blade: 20 Gbps per DP

3.3.6 Circuit latency of a trunk

The resulting RTT for all circuit members in a trunk is that of the longest RTT circuit. If RTT emulation was not done, the result would be the early arrival of some traffic and the later arrival of other traffic. To prevent out of order delivery, the early traffic must wait for the later traffic before being delivered. This process would ultimately slow throughput to the ULP. Therefore, egress scheduling is weighted based on circuit bandwidth and RTT.

A common example is a tunnel that is deployed across two paths. One path is a relatively small RTT and the other path has a longer RTT. It is practical to trunk these two circuits, one across each path. Generally, Asynchronous RDR (Remote Data Replication) applications are not negatively affected when both circuits present the longest path latency. For Synchronous RDR, each path must have the wanted RTT.

3.3.7 Keepalives and circuit timeouts

Circuit failover within a trunk is fully automated with Extension Trunking by using a combination of metrics and failover groups. Active and passive circuits can coexist within the same trunk. Each circuit uses keepalives to constantly reset the keepalive timer.

The interval time between keepalives is a configurable setting on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade. If the keepalive timer expires, the path to the remote side is deemed unreachable and removed from the trunk. In addition, if the Ethernet interface assigned to one or more circuits loses light, those circuits are immediately considered down.

When a circuit goes offline, the egress queue for that circuit is removed from the egress scheduling algorithm. Traffic continues across the remaining circuits, albeit with the reduced available bandwidth due to the removed offline link.

Consider the following questions about keepalives and circuit timeouts:

1. How does the keepalive mechanism work on the trunk/tunnel/circuit?

   A Keepalive Timeout Value (KATOV) is configured for each circuit. The Extension Trunking algorithm uses that value to determine how frequently keepalive frames need to be sent. The math for that is not straightforward and is beyond the scope of this document. The algorithm ensures that multiple keepalive frames are sent within the timeout period. If the receiving side does not receive any keepalives within the timeout period, the circuit is brought down.

   Keep in mind that a heavily congested IP network can buffer traffic in multiple devices along its path. Buffering represents time in which traffic is not moving. If the KATOV is set too small, for example, 1 second, it is possible that excessive buffering might cause the circuit to drop. Additionally, a heavily congested IP network usually has considerable packet loss.

   The combination of buffering with dropped packets can let the keepalive timer expire. Enabling IBM Monitoring Alerting Policy Suite (MAPS) provides alerts for IP network conditions common to circuit drops, including the circuit drops themselves. If MAPS detects any combination of excessive jitter, packet loss, or latency, the IP network is ill-suited for transporting storage over extension.

2. How are keepalives queued and handled through the network layers?

   Transmission of keepalives is guaranteed through the IP network by WO-TCP, and cannot be lost unless a circuit goes down. Keepalives are treated like normal WAN-Optimized TCP (WO-TCP) data and not sent on any special path through the network stack. Therefore, they are intermixed with storage data that is passing through WO-TCP.

   The keepalive process determines true delay in getting a packet through an IP network. This mechanism takes latency, reorder, retransmits, and any other network conditions into account. WO-TCP is the preferred transport for keepalives because it keeps tight control of the time that data is outstanding on the WAN with retransmission if a circuit failure occurs.

   If packets are taking longer than the allotted amount of time to get through the IP network and are exceeding the KATOV, the circuit is torn down and data is requeued on the remaining circuits. Configuring the KATOV based on the storage application's timeout value is important and preferred.

   The default value for Fibre Channel Connection (FICON) is 1 second and should not be changed. The default value for non-FICON circuits is 6 seconds and often can be shortened based on the application and IP network. Such changes must be evaluated on a case-by-case basis and depend on the characteristics of the IP network architecture.

   The KATOV should be less than the application's time out for trunks that contain multiple circuits. This configuration facilitates Lossless Link Loss (LLL) ensuring that application data that is traversing the IP network does not time out.

3. Can having too much data going over a connection bring down a trunk, tunnel, or circuit?

   Yes and No.

   If there are no issues in the IP network and the tunnel is not large enough to handle the workload that is being driven from the application, then no. This situation means there is no congestion. ARL is configured properly, but the bandwidth is just not adequate for the RDR or tape application to succeed.

   When all the available bandwidth is used and Buffer-to-Buffer Credit (BBC) flow control is applying back pressure to incoming flows preventing overflow, the tunnel does not go down due to a keepalive timeout. The time required to get a keepalive through is not affected by large amounts of queued data.

However, keepalives are affected by the WO-TCP data queue, which is limited by the amount of data that can be queued but not yet sent. This amount of data is small (in the range of only a few milliseconds) and certainly not enough time to cause the keepalive timer to expire due to circuit saturation.

The “yes” part: If the circuit is congested because ARL has been misconfigured, allowing more data to enter the IP network than there is available bandwidth, and the network cannot handle the data rate, then network errors result. If these errors become severe enough to lead to buffer overruns and cause packet loss and retransmits, then multiple keepalives can be lost and a circuit drop can result.

3.3.8 Lossless Link Loss

When a connection is lost, data in flight is almost always lost as well. Any frames in the process of being transmitted at the time of the link outage are lost due to partial or no transmission. This situation causes an Out-Of-Order-Frame (OOOF) problem because some frames have already arrived, after which one or more are missing, and frames continue to arrive over the remaining links.

Overall, frames are not in sequence because there are missing frames. This situation is problematic for some storage devices, particularly mainframes, and results in an Interface Control Check (IFCC). In the distributed world, this can result in SCSI errors.

For example, frames 1 and 2 are sent and received. Frame 3 is lost. Frame 4 is sent and received. The receiving side detects 1-2-4, which is an OOO condition because it was expecting three frames but received four.

Extension Trunking resolves this problem with LLL. When a link is lost, inevitably frames inflight are also lost. Those lost frames are retransmitted by LLL. Normally, when a TCP segment is lost due to a dirty link, bit error or congestion, TCP retransmits that segment.

In the case of a broken connection (circuit down), there is no way for TCP to retransmit the lost segment because TCP is no longer operational across the link. If there is an IP network path failure, LLL ensures that all the data is delivered in order to ULP. Extension Trunking, ARL, and LLL are all used to maintain seamless connectivity for storage replication applications between data centers. IP network functions like Port Channeling and LAG are not needed on the WAN side.

Figure 3-2 shows the Lossless Link Loss process.

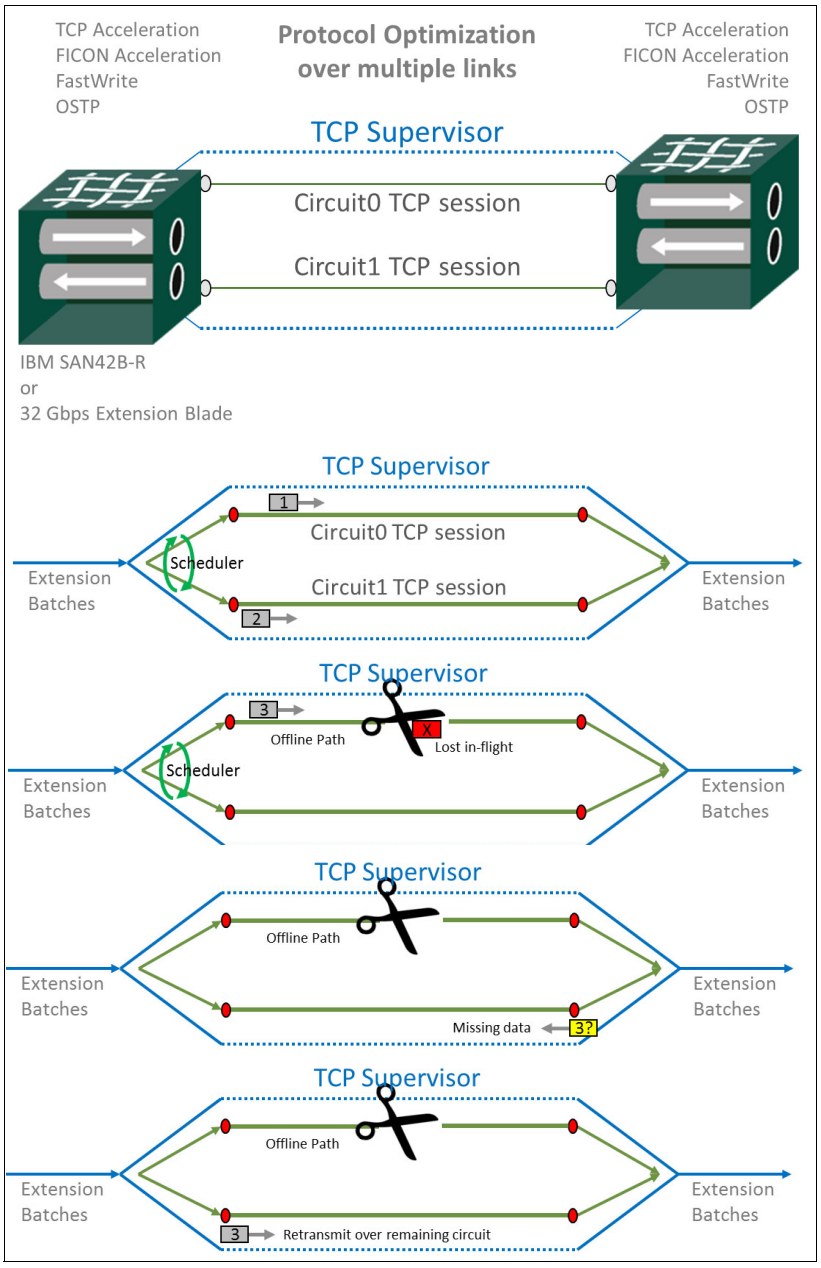


Figure 3-2 Lossless Link Loss (LLL) process

3.3.9 High Efficiency Encapsulation

Using High Efficiency Encapsulation, all the individual TCP sessions that are associated with multiple circuits are supervised within the trunk. This outer TCP session is referred to as the “Supervisor” TCP session, and it feeds each circuit’s TCP session through a DWRR (Deficit Weighted Round Robin) load balancer and a sophisticated egress queuing and scheduling mechanism that compensates for different link bandwidths, latencies, and congestion events.

The TCP Supervisor is an advanced technology, purpose-designed special function algorithm. It operates at the presentation level (Level 6) of the Open Systems Interconnection (OSI) model and does not affect any LAN or WAN network devices that might monitor or provide security at the TCP (Level 4) or lower levels such as firewalls, ACL, or sFlow (RFC 3176). It works at a level above FC and IP such that it can support both FCIP and IPEX across the same extension trunk.

If a connection goes offline and data continues to be sent over remaining connections, noncontiguous header sequence numbers indicate that there are missing frames. To retransmit missing segments, first subsequent segments must arrive indicating out-of-order, which triggers a Duplicate Acknowledgment (DupACK) by the TCP supervisor back to the source.

Even though the lost segments were originally sent by a different link's TCP session, which is no longer online, the TCP supervisor retransmits the lost segments over TCP sessions on the remaining circuits. This process means segments are held in memory that is managed by the TCP supervisor for transmission if a segment must be retransmitted over surviving circuits.

3.3.10 Path MTU

Path MTU (PMTU) is an automated process for determining the IP MTU along the path that a circuit takes. PMTU is configured by indicating "auto" for the MTU in an IP Interface (ipif) configuration. The advantage of using PMTU is that it is automatic.

The process sends out a number of packets of varying sizes. The largest one that makes it to the remote side sets the MTU. The down side to PMTU is that it will not find an MTU over 9100, and the maximum supported MTU is 9216 bytes. In addition, the PMTU process takes some time to complete, which means that each time that a circuit comes up, it has to go through the PMTU process and circuits take a little longer to come up.

Ask your network administrator to ascertain the actual supported MTU of the IP network and configure the actual value in the ipif.

3.3.11 Circuit metrics

Circuit metrics provide alternative circuits that are used for failover if primary circuits go offline. Circuits can be active or passive within a trunk depending on the circuit's metric (0 or 1). Within a failover group, all online circuits with metric 0 are active. Circuits with a value of 1 are passive and only become active after all metric 0 circuits within its failover group go offline. Figure 3-3 on page 56 shows a one-to-one pairing of metric 1 circuits with metric 0 circuits within a failover group. Every metric 0 circuit that goes offline is losslessly replaced by a corresponding metric 1 circuit.

Metric 1 circuits should take a different IP network path because it is likely that an IP network outage caused the metric 0 circuit to go offline. Metrics and failover groups permit circuit configuration over paths not normally used until the production path has gone down.

In a three-site triangular architecture, production traffic normally takes one hop as its primary path, which is usually the shortest in distance and latency. These circuits have a metric set to 0. Sending traffic through the alternate two-hop path, which is typically longer in distance and latency, is prevented unless the primary path is interrupted long enough for the KATOV to expire. This configuration is done by setting a metric of 1 to the alternate path's circuits.

The circuits that are passing through the bunker site do not terminate on the SAN42B-R/IBM b-type Gen 6 Extension Blade. Instead, they merely pass through the IP infrastructure at that location on the way to the final destination. See Figure 3-3.

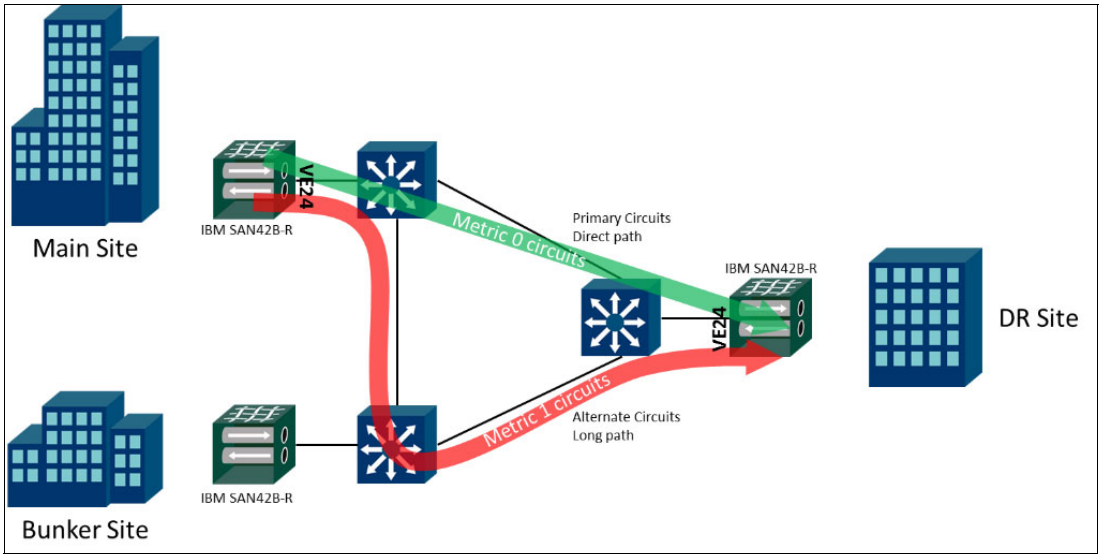


Figure 3-3 Site with Metric 0 and Metric 1 paths

Dormant circuits are still trunk members. Convergence to the secondary path is lossless and there are no out-of-order frames. No mainframe IFCC or SCSI errors are generated. Extension Trunking is required to assign metrics to circuits. See Figure 3-4.

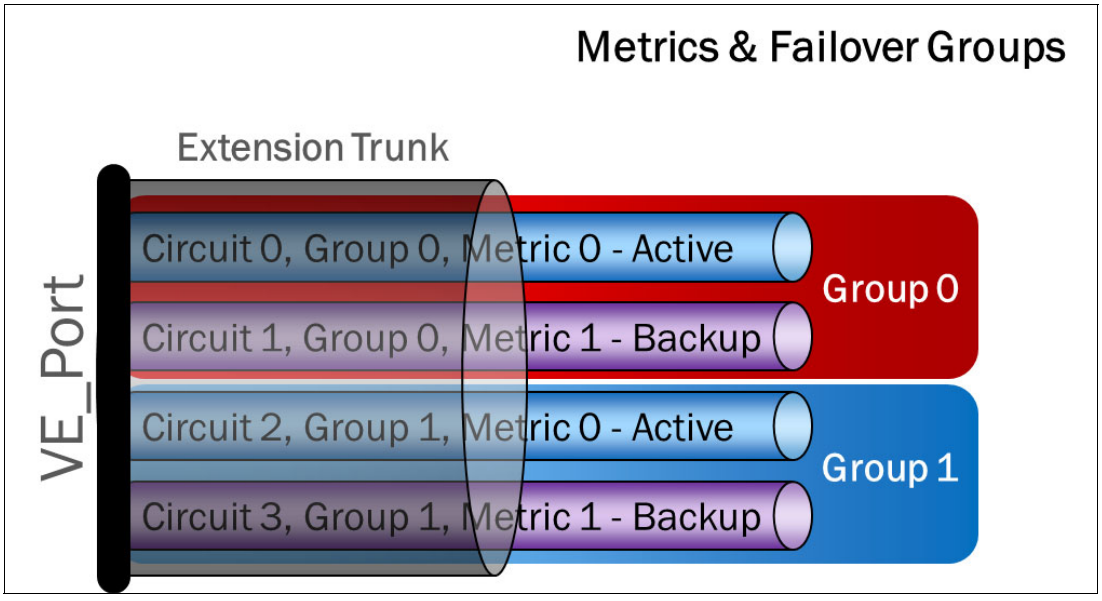


Figure 3-4 Extension Trunking: Circuit metrics and failover groups

3.3.12 Logically a single ISL (protocol optimization)

Extension Trunking provides a single point of termination for multiple circuits. This is important for protocol optimization techniques like FastWrite, Open Systems Tape Pipelining (OSTP), and FICON Acceleration. These protocol optimization techniques are FC/FICON based and not relevant to IPEX. Additionally, only specific FC and FICON applications can use these protocol accelerations. However, IBM Global Mirror (formerly Peer-to-Peer Remote Copy (PPRC)) does not use FastWrite. In a mixed vendor environment, other products might.

Before Extension Trunking, it was not possible to perform protocol acceleration over multiple FCIP connections because each connection was an autonomous tunnel with its own VE_Port. Traffic might take different paths outbound versus inbound, which creates state ambiguity breaking protocol acceleration.

Protocol acceleration requires a deterministic path round-trip. With or without an edge fabric attached, Port Based Routing (PBR), Device Based Routing (DBR), or EBR (Exchange Based Routing) across VE_Ports on the IBM SAN42B-R or IBM b-type Gen 6 Extension Blade cannot guarantee a deterministic bidirectional path when more than one path exists.

Using protocol acceleration, it is necessary to confine traffic to a specific deterministic path in both directions. This can be accomplished in a few ways:

- Use only a single physical path, including a single VE_Port tunnel/trunk.
- Use Virtual Fabrics (VFs) with Logical Switches (LSs) that contain a single VE_Port tunnel/trunk.
- Configure Traffic Isolation Zones (TIZs).

Note: All of these methods prevent the use of any type of load balancing and failover between VE_Ports.

The reason why optimized traffic must be isolated to the same path in both directions is because protocol acceleration requires a state machine. Protocol acceleration needs to know, in the correct order, what happened during a particular exchange. This feature facilitates proper processing of the various sequences that make up the exchange until it is completed, after which the state machine is discarded. These state machines are created and removed with every exchange that passes through the tunnel/trunk.

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade has the capacity for tens of thousands of simultaneous state machines supporting an equivalent number of flows. The ingress FC frames are verified to be from a data flow that can indeed be optimized. If a data flow cannot be optimized, it is merely passed across the trunk without optimization.

These state machines are within each VE_Port endpoint. A state machine cannot exist across more than one VE_Port, and there is no internal communication of state machines between VE_Ports.

As shown in Figure 3-5, if an exchange starts out on tunnel 1 and returns over tunnel 2, and each tunnel uses a different VE_Port, the exchange passes through two different state machines. Each one knows nothing about the other. This situation causes FC Protocol (FCP) and Small Computer Systems Interface (SCSI) to break, producing I/O errors.

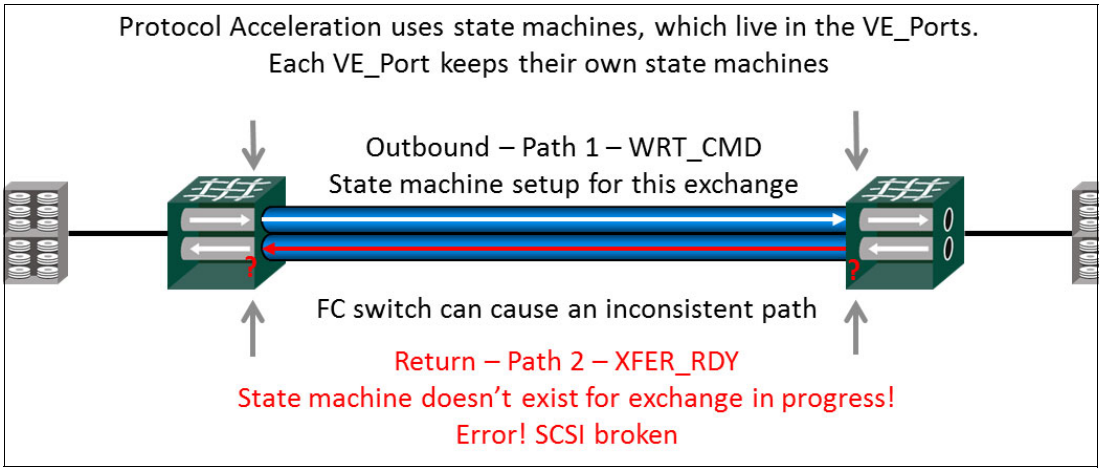


Figure 3-5 SCSI I/O breaks without deterministic path

An advantage of Extension Trunking is that it logically produces a single ISL with a single terminating endpoint (VE_Port). This is true even with multiple circuits. The state machine exists at the VE_Port endpoints and remains consistent even when an exchange uses circuit 0 outbound and circuit 1 inbound. See Figure 3-6. Extension Trunking permits protocol optimization to function across multiple load balanced circuits. Extension Trunking supports failover with error-free operation and without any disruption to the optimization process.

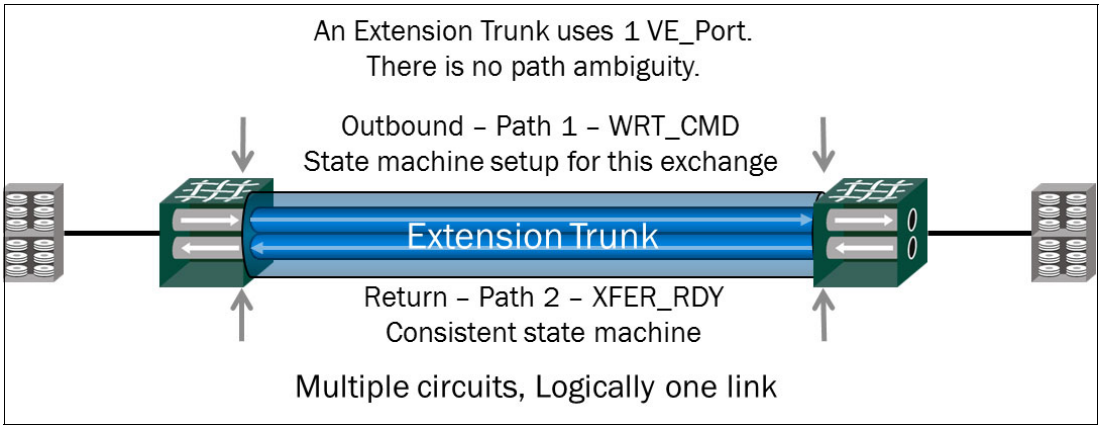


Figure 3-6 Protocol optimization with Extension Trunking as a single logical tunnel

There is one tunnel or trunk per VE_Port. Keep in mind, if more than one VE_Port is used between two domains, the SAN42B-R and IBM b-type Gen 6 Extension Blade can internally route traffic indeterminately to either of the VE_Ports, breaking protocol optimization. In this case, one of the isolation techniques described earlier in this section is required to prevent failure by confining traffic flows to the same tunnel or trunk.

3.3.13 VE_Port load balancing

A key advantage of Extension Trunking is that it offers a superior technology implementation for load balancing, failover, in-order delivery, and protocol optimization. FastWrite, FCR, OSTP, and FICON Acceleration are all supported on the SAN42B-R and IBM b-type Gen 6 Extension Blade (in some cases an optional license might be required).

IBM Storage products do not use FastWrite because the same functionality has been built into the storage arrays. The same Data Processor (DP) that forms tunnels/trunks also performs protocol acceleration for a higher level of integration, efficiency, and utilization of resources, all with lower costs and a reduced number of assets. Additional WAN acceleration is not necessary or supported.

Based on the Fabric OS APTpolicy setting, one of three routing methods is used to route FC between multiple VE_Ports:

- EBR (Exchange Based Routing, the default): Originator Exchange ID/Source ID/Destination ID (OXID/SID/DID)
- DBR (Device-Based Routing = PBR + DLS): SID/DID
- PBR (Port-Based Routing): Source Port

VE_Ports can route FC/FICON flows across different DPs on the same blade or across DPs on different blades. See the IBM support information for your mainframe for information about the support of EBR, DBR, and PBR in the FICON network. This configuration is supported only when protocol acceleration (FastWrite, FICON Accelerator, and OSTP) is not enabled.

Fabric OS performs Exchange Based Routing (EBR) by default internally between the two VE_Ports on each DP internal to the SAN42B-R and IBM b-type Gen 6 Extension Blade. These DPs are engines that process FC frames (FCIP) or TCP flows (IPEX) into extension, performing all the transport functions.

EBR operates between E_Ports, EX_Ports, and VE_Ports. Considering extension trunks with equal-cost FSPF paths, data load shares across multiple VE_Ports. EBR is an exchange-based FC routing technique (do not confuse it with FC Routing (FCR)) directing traffic between DPs based on a SID/DID/OxID (Source ID, Destination ID, and Exchange ID) hash.

If an Ethernet interface, optic, cable, or IP path were to fail, EBR fails over traffic to the remaining equal-cost paths. Routing of FC traffic across multiple equal-cost paths (multiple VE_Ports) is not supported when using any of the protocol accelerations.

3.3.14 FCIP batching

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade use batching to accomplish these goals:

- Improve overall efficiency
- Maintain full utilization of links
- Reduce protocol overhead

Simply put, a batch of frames is formed, after which the batch is compressed and processed as a single unit. This unit is a “compressed byte stream.” TCP transmits a byte stream, not discrete pieces of data. Because it is a stream of bytes, it is not relevant where FC frames begin or end within the stream.

Frame boundaries are arbitrary when transported across the tunnel. A batch is composed of up to 16 FC frames (FCIP) or four FICON frames. All the frames must be from the same nexus, that is, an FCP exchange's DATA_OUT sequence or an IPEX TCP session. SCSI commands, transfer readies, and responses are not batched and immediately expedited for tunnel transmission.

The last frame of a sequence has the end-of-sequence bit set. The batch is now known to be complete and processed immediately, even if fewer than sixteen FC frames have been received. See Figure 3-7. In this example, two open systems batches are created. One is full with sixteen frames and the other has only two FC frames because the End of Sequence bit is set.

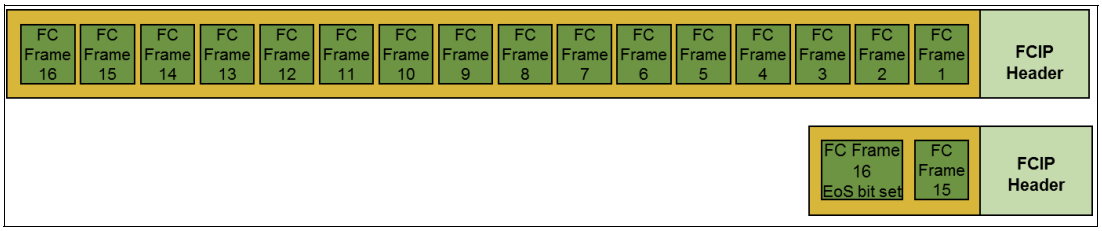


Figure 3-7 FCIP Batch Formation - 1 sequence, 2 batches

3.3.15 IPEX batching

Batching is a form of optimization. Batches are created for both FCIP and IPEX data. IPEX batching is slightly different than FCIP batching, although both are created on a byte stream basis.

IP storage TCP flows coming from the LAN side are called streams. 1024 streams are supported on the SAN42B-R and IBM b-type Gen 6 Extension Blade, 512 streams per DP. A batch fills with IP payloads until it gets to at least 32 KB, after which no more datagrams are added.

In the example shown in Figure 3-8, the 16th IP datagram either meets or exceeds 32 KB making it the last IP datagram. If data is arriving at 10 Gbps, it takes about 25 µs to fill a batch. If datagrams stop arriving from that stream, there is a 2 µs backstop timer that triggers the batch to be processed. After an IPEX batch is formed, the processing of that batch is identical to that of FCIP batches.

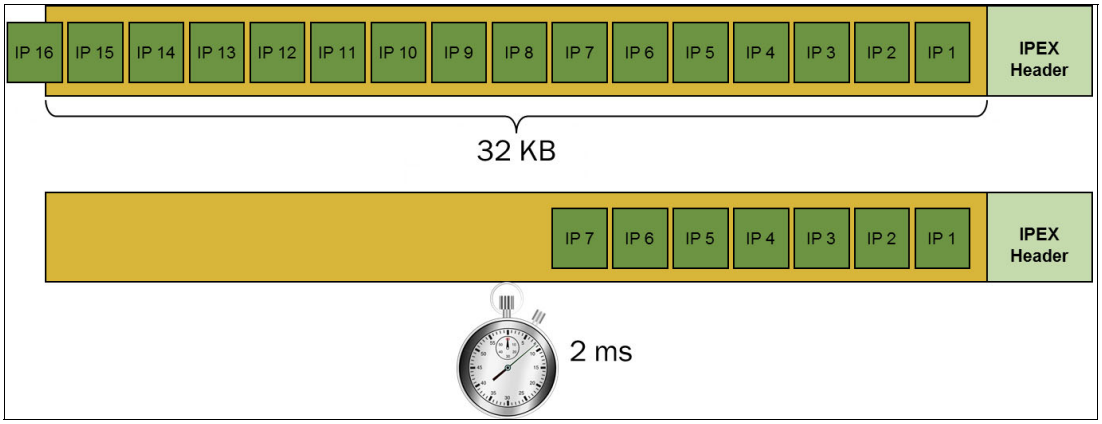


Figure 3-8 IPEX batch formation

A *batch* is a unit of load balancing across circuits within a trunk. Batches are placed in the egress queues by the TCP supervisor using a Deficit Weighted Round Robin (DWRR) algorithm. This process is referred to as scheduling. The scheduler does take into account egress bandwidth, RTT, and queuing levels for each of the circuits.

When a queue becomes full, usually because of disparate circuit characteristics within the IP network, the scheduler skips that queue to permit it to drain, while continuing to service other circuits. This process maintains full link utilization for different bandwidth circuits, circuits with dissimilar latency, and circuits experiencing delay due to retransmits. The TCP supervisor on the receiving side ensures in-order delivery of any batches that arrive out of order due to circuit disparity and encountered path buffering.

As mentioned previously, circuits do not have to have the same bandwidth, ARL rates, or latency (RTT), which permits circuits to simultaneously use a variety of infrastructures like Dense Wavelength-Division Multiplexing (DWDM), Virtual Private LAN Services (VPLS), Multiprotocol Label Switching (MPLS), carrier Ethernet, and so on.

Batches are parsed into TCP segments based on the Maximum Segment Size (MSS), which specifies only the TCP payload size without headers. MSS is not configurable and is based on the IP Maximum Transmission Unit (MTU) minus the IP and TCP headers.

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade support Jumbo Frames with a maximum MTU of 9216 bytes. Depending on the size of the batch, one batch can span multiple TCP segments or it might fit within a single TCP segment. Each TCP segment is filled to its maximum size while there is data to send for that nexus. See Figure 3-9 for an example of the SAN42B-R/IBM b-type Gen 6 Extension Blade 9216-byte MTU using Jumbo Frames. This method generates the lowest amount of overhead resulting in superior performance and highest efficiency.

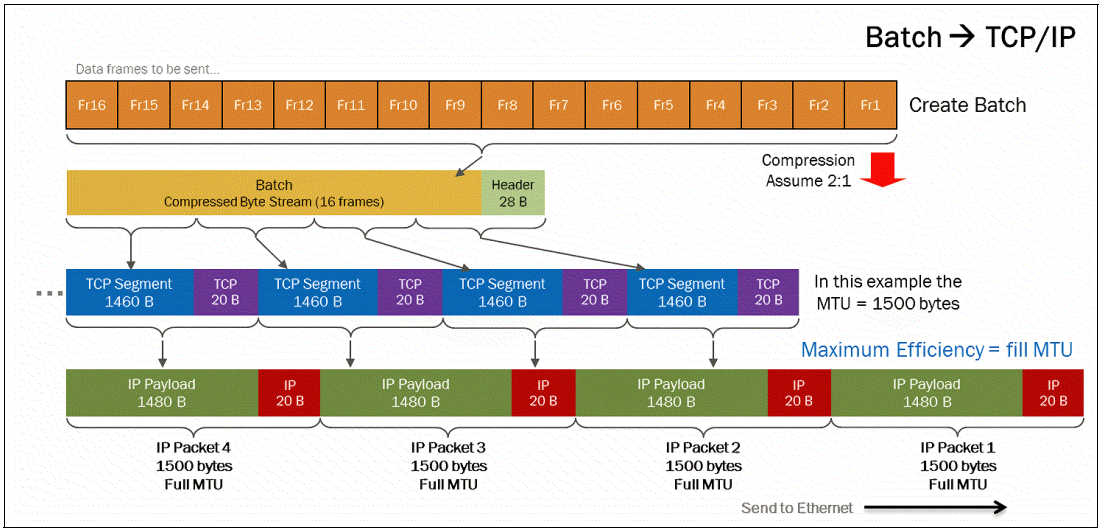


Figure 3-9 A batch into TCP/IP method (1500-byte MTU)

Figure 3-10 shows the batches into TCP/IP method (9216-byte MTU).

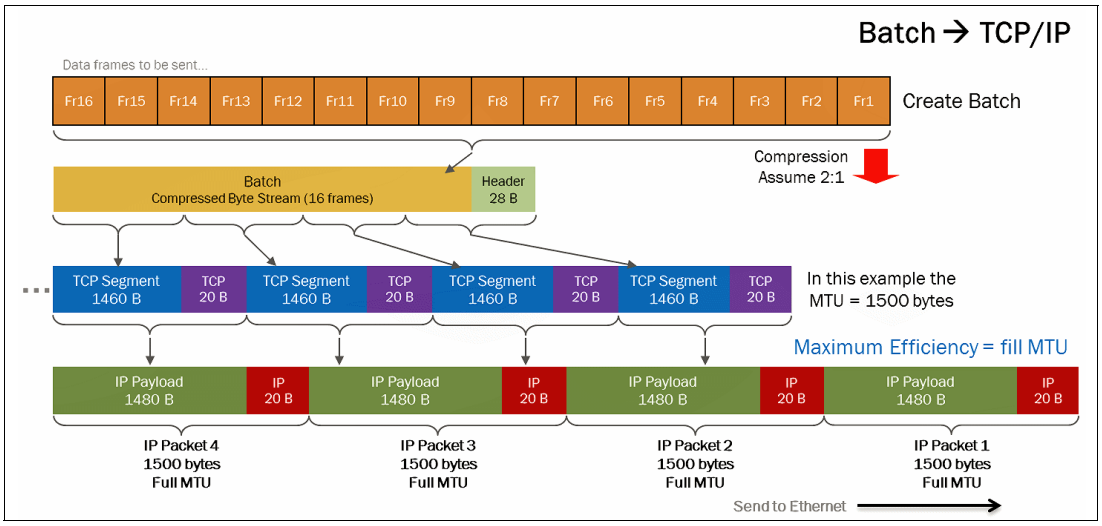


Figure 3-10 Batches into TCP/IP method (9216-byte MTU)

3.3.16 Extension Hot Code Load

Extension Hot Code Load (eHCL) should be enabled when possible. eHCL permits firmware updates without extension tunnels/trunks being interrupted. This functionality has existed on the FC side for many years. With eHCL, the same function exists on the WAN side. eHCL is configured once at deployment, not with each firmware update. eHCL requires two IP addresses for each circuit end instead of just one. Therefore, twice as many IP addresses are needed. Typically, extension environments deploy private IP addresses using RFC 1918 (10.*x.x.x*, 172.16.*x.x*, 192.168.*x.x*), making availability of additional addresses easy.

eHCL requires bandwidth be available on the other DP within the SAN42B-R or IBM b-type Gen 6 Extension Blade. For example, if you are performing a firmware update and eHCL is configured, VE24 (on DP0) has 10 Gbps of circuits configured. During the firmware update process, those circuits are moved to DP1 until DP0 has completed its upgrade, after which, the circuits are returned to DP0.

To maintain full performance, DP1 must have 10 Gbps of bandwidth available to accommodate the circuits that are coming from DP0. If utilization is anticipated to be low during the maintenance window, less bandwidth might be needed. After DP0 completes its upgrade, the mirror process occurs when DP1 performs its upgrade process. DPs have 20 Gbps of capacity each.

For example, each DP has a consumed aggregate bandwidth of normally 10 Gbps (for example, 4 circuits x 2.5 Gbps), leaving 10 Gbps remaining for eHCL. DP0 can use the remaining bandwidth on DP1 and vice versa.

Firmware updates are not typically done during periods of outage or during peak production times. Updates are usually done during maintenance windows in a controlled environment. Upon commencement of a firmware update, circuits are relocated to the other DP non-disruptively, with no data loss and all data is delivered in-order.

eHCL does not cause IFCC in mainframe environments. The advantage of eHCL is that you do not need to disrupt normal traffic flows if you do not have to, which reduces overall risk to applications and the enterprise. Even in consideration of being done during a maintenance window in a controlled environment.

3.3.17 IP network load balancing

Flow-based load balancing within the IP network is fully supported and normally causes no issues. It is used in most IP networks. Nearly all types of IP and WAN architectures are supported.

One type that is not supported is Per-Packet Load Balancing (PPLB). PPLB is not common in modern IP networks. In most cases, PPLB works against TCP by causing excessive and chronic Out-Of-Order-Segments (OOOS), leading to retransmits, delay of data to Upper-Layer Protocols (ULPs), higher response times, excessive CPU utilization, increased overall bandwidth utilization, and lower throughput.

3.3.18 QoS and PTQ

Per-Priority-TCP-QoS (PTQ) is a special QoS function exclusive to IBM SAN42B-R/IBM b-type Gen 6 Extension Blade and developed especially for high-performance tunnels/trunks. For QoS to function properly across an IP network, it is essential that each priority have its own WO-TCP session. Using a single TCP session with multiple QoS priorities is not possible because there would be one merged flow and no granular flow-control per priority, causing severe performance degradation.

PTQ provides a TCP session for each priority. There are seven priorities within each circuit: FC and FICON have four (class F, high, medium, and low) and IPEX has three (high, medium and low). The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade have various QoS functions, including Differentiated Services Code Point (DSCP), 802.1P (L2 CoS), and PTQ. DSCP and L2 CoS perform marking of IP packets (L3) and VLAN tags (L2), respectively.

IPEX batches are stream-specific and fed into a correlating priority's WO-TCP session based on TCL designation. The percentage of bandwidth assigned to IPEX versus FCIP is apportioned during periods of contention. This percentage is configurable. When there is no contention, any flow can use whatever bandwidth is available within the confines of ARL. If the IP network is configured properly, the IP network manages priority traffic based on QoS markings in the headers. All QoS functions are fully compatible with Extension Trunking.

3.3.19 FCIP flow control

On the FC/FICON side, buffer overflow results in lost frames and IBM FC products rarely drop frames. If the queues for all circuits become full and can no longer be serviced by the scheduler, the buffers fill and eventually FC BBC (Buffer-to-Buffer Credits) R_RDYs are withheld by the SAN42B-R/IBM b-type Gen 6 Extension Blade to the end-device, stopping the inflow of data from overflowing the buffers.

There is a tremendous amount of buffer memory within the SAN42B-R and IBM b-type Gen 6 Extension Blade. However, there are advantages to limiting buffering to just little more than the bandwidth-delay product (the amount of data that can be inflight) of the circuit. Robust buffers on the SAN42B-R/IBM b-type Gen 6 Extension Blade can accommodate multiple long fat pipes and a large quantity of simultaneous flows. Additionally, limiting a flow's buffering permits the source device to better know what has or has not been sent by having as little data as possible outstanding.

When queues drain to a particular point, FC frames resume flow from the source and new batches are produced. All the while, there is no idle transmission of any trunk circuit because no queue has ever completely emptied.

On the WAN side, flow-control is done by WO-TCP by using TCP windowing techniques (Figure 3-11).

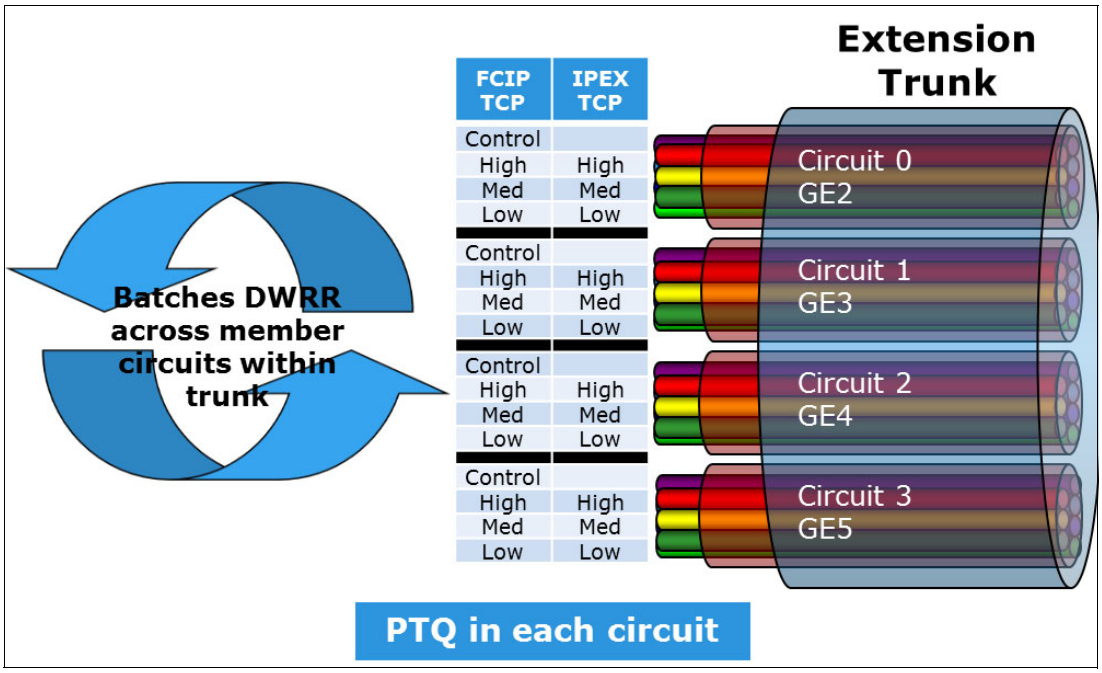


Figure 3-11 Load balancing batches across trunk circuits

3.3.20 IPEX flow control

First, it is important to understand that there are three TCP sessions involved for each and every flow:

- The local TCP session between the originating end-device and the IBM SAN42B-R
- A TCP session across the WAN using WO-TCP
- A TCP session on the remote side between the IBM SAN42B-R and the destination end-device

IPEX flow control requires participation and interaction of all three of these TCP sessions. Because the TCP sessions to the end-devices is typically a standard TCP stack, traditional TCP flow control is used. WO-TCP, however, operates uniquely to prevent inadvertently hindering devices that do not require flow control at that instance.

On the IBM SAN42B-R Extension switch and IBM b-type Gen 6 Extension Blade WAN Optimized TCP (WO-TCP) was introduced. One primary innovation of WO-TCP was IPEX streams. Streams is a technique that prevents collateral effects to other flows by Slow Drain Devices (SDD), which causes Head of Line Blocking (HoLB).

TCP uses send (swnd) and receive (rwnd) windows as its flow-control mechanism. If an IP storage device on the receiving side needs to reduce an incoming flow, it closes the rwnd. Using only one window for all data being transported is deleterious to *all* flows using TCP. Having only one rwnd means that *all* flows would be slowed or halted if just one end-device asserted flow-control. Clearly, this would be a major problem.

The remedy is independent flow-control for every stream. However, it is not practical to create a separate TCP stack for each flow. There are not enough compute or memory resources on the SAN42B-R/IBM b-type Gen 6 Extension Blade for that to be practical.

It is practical to use a single TCP stack that implements virtual TCP windows for each stream. The IBM SAN42B-R/IBM b-type Gen 6 Extension Blade accommodates 512 streams per data processor or 1024 streams per SAN42B-R. WO-TCP streams allows one flow to slow or stop while all other flows go unfettered.

3.3.21 Extension trunk FSPF costs

A trunk has multiple circuits. A logical question is, “How are Fabric Shortest Path First (FSPF) link costs computed with multiple circuits that have variable rate limits?” Furthermore, considering that ARL has a minimum and maximum bandwidth level set for each circuit, how does that affect FSPF costs?

For IPEX, remember that both the control and data planes are separate from the FC/FICON and FCIP control and data planes. All IPEX control is accomplished by using TCL. The TCL configuration defines IPEX behavior only. The Fabric Operating System (FOS) and associated fabric services, routing and switching are not involved with IPEX. FSPF is just one of the fabric services that is specific to FCIP and not relevant to IPEX. IPEX never passes through a FC switching ASIC or a FC fabric.

An FSPF cost is calculated for the whole trunk, which is the same as the aggregate bandwidth of the VE_Port. Circuits are not entities that are recognized by FSPF. Therefore, circuits have no individual FSPF cost associated with them. FSPF cost is determined from the sum of the maximum bandwidth rates configured and only from active circuits within the trunk. For example, all active metric 0 circuits have their maximum ARL value added, and that aggregate value is used in the computation.

FSPF link cost is computed using the following function (see Figure 3-12):

- If the aggregate bandwidth is greater than or equal to 2 Gbps, the cost is 500.
- If the aggregate bandwidth is less than 2 Gbps and greater than 1 Gbps, the cost is 1,000,000 divided by the aggregate bandwidth amount in Mbps (from 500 to 1000).
- If the aggregate bandwidth is less than 1 Gbps, the cost is 2000 minus the aggregate bandwidth amount in Mbps (from 1000 up to 2000).

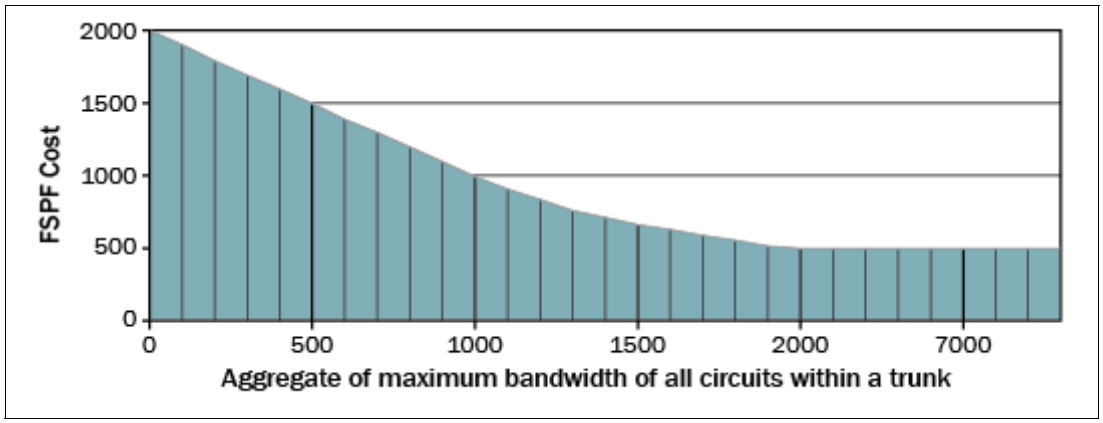


Figure 3-12 FSPF costs for trunks based on maximum aggregate bandwidth

3.3.22 Ethernet interfaces

Each circuit on the SAN42B-R/IBM b-type Gen 6 Extension Blade can be assigned to a WAN-side Ethernet interface. Circuits cannot be assigned to LAN side Ethernet interfaces. Ethernet interfaces should always remain in the Virtual Fabric Default-Switch. Circuit assignments can be from the same or different VE_Ports, and can be from one or more Virtual Fabric logical switches.

Because a VE_Port can have multiple circuits, a VE_Port can use multiple Ethernet interfaces. The endpoint of each circuit terminates at an ipif (IP Interface). The ipif is configured with an Ethernet interface, IP address, subnet mask, and MTU. When the ipif's IP address is used to create a circuit, the IP address in turn associates the Ethernet interface that the circuit uses. The number and types of Ethernet interfaces that are supported on the SAN42B-R/IBM b-type Gen 6 Extension Blade are listed in Figure 3-13, which lists the number of VE_Ports and the maximum bandwidth that is supported per Data Processor.

Hybrid Mode	IBM SAN42B-R Extension Switch and 32 Gbps Extension Blade	
Data Processor	Data Processor 0	Data Processor 1
GE Interfaces	16 GE/10GE interfaces shared by both DPs (GE2 thru GE17)	
10GE	**Note:** 40GE interfaces on the IBM SAN42B-R require an optional license)	
40GE	Two 40GE interfaces shared by both DPs (GE0 & GE1)	
Maximum Bandwidth per VE_Port	20 Gbps	20 Gbps
Maximum WAN Bandwidth per DP	20 Gbps	20 Gbps
Maximum LAN side Bandwidth per DP	20 Gbps	20 Gbps
Maximum FC/FICON side Bandwidth per DP	10 Gbps with 1:1 to 20 Gbps with 2:1 using Fast_Deflate compression	10 Gbps with 1:1 to 20 Gbps with 2:1 using Fast_Deflate compression
Maximum Number of VE_Ports	• No VEX_Ports • 5 per DP at 20 Gbps	• No VEX_Ports • 5 per DP at 20 Gbps

Figure 3-13 Ethernet interfaces on the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade

Note: IBM SAN42B-R/IBM b-type Gen 6 Extension Blade 40GE interfaces cannot connect to local data center switches for IPEX LAN side connectivity. The 40GE interfaces are for WAN side connectivity only.

For Extension Trunking, the DP that owns the VE_Port controls all member circuits. There is no distributed processing, load sharing, or LLL across different DPs in the same box or within the same chassis. Failover between DPs is done at the FC level by the FC switching ASIC provided the configuration and application permit it.

On an Ethernet interface, not all circuits have to be members of the same trunk. There can be multiple trunks on an Ethernet interface. Consider 10 separate 1 Gbps circuits associated with a 10GE interface, each two with their own VE_Port (that is, VE24, VE25, VE26, VE27, and VE28), which would create five separate extension trunks of 2 Gbps each. These might connect five different sites, for example.

The preferred practice is to use 1 VE_Port to connect domains between two sites. Scaling, redundancy and failover are facilitated by member circuits. For example, on a SAN42B-R/IBM b-type Gen 6 Extension Blade, two 10 Gbps circuits are created with a metric of 0 and both stemming from VE24. Circuit 0 is assigned to GE interface 2 and Circuit 1 is assigned to GE interface 3. Over the IP Infrastructure, the circuits take different paths and carriers. The two circuits join up again at the remote side. Logically, this is one 20 Gbps ISL between the two data centers.

The IP network routes circuits over different network paths and different WAN links based on the destination IP addresses and possibly other L2/L3 header attributes. Ethernet interfaces on the IBM SAN42B-R Extension switch and IBM b-type Gen 6 Extension Blade provide physical connections to WAN switches/routers for one or more circuits regardless of the tunnel/trunk, application, or destination.

IBM SAN42B-R/IBM b-type Gen 6 Extension Blade Ethernet interfaces have no specific requirements for connectivity to WAN routers/switches or DWDM devices. Network devices should view SAN42B-R/IBM b-type Gen 6 Extension Blade Ethernet interfaces as though a server network interface card (NIC) is connected and not another Ethernet switch or router. No Ethernet bridging, Spanning Tree, or IP routing is occurring on the SAN42B-R/IBM b-type Gen 6 Extension Blade. The Ethernet interfaces are the origination and termination point of TCP flows, the same as servers.

3.3.23 Extension Trunking use with other features

Features such as compression, IPsec, VLAN tagging (802.1Q), FCR (Fibre Channel Routing), Virtual Fabrics (VF), and quality of service (QoS) all function with Extension Trunking without caveats.

3.3.24 FCIP and IPEX compression

There are three compression algorithms available:

- Fast-Deflate
- Deflate
- Aggressive-Deflate

Selecting the appropriate algorithm is dependent on various things: SAN42B-R/IBM b-type Gen 6 Extension Blade operating mode (FCIP Only or Hybrid), protocol (FCIP or IPEX), and WAN bandwidth. See Figure 3-14.

	Max Rate FCIP Only Mode (per DP – FC side)			Max Rate Hybrid Mode (per DP – FC or LAN side)					
Algorithm	Fast-Deflate	Deflate	Aggressive-Deflate	Fast-Deflate		Deflate		Aggressive-Deflate	
Protocol	FCIP	FCIP	FCIP	FCIP	IPEX	FCIP	IPEX	FCIP	IPEX
FC or LAN side	40 Gbps	16 Gbps	10 Gbps	20 Gbps	n/a	10 Gbps	16 Gbps	10 Gbps	10 Gbp
Ratio	2:1	3:1	4:1	2:1		3:1	3:1	4:1	4:1
WAN side	≈20 Gbps	≈5 Gbps	≈2.5 Gbps	≈10 Gbps		≈3 Gbps	≈5 Gbps	≈2.5 Gbps	≈2.5 Gbps

Figure 3-14 Compression algorithms

Hybrid mode has the following characteristics:

- Fast-Deflate is not available for IPEX.
- For IPEX, there are 20 Gbps in and out of the Deflate/Aggressive-Deflate engine. If IPEX compression is disabled (1:1), 20 Gbps would be the maximum rate.
- For FCIP, there are 20 Gbps into the Fast-Deflate engine and 10 Gbps out. 20 Gbps of FC throughput is only available if the data is compressible to 2:1. For example, if the data was not compressible or FCIP compression was disabled, 10 Gbps would be the maximum rate.

Keep in mind that the compression engines in a DP have a maximum amount of throughput. Fast-Deflate is hardware-based compression with high speed and ultra-low latency. Deflate and Aggressive-Deflate are software-based compression with higher ratios and operate at a slower rate. The Fast-Deflate, Deflate, and Aggressive-Deflate compression engines are internally different engines. Bandwidth used on Fast-Deflate does not consume the bandwidth used on Deflate/Aggressive-Deflate.

Deflate and Aggressive-Deflate use the same engine, but two different algorithms. Being the same engine, if a percentage of bandwidth is consumed for Aggressive-Deflate, then that same percentage of Deflate bandwidth has also been consumed and vice versa. Software compression resources are consumed if any one of the two algorithms are consuming some portion or all of the allotted bandwidth.

For example, Aggressive-Deflate has 10 Gbps and Deflate has 16 Gbps of FC/LAN side bandwidth available. If Aggressive-Deflate consumes 5 Gbps (50%), only 8 Gbps (50%) of Deflate resources remain.

Compression is configured per FCIP and IPEX protocols on a per tunnel/trunk basis. Compression is not configured per circuit. A different compression mode can be selected for FCIP verses IPEX for each tunnel or trunk, based on the bandwidth profile.

For example, for a TS7760 or TS7700 Grid, tape traffic is normally already compressed, so you would not want to compress grid traffic again. In this case, IPEX compression is disabled. The same extension trunk has DS8000 Global Mirror replication across FCIP. FCIP compression is set to the highest-ratio algorithm that is fast enough to accommodate WAN bandwidth.

3.3.25 IPsec

IPsec is used for inflight data encryption and is a framework of various standards:

- IKEv2 Aggressive Mode.
- Suite-B ECDSA P-384 Elliptical certificates.
- PSK Profile: Group 24, SHA-512 HMAC, ESP and IKE: AES-256-GCM.
- PKI Profile: Group 20, SHA-384 HMAC, ESP: AES-256-GCM, IKE: AES-256-CBC.
- IKE SA lifetime is 6 hours.
- ESP SA lifetime is random between 3 to 4 hours.

IPsec requires no additional license and costs only the time and effort to enable it. It is prudent to enable IPsec in nearly every deployment. Extension Trunking is fully compatible with IPsec. IPsec is configured at the tunnel or trunk level, and includes all member circuits. IPsec encrypts/decrypts in hardware adding negligible latency (approximately 5 µs) and runs at maximum DP line rate, which is 20 Gbps on the WAN side. There is negligible performance degradation when using IPsec even for synchronous applications.

3.3.26 WAN optimization

WAN optimization products are not supported or needed with the IBM SAN42B-R/IBM b-type Gen 6 Extension Blade. Generally, WAN optimization products provide little benefit in an extension environment, which is already highly optimized. WAN Optimized TCP (WO-TCP) is not improved upon by using WAN optimization devices. In addition, the nature of most data traversing storage extension does not lend itself to effective data deduplication, for example new transactional DB data. The Aggressive-Deflate algorithm found in the IBM SAN42B-R and IBM b-type Gen 6 Extension Blade is capable of nearly the same data reduction across as WAN optimization products on the market today.

If WAN optimization currently exists in the IP infrastructure, the preferred practice is to either configure those devices to bypass extension traffic or introduce extension traffic upstream from the WAN optimization devices.

3.3.27 VLAN tagging (IEEE 802.1Q)

VLAN tagging (IEEE 802.1Q) inserts a tag into the Ethernet header and is fully supported by Extension Trunking. There is no requirement for circuits to use VLAN tagging. If tagging is needed, circuits can be individually configured into the same or different VLANs. If tagging is enabled on a circuit's ipif, the associated Ethernet interface transmits tagged frames to the Ethernet switch port on the opposite end of that link. Both must be equivalently configured before the Ethernet link can come up.

3.3.28 Extension with Fibre Channel Routing

The SAN42B-R and IBM b-type Gen 6 Extension Blade are not specifically FC Routers (FCR) but do support FCR. FCR functionality has been designed into the switching ASICs found in most B type FC products. There is no requirement for a specialized FCR platform, and FCR is easily enabled on any switch or Director by applying the Integrated Routing (IR) license.

It is easy to build a preferred-practice Edge-Extension Backbone-Edge architecture by enabling FCR on the IBM SAN42B-R or IBM SAN256B-6 and SAN512B-6 Directors with IBM b-type Gen 6 Extension Blade.

Virtual Fabrics are useful when implementing Edge-Backbone-Edge with the IBM b-type Gen 6 Extension Blade, as shown in Figure 3-15. EX_Port connections can be made to an edge Logical-Switch from the backbone Base-Switch. It is not necessary to purchase a separate switch or Director.

It is more cost-effective to put the IBM b-type Gen 6 Extension Blade into an IBM SAN256B-6 or SAN512B-6 director that is part of the edge fabric and create a backbone from the Base-Switch. The EX_Ports and VE_Ports on the IBM b-type Gen 6 Extension Blade are members of the Base-Switch. All other device connection ports (F_Ports) and E_Ports participating in the edge fabric belong to a Logical-Switch other than the Base-Switch.

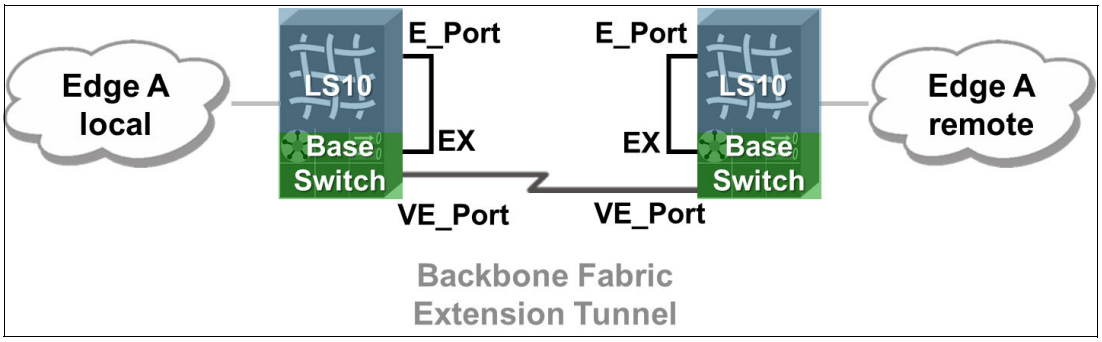


Figure 3-15 Using Virtual Fabric Logical Switches to implement Edge-Backbone-Edge

3.3.29 High availability WAN side architectures

This section describes the following architectures:

- Four box
- Site-to-multisite architecture
- Multiple parallel WAN links
- 3DC triangle with failover metrics
- Ping-ponging

Four box

A popular design is the four IBM SAN42B-R HA (high availability) architecture, as shown in Figure 3-16. This design can easily accommodate two 10 Gbps WAN connections. However, in practice many enterprises suffice with a single WAN connection. The number of WAN connections is a requirement determined by the enterprise on a case-by-case basis and mostly predicated on cost versus risk.

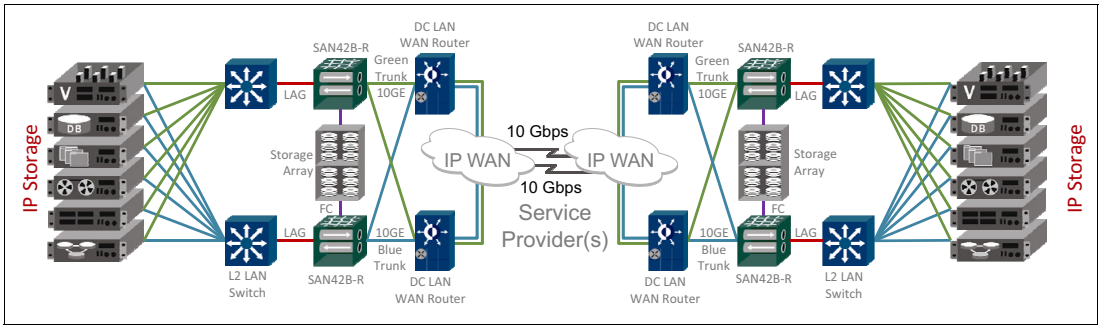


Figure 3-16 Four IBM SAN42B-R high availability architecture with FCIP and IPEX

The IBM SAN42B-R has bandwidth licensing tiers with 5 Gbps as the base tier. The base IBM b-type Gen 6 Extension Blade has no licensing tiers and is capable of operating at full capacity. If future growth dictates or applications such as IPEX or tape are added, the SAN42B-R can be simply upgraded by using software licenses to accommodate larger WAN bandwidths of 10 Gbps (Upgrade1) and 40 Gbps (Upgrade2).

The four IBM SAN42B-R HA architecture uses a single trunk that takes advantage of two paths. The data center IP network has dual WAN routers for redundancy. The trunk has two circuits, one for each WAN router pathway. An extension trunk is equivalent to a single FC ISL between SAN42B-Rs. The array controllers see a connection between corresponding replication ports.

As shown in Figure 3-16 on page 70, there are two sites and each site has two SAN42B-Rs, one attached to controller “A” and one attached to controller “B”. There are two extension trunks, one on each SAN42B-R. Each trunk has two circuits and each circuit has its own dedicated Ethernet interface. The “A” controllers use the green circuit trunk and the “B” controllers use the blue circuit trunk. The circuits are forwarded by the data center switches and routers to specific WAN connections.

Adaptive Rate Limiting (ARL) is used in this architecture because a total of four 10 Gbps Ethernet interfaces from two SAN42B-Rs compete for 20 Gbps of WAN bandwidth. ARL is used to reduce the bandwidth from the SAN42B-Rs to the available capacity of the WAN. This configuration prevents congestion and massive packet loss, which result in degraded performance.

This design has a capacity, from the point of view of the storage array, of twice the bandwidth of the WAN assuming 2:1 compression is achievable using Fast-Deflate. There are other compression algorithms that get higher compression ratios. However, they are not capable of accommodating the 20 Gbps WAN bandwidth in this example. The applicable compression algorithm depends on mostly operating mode and WAN bandwidth. The obtainable compression ratio is specific to the actual data being compressed.

Some portion of the bandwidth is also be used by the IPEX traffic, which can use Deflate or Aggressive-Deflate compression. The same DP manages this bandwidth as well and ARL manages the bandwidth across DPs or across systems.

Note: IBM makes no promises, guarantees, or claims as to the achievable compression ratio for specific customer data.

If data security is needed, IPsec is provided. The preferred practice is to use IPsec between extension platforms. Typically, devices such as firewalls are not capable of maintaining throughput at the demanding rates of storage replication, ultimately becoming a bottleneck. Firewalls must not alter the TCP stream in any way. Altering the TCP stream is not supported.

Site-to-multisite architecture

Figure 3-17 illustrates a Site to Multisite architecture. This architecture uses point-to-point Extension Trunks. Two (VE24 and VE34) originate from the main data center, and one terminates at remote site A (VE24), and one at remote site B (VE34). Both disk RDR and tape flows are traversing the network and a significant amount of bandwidth is required for this purpose. Extension Trunking is used to aggregate bandwidth and provides failover between multiple physical connections to the WAN.

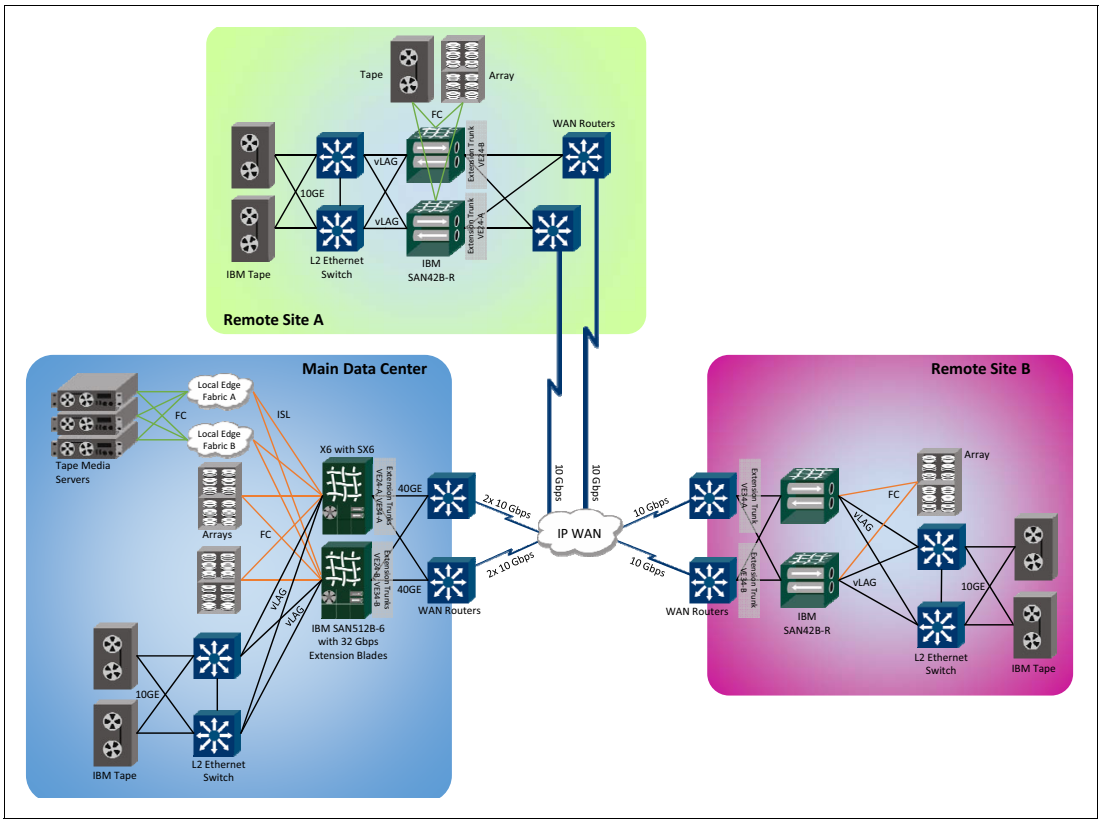


Figure 3-17 Site-to-multisite design using Extension Trunking, fabric A only

Table 3-1 shows failure impact on bandwidth availability for a site-to-multisite dual fabric design that uses Extension Trunking with ARL.

Table 3-1 Failure impact on bandwidth availability

	Normal Operation (Upgrade1 + Upgrade2 Installed)					Failure Analysis		
Sites	Trunk Bandwidth		# Circuits	Min ARL (Gbps)	Max ARL (Gbps)	Main DC Port/Optic/Cable Failure	Site A Port/Optic/Cable Failure	Site B Port/Optic/Cable Failure
	VE24	VE34	Min/Max					
Fabric A (SAN42B-R-A)								
Site A VE24 on DP0 • GE2 circuit0 • GE3 circuit1 • GE4 circuit2 • GE5 circuit3	10 Gbps DP0 is dedicated to site A eHCL BW reserved on DP1		4 (0-3) One circuit to each WAN link Ax2 & Bx2	0-3: 2.5 Normal Operation 10 Gbps to site A	0-3: 5 If SAN42B-R or WAN router goes down	No service degradation	No service degradation	No service degradation
Site B VE34 on DP1 • GE6 circuit0 • GE7 circuit1 • GE8 circuit2 • GE9 circuit3		10 Gbps DP1 is dedicated to site B eHCL BW reserved on DP0	4 (0-3) One circuit to each WAN router x #links Ax2 & Bx2	0-3: 2.5 Normal Operation 10 Gbps to site B	0-3: 5 If SAN42B-R or WAN router goes down	No service degradation	No service degradation	No service degradation
Fabric B (SAN42B-R-B)								
Site A VE24 on DP0 • GE2 circuit0 • GE3 circuit1 • GE4 circuit2 • GE5 circuit3	10 Gbps DP0 is dedicated to site A eHCL BW reserved on DP1		4 (0-3) One circuit to each WAN router x #links Ax2 & Bx2	0-3: 2.5 Normal Operation 10 Gbps to site A	0-3: 5 If SAN42B-R or WAN router goes down	No service degradation	No service degradation	No service degradation
Site B VE34 on DP1 • GE6 circuit0 • GE7 circuit1 • GE8 circuit2 • GE9 circuit3		10 Gbps DP1 is dedicated to site B eHCL BW reserved on DP0	4 (0-3) One circuit to each WAN router x #links Ax2 & Bx2	0-3: 2.5 Normal Operation 10 Gbps to site B	0-3: 5 If SAN42B-R or WAN router goes down	No service degradation	No service degradation	No service degradation
Normal Operation and Failure Analysis								
Total (Gbps)	Site A: 20	Site B: 20		Normal Op Main DC: 40 Site A: 20 Site B: 20	Box down: Main DC: 20 Site A: 10 Site B: 10	20	20	20
WAN Link BW	Main DC: WAN Links-X1&X2: 20	Main DC: WAN Links-Y1&Y2: 20			WAN down: Main DC: 30 Site A: 15 Site B: 15	40	20	20

The Main Data Center's WAN is dedicated to storage applications and accommodates 40 Gbps of bandwidth through 4x 10 Gbps links. Each remote data center (A and B) has 2x 10 Gbps links for redundancy. The 10GE interfaces contain one circuit each. However, because the maximum ARL value is only 5 Gbps, it is possible to converge two circuits onto a single interface to conserve Ethernet interfaces. There are eight 10GE WAN side interfaces available when SAN42B-R/IBM b-type Gen 6 Extension Blade is in Hybrid mode.

Site A uses trunk VE24 on DP0 and site B uses trunk VE34 on DP1. Each fabric (A & B) has four circuits going to Site A and four circuits to site B for a total of eight circuits going to each site. Each circuit is configured with ARL for a minimum bandwidth of 2.5 Gbps and a maximum of 5 Gbps.

Normal operation is 8 circuits x 2.5 Gbps = 20 Gbps. If a SAN42B-R or IBM b-type Gen 6 Extension Blade goes offline, the calculation changes to four circuits x 5 Gbps = 20 Gbps. ARL automatically changes the bandwidth from 2.5 Gbps to 5 Gbps to maintain full service to the application.

Multiple parallel WAN links

Multiple parallel WAN links are a common architecture used for RDR and tape. Two or more 10 Gbps WAN connections are becoming commonplace, satisfying both bandwidth requirements and redundancy. The links can be purchased and provisioned by different service providers and take disparate paths.

Usually, one of three strategies is deployed across multiple parallel WAN connections:

1. Traffic is load balanced across all the links and, if there is a failure, the traffic is rebalanced to the remaining links. This is the default behavior.
2. Traffic is confined to one or some subset of the links until there is a failure, and then traffic uses the designated failover links. This strategy uses metrics 0 and 1 to define active and passive circuits respectively.
3. Traffic prefers a subset of links and, if those links become fully utilized, traffic starts to use the bandwidth of the non-preferred links. This technique is referred to as Fill and Spill.

This example has two data centers with two 10 Gbps WAN connections between them and uses the SAN42B-R/IBM b-type Gen 6 Extension Blade Extension Trunking with ARL solution. As shown in Table 3-2, there is a single extension trunk for the “A” fabric with four circuits and the same for the “B” fabric. Each circuit has been assigned to its own dedicated 10 Gbps Ethernet interface on the SAN42B-R/IBM b-type Gen 6 Extension Blade and the trunk is capable of 20 Gbps of aggregated bandwidth.

Traffic is load balanced across the four circuits on a per-batch basis. The amount of bandwidth on each circuit does not have to be the same, although symmetry is preferred. The tunnel and circuit configuration on both ends must be the same.

Ethernet interfaces should connect directly into the WAN switch/router established by the service provider. This equipment is often referred to as Customer Premise Equipment (CPE). Extension circuits should be statically fixed to an associated WAN connection as shown in Table 3-2.

Table 3-2 Extension circuits fixed to associated WAN connections

Extension Device	VE_Ports	Circuit	Ethernet Interface	ARL Minimum (Gbps)	ARL Maximum (Gbps)	Metric	Failover Group	WAN Router	WAN Link
SAN42B-RA	VE24	Circuit0	GE2	2.5	10	0 (default)	0 (default)	A	1
		Circuit1	GE3	2.5	10	0	0		2
		Circuit2	GE4	2.5	10	0	0	B	1
		Circuit3	GE5	2.5	10	0	0		2
SAN42B-RB	VE24	Circuit0	GE2	2.5	10	0	0	A	1
		Circuit1	GE3	2.5	10	0	0		2
		Circuit2	GE4	2.5	10	0	0	B	1
		Circuit3	GE5	2.5	10	0	0		2

See Figure 3-18.

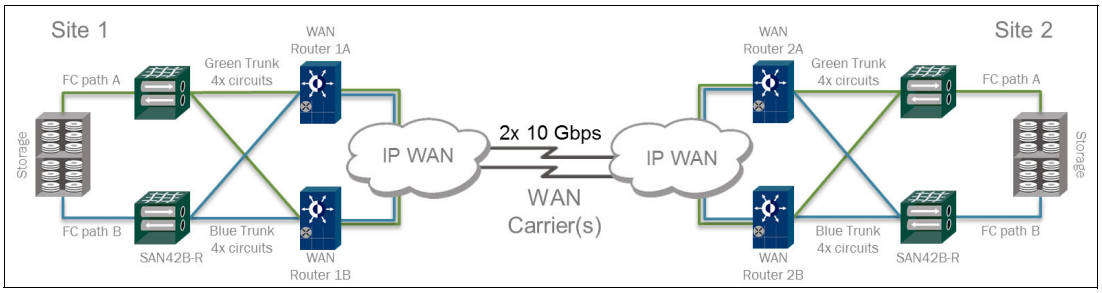


Figure 3-18 Multiple Parallel WAN Links architecture

Statically assigning circuits to WAN connections provides the highest degree of stability and availability. There is no reason for the IP network to perform rerouting, Link Aggregation (LAG), Link Aggregation Control Protocol (LACP), Port Channeling, load balancing, failover/failback, or any similar functions.

LAG is not offered on the SAN42B-R/IBM b-type Gen 6 Extension Blade WAN side. These listed functions are completely taken care of by Extension Trunking. The only thing that needs to be done is statically assigning circuits to specific WAN connections. A wide variety of network problems are eliminated when using static forwarding. There is very little that can go wrong with the IP network when it is static.

If a WAN connection goes down, for each metric 0 circuit, the keepalive timer expires and the circuit changes to a down state. If it goes down and there are any metric 1 circuits across alternative paths, those circuits become active.

If the WAN connections are shared, sometimes full 10 Gbps circuits might not be practical and less bandwidth must be considered. When circuits comes back online, Extension Trunking automatically adds them back into the trunk, resuming transmission and load balancing. The keepalive mechanism is infinitely persistent and detects when a path is available again.

3DC triangle with failover metrics

Many large enterprises, especially financial institutions involved with frequent and sizable transactions, have requirements for both synchronous RDR (RDR/S) and asynchronous RDR (RDR/A). Why use both? RDR/S replicates every transaction in real-time over a limited distance: RPO = 0. RDR/A cannot replicate I/O in real-time, but has no real distance limitations: RPO>0. There are many issues to consider when using this type of architecture.

RDR/S is required to confidently capture all transactions (DB write I/O), except for the write that is in progress as the disaster disables the infrastructure. Other than the last write in progress, every acknowledged write to disk is safely replicated to a location within a synchronous radius of usually 100 km or less. A relatively local (metro area) remote location for backup operations is often referred to as a *bunker site*.

Due to speed of light limitations, RDR/S is limited in distance before causing excessive response times for user applications. 100 km (62 mi.) is about the average maximum distance. However, the distance actually depends on a number of factors not discussed in this paper. RDR/S provides a significantly better recovery point objective (RPO), which might represent a significant amount of transactional revenue.

Considering the types of disasters and acts of man that can occur in today's world, it is prudent for large enterprises to replicate data to a third site more assured of being outside the catastrophe perimeter. Extending data beyond the radius of the bunker site requires RDR/A.

IBM offers Global and Metro Mirror, which manage RDR/A and RDR/S respectively. Loss of any site or connection does not inhibit replication and there is no need to reinitialize any volume pair relationship (which can take a considerable amount of time, depending on the size of the volume).

As shown in Figure 3-19, this example does not use the SAN42B-R/IBM b-type Gen 6 Extension Blade for the RDR/S, which is done across a metro DWDM network connected through B type FC switches providing the needed long-distance Buffer-to-Buffer Credits (BBCs). This mission-critical architecture deploys A and B fabrics for redundancy by way of physically isolated networks.

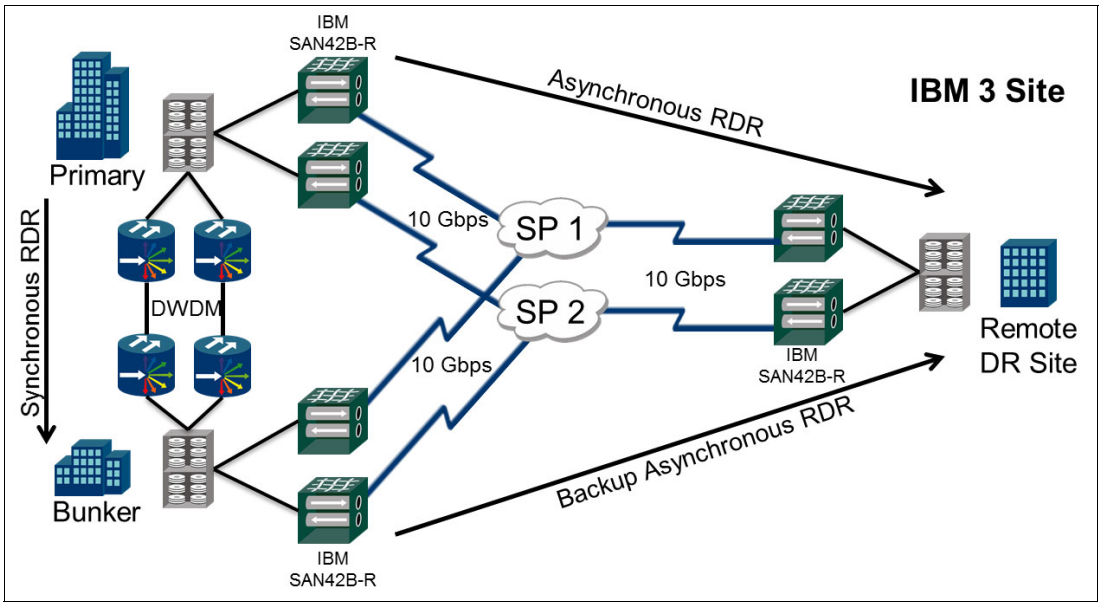


Figure 3-19 Three data center RDR architecture

The infrastructure used to transport RDR/S differs from RDR/A. RDR/S connections are short and many service providers can offer DWDM for metro distances at a reasonable price, although native FC or FICON is often more expensive compared to IP. RDR/A can operate with much less bandwidth relative to RDR/S because it requires the average instead of peak bandwidth to be provisioned. RDR/A needs only the average over a finite period.

Evaluated over a finite period, the period with the most bandwidth is the accepted sample. Samples cannot be too long such that respite periods lower the average. However, samples cannot be so short that averages approximate actual peaks, like in RDR/S. Generally, 15-30 minute periods provide nice approximations.

Consider the following design considerations for RDR/A traffic:

- To maintain primary-to-remote site replication, should RDR/A traffic between the primary and remote site be rerouted across DWDM to the bunker site and forwarded to the remote site or not?
- Alternatively, should rerouting be prevented, allowing failover and forcing the bunker site to take over replication to the remote site?

This decision depends on a few factors:

- Is there enough bandwidth available on the DWDM network to fail over RDR/A traffic and maintain peak RDR/S? A separate channel on the DWDM is preferred.
- What is preferred practice for the RDR application software? Often, preferred practice prevents RDR/A flows from being rerouted to the bunker site and then to the remote site, unless allocated bandwidth is available that is specific to this purpose. This restriction is because often encroachment onto the RDR/S bandwidth is not workable. Circuit metrics best facilitate alternative infrastructure paths.

Ping-ponging

Ping-ponging occurs when a WAN connection goes down and traffic is forced to other remote sites before ultimately find its way to the destination. In Figure 3-20, one of the links (A to C) goes down. However, the FC routing tables in the SAN42B-R and IBM b-type Gen 6 Extension Blade can still reach the final destination by ping-ponging traffic (A to D to B to C). The FSPF cost is greater but is still viable because it is the only remaining path.

Frequently, this configuration introduces considerable latency because a single RTT is now three RTT. Furthermore, it causes undesired WAN utilization effects and breaks protocol acceleration by performing optimization on exchanges that are already being optimized. Repeated protocol acceleration is not supported. Ultimately, it is not a preferred practice architecture.

The preferred practice architecture uses Virtual Fabric Logical Switches (VF LS) to create isolated paths, which does not permit routing between one LS and another. For example, if the VE_Port for circuits crossing WAN A to C are isolated into a Logical Switch, if WAN A to C goes offline, traffic cannot be routed into a different LS containing the VE_Port used for WAN A to D. Alternatively, this can be fixed using Traffic Isolation Zones (TIZs) with failover disabled, but this method has the drawback of more complex configuration and operations.

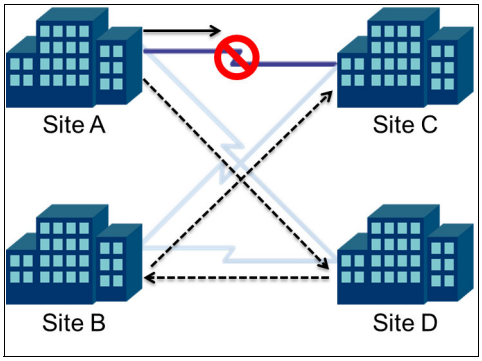


Figure 3-20 Ping-ponging

3.3.30 Dual 40GE links

Mainframe environments can consume large amounts of bandwidth taking advantage of the SAN42B-R/IBM b-type Gen 6 Extension Blade 40GE WAN side interfaces. In this design, VF LSs at each site are connected using a 20 Gbps trunk consisting of 2x 10 Gbps circuits that are assigned to the redundant 40GE interfaces (one 10 Gbps circuit to each).

The “Red” Logical Fabric is dedicated to IBM z/OS Global Mirror (formerly known as XRC) and the “Green” Logical Fabric is dedicated to tape. One IBM SAN42B-R/IBM b-type Gen 6 Extension Blade has a total of 40 Gbps of WAN side capacity, 20 Gbps per DP. 20 Gbps of bandwidth is associated with the Green Logical Fabric and 20 Gbps is associated with the Red Logical Fabric.

Each 20 Gbps trunk is defined by a single VE_Port, allowing protocol acceleration to function if needed. Each VE_Port lives within a Logical Switch. Data flows cannot cross LS boundaries, confining them to a deterministic path. Each trunk is composed of two 10 Gbps circuits taking different WAN and IP network (switches and routers) paths. The circuits from the different Logical-Switches share the same 40GE interfaces. The 40GE interfaces stay in the Default-Switch to accommodate circuits from other Logical-Switches, which is referred to as Ethernet Sharing.

If a WAN outage event occurs, it is possible for either the IP network to failover traffic, or let Extension Trunking perform failover/failback while the IP network remains static. Practical experience has demonstrated that a static IP network is more stable than one attempting to converge. Letting Extension Trunking manage failover and failback is more reliable, lossless, and guarantees in-order delivery. Failover groups and metric 1 circuits have been configured on the opposing 40GE interfaces. The metric 1 circuits travels different IP network paths to maintain as much application bandwidth as possible if a path goes offline.

40GE interfaces on the SAN42B-R and IBM b-type Gen 6 Extension Blade can accommodate 4x 10 Gbps circuits because during normal operation, these interfaces run at 20 Gbps (2x metric0 = 2x 10 Gbps circuits, metric1 are passive = 0), and during failover they operate at 40 Gbps (2x metric0 + 2x metric1 = 4x 10 Gbps circuits). There are eight circuits total between the two DPs and each circuit is assigned to one of the two 40GE interfaces. There are four 10 Gbps circuits on each 40GE interface, two from each trunk (VE24 Red and VE34 Green), two are active and two are passive, as shown in Table 3-3. If one of the WAN links or a WAN switch/router goes down, on the remaining path the metric0 circuits continues to function as normal and the metric1 backup circuits come online. Overall, the same system bandwidth is maintained as during normal operation.

Table 3-3 Circuits and interfaces

Extension Devices	VE_ Ports	10 Gbps Circuits	Logical Switches	40GE Interfaces	Circuit Metric	Failover Groups	WAN Routers	WAN Links
SAN42B-RA	VE24	Circuit0	Red Logical Fabric (Global Mirror)	GE0	0	0	A	A
		Circuit1		GE0	1	1		
		Circuit2		GE1	0	1	B	B
		Circuit3		GE1	1	0		
	VE34	Circuit0	Green Logical Fabric (Tape)	GE0	0	0	A	A
		Circuit1		GE0	1	1		
		Circuit2		GE1	0	1	B	B
		Circuit3		GE1	1	0		
SAN42B-RB	VE24	Circuit0	Red Logical Fabric (Global Mirror)	GE0	0	0	A	A
		Circuit1		GE0	1	1		
		Circuit2		GE1	0	1	B	B
		Circuit3		GE1	1	0		
	VE34	Circuit0	Green Logical Fabric (Tape)	GE0	0	0	A	A
		Circuit1		GE0	1	1		
		Circuit2		GE1	0	1	B	B
		Circuit3		GE1	1	0		

When the circuits enter the WAN switch or router, because each circuit has its own source and destination IP address pair, the individual circuits can be statically routed across various WAN connections as necessary. There are three popular methods for statically confining circuits to a certain path regardless of the path being up or down:

- A pure L2 network using VLANs and no router interface. Data flows are confined to a specific WAN interface.
- Use two narrow static routes. The first route has default administrative distance for directing circuit subnets to the intended interface for the specific WAN path. The second route is an identical route with a larger administrative distance pointing to the NULL interface. If the first route is down, the second route ensures that the flows are not sent to a wider route such as a default route or default gateway.
- Use Policy Based Routing (PBR) and force specific circuit traffic to specific WAN connections.

If VLAN tagging has been chosen as the preferred method for sorting data flows into specific WAN connections, the various circuits use the same physical Ethernet interface (in this case 40GE), it is necessary to identify the different frames with VLAN tagging (802.1Q). This turns the Ethernet link into a trunk (Ethernet trunk is different than an extension trunk).

The SAN42B-R/IBM b-type Gen 6 Extension Blade can VLAN tag individual Ethernet frames from each circuit so that, when it enters the data center's Ethernet switch, the frame is identified and directed toward the correct WAN connection. QoS (802.1P, DSCP, or both) can also be marked on these frames. When it is in the Ethernet switch, the VLAN's path to the WAN is determined and configured by the IP network administrators.

FICON applications require lossless failover to prevent frame loss. When one or more frames are lost in-flight and failover or rerouting subsequently resumes traffic, this situation causes Out Of Order packets (OOO). OOO packets results in a mainframe IFCC, which is an undesirable event. Trunked FICON frames across multiple WAN connections, as shown in this example, are always delivered to the ULP in-order with no missing data. This is a function of the TCP supervisor that is managing data across each circuit in a trunk.

To prevent IFCCs, it is preferred to completely halt traffic rather than letting the IP network reroute traffic, which keeps flows moving. This process is not entirely intuitive. KATOV is set to 1 second on FICON circuits and recovery occurs in 1 second. Virtual Fabric Logical-Switches facilitate deterministic paths with no failover and are preferred over Traffic Isolation Zones (TIZ). If TIZs are used, they are set to not failover when a path becomes available. This is a requirement in mainframe environments and unique to Extension Trunking.

IBM Extension supports Global Mirror (FICON SDM over FCIP) and tape (FICON tape over FCIP, TS7700 Grid over IPEX, or both) across the same circuits simultaneously. There is no requirement to separate the data flows. In fact, commingling different data flows across the same trunk is fully supported and has the benefit of superior bandwidth management. By running the Global Mirror and Grid flows through the same WAN egress scheduler, WAN bandwidth usage is optimized. See Figure 3-21.

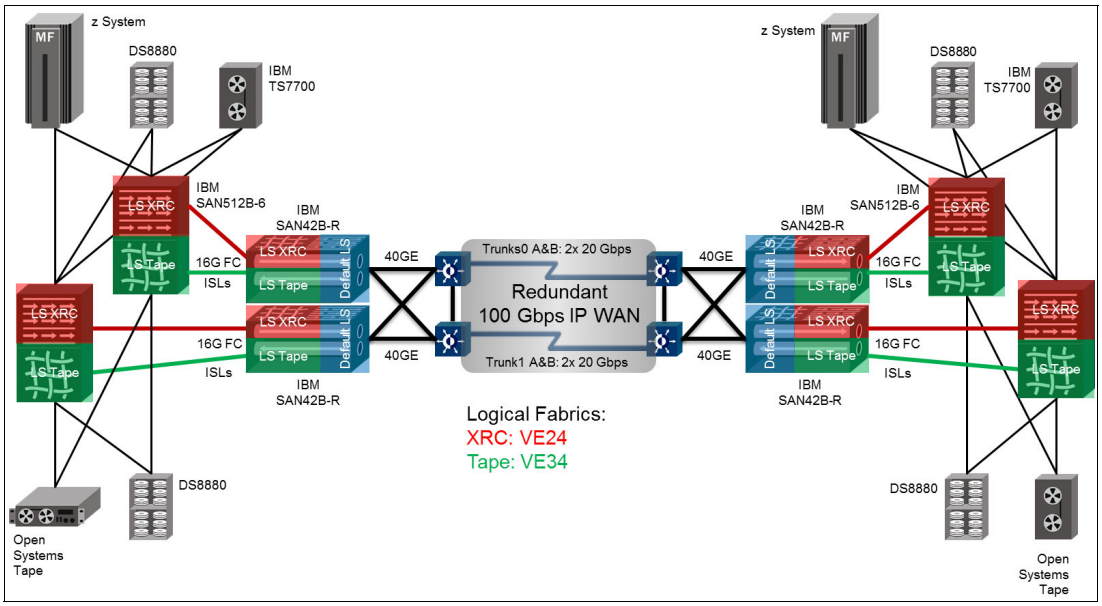


Figure 3-21 Protocol intermix of FICON and Open Systems across 40 Gbps IBM Extension

3.4 The LAN side

The LAN side refers to IPEX only, and it is the local connectivity of IP storage devices from the data center LAN into the SAN42B-R or IBM b-type Gen 6 Extension Blade. IPEX switching and processing is done internally by the Ethernet interfaces, FPGAs, and DPs. IPEX logically uses VE_Ports even when no FC or FICON traffic is being transported. IPEX traffic does not traverse a physical VE_Port. Instead, VE_Ports logically represents the tunnel/trunk that IPEX sends traffic into.

VE_Ports are logical entities, and are better thought of as a tunnel/trunk end-point instead of an FC port. IPEX uses a lan.dpx ipif (LAN, DP = Data Proc #, ipif = IP Interface, x = 0 or 1). There is one lan.dpx ipif for each VLAN coming from the LAN on at least one of the DPs. This lan.dpx ipif acts as the IPEX gateway for the incoming IP storage traffic.

IP storage end-devices communicate with the `lan.dpx` ipif either directly through L2 broadcast domain or indirectly using Policy Based Routing. The `lan.dpx` ipif becomes the gateway for data flows headed toward the remote data center. At the `lan.dpx` ipif, end-device TCP sessions are locally terminated. These terminated TCP sessions are reformed again at the remote `lan.dpx` ipif. This is referred to as TCP Proxy, and has the advantage of local ACK (Local Acknowledgements), which gains great TCP acceleration.

WAN Optimized TCP (WO-TCP) is used as the transport between local and remote SAN42B-R/IBM b-type Gen 6 Extension Blades. Which side is “local” or “remote” is merely a matter of perspective. However, often the primary data center is the local side and the DR site is the remote side.

Of the 16x 10GE interfaces on the SAN42B-R/IBM b-type Gen 6 Extension Blade, up to eight can be used for IPEX LAN side connectivity to either direct end-device connectivity or data center LAN switch LAGs. Eight 10GE interfaces are reserved for WAN side (circuit) connections.

3.4.1 IPEX gateway

The IPEX Gateway is like any other router gateway, except it is limited to a certain number of supported IP storage flows. Those flows are subsequently optimized, encrypted, and use extension trunking to span data centers. The gateway interface is a LAN side IP interface and referred to as “lan.dpx” (x can be either DP0 or DP1). The lan.dpx sits behind up to eight IPEX LAN Ethernet interfaces. All LAN Ethernet interfaces have equal access to any lan.dpx. The Ethernet interface speed can be set to either GE or 10GE. The interface speed requires a matched speed capable optic.

When configuring `lan.dpx` interfaces (ipif), you must select either `lan.dp0` or `lan.dp1`. You would select DP0 if your tunnel/trunk is using VE24-28 and DP1 if your tunnel/trunk is using VE34-38.

3.4.2 Ethernet switching and IP routing

IBM SAN42B-R and IBM b-type Gen 6 Extension Blade are not Ethernet switches or IP routers. They do not use or need Spanning Tree Protocol (STP), and there are no routing protocols. The IPEX LAN side cannot become an active member of an Ethernet Fabric, which involves DCB. IPEX LAN interfaces can indeed connect to any Ethernet Fabric or traditional Ethernet network as a transport, but cannot participate in switching or routing.

One Ethernet interface from each group on the SAN42B-R/IBM b-type Gen 6 Extension Blade should first be used before the second interface from the same group is used. For example, use interfaces GE2 to GE9 first, possibly GE2-GE5 for WAN side and GE6-GE9 for LAN side. If the second interface in a group is used, it must be the same speed as the other interface, so both must be either GE or 10GE. If a group has mixed speeds, there is blocking by the slower interface to the faster interface.

Table 3-4 shows SAN42B-R/IBM b-type Gen 6 Extension Blade Ethernet port grouping (each column is a group). This table does not represent the layout of the interfaces on the front of the SAN42B-R or IBM b-type Gen 6 Extension Blade.

Table 3-4 SAN42B-R/IBM b-type Gen 6 Extension Blade Ethernet port grouping

Group0	Group1	Group2	Group3	Group4	Group5	Group6	Group7
GE2	GE3	GE4	GE5	GE6	GE7	GE8	GE9
GE10	GE11	GE12	GE13	GE14	GE15	GE16	GE17

3.4.3 Direct LAN connections

Direct connects to the LAN side Ethernet interfaces for IPEX can be made. However, 100% of the traffic coming from the end-device, for example DS8000 replication ports, must be destined for the remote data center. No traffic from the local end-device’s Ethernet interface can be destined for other local devices within the same data center.

IBM SAN42B-R and IBM b-type Gen 6 Extension Blade are not Ethernet switches or IP routers. Therefore, it is not possible for any traffic entering the LAN side ports to be rerouted or switched back out another LAN side port, making local data center communications impossible. Any traffic that does enter LAN side ports with a local destination is dropped. This implies 100% of the traffic coming into the LAN side ports must be destined for the remote data center.

If the end-device requires both local and remote communications to go through the same Ethernet interface, direct SAN42B-R/IBM b-type Gen 6 Extension Blade connectivity is not possible. The end-device's Ethernet interface must connect through an intermediate LAN switch and that LAN switch LAGed to the SAN42B-R/IBM b-type Gen 6 Extension Blade.

The LAN switch is responsible for switching traffic either to local data center devices or to the IPEX LAG. The intermediate IP storage LAN switch facilitates local communications as well as remote communications, and sorting various flows that are shared across an end-device's Ethernet interface.

3.4.4 Link Aggregation

LAG is a standard (IEEE 802.1AX). A single Ethernet link can be used, but LAG is preferred because it provides more redundancy using multiple ports, optics and cables. In addition, LAG aggregates multiple links up to 4x 10 Gbps. Two 40-Gbps LAGs can be connected, or four 20-Gbps LAGs, or any variant in between.

When connecting a LAG, it is important that only one path from the data center LAN to the SAN42B-R/IBM b-type Gen 6 Extension Blade exists. Connecting a LAG to two different data center LAN switches that are part of the same LAN forms two paths (path1 = DC-LAN-A, path2 = DC-LAN-B) for data to reach the IPEX GW (lan.dpx ipif), which is not supported.

Address Resolution Protocol (ARP) responses are limited to just one path and the first reply received is the only LAG used. Other LAG pathways are not used. In Figure 3-22, there is an A path and a B path similar to traditional SAN designs. This is supported.

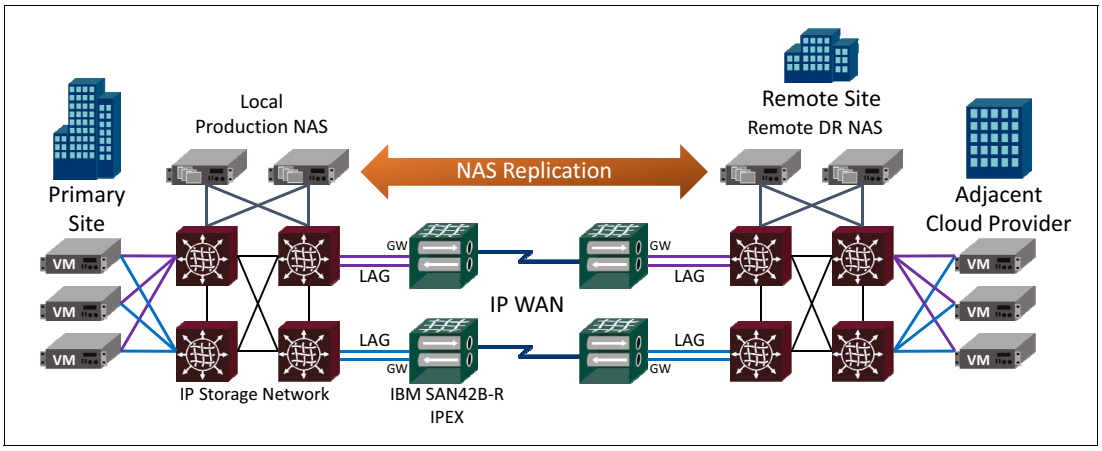


Figure 3-22 Tape Replacement using NAS over IBM SAN42B-R IPEX

In the architecture in Figure 3-23, the IBM SAN42B-R connects to an Ethernet Fabric using Virtual Link Aggregation (vLAG). vLAG is similar to LAG except a vLAG is capable of forming a LAG across multiple switches. Different Ethernet switch manufacturers might refer to this by capability with different names such as Cisco vPC. Because a vLAG spans more than one switch, it has high availability for each path.

The purple vLAG is a single LAG. The difference is that links comprising the LAG are split across two switches. There can be two links to each switch. The same is true with the blue vLAG. It is the job of the Ethernet Fabric to present the vLAG to the SAN42B-R as a single LAG.

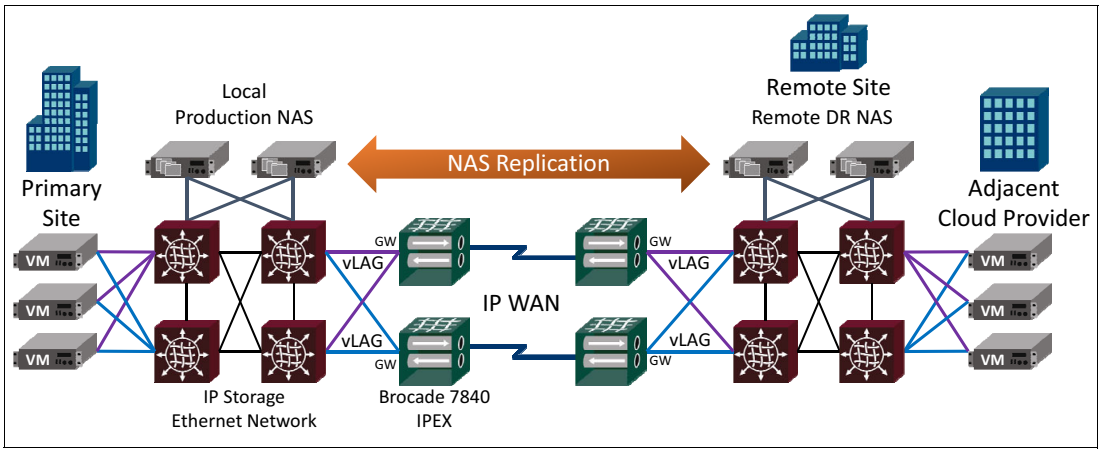


Figure 3-23 VLAG

3.4.5 LAG with VLANs

A LAG/vLAG with VLAN Tagging (802.1Q) enabled carries traffic from multiple VLANs at the same time across the physical connection or LAG of connections. Suppose end-devices are on various different VLANs. There is no reason to create a different LAG for each VLAN. Instead, one LAG is created using VLAN tagging to accommodate all the VLANs.

On the SAN42B-R/IBM b-type Gen 6 Extension Blade, a unique lan.dpx ipif must be created as the gateway for the subnet living in that VLAN. The lan.dpx ipif is configured with the corresponding VLAN ID so that it has visibility into that VLAN and can communicate with the subnet.

If VLAN tagging is configured on one side of an Ethernet link, it must be configured identically on the opposite side because otherwise the link will not come up. This configuration is in addition to forming the LAG. If each individual link cannot come up on its own, the LAG cannot form as a group.

Beware of the difference between an Ethernet interface that is configured to tag Ethernet frames (802.1Q) and an Ethernet interface existing within a VLAN and not tagging frames (normal untagged Ethernet). This disparity often leads to trouble getting Ethernet links to come up.

3.4.6 Link Aggregation Control Protocol

The IBM SAN42B-R and IBM b-type Gen 6 Extension Blade do not support LACP. LACP is a protocol for automating the formation of a LAG when parallel Ethernet links are connected between devices that support it.

To form a LAG on the SAN42B-R/IBM b-type Gen 6 Extension Blade, the links have to be manually added to a LAG group on both ends (the SAN42B-R/IBM b-type Gen 6 Extension Blade side and the data center LAN side). This addition does not happen automatically because LACP is not supported.

3.4.7 Traffic Control List

Before IPEX can function, a TCL must be configured. The default TCL is “Deny All.” Additionally, no traffic has been explicitly directed to any particular VE_Port (tunnel/trunk). It is strongly preferred that you do not configure an “Allow All” TCL.

There are a number of parameters that a TCL can use to narrow and specify the exact traffic assigned to a VE_Port. Often, a proper TCL specifies end-device source and destination subnets or specific host IP addresses. Only specific traffic is permitted. All other traffic is dropped. Traffic is never hair-pinned back out of the LAN interfaces, which prevents loops.

If host traffic is being dropped, consider its arrival at the lan.dpx ipif an error. Unicast traffic not intended to pass through IPEX should never be directed to a lan.dpx ipif gateway. End-device traffic not intended for IPEX should use the data center’s normal gateway as its default route.

End-devices should have a special IP route for any remote subnets or remote host IP addresses in which traffic uses IPEX, and those special IP routes specify IPEX (lan.dpx ipif) as the gateway for that traffic.

TCL syntax is beyond the scope of this publication.

3.4.8 TCL non-terminated TCP traffic

Each DP can accommodate a maximum of 512 TCP sessions. Beyond 512, any new sessions are denied. 512 sessions accommodates nearly all storage environments between data centers, mostly replication. In the specific case of IBM TS7700, the number of control TCP sessions can be significantly high. The actual number generated varies considerably based on clusters and the grid architecture.

TS7700 control sessions do not require IPEX optimization. Control sessions pass very little traffic, do not hinder tape job performance, and often sit idle for long periods. To prevent these control sessions from consuming valuable IPEX TCP slots, the TCL must identify them and assign them to be non-terminated. Non-terminated means that IPEX does not intercept the control TCP session and does not proxy it.

Normally, IPEX intercepts a data TCP session, terminates it locally, and creates a TCP proxy. IPEX provides local acknowledgements to gain TCP acceleration and performance enhancement. WO-TCP transports the data to the remote side. On the remote side, IPEX recreates a local TCP session to the destination end-device. Now both ends are proxied and in the middle WO-TCP is the high-speed transport.

Non-terminated TCP sessions undergo high-efficiency encapsulation and are passed through the extension tunnel/trunk. No compression or IPEX batching is done. If enabled, IPsec is applied. Non-terminated datagrams are each processed one at a time.

Excessive non-terminated traffic can deplete processor resources, negatively affecting extension performance. Non-terminated is intended for TCP sessions not benefiting from IPEX optimizations and generating little traffic. Without the use of non-terminated, the DP ceiling of 512 TCP sessions would be exceeded.

In this example, the TS7700 grid is on a local subnet of 10.10.10.0/24 and a remote subnet of 10.11.11.0/24. IBM TS7700 control TCP sessions use the protocol ports 1415 and 1416. It is possible to specify a range of protocol ports, as in this case 1415 - 1416. All traffic matching the first statement is sent non-terminated (no TCP Proxy). The second statement matches to the tape data flows, which do not use protocol ports 1415 and 1416. Control traffic or not, all traffic is sent into tunnel/trunk VE24. Traffic not matching either statement is dropped.

Example 3-1 shows TCL for IBM TS7700 non-terminated control traffic.

Example 3-1 TCL for IBM TS7700 non-terminated control traffic

```
portcfg tcl IPEX_ctl create --priority 5 --admin enable --action allow --src-addr
10.10.10.0/24  --dst-addr 10.11.11.0/24 --proto-port 1415-1416 --target 24
--non-terminated enable

portcfg tcl IPEX_dat create --priority 6 --admin enable --action allow --src-addr
10.10.10.0/24  --dst-addr 10.11.11.0/24 --target 24
```

UDP Traffic

User Datagram Protocol (UDP) traffic is not a connection-oriented protocol like TCP. Delivery and order of data is best effort. Nevertheless, UDP is not uncommon in networks. UDP relies on upper layers to manage, order, and guarantee data delivery if necessary. Most storage applications do not use UDP as their primary transport. Typically, VPN, real-time voice/video, DHCP, DNS, and RADIUS generate the most UDP.

UDP traffic that arrives at an IPEX gateway (`lan.dpx` ipif) is transported if the TCL allows it. However, UDP cannot be flow-controlled like TCP. Because UDP is not flow-controlled, if UDP traffic overruns the buffers, those datagrams are dropped.

UDP flows are optimized through IPEX batching and make use of compression. UDP flows are then transported across WO-TCP (WAN Optimized TCP) and assured of delivery to the remote SAN42B-R/IBM b-type Gen 6 Extension Blade. UDP delivery is not assured from the SAN42B-R/IBM b-type Gen 6 Extension Blade to the end-device because of UDP's connectionless nature.

3.4.9 Broadcast, Unknown, and Multicast traffic

Broadcast, Unknown, and Multicast (BUM) traffic is generated for all kinds of reasons by various end and network devices. This traffic is not TCP-based and has no single known destination. BUM traffic is not IPEX batched.

Every BUM datagram must be processed individually, so excessive amounts consume considerable processing resources, possibly resulting in performance loss. Do not send large quantities of ongoing BUM traffic across IPEX, such as multicast and broadcasts. BUM traffic is not directed through any compression engine, so BUM is not compressed.

It is typically not preferred to transmit BUM traffic across a WAN. However, it depends on the application. Connecting a SAN42B-R/IBM b-type Gen 6 Extension Blade to a data center LAN switch allows BUM traffic to appear at the IPEX gateway. Unless BUM traffic is an operational part of the storage application, it is not prudent to enable BUM traffic transport through an IPEX TCL.

By default BUM traffic is denied. In fact, all traffic is denied by default (“Deny All”). Purposeful TCL configuration or an “Allow All” configuration would be necessary before BUM traffic could traverse the WAN. However, “Allow All” is not recommended.

3.4.10 IPsec and IPEX

IPsec is a WAN side encryption technology applicable to the tunnel/trunk level, including all member circuits and applicable to both FCIP and IPEX traffic. If IPsec is enabled, all data is encrypted, including application TCP/UDP traffic and BUM traffic. IPsec is not supported on the LAN side.

3.4.11 IPEX LAN connectivity

Connecting end devices to IPEX is frequently done through a dedicated IP Storage Network (IPSN) or the existing data center LAN. This LAN must be Layer 2 from the end device to the IPEX gateway (`lan.dpx`), which is sometime referred to as a *broadcast domain*. In other words, the end-device Ethernet interface and the SAN42B-R/IBM b-type Gen 6 Extension Blade Ethernet interface are both in the same broadcast domain.

If there are multiple subnets within the same VLAN, multiple `lan.dpx` interfaces can be configured as IPEX gateways, one for each subnet. This architecture requires data flows destined for the remote data center to be sent directly to the IPEX gateway from the end-device. This decision can be done in various ways depending on the situation. These situations assume there is an intermediate LAN switch between the end-device and the SAN42B-R/IBM b-type Gen 6 Extension Blade.

If all end-device Ethernet interface communications are within the local subnet, or to the remote subnet, the end-device’s default-route can be the IPEX gateway IP address (`lan.dpx` ipif). For example, isolated subnets in both the local and remote data centers might be deployed specifically for IBM TS7700 grid clusters.

Any communication outside of a subnet first requires sending the data to a gateway to be forwarded. In this case, that is IPEX. Routing is not involved for communications within a subnet.

Communications in the same subnet are done directly with the other device and no gateway is involved. In this case, the normal data center default-route is not used because no traffic is destined for any location other than the local subnet or the remote subnet. In other words, no data is being sent outside of the local subnet at each data center, so there is no need for the data center gateway.

IPEX requires that the subnets be different in each data center.

In some cases, communications are *not* confined to isolated subnets, and the traditional default-route is required. In this case, a special IP route must be added to the end-device in addition to the default-route. The data center’s default-route is used for all local traffic outside of the end-device’s subnet.

The special IP route is used for traffic destined for the remote data center through IPEX. This traffic must be directed to the IPEX gateway (`lan.dpx` ipif). There is now more than one gateway on the same subnet. The intermediate LAN switch directs traffic to the correct gateway depending on the destination of the traffic.

The end-device first tries to match the traffic destination to the special route because it is more specific. If not, the default-route is used because it is more general.

It should be noted that deployment of IPEX can be done over existing extension tunnels. For example, you have array to array replication and currently extension is being used for FCIP. Adding IPEX does not require changing the existing tunnel/trunk, other than possibly enabling IPEX and adjusting ARL to accommodate the increased bandwidth.

In addition, the deployment of IPEX can be pre-setup and configured in its entirety except for the last step of adding the special IP route to the end-device or enabling PBR. Traffic is not diverted to IPEX until that last step is complete. Removing the special IP route or disabling PBR lets traffic resume its normal pathway providing easy and quick rollback and facilitating proper change control. See Figure 3-24.

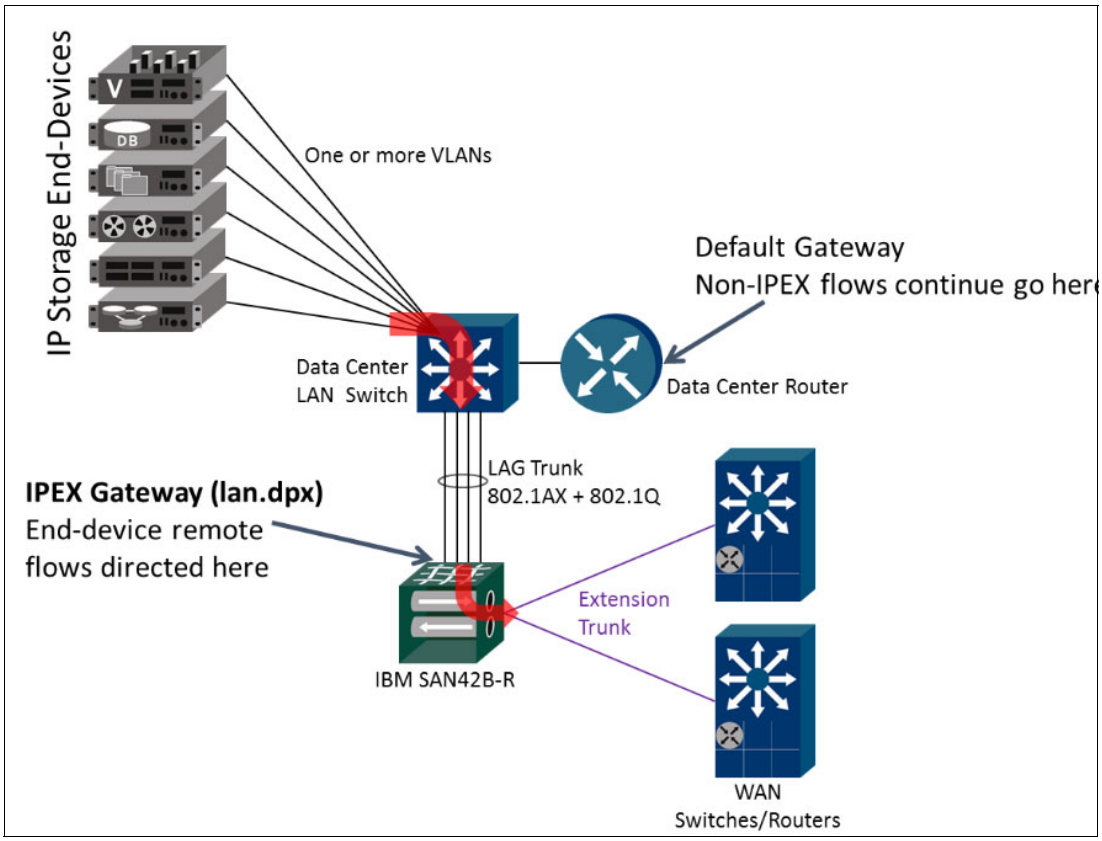


Figure 3-24 Traffic is not diverted to IPEX until the last step is complete

3.4.12 IPEX PBR connectivity

Policy Based Routing (PBR) is used to preempt normal router logic to direct certain predefined traffic to a certain predefined destination. PBR is supported on all major router brands. With IPEX, the idea is to not disrupt or reconfigure end-devices by having to add or change special routes. In addition, special routes might not be supported or practical on some end-devices.

Assuming the IP network operates at layer 3 between data centers, an end-device sends data to the remote data center in a normal manner. Inherent to that process, the data must pass through a local router on the way to the WAN, which is the gateway IP address of the end-device's already configured default-route. Likely, this is the router that should implement PBR for IPEX. However, technically PBR can be configured on any router along the IP storage pathway before entering the WAN. Figure 3-25 shows this configuration.

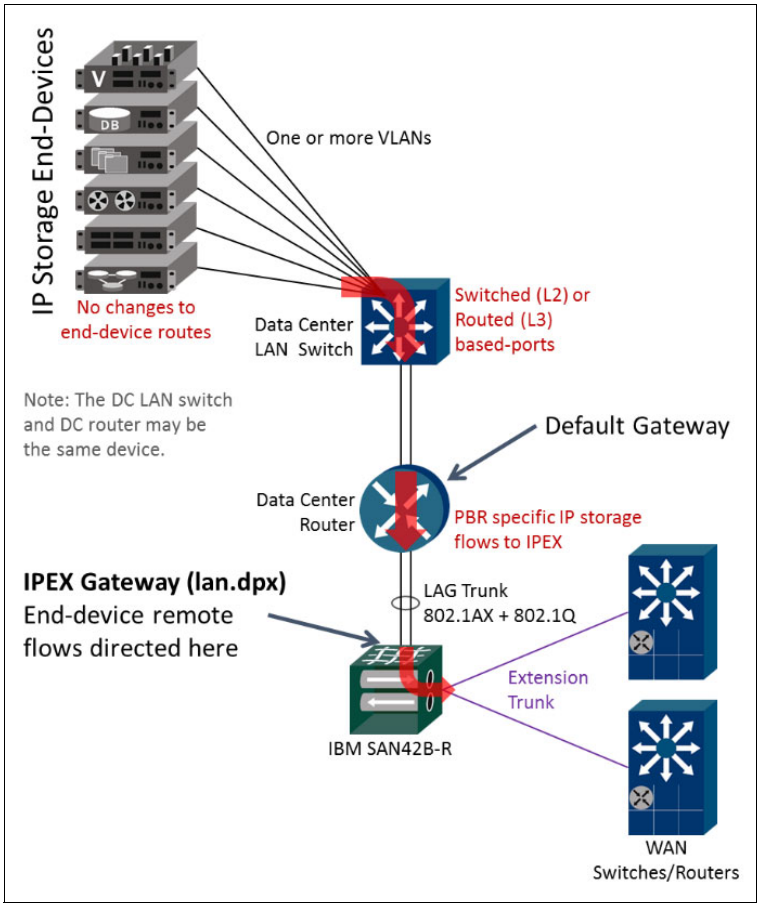


Figure 3-25 IPEX PBR connectivity

PBR would be configured to match all traffic from some set of specific local end-devices headed to some set of specific remote end-devices. PBR would divert traffic to the IPEX gateway. The IPEX gateway is an IP address configured on the SAN42B-R/IBM b-type Gen 6 Extension Blade `lan.dpx` ipif.

IPEX can be placed into-path or out-of-path simply by enabling or disabling PBR. When PBR is disabled, normal routing logic persists and all traffic takes its normal pathway over the WAN.

It is not necessary to have layer 2 connectivity from the end-device to the IPEX LAN Ethernet interfaces when using PBR. If your data center switch implementation is such that every port on the switch is a routed port, PBR is a good option.

Certain traffic flows can easily bypass IPEX using PBR. For example, TS7700 control TCP sessions can be ignored by PBR, which forces them to take their normal pathway through the IP network and WAN. Essentially nothing changes for the control sessions. By not going through IPEX, those control sessions do not consume limited TCP slots.

The TS7700 tape data TCP sessions would be specifically recognized by the PBR configuration and redirected to the IPEX lan.dpx gateway. PBR not only redirects traffic without the need for making alterations on the end-device, but also only redirects TCP flows actually benefiting from IPEX optimization.

3.4.13 IPEX LAN side architectures

There are important LAN side considerations that must be made when implementing IPEX. This section describes a few example architectures:

- IBM z System Global Mirror and TS7760/7700 Grid connectivity
- Cloud DR

IBM z System Global Mirror and TS7760/7700 Grid connectivity

Figure 3-26 shows the IPEX LAN side of a 3-site mainframe architecture. This architecture supports both FCIP for Global Mirror between DS8880 storage arrays and IPEX for TS7700 Grid between clusters. There are three sites:

- Primary production site
- Metro bunker site
- Remote DR site

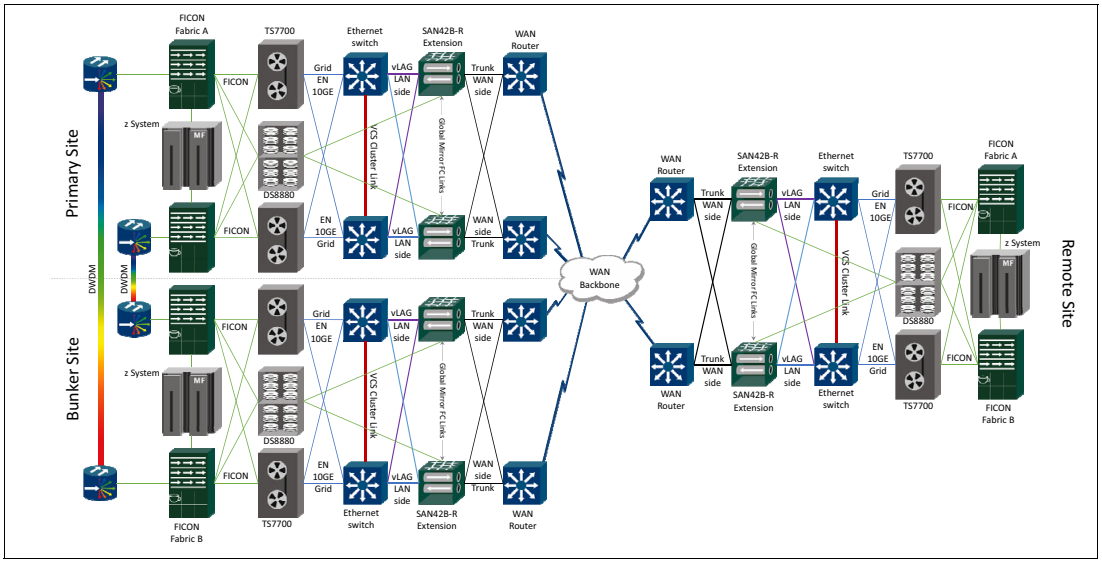


Figure 3-26 IBM z System Global Mirror and TS7700 Grid connectivity

A DWDM metro optical network connects the Primary Site and the Bunker Site. Metro Mirror synchronous replication between arrays takes this path due to the large bandwidth (10 Gbps or 100 Gbps) and low latency that metro DWDM offers. This is native FC replication, Metro Mirror uses FC and not FICON, even though the volumes are written with FICON. Native FC replication and DWDM are not within the scope of this document.

In this architecture, there are two FICON fabrics at each site: A and B. Many mainframe shops have decided to use more than two fabrics to maintain as much service as possible if a fabric goes offline. For example, if one of two fabrics fails, 50% of the bandwidth and connectivity is lost and the operational risk of no redundancy is created. If there were four fabrics, only 25% of the bandwidth and connectivity would be lost and no significant operational risk is created.

The number of FICON fabrics depends on the mission criticality of the environment and the potential risk to business revenue when a fabric goes offline.

The “A” and “B” FICON Directors connect to the DASD, VTL, and host. The FICON Directors do not connect to each other and maintain “air-gap” separation. Setup and configuration of FICON Directors is not within the scope of this document.

The Global Mirror FC connections are direct from the DS8880 replication ports to the FC ports on the SAN42B-R. The preferred practice is to not run Global Mirror connections through a production fabric. DS8880 Global Mirror connections are dedicated to only Global Mirror, and those array ports cannot be used for other traffic than Global Mirror. Therefore, there is no reason to connect through the production fabric unless the SAN42B-R does not have an adequate number of FC ports.

The SAN42B-R has 24 FC ports. Connecting directly to the SAN42B-R prevents any issues with slow-drain device or head of line blocking issues caused by differences in WAN speed, latency, and responsiveness of remote replication ports versus the production fabric’s speed, latency, and responsiveness of local ports.

IBM TS7700 Grid connections are made through the EN interfaces. Depending on the TS7700 model, the number and speed of these interfaces varies:

- 4x GE
- 2x 10GE
- 4x 10GE

EN ports connect to an intermediate data center LAN switch to enable communications with other EN ports within the data center. If there are no other EN ports in the data center or if EN communications are limited to only remote EN ports, connectivity can be made directly to the SAN42B-R or IBM b-type Gen 6 Extension Blade.

If all Grid traffic is contained within the local EN subnets or to the remote EN subnets, the TS7700 EN port default gateways would be the corresponding IPEX `lan.dpx` (ipif) IP address that is on the same subnet. Often, each EN port has its own IP subnet and probably its own VLAN. This means that in a cluster of multiple TS7700 units, if all the EN0 ports are on the same subnet in the same VLAN, those ports can communicate without the need for an intervening router. This is the nature of pure Ethernet (layer 2) connectivity.

In this example, the TS7700 VTLs at each site, EN0 ports can communicate with each other in their own VLAN, and EN1 ports can communicate with each other in their own VLAN.

The Ethernet switches provide Ethernet connectivity within the data center and vLAG (also known as vPC) support to the SAN42B-R. It is preferred that TS7700 Grid be implemented on a dedicated isolated IP Storage Network - Ethernet Fabric (IPSN), which enables the use of Virtual LAGs (vLAGs). Often Ethernet switch interfaces are capable of either GE or 10GE speeds, which accommodate different configurations of TS7700.

As shown in Figure 3-26 on page 89, there are two vLAGs (purple and blue). A typical configuration is to configure the EN0 ports to use the `lan.dp0` ipif behind the purple vLAG and configure the EN1 ports to use the `lan.dp0` ipif behind the blue vLAG.

Additionally, if the vLAGs are configured as 802.1Q Trunks, they can carry traffic from multiple VLANs. If the TS7700 environment is using multiple VLANs and each must reach its lan.dp0 interface, using vLAG trunks (802.1Q enabled) is preferred.

When sending traffic outside of the local subnet, the end-device is configured to send to a specific gateway IP address and that gateway must be on either DP0 or DP1. EN traffic is switched to the appropriate vLAG and upon reaching the SAN42B-R/IBM b-type Gen 6 Extension Blade, the traffic is directed to the appropriate DP.

Remotely destined traffic is sent by the TS7700 towards the IPEX gateway IP address through an intermediate Ethernet switch. The Ethernet switch sends it across the vLAG to the SAN42B-R LAN side, VLAN tagging the traffic if applicable. The traffic is directed to the corresponding `lan.dpx` (ipif) on a particular DP, which is based on both the IP address and if the traffic is VLAN tagged or not.

Upon reaching the DP, traffic is processed by the TCL and either placed into a specified tunnel (VE_Port) or dropped. Traffic passing the TCL is now managed by the WAN side for transport.

For information about WAN side settings and architecture, see 3.3, “The WAN side” on page 47.

Cloud DR

In the IPEX architecture that is described in this section, the user has a primary data center and a collocation site that is in or adjacent to (dark fiber) a cloud service provider.

Normally the propagation delay and quality of WAN connections make remote NAS replication impractical as a backup modality. The speeds at which NAS replication can perform locally in a data center versus across the WAN are quite different. NAS replication across a WAN tends to be considerably slower because it is hampered by latency and possibly congestion and packet loss. Using the IBM SAN42B-R or IBM b-type Gen 6 Extension Blade with WO-TCP, which includes Streams, Virtual Windows, and TCP Acceleration, results in a “local performance” experience over long distance.

NAS is used as backend storage for various servers in the primary data center. There is an isolated Brocade VDX 6740 backend IPSN Ethernet Fabric used for NAS connectivity between VMs, the NAS heads, and the IBM SAN42B-Rs. Other IP storage applications might use this fabric as well.

At the remote side is the same architecture of IPSN Ethernet Fabric and NAS heads. The remote VMs at the cloud provider only come online when needed. The remote VMs can access NAS backup data bringing the enterprise applications back online when needed.

IPEX provides the following functions to storage administrators:

- Increased NAS replication performance across distance
- Encryption of NAS data inflight
- Compression of NAS data Inflight
- High availability multipath connectivity using IBM Extension Trunking and ARL
- Operational excellence by using IBM Network Advisor, MAPS, Flow Vision and Extension dashboard
- Diagnostic and Troubleshooting tools with Wtool flow generator

Figure 3-27 shows the Cloud DR diagram.

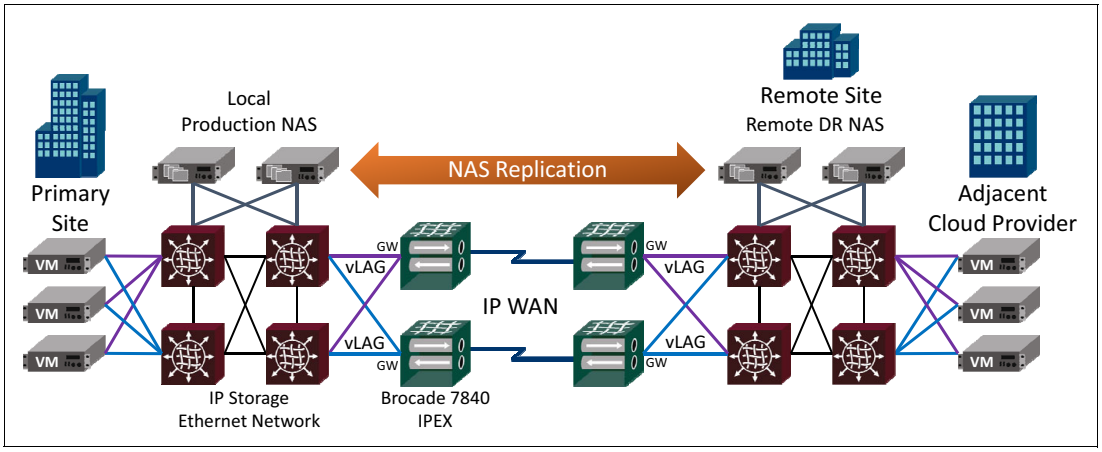


Figure 3-27 Cloud DR

4

FCIP replication

This chapter describes the implementation of an FCIP replication solution and discusses the following topics:

- Overview of all tasks
- Current lab configuration
- Prerequisites
- Creating IP interfaces with the command line
- Creating IP routes with the command line
- Validating connectivity with the command line
- Creating an IP tunnel with the command line
- Verifying the tunnel configuration with the command line
- Setting up storage replication in the lab
- Modifying or deleting an IP tunnel configuration with the command line
- Configuring FCIP with the IBM Network Advisor GUI

4.1 Overview of all tasks

The following steps are required to implement an Fibre Channel over IP (FCIP) solution:

1. Creating IP interfaces
2. Creating IP Routes
3. Validating connectivity
4. Creating IP tunnels
5. Creating additional circuits
6. Verifying tunnel configuration
7. Modifying or deleting FCIP tunnels

In addition, Extension Hot Code Load is implemented.

4.2 Current lab configuration

This section describes the configuration of our lab environment, which is used to demonstrate the FCIP feature for storage replication.

In the following scenario, we set up an inter-cluster Metro Mirror relationship between two IBM Storwize V7000 Storage Systems, which are *ITSO_V7000_Gen2_SiteA* at the primary site and *ITSO_V7000_Gen2_SiteB* at the secondary site. Each V7000 is connected to a SAN42B-R extension switch with FC-connection. The traffic between the two SAN42B-Rs is routed over Internet Protocol network, not FC. In our lab, the IP network is a direct connection between the LAN ports of the two SAN42B-R extension switches.

Figure 4-1 shows the configuration of our lab environment for FCIP.

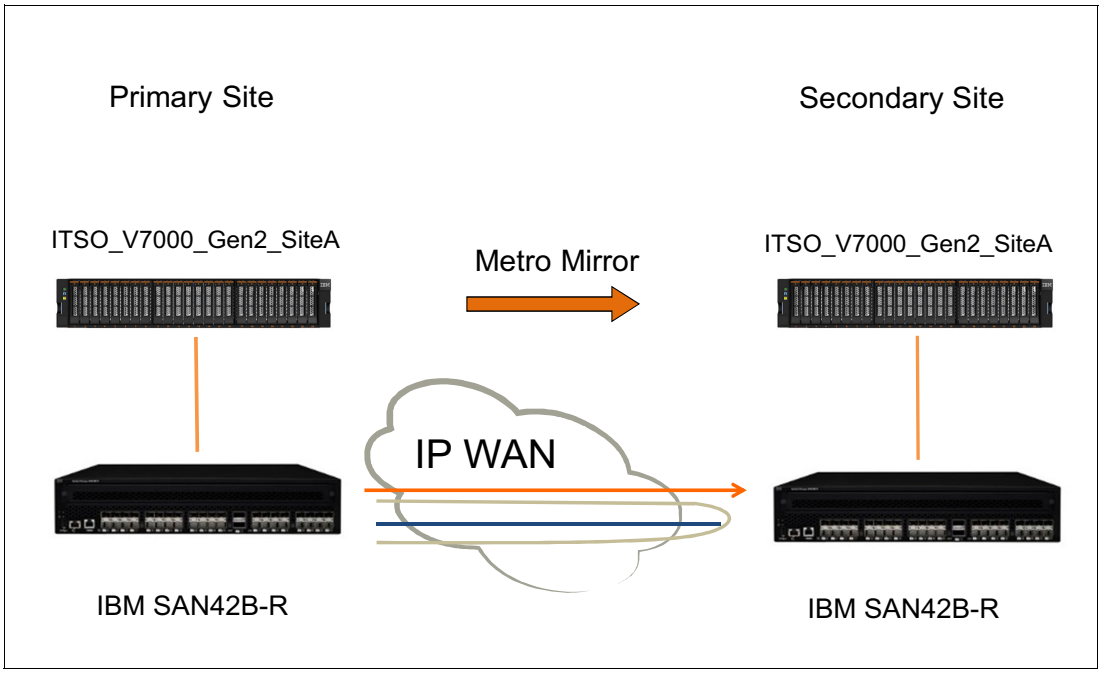


Figure 4-1 Lab setup: FCIP for V7000 replication

Table 4-1 and Table 4-2 on page 96 show the IP addresses that we use to configure IBM SAN42B-R extension switches in the following examples. The following IP addresses are included:

- MT: The main tunnel interface
- LBT: The local backup tunnel interface for Hot Code Load (HCL)
- RBT: The remote backup tunnel interface for HCL

Table 4-1 shows the IP address table for 40 GB ports.

Table 4-1 The IP address table of 40 Gb ports for FCIP replication

Site	IPIF Name	Function	IP address	VE port	Failure Group	Metric
ITSO_V7000_ Gen2_SiteA	ge0.dp0	MT	10.1.1.10	24	0	0
	ge0.dp0	MT	10.1.1.11	24	1	0
	ge1.dp0	MT	10.1.1.12	24	0	1
	ge1.dp0	MT	10.1.1.13	24	1	1
	ge0.dp1	LBT	10.1.1.14	-	-	-
	ge0.dp1	LBT	10.1.1.15	-	-	-
	ge1.dp1	LBT	10.1.1.16	-	-	-
	ge1.dp1	LBT	10.1.1.17	-	-	-
ITSO_V7000_ Gen2_SiteB	ge0.dp0	MT	10.1.1.30	24	0	0
	ge0.dp0	MT	10.1.1.31	24	1	0
	ge1.dp0	MT	10.1.1.32	24	0	1
	ge1.dp0	MT	10.1.1.33	24	1	1
	ge0.dp1	LBT	10.1.1.34	-	-	-
	ge0.dp1	LBT	10.1.1.35	-	-	-
	ge1.dp1	LBT	10.1.1.40	-	-	-
	ge1.dp1	LBT	10.1.1.41	-	-	-

Table 4-2 shows the IP address table for 10 GB ports.

Table 4-2 The P address table of 10 Gb ports for FCIP replication

Site	IPIF Name	Function	IP address	VE port	Failure Group	Metric
ITSO_V7000_Gen2_SiteA	ge2.dp1	MT	10.1.1.50	34	0	0
	ge3.dp1	MT	10.1.1.51	34	1	0
	ge6.dp1	MT	10.1.1.52	34	0	1
	ge7.dp1	MT	10.1.1.53	34	1	1
	ge2.dp0	LBT	10.1.1.54	-	-	-
	ge3.dp0	LBT	10.1.1.55	-	-	-
	ge6.dp0	LBT	10.1.1.56	-	-	-
	ge7.dp0	LBT	10.1.1.57	-	-	-
ITSO_V7000_Gen2_SiteB	ge2.dp1	MT	10.1.1.60	34	0	0
	ge3.dp1	MT	10.1.1.61	34	1	0
	ge6.dp1	MT	10.1.1.62	34	0	1
	ge7.dp1	MT	10.1.1.63	34	1	1
	ge2.dp0	LBT	10.1.1.64	-	-	-
	ge3.dp0	LBT	10.1.1.65	-	-	-
	ge6.dp0	LBT	10.1.1.66	-	-	-
	ge7.dp0	LBT	10.1.1.67	-	-	-

4.2.1 Lab configuration: IP tunnel

Figure 4-2 shows the IP tunnel configuration.

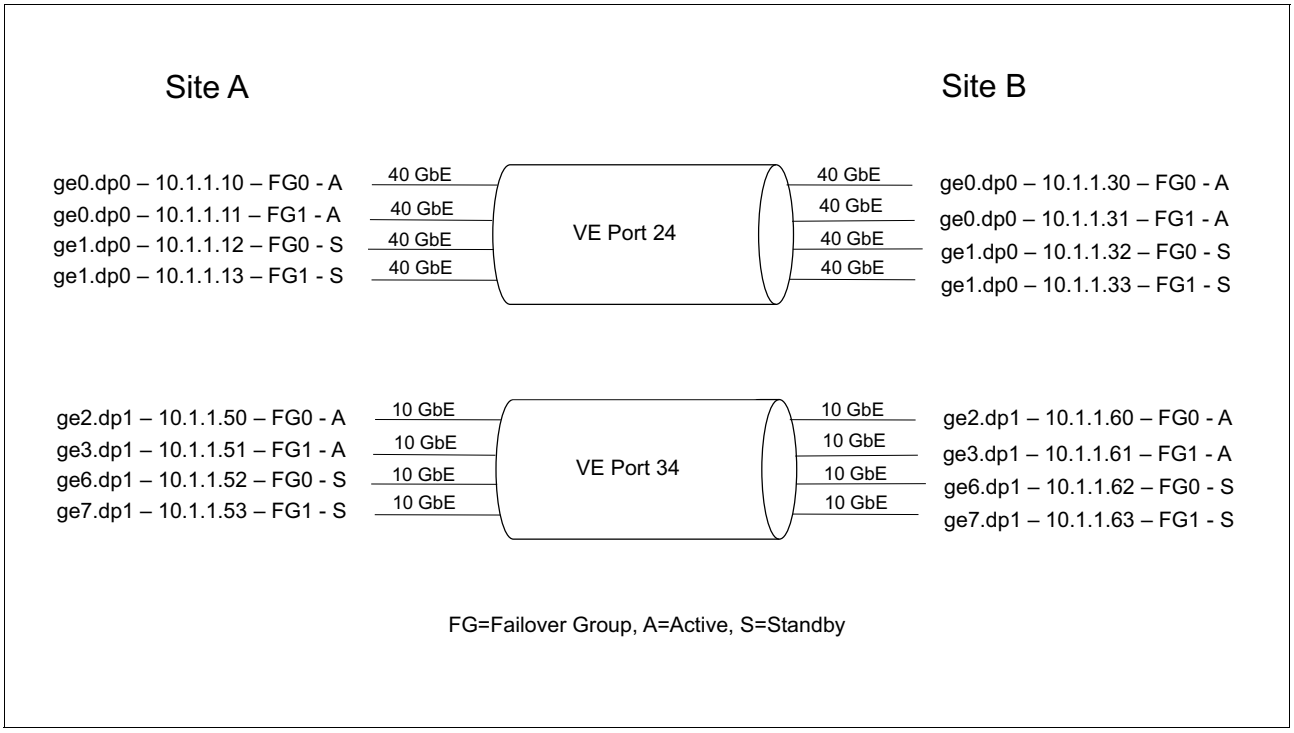


Figure 4-2 IP tunnel configuration

4.3 Prerequisites

The following tasks must be completed before you begin the installation:

1. Configuring the SAN
2. Verifying licenses
3. Verifying the IP WAN network configuration

4.3.1 Configuring the SAN

The configuration of the SAN42B-R product must follow the IBM Implementation guidelines described in *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331.

4.3.2 Verifying licenses

Example 4-1 shows the licenses installed on our lab switch.

Example 4-1 Example licenseshow output

```
ITSO_7840_SiteA:FID128:admin> licenseshow
S4TDrJLJPKHDmXXGtFESSZNrZMPACARXfPYQNWEAQaSA:
    Advanced Extension license
WaDrEXWRLKWrC4LPC4JTraKTESTYNFCRBJN9L:
    Extended Fabric license
```

```
    Trunking license
    Fabric Vision license
L77FYrAQQR9gLf9EXLTrRDrXWrJaF47GCtXfZCEAWXXB:
    Advanced FICON Acceleration license
XJgP3tGPrPNgCHS4CDYaDAXX3NXZSmJBXagJM9FACWrA:
    WAN Rate Upgrade 1 license
MgGGM9ZHTYXPaP7FZCQH4TgfF7CRKtHtSYFRfEEAB9XB:
    WAN Rate Upgrade 2 license
HPHRJ9NBgWGLJ7Xm9mYNWZaLgrrLP9MLB7GSD:
    FICON_CUP license
EQAGMXgg4QSFgK9tRJMJfZmME3aHf4aKBJtEL:
    Integrated Routing license
```

Note: The license for the Integrated routing, WAN Rate Upgrade 1 (10 Gigabit Ethernet), WAN Rate Upgrade 2 (40 Gigabit Ethernet), and the Advanced Ficon Acceleration are optional. For more information, see 2.4.13, “Licensing” on page 31.

4.3.3 Verifying the IP WAN network configuration

Routers and firewalls that are located between FCIP Routers must be configured to pass traffic through specific TCP ports that are used for extension and IPsec traffic. The following ports must be enabled:

- TCP Port 3225 and 3226
- If IPsec is used, UDP 1500 must be configured

4.4 Creating IP interfaces with the command line

In our example, the IP interfaces on both SAN42B-R products must be defined, as shown in Example 4-2 and Example 4-3 on page 99.

Example 4-2 shows the creation of IP interfaces with the default MTU size of 1500k and the netmask 255.255.255.0.

Example 4-2 Site A: Create IP Interface for the main tunnel

```
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp0 create 10.1.1.10/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp0 create 10.1.1.11/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp0 create 10.1.1.12/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp0 create 10.1.1.13/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.20/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.21/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.22/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.23/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

```
ITSO_7840_SiteA:FID128:admin> portshow ipif

 Port         IP Address                    / Pfx  MTU   VLAN  Flags
-------------------------------------------------------------------------------
 ge0.dp0      10.1.1.10                     / 24   1500  0     U R M
 ge0.dp0      10.1.1.11                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.22                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.23                     / 24   1500  0     U R M
 ge1.dp0      10.1.1.12                     / 24   1500  0     U R M
 ge1.dp0      10.1.1.13                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.20                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.21                     / 24   1500  0     U R M
-------------------------------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
       N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport

ITSO_7840_SiteA:FID128:admin>
```

Example 4-3 shows the command to create the IP interface for the main tunnel.

Example 4-3 Site B: Create IP Interface for the main tunnel

```
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge0.dp0 create 10.1.1.31/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp0 create 10.1.1.32/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp0 create 10.1.1.33/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.40/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.41/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.42/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.43/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portshow ipif

 Port         IP Address                    / Pfx  MTU   VLAN  Flags
-------------------------------------------------------------------------------
 ge0.dp0      10.1.1.30                     / 24   1500  0     U R M
 ge0.dp0      10.1.1.31                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.42                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.43                     / 24   1500  0     U R M
 ge1.dp0      10.1.1.32                     / 24   1500  0     U R M
 ge1.dp0      10.1.1.33                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.40                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.41                     / 24   1500  0     U R M
-------------------------------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
       N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport

ITSO_7840_SiteB:FID128:admin>
```

4.5 Creating IP routes with the command line

An IP route must be defined if the destination address is located in another subnet.

In our lab environment, it is not necessary to define IP routes because the interfaces are direct connected. To create an IP route, run the **portcfg iproute create** command.

Example 4-4 creates an IP route for the Network ID 10.1.142.0 and the gateway 10.1.42.1.

Example 4-4 Creating an IP route

```
portcfg iproute ge2.dp0 create 10.1.142.0/24 10.1.42.1
```

4.6 Validating connectivity with the command line

Before the circuits and tunnels can be configured, the connectivity between IP interfaces belonging to the same circuit must be verified, as shown in the Example 4-5.

Example 4-5 Site A: Verify IP connection

```
ITSO_7840_SiteA:FID128:admin> portcmd --ping ge0.dp0 -s 10.1.1.10 -d 10.1.1.30

PING 10.1.1.30 (10.1.1.10) with 64 bytes of data.
64 bytes from 10.1.1.30: icmp_seq=1 ttl=64 time=1 ms
64 bytes from 10.1.1.30: icmp_seq=2 ttl=64 time=1 ms
64 bytes from 10.1.1.30: icmp_seq=3 ttl=64 time=1 ms
64 bytes from 10.1.1.30: icmp_seq=4 ttl=64 time=1 ms

--- 10.1.1.30 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 708 ms
rtt min/avg/max = 1/1/1 ms

ITSO_7840_SiteA:FID128:admin> portcmd --ping ge0.dp0 -s 10.1.1.11 -d 10.1.1.31

PING 10.1.1.31 (10.1.1.11) with 64 bytes of data.
64 bytes from 10.1.1.31: icmp_seq=1 ttl=64 time=1 ms
64 bytes from 10.1.1.31: icmp_seq=2 ttl=64 time=1 ms
64 bytes from 10.1.1.31: icmp_seq=3 ttl=64 time=1 ms
64 bytes from 10.1.1.31: icmp_seq=4 ttl=64 time=1 ms

--- 10.1.1.31 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 729 ms
rtt min/avg/max = 1/1/1 ms

ITSO_7840_SiteA:FID128:admin> portcmd --ping ge0.dp0 -s 10.1.1.12 -d 10.1.1.32

PING 10.1.1.32 (10.1.1.12) with 64 bytes of data.

Object does not exist
IP Address 10.1.1.12 is configured on a different port

ITSO_7840_SiteA:FID128:admin> portcmd --ping ge1.dp0 -s 10.1.1.12 -d 10.1.1.32

PING 10.1.1.32 (10.1.1.12) with 64 bytes of data.
```

```
64 bytes from 10.1.1.32: icmp_seq=1 ttl=64 time=1 ms
64 bytes from 10.1.1.32: icmp_seq=2 ttl=64 time=1 ms
64 bytes from 10.1.1.32: icmp_seq=3 ttl=64 time=1 ms
64 bytes from 10.1.1.32: icmp_seq=4 ttl=64 time=1 ms

--- 10.1.1.32 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 736 ms
rtt min/avg/max = 1/1/1 ms

ITSO_7840_SiteA:FID128:admin> portcmd --ping ge1.dp0 -s 10.1.1.13 -d 10.1.1.33

PING 10.1.1.33 (10.1.1.13) with 64 bytes of data.
64 bytes from 10.1.1.33: icmp_seq=1 ttl=64 time=1 ms
64 bytes from 10.1.1.33: icmp_seq=2 ttl=64 time=1 ms
64 bytes from 10.1.1.33: icmp_seq=3 ttl=64 time=1 ms
64 bytes from 10.1.1.33: icmp_seq=4 ttl=64 time=1 ms

--- 10.1.1.33 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 754 ms
rtt min/avg/max = 1/1/1 ms

ITSO_7840_SiteA:FID128:admin>
```

4.7 Creating an IP tunnel with the command line

In our example, we are going to create one 40 Gigabit Ethernet tunnel with four circuits.

For tunnel and circuit requirements, see the multigigabit circuits and tunnel requirements in the *Brocade Fabric OS Extension Configuration Guide, 8.0.1*:

http://www.brocade.com/content/html/en/configuration-guide/fos-801-extension/GUID-4FBA51F0-F39A-478F-9CC7-8F8A84015F3C.html

4.7.1 Creating an IPsec policy

Before you create an IP tunnel, you must define an IPsec policy as shown in Example 4-6 and Example 4-7.

Example 4-6 Site A: Create an IPsec policy named myPolicy1

```
ITSO_7840_SiteA:FID128:admin> portcfg ipsec-policy myPolicy1 create  -k
"DbnQh4wQlI8b7QtZRgIt"
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-7 shows how to create a policy for site B.

Example 4-7 Site B: Create an IPsec policy named myPolicy1

```
ITSO_7840_SiteB:FID128:admin> portcfg ipsec-policy myPolicy1 create  -k
"DbnQh4wQlI8b7QtZRgIt"
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

4.7.2 Creating a 40 Gigabit Ethernet IP tunnel

This section describes the implementation of one 40 Gigabit Ethernet tunnel with four circuits. Creating an additional circuit is described in 4.7.4, “Creating an additional circuit” on page 104.

Example 4-8 and Example 4-9 show how to create an FCIP tunnel on VE Port 24 with a minimum and maximum committed rate of 1000000 kbps, with compression setting fast-deflate and IPsec enabled. The failover group and metric are 0.

Example 4-8 Site A: Create FCIP tunnel 1 on VE Port 24

```
ITSO_7840_SiteA:FID128:admin> portcfg fciptunnel 24 create --local-ip 10.1.1.10
--remote-ip 10.1.1.30 -b 10000000 -B 10000000 -k 1000 -i myPolicy1 -c fast-deflate
-g 0 -x 0

!!!! WARNING !!!!
Fast deflate compression cannot be applied as IP-compression. Will be taken as no
compression

Continue with operation (Y,y,N,n): [ n] y
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-9 shows an example for Site B.

Example 4-9 Site B: Create FCIP tunnel 1 on VE Port 24

```
ITSO_7840_SiteB:FID128:admin> portcfg fciptunnel 24 create --local-ip 10.1.1.30
--remote-ip 10.1.1.10 -b 10000000 -B 10000000 -k 1000 -i myPolicy1 -c fast-deflate
-g 0 -x 0

!!!! WARNING !!!!
Fast deflate compression cannot be applied as IP-compression. Will be taken as no
compression

Continue with operation (Y,y,N,n): [ n] y
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

Example 4-10 shows the configuration settings for the IP tunnel that is defined on VE Port 24 and on the circuit with index 0, which is configured as active link within failover group 0.

Example 4-10 Verify FCIP tunnel status

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel 24 --circuit

 Tunnel: VE-Port:24 (idx:0, DP0)
 ================================================
  Oper State           : Disabled
  TID                  : 24
  Flags                : 0x00000000
  IP-Extension         : Disabled
  Compression          : Fast Deflate
  QoS BW Ratio         : 50% / 30% / 20%
  Fastwrite            : Disabled
  Tape Pipelining      : Disabled
```

```
 IPSec                : Enabled
 IPSec-Policy         : myPolicy1
 Load-Level (Cfg/Peer): Failover (Failover / Failover)
 Local WWN            : 10:00:50:eb:1a:d7:83:80
 Peer WWN             : 00:00:00:00:00:00:00:00
 RemWWN (config)      : 00:00:00:00:00:00:00:00
 cfgmask              : 0x0000001f 0x4000020c
 Uncomp/Comp Bytes    : 0 / 0 / 1.00 : 1
 Uncomp/Comp Byte(30s): 0 / 0 / 1.00 : 1
 Flow Status          : 0
 ConCount/Duration    : 0 / 9m14s
 Uptime               : 0s
 Stats Duration       : 0s
 Receiver Stats       : 0 bytes / 0 pkts /    0.00 Bps Avg
 Sender Stats         : 0 bytes / 0 pkts /    0.00 Bps Avg
 TCP Bytes In/Out     : 0 / 0
 ReTx/OOO/SloSt/DupAck: 0 / 0 / 0 / 0
 RTT (min/avg/max)    : 0 / 0 / 0 ms
 Wan Util             : 0.0%
 TxQ Util             : 0.0%

Circuit 24.0 (DP0)
==================================================
 Admin/Oper State     : Enabled / Disabled
 Flags                : 0x00000000
 IP Addr (L/R)        : 10.1.1.10 ge0 <-> 10.1.1.30
 HA IP Addr (L/R)     : 0.0.0.0 ge0 <-> 0.0.0.0
 Configured Comm Rates: 10000000 / 10000000 kbps
 Peer Comm Rates      : 0 / 0 kbps
 Actual Comm Rates    : 3000000 / 1000000 kbps
 Keepalive (Cfg/Peer) : 1000 (1000 / 0) ms
 Metric               : 0
 Connection Type      : Default
 ARL-Type             : Auto
 PMTU                 : Disabled
 SLA                  : (none)
 Failover Group       : 0
 VLAN-ID              : NONE
 L2Cos (FC:h/m/l)     : 0 / 0 / 0 (Ctrl:0)
 L2Cos (IP:h/m/l)     : 0 / 0 / 0
 DSCP (FC:h/m/l)      : 0 / 0 / 0 (Ctrl:0)
 DSCP (IP:h/m/l)      : 0 / 0 / 0
 cfgmask              : 0x40000000 0x00000caf
 Flow Status          : 1
 ConCount/Duration    : 0 / 9m14s
 Uptime               : 0s
 Stats Duration       : 0s
 Receiver Stats       : 0 bytes / 0 pkts /    0.00 Bps Avg
 Sender Stats         : 0 bytes / 0 pkts /    0.00 Bps Avg
 TCP Bytes In/Out     : 0 / 0
 ReTx/OOO/SloSt/DupAck: 0 / 0 / 0 / 0
 RTT (min/avg/max)    : 0 / 0 / 0 ms
 Wan Util             : 0.0%
```

Note: The Hot Code Load feature requires the implementation of local backup tunnel IP interfaces, which is described in 4.10.1, “Configuring the hot code load extension feature” on page 107.

4.7.3 Creating a 10 Gigabit Ethernet IP tunnel

See 4.7.2, “Creating a 40 Gigabit Ethernet IP tunnel” on page 102 for all required steps to create a tunnel.

4.7.4 Creating an additional circuit

In our example, three additional 40 Gigabit Ethernet circuits must be added to the tunnel that is defined on VE Port 24.

Example 4-11 shows the creation of an additional circuit with index 1 to the existing tunnel defined on VE Port 24 on ITSO_7840_SiteA switch. The circuit is assigned to failover group 1 with metric 0.

Example 4-11 Site A: Create circuit 1 on VE Port 24

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 create 1 --remote-ip
10.1.1.31 --local-ip  10.1.1.11 -b 10000000 -B 10000000 -g 1 -x 0
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-12 shows the creation of an additional circuit with index 1 to the existing tunnel that is defined on VE Port 24 on the ITSO_7840_SiteB switch. The circuit is assigned to failover group 1 with metric 0.

Example 4-12 Site B: create circuit 1 on VE Port 24

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 create 1 --remote-ip
10.1.1.11 --local-ip  10.1.1.31 -b 10000000 -B 10000000 -g 1 -x 0
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

Example 4-13 shows the creation of an additional circuit with index 2 to the existing tunnel that is defined on VE Port 24 on ITSO_7840_SiteA switch. The circuit is assigned to failover group 0 with metric 1.

Example 4-13 Site A: Add circuit 2 on VE Port 24

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 create 2 --remote-ip
10.1.1.32 --local-ip  10.1.1.12 -b 10000000 -B 10000000 -g 0 -x 1
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-14 shows the creation of an additional circuit with index 2 to the existing tunnel that is defined on VE Port 24 on ITSO_7840_SiteB switch. The circuit is assigned to failover group 0 with metric 1.

Example 4-14 Site B: Add circuit 2 on VE Port 24

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 create 2 --remote-ip
10.1.1.12 --local-ip  10.1.1.32 -b 10000000 -B 10000000 -g 0 -x 1
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

Example 4-15 shows the creation of an additional circuit with index 3 to the existing tunnel that is defined on VE Port 24 on the ITSO_7840_SiteA switch. The circuit is assigned to failover group 1 with metric 1.

Example 4-15 Site A: Add circuit 3 on VE Port 24

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 create 3 --remote-ip
10.1.1.33 --local-ip  10.1.1.13 -b 10000000 -B 10000000 -g 1 -x 1
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-16 shows the creation of an additional circuit with index 3 to the existing tunnel that is defined on VE Port 24 on the ITSO_7840_SiteB switch. The circuit is assigned to failover group 1 with metric 1.

Example 4-16 Site B: Add circuit 3 on VE Port 24

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 create 3 --remote-ip
10.1.1.13 --local-ip  10.1.1.33 -b 10000000 -B 10000000 -g 1 -x 1
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

4.8 Verifying the tunnel configuration with the command line

Example 4-17 shows that the active circuits defined ge0, index 0 and ge0, and index 1 are used while the circuits defined on ge1, index 2 and ge1, and index 3 do not show any traffic and act as a standby.

Example 4-17 Tunnel verification with the fciptunnel command

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel --circuit

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 24    -         Up      --i----a-     1m5s    0.67    0.67   1      -       -
 24    0 ge0     Up      ----a---4     1m6s    0.34    0.34   1 10000/10000 0/-
 24    1 ge0     Up      ----a---4     1m4s    0.33    0.33   1 10000/10000 0/1
 24    2 ge1     Up      ----a---4     1m4s    0.00    0.00   1 10000/10000 1/-
 24    3 ge1     Up      ----a---4     1m4s    0.00    0.00   1 10000/10000 1/1
--------------------------------------------------------------------------------
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
```

```
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
ITSO_7840_SiteA:FID128:admin>
```

4.9 Setting up storage replication in the lab

After the tunnel is successfully defined, the configuration for the storage replication must be done. For the configuration of partnership of two storage systems over Fibre Channel, such as IBM Storwize V7000 storage, see IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/en/ST3FR7_7.6.0/com.ibm.storwize.v7000.760.doc/svc_remotecopypartovr_21iakm.html

In our lab configuration example, we defined a partnership between two V7000 Gen 2 storage systems over Fibre Channel.

4.10 Modifying or deleting an IP tunnel configuration with the command line

This section shows the modification and deletion of an IP tunnel.

Example 4-18 shows the modification of the circuit with index 0 that is defined on VE Port 24.

Example 4-18 Modify the circuit with index 0 on VE-Port 24

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 modify 0 -b 5000000
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 modify 0 -b 5000000
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

Example 4-19 shows the new value set for the **-b** parameter.

Example 4-19 Site A: Show IP tunnel

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel --circuit

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 24    -         Up      --i----a-   37m17s    0.66    0.66   1      -       -
 24    0 ge0     Up      ----a---4   37m18s    0.33    0.33   1  5000/10000 0/-
 24    1 ge0     Up      ----a---4   37m16s    0.33    0.33   1 10000/10000 0/1
 24    2 ge1     Up      ----a---4   37m16s    0.00    0.00   1 10000/10000 1/-
 24    3 ge1     Up      ----a---4   37m16s    0.00    0.00   1 10000/10000 1/1
--------------------------------------------------------------------------------
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

Example 4-20 shows the deletion of the IP tunnel that is defined on VE Port 24.

Example 4-20 Site A: Delete IP tunnel

```
ITSO_7840_SiteA:FID128:admin> portcfg fciptunnel 24 delete
Operation Succeeded
```

4.10.1 Configuring the hot code load extension feature

The hot code load extension feature requires additional IP interfaces for the configuration of the local and remote backup tunnel. These additional IP interfaces are created in the following examples.

The hot code load feature is described in 2.4.12, “Extension Hot Code Load” on page 30.

Adding IP interfaces for the local and remote backup tunnels

Example 4-21 shows how to configure the IP interfaces for Site A for the local backup tunnel.

Example 4-21 Site A: Configure IP interfaces for Local backup tunnel

```
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.14/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.15/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.16/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.17/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteA:FID128:admin> portsow ipif
ITSO_7840_SiteA:FID128:admin> portshow ipif

 Port         IP Address                     / Pfx  MTU   VLAN  Flags
-------------------------------------------------------------------------------
 ge0.dp0      10.1.1.10                      / 24   1500  0     U R M I
 ge0.dp0      10.1.1.11                      / 24   1500  0     U R M I
 ge0.dp1      10.1.1.22                      / 24   1500  0     U R M
 ge0.dp1      10.1.1.23                      / 24   1500  0     U R M
 ge0.dp1      10.1.1.14                      / 24   1500  0     U R M
 ge0.dp1      10.1.1.15                      / 24   1500  0     U R M
 ge1.dp0      10.1.1.12                      / 24   1500  0     U R M I
 ge1.dp0      10.1.1.13                      / 24   1500  0     U R M I
 ge1.dp1      10.1.1.20                      / 24   1500  0     U R M
 ge1.dp1      10.1.1.21                      / 24   1500  0     U R M
 ge1.dp1      10.1.1.16                      / 24   1500  0     U R M
 ge1.dp1      10.1.1.17                      / 24   1500  0     U R M
 ge2.dp0      10.1.1.50                      / 24   1500  0     U R M
 ge2.dp1      10.1.1.54                      / 24   1500  0     U R M
 ge3.dp0      10.1.1.51                      / 24   1500  0     U R M
 ge3.dp1      10.1.1.55                      / 24   1500  0     U R M
 ge5.dp0      10.1.1.52                      / 24   1500  0     U R M
 ge5.dp1      10.1.1.56                      / 24   1500  0     U R M
 ge6.dp0      10.1.1.53                      / 24   1500  0     U R M
```

```
 ge6.dp1      10.1.1.57                     / 24   1500  0     U R M
-----------------------------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
       N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport
```

Example 4-22 shows how to configure the IP interfaces for Site A for the local backup tunnel.

Example 4-22 Site B: Configure IP Interface for local backup tunnel

```
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.34/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge0.dp1 create 10.1.1.35/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.36/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcfg ipif  ge1.dp1 create 10.1.1.37/24 mtu 1500
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portshow ipif

 Port         IP Address                    / Pfx  MTU   VLAN  Flags
-----------------------------------------------------------------------------
 ge0.dp0      10.1.1.30                     / 24   1500  0     U R M I
 ge0.dp0      10.1.1.31                     / 24   1500  0     U R M I
 ge0.dp1      10.1.1.42                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.43                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.34                     / 24   1500  0     U R M
 ge0.dp1      10.1.1.35                     / 24   1500  0     U R M
 ge1.dp0      10.1.1.32                     / 24   1500  0     U R M I
 ge1.dp0      10.1.1.33                     / 24   1500  0     U R M I
 ge1.dp1      10.1.1.40                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.41                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.36                     / 24   1500  0     U R M
 ge1.dp1      10.1.1.37                     / 24   1500  0     U R M
 ge2.dp0      10.1.1.60                     / 24   1500  0     U R M
 ge2.dp1      10.1.1.64                     / 24   1500  0     U R M
 ge3.dp0      10.1.1.61                     / 24   1500  0     U R M
 ge3.dp1      10.1.1.65                     / 24   1500  0     U R M
 ge5.dp0      10.1.1.62                     / 24   1500  0     U R M
 ge5.dp1      10.1.1.66                     / 24   1500  0     U R M
 ge6.dp0      10.1.1.63                     / 24   1500  0     U R M
 ge6.dp1      10.1.1.67                     / 24   1500  0     U R M
-----------------------------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
       N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport

ITSO_7840_SiteB:FID128:admin>
```

Adding local and remote IP addresses to the current tunnel configuration

The `local-ha-ip` address and the `remote-ha-ip` address must be added to the current tunnel configuration, as shown in the following examples.

Example 4-23 shows how to add the `local-ha-ip` and `remote-ha-ip` addresses to the current tunnel configuration on VE Port 24, circuit 0 at Site A.

Example 4-23 Site A: Add the local-ha-ip and remote-ha-ip address to current tunnel configuration on VE Port 24, circuit 0

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 modify 0 --local-ha-ip
10.1.1.14  --remote-ha-ip 10.1.1.34

!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.

Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-24 shows how to add the `local-ha-ip` and `remote-ha-ip` addresses to the current tunnel configuration on VE Port 24, circuit 0 at site B.

Example 4-24 Site B: Add the local-ha-ip and remote-ha-ip address to current tunnel configuration on VE Port 24, circuit 0

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 modify 0 --local-ha-ip
10.1.1.34  --remote-ha-ip 10.1.1.14

!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.

Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteB:FID128:admin
```

Example 4-25 shows how to add the `local-ha-ip` and `remote-ha-ip` addresses to the current tunnel configuration on VE Port 24, circuit 1 at Site A.

Example 4-25 Site A: Add the local-ha-ip and remote-ha-ip address to current tunnel configuration on VE Port 24, circuit 1

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 modify 1 --local-ha-ip
10.1.1.15  --remote-ha-ip 10.1.1.35

!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.

Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 4-26 shows how to add the `local-ha-ip` and `remote-ha-ip` addresses to the current tunnel configuration on VE Port 24, circuit 1 at Site B.

Example 4-26 Site B: Add the local-ha-ip and remote-ha-ip address to current tunnel configuration on VE Port 24, circuit 1

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 modify 1 --local-ha-ip
10.1.1.35  --remote-ha-ip 10.1.1.15

!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.

Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded
```

Verifying the hot code load feature configuration

This section shows that the hot code load feature configuration is completed and that the firmware update scan can be performed without affecting the traffic.

Example 4-27 shows how to verify the configuration for Site A.

Example 4-27 Site A: Verfiy configuration

```
ITSO_7840_SiteA:FID128:admin>  portshow fciptunnel --ha --circuit

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 24    -         Up      -Mi----a- 1h16m36s    0.67    0.67   1      -       -
 24    0 ge0     Up      ----ah--4 1h16m36s    0.33    0.33   1 10000/10000 0/-
 24    1 ge0     Up      ----ah--4   47m21s    0.33    0.33   1 10000/10000 0/1
 24    2 ge1     Up      ----ah--4 1h12m31s    0.00    0.00   1 10000/10000 1/-
 24    3 ge1     Up      ----ah--4   50m40s    0.00    0.00   1 10000/10000 1/1
 24    -         Up      -Ri----a-    7m34s    0.00    0.00   1      -       -
 24    0 ge0     Up      ----ah--4    7m35s    0.00    0.00   1 10000/10000 0/-
 24    1 ge0     Up      ----ah--4    6m25s    0.00    0.00   1 10000/10000 0/1
 24    2 ge1     Up      ----ah--4    2m12s    0.00    0.00   1 10000/10000 1/-
 24    3 ge1     Up      ----ah--4    1m31s    0.00    0.00   1 10000/10000 1/1
 24    -         Up      -Li----a-    7m34s    0.00    0.00   1      -       -
 24    0 ge0     Up      ----ah--4    7m34s    0.00    0.00   1 10000/10000 0/-
 24    1 ge0     Up      ----ah--4    6m24s    0.00    0.00   1 10000/10000 0/1
 24    2 ge1     Up      ----ah--4    2m11s    0.00    0.00   1 10000/10000 1/-
 24    3 ge1     Up      ----ah--4    1m30s    0.00    0.00   1 10000/10000 1/1
--------------------------------------------------------------------------------
 Flags (tunnel): M=MainTunnel  L=LocalBackup  R=RemoteBackup
                 i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA

ITSO_7840_SiteA:FID128:admin>
```

Example 4-28 shows how to check the Site A FCIP tunnel status.

Example 4-28 Site A: Checking FCIP tunnel status

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel --hcl-status

Checking FCIP Tunnel HA Status.

Current Status       : Ready
CP Version           : v8.0.1a
DP0 Status:
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE
DP1 Status:
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE

 Tunnel 24 (FID:128) HA configured and HA Online. Traffic will not be disrupted.

ITSO_7840_SiteA:FID128:admin>
```

Example 4-29 shows the Site A hot code load configuration.

Example 4-29 Site A: Show hot code load configuration

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel --hcl-status

Checking FCIP Tunnel HA Status.

Current Status       : Ready
CP Version           : v8.0.1a
DP0 Status:
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE
DP1 Status:
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE

 Tunnel 24 (FID:128) HA configured and HA Online. Traffic will not be disrupted.

ITSO_7840_SiteA:FID128:admin>
```

Example 4-30 shows the Site B hot code load configuration.

Example 4-30 Site B: Show hot code load configuration

```
ITSO_7840_SiteB:FID128:admin> portshow fciptunnel --hcl-status

Checking FCIP Tunnel HA Status.

Current Status       : Ready
CP Version           : v8.0.1a
DP0 Status:
```

```
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE
DP1 Status:
   State             : Online - Inactive
   Version           : v8.0.1a
   Current HA Stage  : IDLE

 Tunnel 24 (FID:128) HA configured and HA Online. Traffic will not be disrupted.
```

Example 4-31 shows how to verify the Site B configuration.

Example 4-31 Site B: Verify configuration

```
ITSO_7840_SiteB:FID128:admin> portshow fciptunnel --ha --circuit

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 24    -         Up      -M------- 4h42m16s    0.34    0.34   2      -       -
 24    0 ge0     Up      ----ah--4 4h42m17s    0.17    0.17   2  5000/10000 0/-
 24    1 ge1     Up      ----ah--4 4h37m17s    0.17    0.17   1  5000/10000 0/-
 24    -         Up      -R------- 4h16m19s    0.00    0.00   1      -       -
 24    0 ge0     Up      ----ah--4 4h16m19s    0.00    0.00   1  5000/10000 0/-
 24    1 ge1     Up      ----ah--4 4h10m30s    0.00    0.00   1  5000/10000 0/-
 24    -         Up      -L------- 4h16m20s    0.00    0.00   1      -       -
 24    0 ge0     Up      ----ah--4 4h16m20s    0.00    0.00   1  5000/10000 0/-
 24    1 ge1     Up      ----ah--4 4h10m29s    0.00    0.00   1  5000/10000 0/-
 34    -         Up      -M------- 4h39m47s    0.33    0.33   1      -       -
 34    0 ge2     Up      ----ah--4 4h39m47s    0.16    0.16   1  5000/10000 0/-
 34    1 ge3     Up      ----ah--4 4h35m32s    0.00    0.00   1  5000/10000 1/-
 34    2 ge6     Up      ----ah--4 4h32m45s    0.16    0.16   1  5000/10000 0/-
 34    3 ge7     Up      ----ah--4 4h32m14s    0.00    0.00   1  5000/10000 1/-
 34    -         Up      -R-------  4h9m21s    0.00    0.00   1      -       -
 34    0 ge2     Up      ----ah--4  4h9m21s    0.00    0.00   1  5000/10000 0/-
 34    1 ge3     Up      ----ah--4  4h7m36s    0.00    0.00   1  5000/10000 1/-
 34    2 ge6     Up      ----ah--4  4h6m21s    0.00    0.00   1  5000/10000 0/-
 34    3 ge7     Up      ----ah--4  4h5m16s    0.00    0.00   1  5000/10000 1/-
 34    -         Up      -L-------  4h9m21s    0.00    0.00   1      -       -
 34    0 ge2     Up      ----ah--4  4h9m21s    0.00    0.00   1  5000/10000 0/-
 34    1 ge3     Up      ----ah--4  4h7m37s    0.00    0.00   1  5000/10000 1/-
 34    2 ge6     Up      ----ah--4  4h6m20s    0.00    0.00   1  5000/10000 0/-
 34    3 ge7     Up      ----ah--4  4h5m16s    0.00    0.00   1  5000/10000 1/-
--------------------------------------------------------------------------------
 Flags (tunnel): M=MainTunnel  L=LocalBackup  R=RemoteBackup
                 i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

4.11 Configuring FCIP with the IBM Network Advisor GUI

Here we use IBM Network Advisor that is installed on a Microsoft Windows server for our SAN42B-R Extension Switch configuration.

4.11.1 Creating a FCIP Tunnel

To create a FCIP Tunnel, complete the following steps:

1. To log in to the GUI, double-click the desktop icon or open the application from the windows **Start** menu. The Log in window is displayed (Figure 4-3).

Figure 4-3 IBM Network Advisor Login

After log in, the main panel is displayed (Figure 4-4).

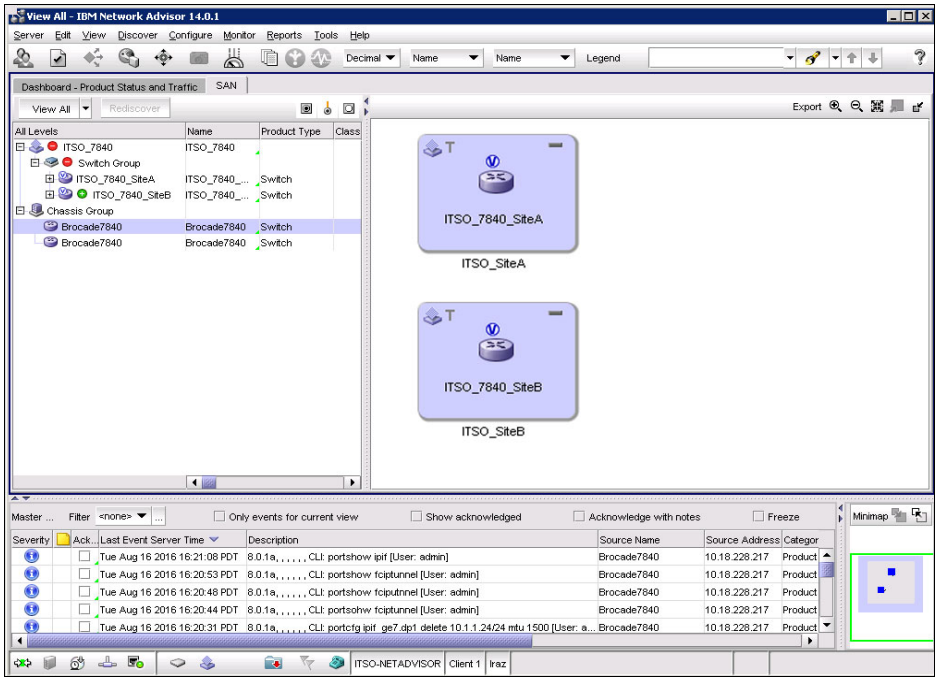


Figure 4-4 Overview of IBM Network Advisor GUI

2. From the Overview window, click **Configure** → **FCIP Tunnels** as shown in Figure 4-5**.**

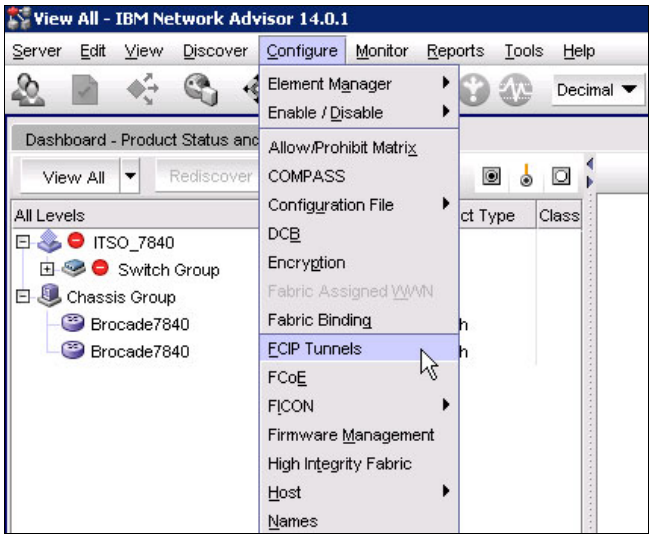


Figure 4-5 FCIP Tunnels menu

3. Select the extension switch for which you want to perform the action. In this case, we select the first switch, **ITSO_7840_SiteA**, and then click **Add** to add an FCIP tunnel (Figure 4-6).

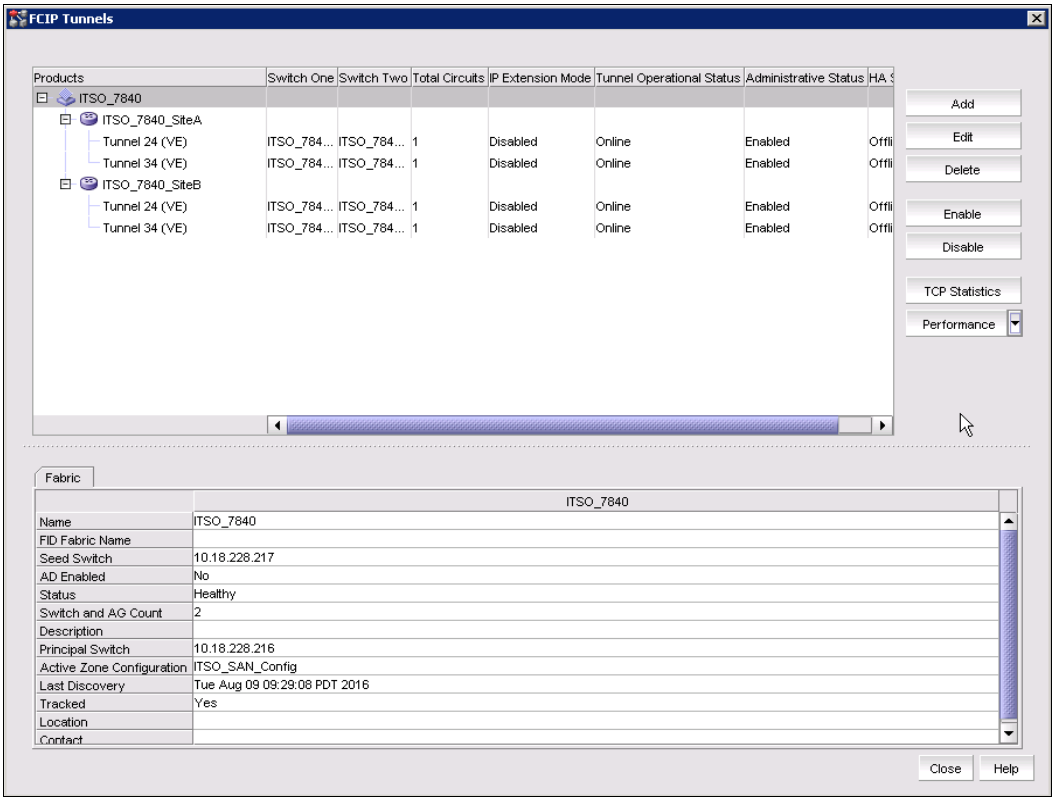


Figure 4-6 FCIP tunnel creation window

4. Click **Select Switch Two** to select the extension switch that is at site B (Figure 4-7).

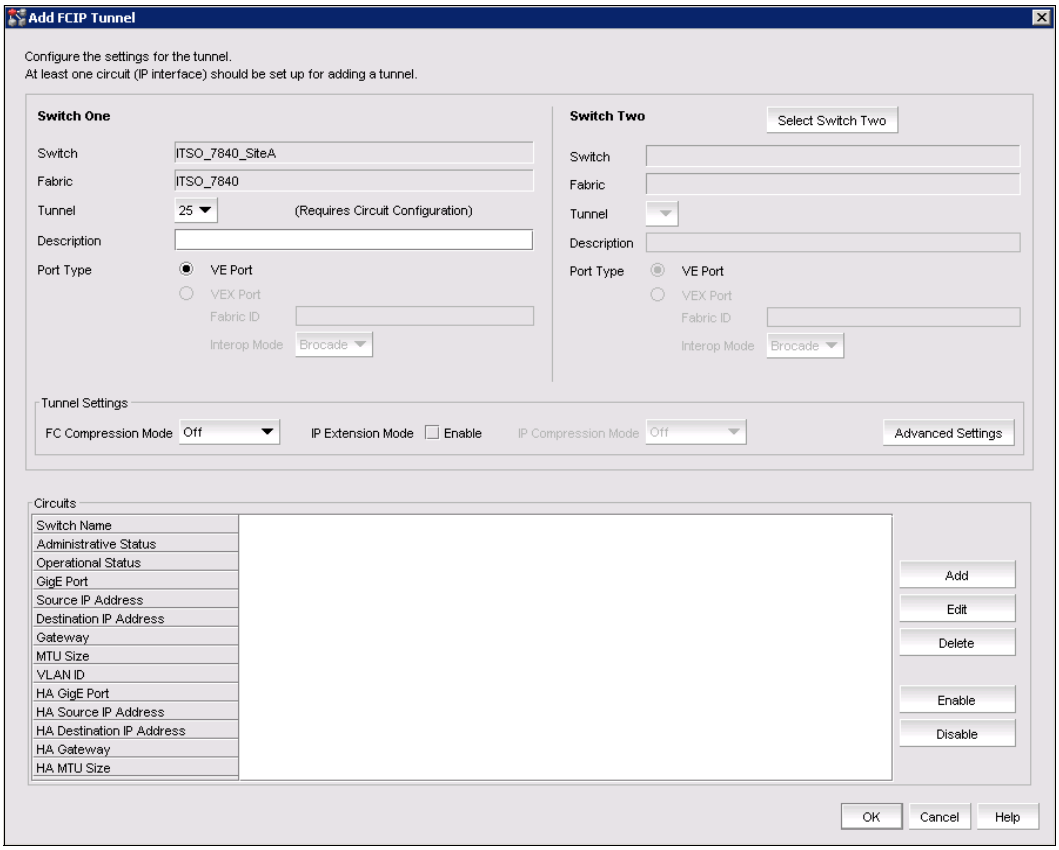


Figure 4-7 Select Switch Two

5. Click **OK** (Figure 4-8).

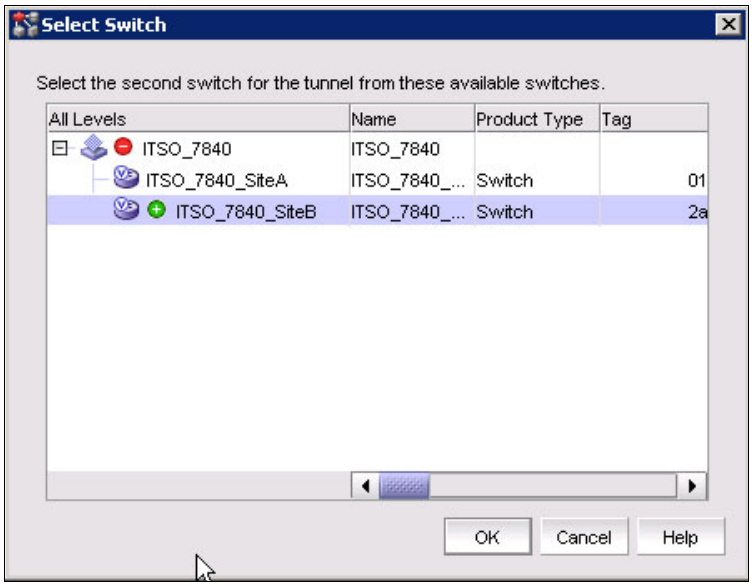


Figure 4-8 Select the extension switch on the site B

6. Select a value for **Tunnel** for both of the extension switches. Here, we select 34 for the new tunnel. Go to the circuits window and click **Add** to add circuits for the tunnel.

7. Here, we created four FCIP circuits for tunnel 34. To add the first circuit (Figure 4-9 on page 117), use the following procedure:
   a. Select the **GigE Port** that is used for the Ethernet connection on each extension switch. Here we select **ge2** for the first circuit.
   b. Select the **IP Address Type**. Here we select **IPv4**.
   c. Select the **IP Address** for each port. If no IP address has been defined for this port before, a new IP address must be entered. Here we enter 10.1.1.50 to create an IP address for the port ge2 on site A, and 10.1.1.60 for ge2 on site B.
   d. For IPv4 addresses, specify the **Subnet Mask**. For IPv6 addresses, specify the prefix length. Here we enter 255.255.255.0 for the subnet mask.
   e. If the IP addresses of the two sites are not in the same subnet, a specific route gateway is needed. To create a route through a gateway, click **Create Non-Default Route**, and select a **Gateway** address. Here we do not need to configure the gateway because the IP addresses from two sites are in the same subnet.
   f. Enter the **MTU size** or click **Auto MTU Size** for both extension switches. This is mandatory. The MTU size must match on both ends of the tunnel.
   g. If a VLAN ID is used to route frames between extension switches, input the **VLAN ID** value for both of the extension switches. Here we do not need it.
   h. To enable the Hot Code Load (HCL) feature, **GigE Port** for **HA Connectivity** should be specified. Here we select **ge2** for HCL backup usage on both sites. For more information about HCL, see 2.4.12, “Extension Hot Code Load” on page 30.

      Notes:
      - For configuration with the IBM Network Advisor GUI, you must input the HA connectivity information that is used for HCL, including HA IP addresses, net mask, and MTU size.
      - For the IBM SAN42B-R extension switch, the IP MTU size must be at least 1280. If the supported maximum IP MTU size in the network is larger than 9216, the IP MTU must be 9216.

   i. Enter the **IP Address**, **Subnet Mask,** and **MTU Size** of the HCL network port.

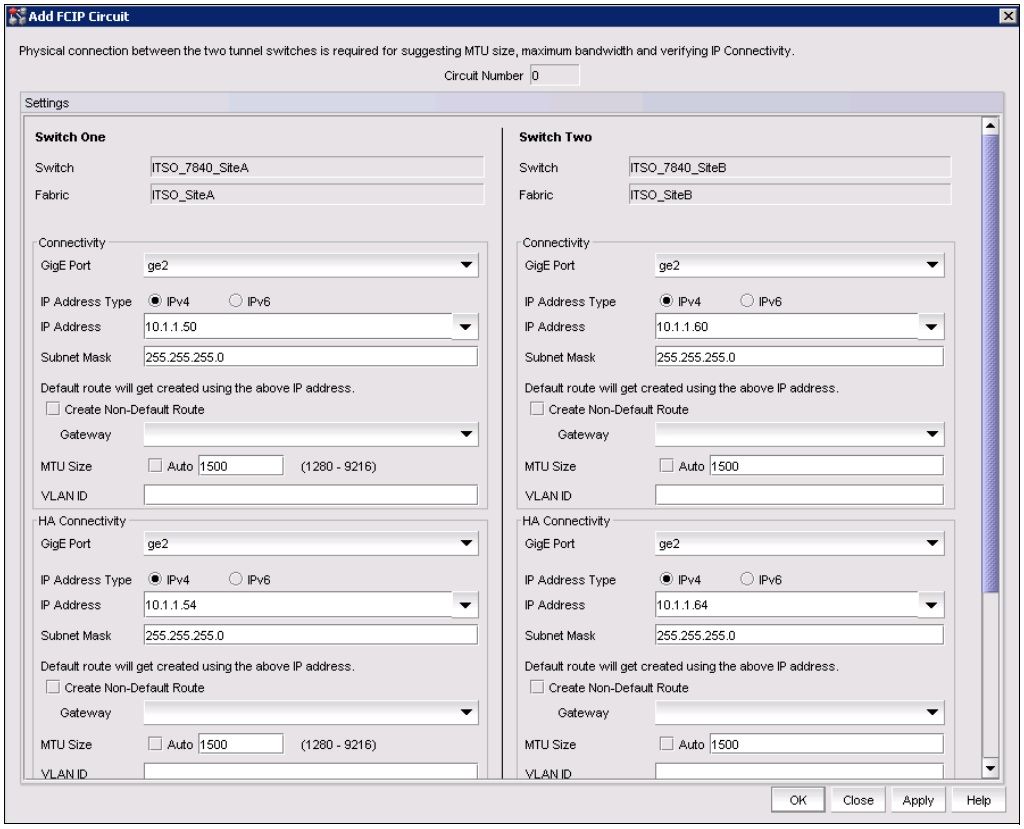


Figure 4-9 Add the first circuit for FCIP Tunnel

j. The **Metric** option is used to identify the failover circuit (Figure 4-10). If the value 0 (default value) is assigned, this circuit is used as the active circuit for load balancing. If the value 1 is assigned, it is used as the standby circuit to operate during circuit failover. In our scenario, we assign value 0 for the circuits 34.0 and 34.2 and value 1 for 34.1 and 34.3.

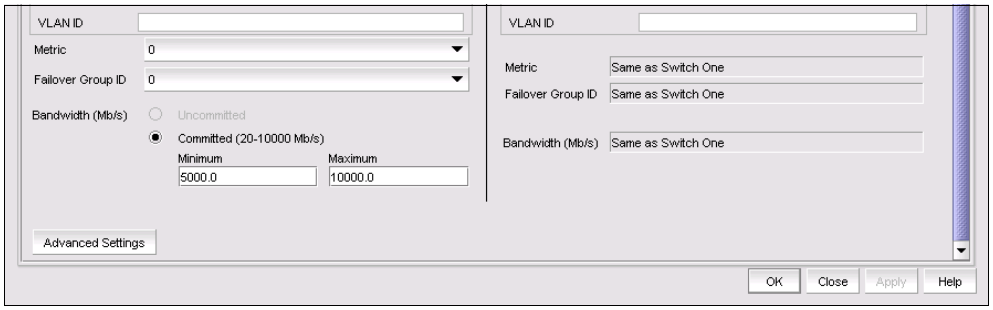


Figure 4-10 Input Metric and Failover Group ID for circuits

k. Specify the **Failover Group ID** for the circuit. With circuit failure groups, we can get better control for metric 1 circuits when one of the metric 0 circuits fails. For more information about circuit failover group feature, see 2.4.5, “Circuit failover/failback and failover groups” on page 21.

l. Enter the values of **Bandwidth** settings. It is mandatory to assign committed bandwidth for each circuit. Also, ensure that the values from both site are identical.

8. Click **OK** to confirm the setting of the first circuit and return to the Add FCIP Tunnel window.
9. Click **Add** to add more FCIP circuits for this FCIP tunnel. Here we create the second one (Figure 4-11).

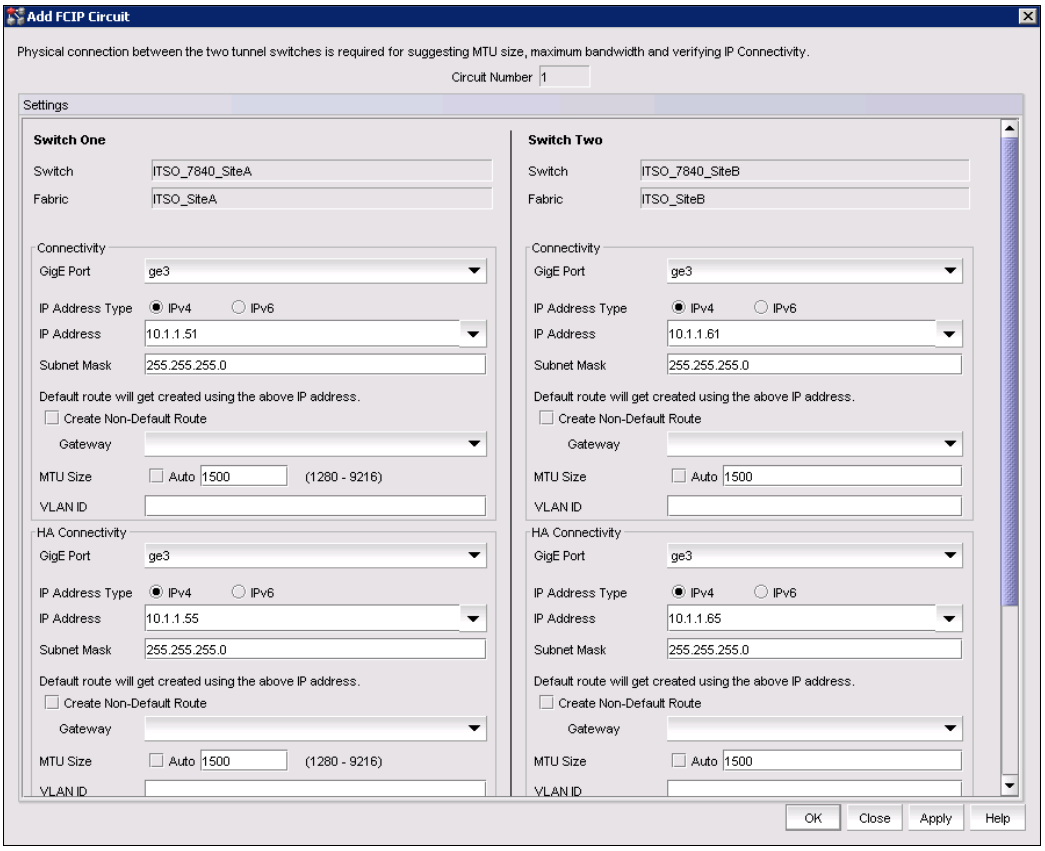


Figure 4-11 Add the second circuit

10. Repeat this action to create the other FCIP circuits. We created a total of four FCIP circuits for FCIP tunnel 35.

11. After four FCIP circuits are configured, click **OK** to begin the creation process and monitor the status from the configuration report until it completes (Figure 4-12).

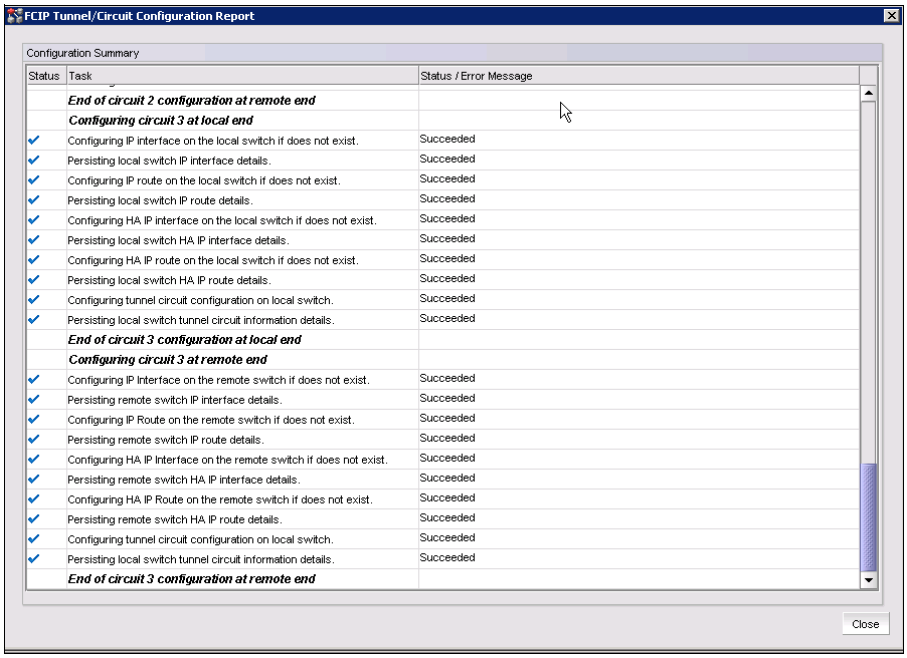


Figure 4-12 FCIP Tunnel Configuration report

12. After the process, the FCIP tunnel has been created successfully. Click **Configure** → **FCIP Tunnels** to view general FCIP properties of each FCIP Tunnel (Figure 4-13).

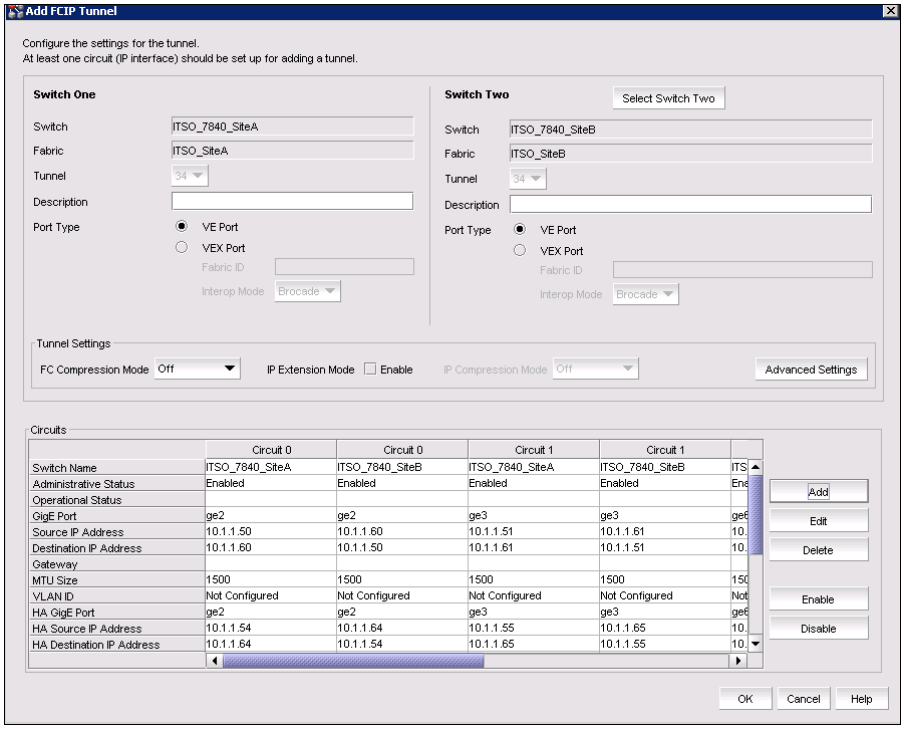


Figure 4-13 View the FCIP Tunnel properties

4.11.2 Configuring QoS for the FCIP tunnel

Quality of service (QoS) refers to policies for handling differences of data traffic. QoS for FC traffic is enabled through internal QoS priorities that can be mapped to IP network priorities by creating zone name prefixes (**`QosH_, QosM_, QosL_`**). For more detailed information about QoS, see 2.4.9, “Compression” on page 25.

To configure the QoS setting for a FCIP tunnel, complete the following steps:

1. Click **Configure** → **FCIP Tunnels** from the Overview window (Figure 4-5 on page 114).
2. Select the FCIP Tunnel to configure QOS, and then click **EDIT**.
3. Click **Advanced Settings**.

   There are three fabric QOS priorities (**High**, **Medium**, and **Low**) that are supported for FC, which are shown in Figure 4-14. The default distribution is 50% for high, 30% for medium and 20% for low. To modify the default value, enter the new value that you want to assign to the three priorities and click **OK**.

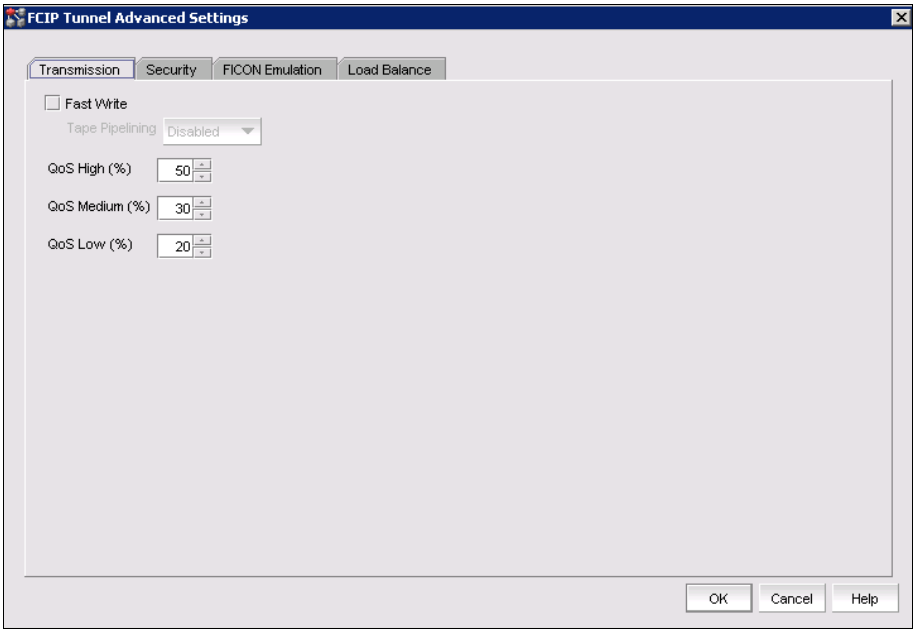


Figure 4-14 QoS configuration

4. Click **OK** to confirm the setting (Figure 4-15).

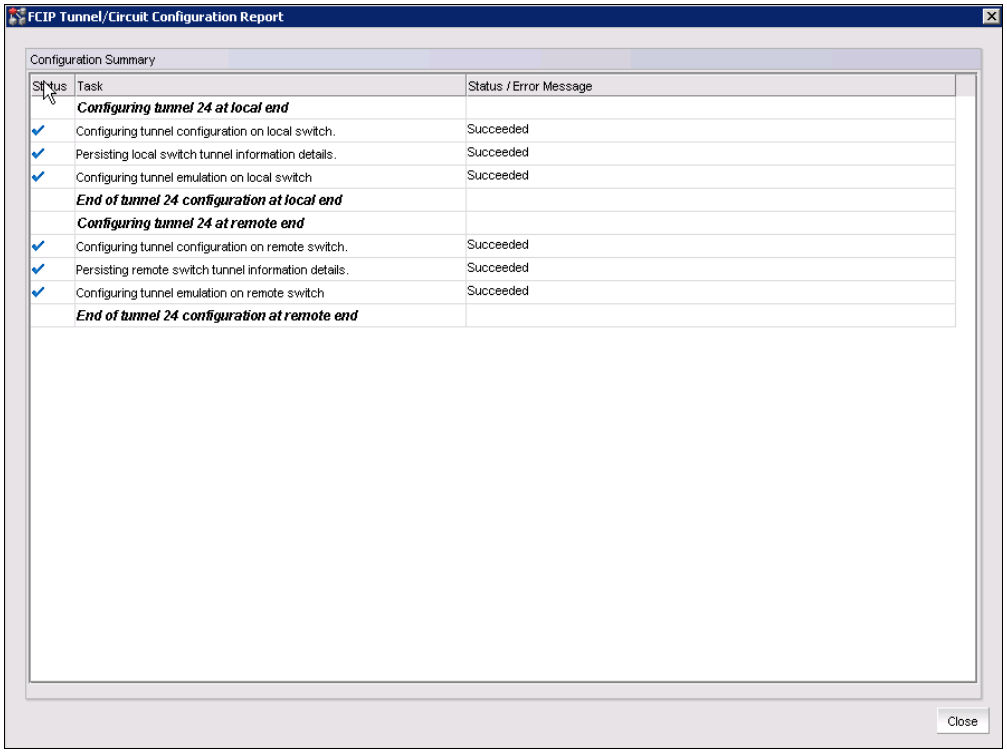


Figure 4-15 QoS configuration is complete

4.11.3 Configuring IPsec for the FCIP tunnel

Internet Protocol Security (IPsec) is a feature to ensure secure and private communications over Internet networks. For more information about IPsec, see 2.4.7, “IPsec” on page 24.

To configure IPsec for the FCIP tunnel, complete the following steps:

1. Click **Configure** → **FCIP Tunnels** from the Overview window (Figure 4-5 on page 114).
2. Select the FCIP Tunnel for which to configure IPsec, and then click **Edit**.
3. Click **Advanced Settings** and select the Security tab.

4. Select **Enable IPSec** (Figure 4-16).

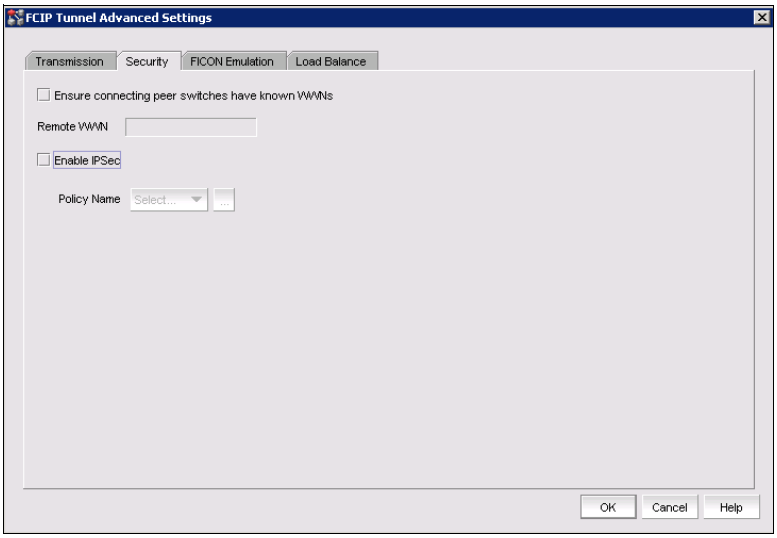


Figure 4-16 IPsec setting for FCIP

5. Click **Select** to choose one IPsec policy from the list. If there is no IPsec policy defined before, click the button on the right to create one.
6. Enter the name and preshared key for the new policy. The preshared key should be 16 - 64 characters (Figure 4-17).

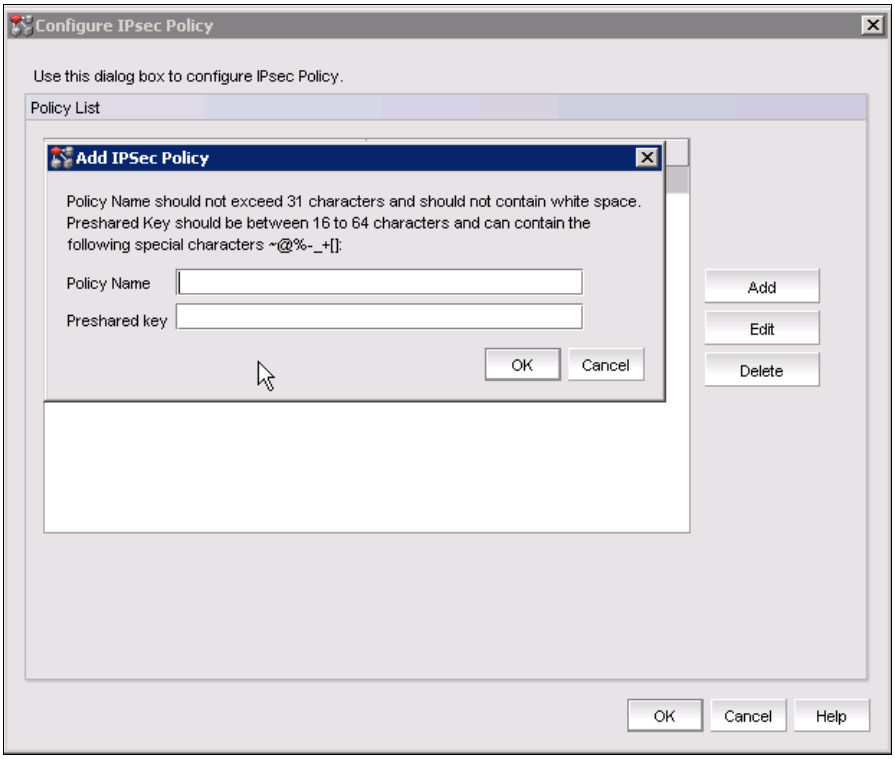


Figure 4-17 Create an IPsec policy

7. Click **OK** to confirm the settings for the IPsec policy (Figure 4-18).

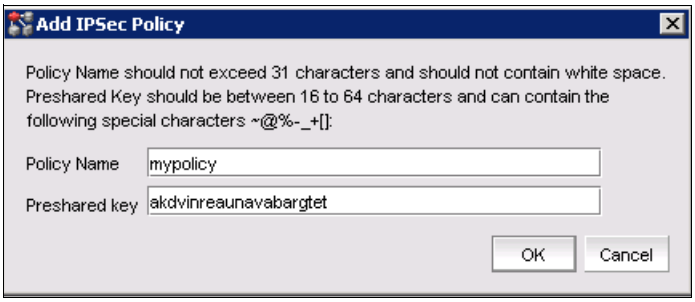


Figure 4-18 Enter policy key

8. Select the policy that was created and click **OK** to complete the task (Figure 4-19).

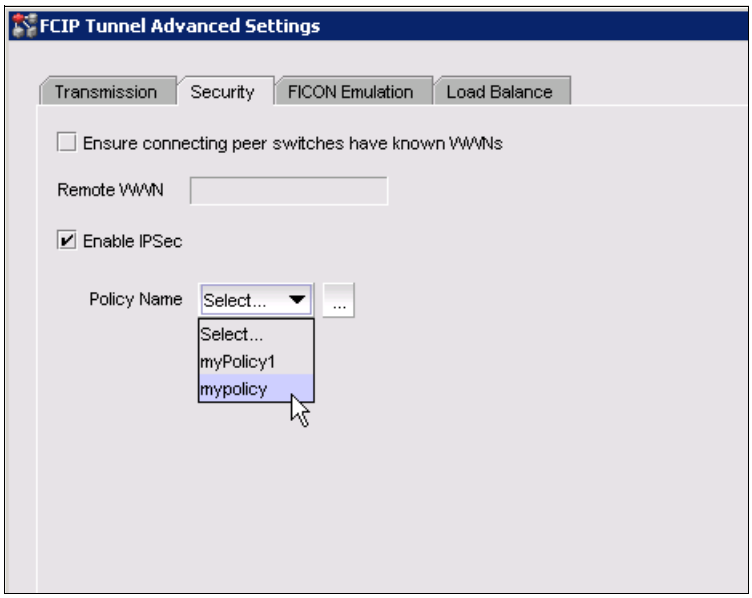


Figure 4-19 Select an IPsec policy

5

IP Extension with the TS7760/TS7700 Grid

This chapter discusses the implementation of a TS7760 or TS7700 Grid with SAN42B-R extension switches.

It provides the following information:

- Introduction
- TS7760 and TS7700 overview
- Implementation of IP Extension
- Configuring the TS7700 Grid cluster

5.1 Introduction

For an overview of TS7760/7700 Grid extension technology, see 1.1.4, “IBM TS7760/7700, business continuity, and grid basics” on page 5.

5.2 TS7760 and TS7700 overview

The TS7700 is a modular, scalable, and high-performing architecture for mainframe tape virtualization. This is a fully integrated, tiered storage hierarchy of disk and tape. It incorporates extensive self-management capabilities consistent with IBM Information Infrastructure initiatives. The examples in this chapter describe a TS7700 implementation. See your product documentation for the most current information.

For information about architecture, planning, and implementation, see *IBM TS7700 Release 3.3*, SG24-8122:

http://www.redbooks.ibm.com/abstracts/sg248122.html?Open

Also, see your product documentation and IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/STFS69

5.3 Implementation of IP Extension

This section describes the implementation steps on the SAN42B-R switches to use the IP extension feature for TS7760 and TS7700 products.

5.3.1 Lab configuration

The following example provides an overview of our lab configuration that is used to describe all further implementation steps within this chapter.

Figure 5-1 shows our lab configuration with a TS7700 cluster, a SAN42B-R switch, and a layer 2 switch on sites A and B. Cluster A acts as primary TS7700 cluster on site A, Cluster B acts as a secondary TS7700 cluster on site B. Hydswitchb1 is the SAN42B-R switch on site A, and hydswitchb2 is the SAN42B-R switch on site B.

Gigabit Ethernet ports ge10 and ge11 defined on SAN42B-R switches are configured as a LAN gateway, while ge2 and ge3 are used for IP extension tunnel configuration.

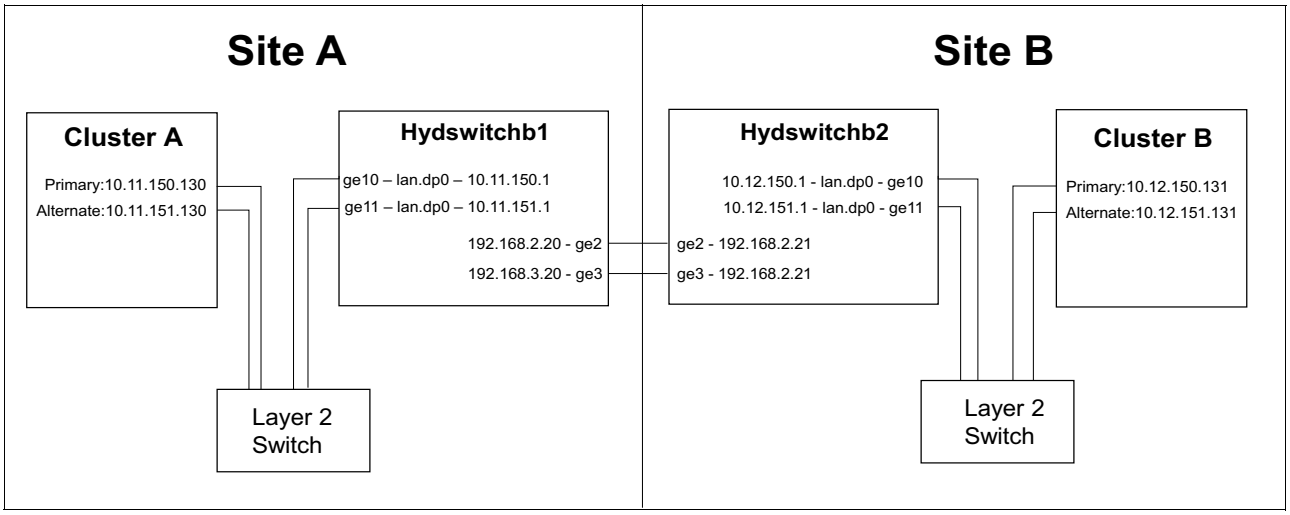


Figure 5-1 ITSO lab configuration overview with TS7700 clusters, SAN42B-R switches, and layer 2 switches

Figure 5-2 shows the IP extension tunnel configuration. The circuit defined between Gigabit Ethernet ports ge2.dp0 on site A and ge2.dp0 on site B acts as an active circuit, while the circuit defined between Gigabit Ethernet ports ge3.dp0 on site A and ge3.dp0 on site B acts as a standby circuit. Both circuits are configured to failover group 0. See Chapter 4, “FCIP replication” on page 93 for more information about IP tunnels.

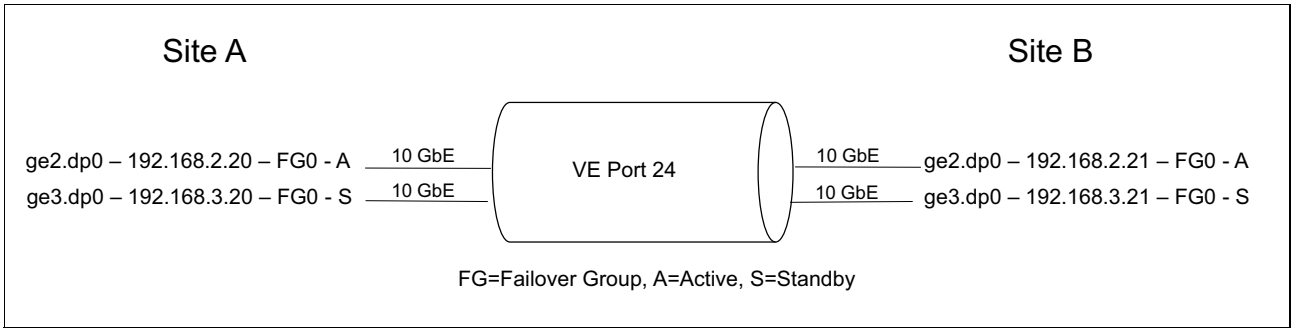


Figure 5-2 ITSO lab IP extension tunnel configuration.

5.3.2 Overview of configuration steps

The following list provides an outline of the required implementation steps:

- Hybrid mode configuration
- Ethernet interface mode configuration
- Ethernet link aggregation group (LAG) configuration
- IPEX LAN gateway configuration
- IP extension tunnel configuration
- Traffic control list

5.3.3 Hybrid mode configuration

Hybrid mode configuration allows you to operate a SAN42B-R switch with FCIP tunnel and the IP extension feature. The SAN42B-R switch must be configured to hybrid mode with the **extncfg** command before the IPEX configuration. Note that 20VE mode is not supported when using hybrid mode.

CLI

In our lab configuration, hybrid mode had to be enabled on SAN42B-R on site A and site B. Example 5-1 shows how to enable hybrid mode with the **extncfg** command. Note that this step requires a reboot of the switch.

Example 5-1 Enable hybrid mode

```
hydswitchb1:FID128:admin>extncfg --app-mode hybrid
This action will configure the system for Hybrid (FCIP/IPExt) mode.
WARNING: This is a disruptive operation that requires a reboot to take effect.
Would you like to continue (Y,y,N,n): [ n] y
Operation succeeded. Rebooting the system
```

Example 5-2 shows the current switch mode setting on the switch.

Example 5-2 Show current switch setting

```
hydswitchb1:FID128:admin> extncfg --show
APP Mode is HYBRID (FCIP with IPEXT)
VE-Mode: configured for 10VE mode.
hydswitchb1:FID128:admin>
```

5.3.4 Ethernet Interface Mode configuration

IPEX configuration requires to you set Gigabit Ethernet ports to LAN mode prior link aggregation group configuration. The following points must be noted before configuration:

- The switch must be in hybrid mode
- Only Gigabit Ethernet ports ge2-ge17 can be configured as LAN Ports
- Gigabit Ethernet ports ge0 and ge1 used for 40 Gigabit Ethernet do not support LAN mode
- Up to eight 10 Gigabit Ethernet ports can be configured to LAN mode

In our lab configuration, ge10 and ge11 on SAN42B-R switches on site A and B must be set to LAN mode. Example 5-3 shows the configuration of ports ge10 and ge11 on site A to LAN mode.

Example 5-3 Site A: Create a new link aggregation group

```
hydswitchb1:FID128:admin> portcfgge ge10 --set -lan
Operation Succeeded
hydswitchb1:FID128:admin> portcfgge ge11 --set -lan
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-4 shows the settings for port ge10 and ge11 that are displayed by using the **portshow** command.

Example 5-4 Verify port status

```
hydswitchb1:FID128:admin> portshow ge10
Eth Mac Address: 50.eb.1a.d7.83.8a
Port State: 1   Online
Port Phys:  6   In_Sync
Port Flags: 0x4003       PRESENT ACTIVE LED
Port Speed: 10G
hydswitchb1:FID128:admin> portshow ge11
Eth Mac Address: 50.eb.1a.d7.83.8b
Port State: 1   Online
Port Phys:  6   In_Sync
Port Flags: 0x4003       PRESENT ACTIVE LED
Port Speed: 10G
hydswitchb1:FID128:admin>
```

5.3.5 Configuring Ethernet link aggregation groups

A link aggregation group consist of multiple physical Ethernet interfaces and allows you to form a single logical link that provides redundant cables and optics.

The following points must be noted before implementing link aggregation groups:

- A link aggregation group must have a name and ID.
- Port speed and auto-negotiate setting must match for all members within a link aggregation group.
- Up to four links can be assigned to one link aggregation group.
- A total of eight link aggregation groups are supported on SAN42B-R.
- Link aggregation groups are supported between LAN ports of SAN42B-R switch and a layer 2 switch.
- 10 Gigabit Ethernet ports and 40 Gigabit Ethernet ports do not have an auto-negotiation setting.
- On 1 Gbps interfaces, auto-negotiation is the default setting.
- The port setting on SAN42B-R router and the direct connect LAN switch must match.

The following steps show the configuration of link aggregation group with the CLI:

1. Create a link aggregation group.

   You must create a new link aggregation group with the **portcfg lag name --create** command.

   Example 5-5 shows the creation of link aggregation group lag1 on Site A.

   Example 5-5 Site A: Create a new link aggregation group lag1

   ```
   hydswitchb1:FID128:admin> portcfg lag lag1 --create
   Operation Succeeded
   hydswitchb1:FID128:admin>
   ```

Example 5-6 shows the creation of link aggregation group lag1 on Site B.

Example 5-6 Site B: Create a new link aggregation group lag1

```
hydswitchb2:FID128:admin> portcfg lag lag1 --create
Operation Succeeded
hydswitchb2:FID128:admin>
```

2. Add Gigabit Ethernet ports to the link aggregation group.

   In our lab configuration, ge10 and ge11 must be assigned to link aggregation group lag1. Example 5-7 shows the assignment of ports ge10 and ge11 to link aggregation group lag1 on Site A.

   Example 5-7 Site A: Assign ge10 and ge11 to link aggregation group lag1

```
hydswitchb1:FID128:admin> portcfg lag lag1 --add ge10-ge11

WARNING: While making configuration changes the modified LAN GE ports will be
disabled. Please run "portenable" command to manually enable the modified LAN
GE ports after completing all the  configuration changes.

Would you like to continue  (Y,y,N,n): [ n]     y
Operation Succeeded
hydswitchb1:FID128:admin>
```

   Ge10 and ge11 must be enabled with the **portenable** command after assignment to link aggregation group lag1. To remove Gigabit Ethernet ports from a link aggregation group, you must use the **portcfg lag name --remove** command.

   Example 5-8 shows details of link aggregation group lag1.

   Example 5-8 Show details of link aggregation group lag1

```
hydswitchb1:FID128:admin> portshow lag --detail
LAG : lag1
------------------------------------------------------------
Oper State : Offline
Port Count : 2
port      AdminSt   Oper state     Speed     AutoNeg
ge10      Enabled   OFFLINE        10G       Disabled
ge11      Enabled   OFFLINE        10G       Disabled
```

   Note: Auto-negotiation applies only to 1 Gbps Ethernet interface mode, and cannot be set on 10GE and 40GE interfaces.

   Auto-negotiation allows you to determine Ethernet pause frame flow-control and full/half duplex link settings between two endpoints. This setting must match between DC LAN switch and the SAN42B-R switch interfaces.

3. Set the speed on Ethernet ports within link aggregation group lag1.

   The default speed of 10 Gigabit Ethernet interfaces is 10 Gbps. You can use the **portcfg lag** command to modify the speed of interfaces within a link aggregation group.

Example 5-9 shows the speed of interfaces ge10 and ge11 within the link aggregation group lag1.

Example 5-9 Display speed of interfaces ge10 and ge11 within link aggregation group lag1

```
hydswitchb1:FID128:admin> portshow lag --detail

LAG : lag1
------------------------------------------------------------
Oper State : Online
Port Count : 2
port      AdminSt   Oper state     Speed     AutoNeg
ge10      Enabled   ONLINE         10G       Disabled
ge11      Enabled   ONLINE         10G       Disabled
hydswitchb1:FID128:admin>
```

5.3.6 IPEX LAN gateway configuration

For IPEX LAN gateway configuration, the switch virtual interface (SVI) IP interface (IPIF) must be configured to allow redirection of end-devices over IP extension. The following points must be noted before configuration:

- There is only one switch virtual interface IP interface with one LAN-side Ethernet device and MAC address per data processor complex.
- Up to eight switch virtual interfaces IP interfaces can be defined per data processor, but all use the same single switch virtual interface IP interface.
- The switch virtual interface IP interface (SVI IPIF) is referred to as a LAN gateway.
- SVI IPIF must be configured for each subnet that end devices are connecting from.
- An IP route must be added on end-devices for each different IP subnet used to connect to the SAN42B-R router.
- Local end devices and remote end devices must be in different subnets.
- A gateway must be in the same subnet as local connected end-device interfaces.
- Multiple LAN gateways belonging to the same IP subnet cannot be configured on the same DP.
- Each VLAN that is used in a link aggregation group requires a separate IPIF gateway.
- LAN gateway interfaces do not support PMTU auto-discovery.

CLI

The configuration of LAN gateway must be done on SAN42B-R switches site A and B. Example 5-10 shows the configuration of a new LAN gateway on data processor 0 with the **`portcfg ipif create`** command on site A.

Example 5-10 Site A: Configuration of a LAN gateway

```
hydswitchb1:FID128:admin> portcfg ipif lan.dp0 create  10.11.150.1/24 mtu 1500
Operation Succeeded
hydswitchb1:FID128:admin> portcfg ipif lan.dp0 create  10.11.151.1/24 mtu 1500
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-11 shows the configuration of a new LAN gateway on data processor 0 with the **portcfg ipif create** command on site B.

Example 5-11 Site B: Configuration of a LAN gateway

```
hydswitchb2:FID128:admin> portcfg ipif lan.dp0 create  10.12.150.1/24 mtu 1500
Operation Succeeded
hydswitchb2:FID128:admin> portcfg ipif lan.dp0 create  10.12.151.1/24 mtu 1500
Operation Succeeded
hydswitchb2:FID128:admin>
```

5.3.7 Configuring the extension tunnel

This section describes all required steps for implementing an extension tunnel on the SAN42B-R switch.

Disabling the VE Port

The VE Port must be disabled with the **portcfgpersistentdisable** command. See Chapter 4, “FCIP replication” on page 93.

Configuring IPIF

LAN interfaces used for IPEX tunnel configuration must belong to the default logical switch. See 4.4, “Creating IP interfaces with the command line” on page 98 for configuration information.

Configuring IP Routes

See 4.5, “Creating IP routes with the command line” on page 100.

IPsec configuration

See 4.7.1, “Creating an IPsec policy” on page 101.

Creating an IPEX tunnel

This section provides all required configuration steps for implementing an IPEX tunnel. For more information about tunnels, see Chapter 4, “FCIP replication” on page 93.

Example 5-12 shows the creation of a tunnel within failover group 0, with metric 0 and IP extension enabled on the switch `hydswitchb1`.

Example 5-12 Site A: Create a tunnel

```
hydswitchb1:FID128:admin> portcfg fciptunnel 24 create --local-ip 192.168.2.20
--remote-ip 192.168.2.21 -b 5000000 -B 10000000 -k 1000 -i none -c fast-deflate -g
0 -x 0 --ipext enable

!!!! WARNING !!!!
Fast deflate compression cannot be applied as IP-compression. Will be taken as no
compression

Continue with operation (Y,y,N,n): [ n] y
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-13 shows the creation of a tunnel within failover group 0, metric 0 with IP extension enabled on switch `hydswitchb2`.

Example 5-13 Site B: Create a tunnel

```
hydswitchb2:FID128:admin> portcfg fciptunnel 24 create --local-ip 192.168.2.21
--remote-ip 192.168.2.20 -b 5000000 -B 10000000 -k 1000 -i none -c fast-deflate -g
0 -x 0 --ipext enable

!!!! WARNING !!!!
Fast deflate compression cannot be applied as IP-compression. Will be taken as no
compression

Continue with operation (Y,y,N,n): [ n] y
Operation Succeeded
hydswitchb2:FID128:admin>
```

Note: Extension tunnels are not enabled by default. When you use the **`portcfg fciptunnel modify`** command, an existing FCIP tunnel can be configured to be IPEX capable.

Example 5-14 shows the configuration details of the tunnel that is defined on VE port 24.

Example 5-14 Display details of tunnel configuration

```
hydswitchb1:FID128:admin> portshow fciptunnel 24 --detail --circuit

 Tunnel: VE-Port:24 (idx:0, DP0)
 ==================================================
  Oper State           : Online
  TID                  : 24
  Flags                : 0x00000000
  IP-Extension         : Enabled
  Compression          : Fast Deflate
  FC-Compression       : Fast Deflate (Inherited)
  IP-Compression       : None (Inherited)
  QoS Distribution     : Protocol (FC:50% / IP:50%)
  FC QoS BW Ratio      : 50% / 30% / 20%
  IP QoS BW Ratio      : 50% / 30% / 20%
  Fastwrite            : Disabled
  Tape Pipelining      : Disabled
  IPSec                : Disabled
  Load-Level (Cfg/Peer): Failover (Failover / Failover)
  Local WWN            : 10:00:50:eb:1a:d7:83:80
  Peer WWN             : 10:00:50:eb:1a:36:1d:38
  RemWWN (config)      : 00:00:00:00:00:00:00:00
  cfgmask              : 0x0000001f 0x40000248
  Uncomp/Comp Bytes    : 362911132 / 362911132 / 1.00 : 1
  Uncomp/Comp Byte(30s): 15404740 / 15404740 / 1.00 : 1
  Flow Status          : 0
  ConCount/Duration    : 1 / 15m10s
  Uptime               : 13m11s
  Stats Duration       : 13m11s
  Receiver Stats       : 362870492 bytes / 2554784 pkts /  512.92 KBps Avg
  Sender Stats         : 362957888 bytes / 2554913 pkts /  513.51 KBps Avg
  TCP Bytes In/Out     : 608332544 / 611736316
```

```
  ReTx/OOO/SloSt/DupAck: 0 / 0 / 0 / 0
  RTT (min/avg/max)    : 1 / 1 / 48 ms
  Wan Util             : 0.9%
  TxQ Util             : 0.0%

 Circuit 24.0 (DP0)
 ==================================================
  Admin/Oper State     : Enabled / Online
  Flags                : 0x00000000
  IP Addr (L/R)        : 192.168.2.20 ge2 <-> 192.168.2.21
  HA IP Addr (L/R)     : 0.0.0.0 ge0 <-> 0.0.0.0
  Configured Comm Rates: 5000000 / 10000000 kbps
  Peer Comm Rates      : 5000000 / 10000000 kbps
  Actual Comm Rates    : 5000000 / 10000000 kbps
  Keepalive (Cfg/Peer) : 1000 (1000 / 1000) ms
  Metric               : 0
  Connection Type      : Default
  ARL-Type             : Auto
  PMTU                 : Disabled
  SLA                  : (none)
  Failover Group       : 0
  VLAN-ID              : NONE
  L2Cos (FC:h/m/l)     : 0 / 0 / 0 (Ctrl:0)
  L2Cos (IP:h/m/l)     : 0 / 0 / 0
  DSCP (FC:h/m/l)      : 0 / 0 / 0 (Ctrl:0)
  DSCP (IP:h/m/l)      : 0 / 0 / 0
  cfgmask              : 0x40000000 0x00000caf
  Flow Status          : 0
  ConCount/Duration    : 1 / 15m9s
  Uptime               : 13m11s
  Stats Duration       : 13m11s
  Receiver Stats       : 362870492 bytes / 2554784 pkts /  515.61 KBps Avg
  Sender Stats         : 362957888 bytes / 2554913 pkts /  515.07 KBps Avg
  TCP Bytes In/Out     : 608336352 / 611740012
  ReTx/OOO/SloSt/DupAck: 0 / 0 / 0 / 0
  RTT (min/avg/max)    : 1 / 1 / 48 ms
  Wan Util             : 0.9%

hydswitchb1:FID128:admin>
```

Example 5-15 shows the modification of the distribution between FCIP and IPEX traffic. The distribution must match with the tunnel settings of `hydswitchb2`. The first parameter of the distribution setting applies to FCIP, and the second to IPEX.

Example 5-15 Modify the distribution

```
hydswitchb1:FID128:admin> portcfg fciptunnel 24 modify --distribution 40,60

Warning: Modification of the distribution ratio will reset the QoS ratio values to
default.

!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a brief
period of time. This operation will bring the existing tunnel down (if tunnel is
up) before applying new configuration.
```

```
Continue with Modification (Y,y,N,n): [ n]      y
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-16 shows the tunnel details with the new distribution settings.

Example 5-16 Show tunnel details

```
hydswitchb1:FID128:admin> portshow fciptunnel --detail

 Tunnel: VE-Port:24 (idx:0, DP0)
 ================================================
  Oper State           : Online
  TID                  : 24
  Flags                : 0x00000000
  IP-Extension         : Enabled
  Compression          : Fast Deflate
  FC-Compression       : Fast Deflate (Inherited)
  IP-Compression       : None (Inherited)
  QoS Distribution     : Protocol (FC:40% / IP:60%)
  FC QoS BW Ratio      : 50% / 30% / 20%
  IP QoS BW Ratio      : 50% / 30% / 20%
  Fastwrite            : Disabled
  Tape Pipelining      : Disabled
  IPSec                : Disabled
  Load-Level (Cfg/Peer): Failover (Failover / Failover)
  Local WWN            : 10:00:50:eb:1a:d7:83:80
  Peer WWN             : 10:00:50:eb:1a:36:1d:38
  RemWWN (config)      : 00:00:00:00:00:00:00:00
  cfgmask              : 0x0000001f 0x40000248
  Uncomp/Comp Bytes    : 24672782268 / 24672782268 / 1.00 : 1
  Uncomp/Comp Byte(30s): 15403296 / 15403296 / 1.00 : 1
  Flow Status          : 0
  ConCount/Duration    : 1 / 13h29m
  Uptime               : 13h27m
  Stats Duration       : 13h27m
  Receiver Stats       : 24674789512 bytes / 173758749 pkts /  513.77 KBps Avg
  Sender Stats         : 24673783528 bytes / 173758619 pkts /  513.46 KBps Avg
  TCP Bytes In/Out     : 41889885788 / 41947553248
  ReTx/OOO/SloSt/DupAck: 0 / 0 / 0 / 0
  RTT (min/avg/max)    : 1 / 1 / 48 ms
  Wan Util             : 1.0%
  TxQ Util             : 0.0%
```

Configuring circuits

In our example, a second circuit on ge3 on both switches was added to failover group 0 as a standby circuit. See 4.7.4, “Creating an additional circuit” on page 104 for information about adding additional circuits to an existing tunnel.

Also, see the *Brocade Fabric OS Extension Configuration Guide 8.0.1:*

http://www.brocade.com/content/html/en/configuration-guide/fos-801-extension/GUID-647FE9BE-0AA2-4774-B8E2-3BB2986D74F1-homepage.html

Configuring compression

See Chapter 2, “The IBM System Storage SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade” on page 13 for more comprehensive information and Chapter 6, “FCIP and integrated routing” on page 153 for configuration steps.

Configuring quality of service

The configuration of quality of service is optional and can be used for configuring bandwidth distribution between IPEX and FCIP. By default, the distribution between FC and IP traffic is set to 50,50 when IPEX is enabled. This setting can be modified by using the **portcfg fciptunnel modify** command. The first value of the **--distribution** parameter applies to FCIP traffic, and the second value applies to IPEX traffic.

The bandwidth received for FCIP and IPEX can be divided with the parameter **-- fc-qos-ratio** for FCIP and **-- ip-qos-ratio** for IP extension. See the *Brocade Fabric OS Extension Configuration Guide 8.0.1* for more comprehensive information:

http://www.brocade.com/content/html/en/configuration-guide/fos-801-extension/GUID-647FE9BE-0AA2-4774-B8E2-3BB2986D74F1-homepage.html

Note: Protocol distribution allows a protocol to use more bandwidth than is specified by the **--distribution** command because the other protocol is not using its allocated bandwidth.

5.3.8 Traffic control list

The configuration of a traffic control list provides a mapping of LAN traffic to specific IP tunnels by creating rules of LAN characteristics including but not limited to IP addresses, VLAN tag, layer 4 protocols, and ports. Traffic control list (TCL) rules are organized by priority and act as an input filter for the data processor that can allow or deny certain traffic from being used by an IP extension tunnel.

TCL rules are identified by a unique name and can contain 31 or fewer characters. There is one default TCL rule defined that denies all traffic. This default rule cannot be removed or modified. A new TCL rule must be created.

CLI

In our lab configuration, a TCL rule must be created with the **portcftg tcl create** command. Example 5-17 shows the creation of a new TCL rule for subnet 10.12.150.0/24 and for ports 1415 - 1416. Non-terminated must be enabled.

Example 5-17 Site A: Create a new TCL Rule for subnet 10.11.150.0/24 and ports 1415-1416

```
hydswitchb1:FID128:admin>  portcfg tcl IPEX_150_CTRL_1415-1416 create --admin
enable --action allow --src-addr 10.11.150.0/24 --dst-addr 10.12.150.0/24
--proto-port 1415-1416 --priority 10 --target 24 --non-terminated enable
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-18 shows the creation of a new TCL rule for TCP Ports 1415 - 1416 and for the subnet 151 on site A. Non-terminated must be enabled.

Example 5-18 Site A: Create a TCL Rule for subnet 10.11.151.0/24 and ports 1415-1416

```
hydswitchb1:FID128:admin>  portcfg tcl IPEX_151_CTRL_1415-1416 create --admin
enable --action allow --src-addr 10.11.151.0/24 --dst-addr 10.12.151.0/24
--proto-port 1415-1416 --priority 11 --target 24 --non-terminated enable
Operation Succeeded
```

Example 5-19 shows the creation of a new TCL rule for data traffic on site A. Note that TCL rule defined for data traffic must have a lower priority (a higher priority number) than TCL rules defined for ports 1415 - 1416.

Example 5-19 Site A: Create TCL rule for data traffic

```
hydswitchb1:FID128:admin> portcfg tcl IPEX_data_150 create --admin enable --action
allow --src-addr 10.11.150.0/24 --dst-addr 10.12.150.0/24 --priority 12 --target
24
Operation Succeeded
hydswitchb1:FID128:admin>
hydswitchb1:FID128:admin> portcfg tcl IPEX_data_151 create --admin enable --action
allow --src-addr 10.11.151.0/24 --dst-addr 10.12.151.0/24 --priority 13 --target
24
Operation Succeeded
hydswitchb1:FID128:admin>
```

Example 5-20 shows the creation for TCP Ports 1415 - 1416 and for the subnet 150 on site B.

Example 5-20 SiteB: Create a TCR Rule for TCP Ports 1415 and 1416 within subnet 150

```
hydswitchb2:FID128:admin> portcfg tcl IPEX_150_CTRL_1415-1416 create --admin
enable --action allow --src-addr 10.12.150.0/24 --dst-addr 10.11.150.0/24
--proto-port 1415-1416 --priority 10 --target 24 --non-terminated enable
Operation Succeeded
```

Example 5-21 shows the creation for TCP Ports 1415 - 1416 and for the subnet 151 on site B.

Example 5-21 Site B: Create a TCR Rule for TCP Ports 1415 and 1416 within subnet 151

```
hydswitchb2:FID128:admin> portcfg tcl IPEX_151_CTRL_1415-1416 create --admin
enable --action allow --src-addr 10.12.151.0/24 --dst-addr 10.11.151.0/24
--proto-port 1415-1416 --priority 11 --target 24 --non-terminated enable
Operation Succeeded
hydswitchb2:FID128:admin>
```

Example 5-22 shows the creation of a TCL rule for data traffic on site B.

Example 5-22 Site B: Create TCL rule for data traffic

```
hydswitchb2:FID128:admin> portcfg tcl IPEX_data create --admin enable --action
allow --src-addr 10.12.150.0/24 --dst-addr 10.11.150.0/24 --priority 12 --target
24
Operation Succeeded
```

```
hydswitchb2:FID128:admin> portcfg tcl IPEX_data create --admin enable --action
allow --src-addr 10.12.151.0/24 --dst-addr 10.11.151.0/24 --priority 12 --target
24
TCL already exists.
hydswitchb2:FID128:admin>
```

Example 5-23 shows all traffic control list rules created on site A.

Example 5-23 View traffic control list rules

```
hydswitchb1:FID128:admin> portshow tcl

 Pri    Name                 Flgs   Target     L2COS  VLAN DSCP Proto Port Hit
                             Src-Addr                 Dst-Addr
-------------------------------------------------------------------------------
 *10    IPEX_150_CTRL_1415-.. AI--N  24-Med     ANY    ANY  ANY  ANY
                                                              1415-1416 0
                             10.11.150.0/24           10.12.150.0/24

 *11    IPEX_151_CTRL_1415-.. AI--N  24-Med     ANY    ANY  ANY  ANY
                                                              1415-1416 0
                             10.11.151.0/24           10.12.151.0/24

 *12    IPEX_data_150        AI---  24-Med     ANY    ANY  ANY  ANY   ANY  0
                             10.11.150.0/24           10.12.150.0/24

 *13    IPEX_data_151        AI---  24-Med     ANY    ANY  ANY  ANY   ANY  0
                             10.11.151.0/24           10.12.151.0/24

 *65535 default              D----  -          ANY    ANY  ANY  ANY   ANY  0
                             ANY                      ANY

-------------------------------------------------------------------------------
Flags: *=Enabled ..=Name Truncated (see --detail for full name)
       A=Allow D=Deny I=IP-Ext P=Segment Preservation
       R=End-to-End RST Propagation N=Non Terminated.

Active TCL Limits:   Cur / Max
---------------------------------------------
  DP0                  5 / 128
  DP1                  1 / 128
---------------------------------------------
Configured Total:      5 / 1024

hydswitchb1:FID128:admin>
```

The traffic control list rules defined for ports 1415 and1416 allow the SAN42B-R switch to enable control connections for GRID and bypass TCP acceleration. GRID can create a large number of these control connections and are used for control operations.

When control connections are accelerated, they count against the TCP connection limit in the SAN42B-R. By creating the bypass, control connections do not count against the limit, allowing for more GRID data connections (port 350) to be accelerated.

Note: The non-terminated rules must have a higher priority (lower priority number) than the data rule. When a frame needs to be match against a TCL, the code looks at the TCL list one at a time. It starts with the lower number first and works its way to the higher numbers.

The first rule that matches with the packet is picked. In this case, we want the TCP port specific rules that call out non-terminated to be hit first. They must be the lowest priority number.

See the *Brocade Fabric OS Extension Configuration Guide, 8.0.1* for more information:

http://www.brocade.com/content/html/en/configuration-guide/fos-801-extension/GUID-647FE9BE-0AA2-4774-B8E2-3BB2986D74F1-homepage.html

5.4 Configuring the TS7700 Grid cluster

This section discusses all required configuration steps for implementing TS7700 Grid cluster. Be sure to consider restrictions and prerequisites before beginning.

5.4.1 Restrictions for Grid joins

The following points must be noted for Grid joins:

- Cluster join and merge are supported on microcode level 8.21.x.x and later. This information can be viewed in the second line of the **q_node** command output (8.40.1.x).
- The final grid cannot have more than three different microcode levels.
- If the joining cluster has FC “5275 - Additional Virtual Devices” installed and it is joined to clusters with microcode level earlier than 8.32.0.x, the additional logical devices on the joining cluster is not accessible to the host until all clusters in the Grid are at 8.32.0.x or later.
- For a joining process, the joining cluster must not have any logical volumes in db. This setting can be verified by the **q_node** command at the line “Logical volumes actual=0.”
- If the primary cluster is already a member of a grid, then it must be the latest microcode level of any member in the Grid.
- The cluster that is joining must be at an equal or later microcode level than the primary cluster.
- The joining and primary clusters must have feature code “4017:Grid Enablement” installed.
- If the primary cluster has feature code “5270: 200 K Logical Volume Increment” installed, then the joining cluster must have the same number or more of this feature.
- The joining cluster must have feature code “Enablement - Selective Device Access Enablement” if that Primary code has it installed.
- If the feature code “1035 - 10 Gb Optical LW Connection” is installed in the joining or primary cluster, the network infrastructure must support 10 Gb.
- If the primary cluster has installed feature code “0001 - 6/25 GB Logical Volume Enablement” and the joining cluster is at microcode level 8.32.0.x or later, the joining cluster does not require this feature code.

5.4.2 Overview of configuration steps

This section provides an outline of all required TS7700 grid/cluster configuration steps:

- Configuring the local and remote TS7700 on the primary cluster
- Configuring the local and remote TS7700 on the remote cluster
- Verifying the network configuration and feature codes
- Joining the local grid and cluster system with remote grid and cluster
- Verifying the cluster configuration
- Varying the primary cluster online
- Displaying the final cluster configuration

In our configuration, the name *Cluster A* is used for the primary cluster and *Cluster B* is used for the secondary cluster.

5.4.3 Configuring the local and remote TS7700 on the primary cluster

This section provides the configuration of the local and remote TS7700 on primary cluster Cluster A.

Example 5-24 shows the status of Cluster A before configuration. The cluster must be offline.

Example 5-24 Verify Cluster A is offline

```
root@ClusterA[/home/pfe] q_node
2016.08-29.17:42:51. Node ClusterA (c0f-)
2016.08-29.17:42:53. VE 8.40.1.7  AIX 7100-04-01-1543  MQ 8.0.0.3
2016.08-29.17:42:53. DB2 10.5.0.5  Atape 13.0.4.0  Atldd 6.8.8.0  GPFS 4.1.1.2
2016.08-29.17:42:53. ------------------------------------------------------------
2016.08-29.17:42:53. c0*  ClusterA h0-OFFline  v0-OFFline
2016.08-29.17:42:53. ------------------------------------------------------------
2016.08-29.17:42:53. Composite_lib_sequence-BA082   name-
2016.08-29.17:42:53. Logical volumes   actual=0              licensed=1,000,000
2016.08-29.17:42:53.
2016.08-29.17:42:53. c0f-* ClusterA VEC--8.40.1.7  ram(32GiB)  [disk-only]
2016.08-29.17:42:53.     *             dist_lib_seq-BA82A  serial_num-EFC6V
2016.08-29.17:42:53.     *             Encrypt Dsk-disabled
2016.08-29.17:42:53.     *             Cache[CSA]  [tot/usd](93.96TB/0%)
2016.08-29.17:42:53.     *             Drives  Virt(256)
```

Example 5-25 shows the configuration of the local and remote TS7700 grid network on the primary node Cluster A.

Example 5-25 Cluster A: Configure local and remote TS7700 GRID networks

```
Configure Local and Remote TS7700 GRID Network IPs (IPv4)

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

[TOP]                                                  [Entry Fields]
  Local Cluster ID:                                  0
  ************************************************
  * NOTICE:  The IP addresses to be entered in the *
  * fields below, should have been supplied by the *
  * customer.                                      *
```

```
  ************************************************
* Enter the IP Address for                        [10.11.150.130]
    the Primary Local Interface
* Enter the Network Mask Address for              [255.255.255.0]
    the Primary Local Interface
  Enter the Gateway Address for                   [10.11.150.1]
    the Primary Local Interface

* Enter the IP Address for                        [10.11.151.130]
    the Alternate Local Interface
* Enter the Network Mask Address for              [255.255.255.0]
    the Alternate Local Interface
  Enter the Gateway Address for                   [10.11.151.1]
    the Alternate Local Interface

* Grid Network Speed/Duplex Setting:          Auto_Negotiation
+

 ************************************************
  * NOTICE:  The information to be entered in the  *
  * fields below, should have been supplied by the *
  * customer and are for the remote cluster        *
  * that will be joined to this TS7700             *
  ************************************************
 Enter the Cluster Number for               []
+#
    the Remote TS7700 System
  Enter the IP Address for                        []
    the Remote Cluster Primary Interface
  Enter the IP Address for                        []
    the Remote Cluster Alternate Interface
  Enter the IP Address for                        []
    the Remote Cluster 2nd Primary Interface
  Enter the IP Address for                        []
    the Remote Cluster 2nd Alternate Interface

  ************************************************
  * NOTICE:  It is only necessary to enter either  *
  * the IP Addresses for the Primary and Alternate *
  * Interfaces when only 2 GRID links are present  *
  * on the remote cluster, or enter IP Addresses   *
  * for the Primary, Alternate, 2nd Primary and    *
  * 2nd Alternate Interfaces when 4 GRID links are *
  * present on the remote cluster.                 *
  ************************************************
Enter the Cluster Number for                [0]
+#
    the Remote TS7700 System
  Enter the IP Address for                        [10.12.150.130]
    the Remote Cluster Primary Interface
  Enter the IP Address for                        [10.12.151.130]
    the Remote Cluster Alternate Interface
  Enter the IP Address for                        []
    the Remote Cluster 2nd Primary Interface
  Enter the IP Address for                        []
```

```
    the Remote Cluster 2nd Alternate Interface

  ************************************************
  * NOTICE:  It is only necessary to enter either  *
  * the IP Addresses for the Primary and Alternate *
  * Interfaces when only 2 GRID links are present  *
  * on the remote cluster, or enter IP Addresses   *

  * for the Primary, Alternate, 2nd Primary and    *
  * 2nd Alternate Interfaces when 4 GRID links are *
  * present on the remote cluster.                 *
  ************************************************
```

5.4.4 Configuring the local and remote TS7700 on the remote cluster

This section provides the configuration steps required on remote cluster Cluster B.

Example 5-26 shows the configuration of local and remote TS7700 grid network on the secondary cluster, Cluster B.

Example 5-26 Cluster B: Configure local and remote TS7700 GRID networks

```
Configure Local and Remote TS7700 GRID Network IPs (IPv4)

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

[TOP]                                                 [Entry Fields]
  Local Cluster ID:                                  1
  ************************************************
  * NOTICE:  The IP addresses to be entered in the *
  * fields below, should have been supplied by the *
  * customer.                                      *
  ************************************************
* Enter the IP Address for                          [10.12.150.131]
    the Primary Local Interface
* Enter the Network Mask Address for                [255.255.255.0]
    the Primary Local Interface
  Enter the Gateway Address for                     [10.12.150.1]
    the Primary Local Interface

* Enter the IP Address for                          [10.12.151.131]
    the Alternate Local Interface
* Enter the Network Mask Address for                [255.255.255.0]
    the Alternate Local Interface
  Enter the Gateway Address for                     [10.12.151.1]
    the Alternate Local Interface

* Grid Network Speed/Duplex Setting:           Auto_Negotiation
+

 ************************************************
  * NOTICE:  The information to be entered in the  *
  * fields below, should have been supplied by the *
  * customer and are for the remote cluster        *
```

```
  * that will be joined to this TS7700             *
  ************************************************
 Enter the Cluster Number for                [0]
+#
    the Remote TS7700 System
  Enter the IP Address for                          [10.11.150.130]
    the Remote Cluster Primary Interface
  Enter the IP Address for                          [10.11.151.130]
    the Remote Cluster Alternate Interface
  Enter the IP Address for                          []
    the Remote Cluster 2nd Primary Interface
  Enter the IP Address for                          []
    the Remote Cluster 2nd Alternate Interface

  ************************************************
  * NOTICE:  It is only necessary to enter either  *
  * the IP Addresses for the Primary and Alternate *
  * Interfaces when only 2 GRID links are present  *
  * on the remote cluster, or enter IP Addresses   *

  * for the Primary, Alternate, 2nd Primary and    *
  * 2nd Alternate Interfaces when 4 GRID links are *
  * present on the remote cluster.                 *
  ************************************************
```

5.4.5 Verifying the network configuration and feature codes

This section provides the network configuration verification result of both clusters: Cluster A and Cluster B.

Verifying the network configuration

Example 5-27 shows the result of the network configuration on primary cluster Cluster A.

Example 5-27 Verification on primary cluster Cluster A

```
COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
Verifying 10.11.150.130 is valid.
IP checking for value 10.11.150.130 complete.  A VALID IP was entered.

Verifying 255.255.255.0 is valid.
IP checking for value 255.255.255.0 complete.  A VALID IP was entered.

Verifying 10.11.150.1 is valid.
IP checking for value 10.11.150.1 complete.  A VALID IP was entered.

Verifying 10.11.151.130 is valid.
IP checking for value 10.11.151.130 complete.  A VALID IP was entered.
```

```
Verifying 255.255.255.0 is valid.
IP checking for value 255.255.255.0 complete.  A VALID IP was entered.

Verifying 10.11.151.1 is valid.
IP checking for value 10.11.151.1 complete.  A VALID IP was entered.

Verifying 10.12.150.131 is valid.
IP checking for value 10.12.150.131 complete.  A VALID IP was entered.

[MORE...19]

F1=Help                    F2=Refresh                 F3=Cancel
F6=Command
F8=Image                   F9=Shell                   F10=Exit
/=Find
```

Example 5-28 shows the result of the network configuration on remote cluster Cluster B.

Example 5-28 Verification on remote cluster Cluster B

```
COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
Verifying 10.12.150.131 is valid.
IP checking for value 10.12.150.131 complete.  A VALID IP was entered.

Verifying 255.255.255.0 is valid.
IP checking for value 255.255.255.0 complete.  A VALID IP was entered.

Verifying 10.12.150.1 is valid.
IP checking for value 10.12.150.1 complete.  A VALID IP was entered.

Verifying 10.12.151.131 is valid.
IP checking for value 10.12.151.131 complete.  A VALID IP was entered.

Verifying 255.255.255.0 is valid.
IP checking for value 255.255.255.0 complete.  A VALID IP was entered.

Verifying 10.12.151.1 is valid.
IP checking for value 10.12.151.1 complete.  A VALID IP was entered.

Verifying 10.11.150.130 is valid.
IP checking for value 10.11.150.130 complete.  A VALID IP was entered.
```

Verifying feature codes

This part shows the feature codes installed on Cluster A and Cluster B. Feature code “4015 - Grid Enablement” and “5271 - Selective Device Access Enablement” must be installed.

Example 5-29 shows all feature codes installed on hardware node and virtual node on Cluster A.

Example 5-29 Cluster A: Show feature codes

```
fc_maint
Node Serial Number:        78-EFC6V
Cache Capacity Configured: N/A
Peer-to-Peer Enabled:      Yes
Encryption Enabled  :      N/A
-------------------------------------

Node Features:

Feature Code : 1038 - 10 Gb Optical LW Dual Connection
Feature Code : 1038 - 10 Gb Optical LW Dual Connection
Feature Code : 4015 - Grid Enablement
d07714e3-3124578b-30af0ba1-3ef69365
Feature Code : 1035 - 10 Gb Optical LW Connection
Feature Code : 5277 - External Key Management Certificate
Feature Code : 5272 - Disk Encryption with Local Key Management
d07714e3-3124578b-d4d58105-c910d40e
Feature Code : 4017 - Enable Cluster Cleanup
d07714e3-3124578b-9b7a3d03-3902422f


VNode:

Node Serial Number:        78-EFC6V
Vnode Throughput:          1100 MB/s
Virtual Drives Enabled:    N/A
Peer-to-Peer Enabled:      Yes
-------------------------------------

Node Features:

Feature Code : 1038 - 10 Gb Optical LW Dual Connection
Feature Code : 1038 - 10 Gb Optical LW Dual Connection
Feature Code : 9000 - Mainframe Attachment
Feature Code : 3439 - 8Gb FICON Long Wavelength Attachment
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-7baf4e24-d52c42b6
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-199caf31-39a175bb
Feature Code : 4015 - Grid Enablement
d07714e3-3124578b-30af0ba1-3ef69365
Feature Code : 3438 - 8Gb FICON Short Wavelength Attachment
Feature Code : 3438 - 8Gb FICON Short Wavelength Attachment
Feature Code : 3401 - 8Gb FICON Dual Port Enablement
d07714e3-3124578b-7decef3e-38352c4f
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-d868ddab-c289aecc
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-c64f54e0-3aed3115
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-1bbc2f78-04646e31
```

```
Feature Code : 1035 - 10 Gb Optical LW Connection
Feature Code : 5277 - External Key Management Certificate
Feature Code : 5271 - Selective Device Access Enablement
d07714e3-3124578b-d3492fed-6cee30c5
Feature Code : 5272 - Disk Encryption with Local Key Management
d07714e3-3124578b-d4d58105-c910d40e
Feature Code : 5271 - Selective Device Access Enablement
d07714e3-3124578b-cff41e59-da1a9c4c
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-c4ac41e0-604f009f
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-fbfc8921-7137a0b4
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-0531b340-0bc10a68
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-30de3874-08962599
Feature Code : 5268 - 100 MB/s Increment
d07714e3-3124578b-789a4be6-90506eb5
Feature Code : 9268 - 100 MB/s Throughput
d07714e3-3124578b-308d0169-b13492a
```

Example 5-30 shows all feature codes installed on the hardware node and virtual node on Cluster B.

Example 5-30 Cluster B: Show feature codes

```
root@ClusterB[/home/pfe] fc_maint
HNode:

Node Serial Number:        78-EFC2V
Cache Capacity Configured: N/A
Peer-to-Peer Enabled:      Yes
Encryption Enabled  :      N/A
-------------------------------------

Node Features:

Feature Code : 4015 - Grid Enablement
eccb135a-a6378c01-30af0ba1-3ef69365
Feature Code : 9277 - External Key Management Certificate
Feature Code : 5273 - Enable TS7720 Tape Attachment
eccb135a-a6378c01-2815ad21-1d114915
Feature Code : 5274 - Enable 1 TB Pending Tape Capacity
eccb135a-a6378c01-868084b9-09dd6dd8
Feature Code : 9900 - Tape Encryption Configuration
eccb135a-a6378c01-7df0f48d-aec7f400
Feature Code : 5276 - Disk Encryption with External Key Management
eccb135a-a6378c01-2be20834-a2ca351f
Feature Code : 1037 - 1 Gb Optical Dual Connection
Feature Code : 1037 - 1 Gb Optical Dual Connection
Feature Code : 5274 - Enable 1 TB Pending Tape Capacity
eccb135a-a6378c01-849de285-b5767f26
Feature Code : 5274 - Enable 1 TB Pending Tape Capacity
eccb135a-a6378c01-5c0ed34e-eddc065c
Feature Code : 1035 - 10 Gb Optical LW Connection
```

```
VNode:

Node Serial Number:        78-EFC2V
Vnode Throughput:          1700 MB/s
Virtual Drives Enabled:    N/A
Peer-to-Peer Enabled:      Yes
-------------------------------------

Node Features:

Feature Code : 9000 - Mainframe Attachment
Feature Code : 9268 - 100 MB/s Throughput
eccb135a-a6378c01-308d0169-b13492a8
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-7baf4e24-d52c42b6
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-199caf31-39a175bb
Feature Code : 4015 - Grid Enablement
eccb135a-a6378c01-30af0ba1-3ef69365
Feature Code : 9277 - External Key Management Certificate
Feature Code : 5273 - Enable TS7720 Tape Attachment
eccb135a-a6378c01-2815ad21-1d114915
Feature Code : 3438 - 8Gb FICON Short Wavelength Attachment
Feature Code : 3438 - 8Gb FICON Short Wavelength Attachment
Feature Code : 3401 - 8Gb FICON Dual Port Enablement
eccb135a-a6378c01-7decef3e-38352c4f
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-d868ddab-c289aecc
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-c64f54e0-3aed3115
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-1bbc2f78-04646e31
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-c4ac41e0-604f009f
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-f7697e79-873f997c
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-fbfc8921-7137a0b4
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-d393bb33-aeeccbae
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-0531b340-0bc10a68
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-30de3874-08962599
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-789a4be6-90506eb5
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-28e67a9c-ccc2e189
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-eb670f26-0e78d145
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-1c750481-8483d4ad
Feature Code : 5268 - 100 MB/s Increment
eccb135a-a6378c01-1f297b3a-21e3898c
```

```
Feature Code : 5271 - Selective Device Access Enablement
eccb135a-a6378c01-d3492fed-6cee30c5
Feature Code : 5271 - Selective Device Access Enablement
eccb135a-a6378c01-cff41e59-da1a9c4c
Feature Code : 5274 - Enable 1 TB Pending Tape Capacity
eccb135a-a6378c01-868084b9-09dd6dd8
Feature Code : 1035 - 10 Gb Optical LW Connection
```

5.4.6 Joining the local grid and cluster system with remote grid and cluster

This section shows the join of primary cluster Cluster A with secondary cluster Cluster B.

Example 5-31 shows the required configuration steps for cluster build on local cluster Cluster A with remote cluster Cluster B.

Example 5-31 Cluster A: Join local grid/cluster system with remote grid/cluster

```
TS7700 GRID Configuration Menus

Move cursor to desired item and press Enter.

  Display and Test Current TS7700 GRID Configuration
  Configure Local and Remote TS7700 GRID Network IPs
  Configure Local TS7700 Cluster with Remote Cluster Grid Network IPs
  Configure IP Security (IPsec) on the Grid Network Menus
  Test TS7700 GRID Communication
  Synchronize Time Between Clusters
  Join Local Grid/Cluster Pre-Checks
  Join Local Grid/Cluster System with Remote Grid/Cluster
  Verify Peer-to-Peer Heartbeat
  Remove a Cluster from a Grid
  Autonomic Ownership Takeover Manager Menus
  Renumber Clusters in the GRID
  Reset Grid Network Routes
  Display Cluster/Grid Configuration Information (q_node)



tlqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqk
  Reset Grid Netx                    Is the Remote Site Online?
x
  Display Clustex
x
             x Move cursor to desired item and press Enter.
x
              x
x
             x   # WARNING: In order to join this system with a remote
x
             x   #  site, the remote site must be Online.
x
             x   #  AN EXTENDED CUSTOMER OUTAGE COULD OCCUR
x
             x   #  IF THIS REQUIREMENT IS NOT MET.
x
```

```
              x
x
             x   #Is the Remote Site Online?
x
              x   Yes
x
              x   No
x
              x
x
             x F1=Help              F2=Refresh            F3=Cancel
x
             x F8=Image             F10=Exit              Enter=Do
x
F1=Help        x /=Find              n=Find Next
x
F9=Shell
mqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq


Join Local Grid/Cluster System with Remote Grid/Cluster

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                       [Entry Fields]
  Local Cluster ID:                                   0
  **************************************************
  * NOTICE:  The information to be entered in the  *
  * fields below, is for the remote cluster        *
  * that you are attempting to join this TS7700 to *
  **************************************************
* Enter the Cluster Number for                [1]
+#
    the Remote TS7700 Cluster


COMMAND STATUS

Command: running       stdout: yes          stderr: no

Before command completion, additional instructions may appear below.

[MORE...12]
2016.08-29.18:18:17. - MSG: Test port 1416 to remote cluster - PASS
2016.08-29.18:18:17. -
2016.08-29.18:18:17. - MSG: Checking ntp server on remote system.
2016.08-29.18:18:17. - MSG: Test ntp server for primary ip - PASS
2016.08-29.18:18:17. - MSG: Test ntp server for alternate ip - PASS
2016.08-29.18:18:17. - MSG: test for checkpoint restart.
2016.08-29.18:18:20. - CHECKPOINT:0  - Complete.
2016.08-29.18:18:20. - MSG: Join pre-checks.
2016.08-29.18:18:24. - CHECKPOINT:1  - Complete.
2016.08-29.18:18:24. - MSG: Feature code 4015 installed on local cluster.
```

```
2016.08-29.18:18:24. - MSG: Feature code 4015 installed on remote cluster.
2016.08-29.18:18:25. - MSG: Remote cluster installed 0 feature code 5270.
2016.08-29.18:18:25. - MSG: Local cluster installed 0 feature code 5270.
2016.08-29.18:18:25. - CHECKPOINT:2  - Complete.
2016.08-29.18:18:25. - MSG: Local and remote composite_lib_sequence numbers match.
BA082
2016.08-29.18:18:26. - MSG: No touch file is found, a refresh join process.
2016.08-29.18:18:26. - CHECKPOINT:3  - Complete.
2016.08-29.18:18:26. - CHECKPOINT:4  - Complete.
2016.08-29.18:18:26. - MSG: The local cluster list (0) is open in the remote
cluster list (1)
2016.08-29.18:18:27. - CHECKPOINT:5  - Complete.
2016.08-29.18:18:29. - MSG: Local cluster 0 is offline.
2016.08-29.18:18:29. - MSG: Checking file systems on cluster 0.
```

5.4.7 Verifying the cluster configuration

Example 5-32 shows the result of the join cluster configuration on primary cluster Cluster A.

Example 5-32 Cluster A: Verify join cluster process completed

```
COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
Beginning Join process...
Issuing command: /usr/vtd/scripts/vtd_domainConfig -j -R1:0
2016.08-29.18:18:13. - MSG : db2profile, rc = 0

2016.08-29.18:18:14. - MSG: Checking remote cluster.
2016.08-29.18:18:16. - MSG: Ping 10.11.151.1 - PASS
2016.08-29.18:18:16. - MSG: Ping 10.12.151.131 - PASS
2016.08-29.18:18:16. - MSG: Ping 10.11.150.1 - PASS
2016.08-29.18:18:16. - MSG: Ping 10.12.150.131 - PASS
2016.08-29.18:18:16. -
2016.08-29.18:18:16. - MSG: Checking port 1415 and 1416
2016.08-29.18:18:16. - MSG: Test port 1415 to remote cluster - PASS
2016.08-29.18:18:17. - MSG: Test port 1416 to remote cluster - PASS
2016.08-29.18:18:17. -
2016.08-29.18:18:17. - MSG: Checking ntp server on remote system.
2016.08-29.18:18:17. - MSG: Test ntp server for primary ip - PASS
2016.08-29.18:18:17. - MSG: Test ntp server for alternate ip - PASS

2016.08-29.18:18:17. - MSG: test for checkpoint restart.
2016.08-29.18:18:20. - CHECKPOINT:0  - Complete.
2016.08-29.18:18:20. - MSG: Join pre-checks.
[MORE...178]

F1=Help                    F2=Refresh                F3=Cancel
F6=Command
```

```
F8=Image                    F9=Shell                    F10=Exit
/=Find
n=Find Next
```

5.4.8 Varying the primary cluster online

After completing cluster configuration, Cluster A must be varied on, as shown in Example 5-33.

Example 5-33 Vary on Cluster A

```
Sending ClusterA online by SMIT
                                        IBM TS7700 Maintenance

Move cursor to desired item and press Enter.

  System Checkout Menus
  SIM, Error Log, Diagnostics, Trace/Dump Menus
  Utility Menus
  Microcode Maintenance Menus
  Subsystem Configuration Menus
  Removal/Replacement Menus
  3957-Vxx Online/Offline Menus
  3957-Vxx Shutdown and Restart Menus
  PFE/Support Tools Menus
  TS7700 GRID Menus
[BOTTOM]
                                          3957-Vxx Online/Offline Menus

Move cursor to desired item and press Enter.
  Display Current Status of 3957-Vxx
  Display Current Code Component Status (LOCAL & REMOTE Clusters)
  Vary 3957-Vxx Online
  Vary 3957-Vxx Offline
  Force a 3957-Vxx Offline
  Online/Offline Utility Menus

 COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

[TOP]
It may take several minutes before the system transitions
to the Online State.

Onlining local cluster...
BEG : Script(online) Function(online) Node(Gnode) Mon 08/29/16 18:52:05
MSG : verify vpd can be accessed
MSG : verify the minimum required feature codes are installed
MSG : check requirements IF this is a 7760 with tape attached
MSG : start websphere if needed
MSG : verify Websphere MQ successfully started
MSG : verify required processes are running
```

```
MSG : starting execution of fs_notify
MSG : Verifying that the filesystem is configured properly
MSG : start pdsd if it is not running
MSG : start micd if it is not running
MSG : start the node processes
MSG : Starting the AIX error daemon
MSG : verify message versions on nodes in the domain are compatible
MSG : verify a formatted DVD-RAM is available for mounting
MSG : Failed to detect a formatted backup DVD-ROM in drive. SIM Posted
MSG : start the Vnode
[MORE...10]

F1=Help                    F2=Refresh                 F3=Cancel
F6=Command
F8=Image                   F9=Shell                   F10=Exit
/=Find
```

5.4.9 Displaying the final cluster configuration

Example 5-34 shows the final cluster configuration settings of Cluster A and Cluster B.

Example 5-34 Show cluster configuration on Cluster A and Cluster B

```
2016.08-29.18:59:25. Node ClusterA (c0f- of grid c0,c1)
2016.08-29.18:59:27. VE 8.40.1.7  AIX 7100-04-01-1543  MQ 8.0.0.3
2016.08-29.18:59:27. DB2 10.5.0.5  Atape 13.0.4.0  Atldd 6.8.8.0  GPFS 4.1.1.2
2016.08-29.18:59:27. ------------------------------------------------------------
2016.08-29.18:59:27. c0*  ClusterA    h0-ONline   v0-ONline
2016.08-29.18:59:27. c1   ClusterB h0-ONline   v0-ONline
2016.08-29.18:59:27. ------------------------------------------------------------
2016.08-29.18:59:27. Composite_lib_sequence-BA082   name-
2016.08-29.18:59:27. Logical volumes   actual=0              licensed=1,000,000
2016.08-29.18:59:27.
2016.08-29.18:59:27. c0f-* ClusterA     VEC--8.40.1.7  ram(32GiB)  [disk-only]
2016.08-29.18:59:27.     *             dist_lib_seq-BA82A  serial_num-EFC6V
2016.08-29.18:59:27.     *             Encrypt Dsk-disabled
2016.08-29.18:59:27.     *             Cache[CSA]  [tot/usd](93.96TB/0%)
2016.08-29.18:59:27.     *             Drives  Virt(256)
2016.08-29.18:59:27.     *  GRIDip                           P1( 1Gb)  A1( 1Gb)
2016.08-29.18:59:27.     *             P1/A1      10.11.150.130/10.11.151.130
2016.08-29.18:59:27.
2016.08-29.18:59:27. c1f-  ClusterB  VEC--8.40.1.7  ram(32GiB)  [disk-only]
2016.08-29.18:59:27.                   dist_lib_seq-BA82B  serial_num-EFC2V
2016.08-29.18:59:27.                   Encrypt Dsk-disabled
2016.08-29.18:59:27.                   Cache[CSA]  [tot/usd](62.53TB/0%)
2016.08-29.18:59:27.                   Drives  Virt(256)
2016.08-29.18:59:28.        GRIDip                           P1( 1Gb)  A1( 1Gb)
2016.08-29.18:59:28.                   P1/A1      10.12.150.131/10.12.151.131
```

6

FCIP and integrated routing

This chapter provides the fundamental steps for using FCIP replication in virtual fabrics and the integrated routing function.

It provides the following information:

- Routing in virtual fabrics
- Implementing XISL
- Implementing the Integrated Routing concept

6.1 Routing in virtual fabrics

The virtual fabric concept allows you to partition a physical switch into multiple logical switches with their own data, control, and management paths. The purpose of this chapter is to show a possible implementation of Fibre-to-Fibre Channel routing based on an FCIP solution.

For more comprehensive information about configuring virtual fabrics, see *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331.

6.2 Implementing XISL

In this section, we describe the implementation of virtual fabrics based on the XISL concept.

6.2.1 Lab configuration overview - XISL implementation

This section describes the implementation of an XISL concept based on a 10 Gigabit Ethernet tunnel. All configuration steps are shown with both the command-line interface (CLI) and the IBM Network Advisor GUI.

This chapter describes all of the required configuration steps for implementing the solution that is shown in Figure 6-1.

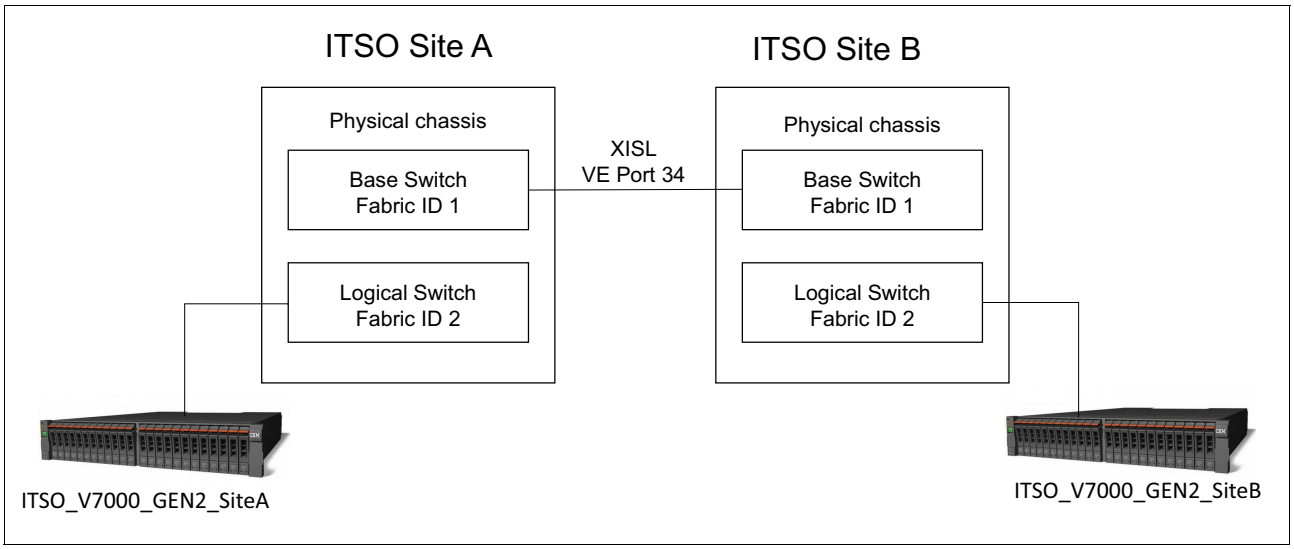


Figure 6-1 Overview of lab configuration

Based on the concept that is described in Figure 6-1, the virtual fabric concept is implemented in the following examples.

CLI

This section shows the implementation steps on the CLI.

1. Ensure that virtual fabric is enabled.

   By default, virtual fabric is enabled. This setting can be verified with the **fosconfig --show** command, as shown in Example 6-1.

   Example 6-1 Verify if virtual fabric is enabled

```
ITSO_7840_SiteA:FID128:admin> fosconfig --show
FC Routing service:             disabled
Virtual Fabric:                 enabled
ITSO_7840_SiteA:FID128:admin>
```

2. Create a virtual switch.

   First, a new logical base switch must be created with fabric ID 1, as shown in Example 6-2.

   Example 6-2 SiteA: Create a base switch and assign a fabric ID 1

```
ITSO_7840_SiteA:FID128:admin> lscfg --create 1 -base
Creation of a base switch requires that the proposed new base switch on this
system be disabled.
Would you like to continue [y/n]?: y
About to create switch with fid=1. Please wait...
Logical Switch with FID (1) has been successfully created.

Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
```

   The logical switch configuration can be verified with the **switchshow** command.

3. Configure a virtual switch.

   Example 6-3 shows the configuration of the domain ID, the switch name, and the IP address. Before you update the fabric parameters, the logical switch must be disabled.

   Example 6-3 Configuration of new virtual switch with fabric ID 1:

```
switch_1:FID1:admin> switchdisable
switch_1:FID1:admin> switchname "ITSO_7840_SiteA_FID1"
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
ITSO_7840_SiteA_FID1:FID1:admin>
ITSO_7840_SiteA_FID1:FID1:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 10
WARNING: The domain ID will be changed. The port level zoning may be affected
ITSO_7840_SiteA_FID1:FID1:admin> switchenable
ITSO_7840_SiteA_FID1:FID1:admin> ipaddrset -ls 1 --add 10.18.228.95/24
IP address is being changed...
ITSO_7840_SiteA_FID1:FID1:admin>
```

4. Assign ports to the base switch.

   VE Port and the IP interfaces must be assigned to the base switch with fabric ID 1 with the **lscfg** command, as shown in Example 6-4.

   Example 6-4 Assign VE Port 24 and IP interfaces ge2,ge3, ge6 and ge7 to virtual switch with FID 1

```
ITSO_7840_SiteA_FID1:FID1:admin>
ITSO_7840_SiteA_FID1:FID1:admin> lscfg --config 1 -port 34
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_FID1:FID1:admin>
ITSO_7840_SiteA_FID1:FID1:admin> lscfg --config 1 -port ge2-3
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_FID1:FID1:admin> lscfg --config 1 -port ge6-7
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_FID1:FID1:admin> switchenable
```

5. Create the IP Interface and IP Tunnel.

   In our example, the VE Ports of Site A and B are direct connected. The tunnel is based on four circuits with two failover groups. See Chapter 4, “FCIP replication” on page 93 for information about creating IP interfaces and IP tunnels with failover groups.

6. Verify the status of the fabric after the tunnel is created.

   After enabling the tunnel on VE Port 34, the base fabric is formed. This configuration can be verified with the **fabricshow** command.

7. Create a virtual switch with fabric ID 2.

   Example 6-5 shows the configuration of the domain ID, switch name, and the IP address of the virtual switch with fabric ID 2. Before updating fabric parameters, the logical switch must be disabled.

   Example 6-5 Configure logical switch with fabric ID 2 on each site

```
ITSO_7840_SiteA:FID128:admin> lscfg --create 2
A Logical switch with FID 2 will be created with default configuration.
Would you like to continue [y/n]?: y
About to create switch with fid=1. Please wait...
Logical Switch with FID (1) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
ITSO_7840_SiteA:FID128:admin> setcontext 2
switch_2:FID2:admin> switchdisable
switch_2:FID2:admin> switchname "ITSO_7840_SiteA_FID2"
Done.
```

```
Switch name has been changed.Please re-login into the switch for the change to
be applied.
ITSO_7840_SiteA_FID2:FID2:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 11
WARNING: The domain ID will be changed. The port level zoning may be affected
ITSO_7840_SiteA_FID2:FID2:admin> ipaddrset -ls 2 --add 10.18.228.97/24
ITSO_7840_SiteA_FID2:FID2:admin>
```

8. Assign physical ports to the logical switch.

   Ports connected to ITSO_V7000_GEN2_SIteA must be assigned to logical switch ITSO_7840_SiteA_FID2. Ports connected to ITSO_V7000_GEN2_SIteB must be assigned to logical switch ITSO_7840_SiteB_FID2, as shown in Example 6-6.

   Example 6-6 Assign port 0-3 and 8-11

```
ITSO_7840_SiteA_FID2:FID2:admin> lscfg --config 2 -port 0-3
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_FID2:FID2:admin>
ITSO_7840_SiteA_FID2:FID2:admin> lscfg --config 2 -port 8-11
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
```

   After moving ports to logical switch, ports must be enabled with the **portcfgpersistentenable** command.

9. Verify the switch settings.

   The status of the ports can be verified with the **switchshow** command, as shown in Example 6-7.

   Example 6-7 Verification

```
ITSO_7840_SiteA_FID2:FID2:admin> switchshow
switchName:     ITSO_7840_SiteA_FID2
switchType:     148.0
switchState:    Online
switchMode:     Native
switchRole:     Principal
switchDomain:   11
switchId:       fffc0b
switchWwn:      10:00:50:eb:1a:d7:83:82
zoning:         OFF
switchBeacon:   OFF
FC Router:      OFF
HIF Mode:       OFF
Allow XISL Use: ON
LS Attributes:  [FID: 2, Base Switch: No, Default Switch: No, Address Mode 0]
```

```
Index Port Address  Media Speed   State       Proto
==================================================
   0   0   0b0000   id    N8      Online      FC  F-Port
50:05:07:68:0b:21:21:7b
   1   1   0b0100   id    N8      Online      FC  F-Port
50:05:07:68:0b:22:21:7b
   2   2   0b0200   id    N8      Online      FC  F-Port
50:05:07:68:0b:23:21:7b
   3   3   0b0300   id    N8      Online      FC  F-Port
50:05:07:68:0b:21:21:7a
   8   8   0b0400   id    N8      Online      FC  F-Port
50:05:07:68:0b:22:21:7a
   9   9   0b0500   id    N8      Online      FC  F-Port
50:05:07:68:0b:23:21:7a
  10  10   0b0600   id    N8      Online      FC  F-Port
50:05:07:68:0b:24:21:7a
  11  11   0b0700   id    N8      Online      FC  F-Port
50:05:07:68:0b:24:21:7b
  45  45   ------   --    --       Online      FC  E-Port
10:00:50:eb:1a:36:1d:3a "ITSO_7840_SiteB_FID2" (downstream)
ITSO_7840_SiteA_FID2:FID2:admin>
```

10. Verify that the fabric with fabric ID 2 is built correctly for the virtual switches.

    Example 6-8 shows that a fabric between the logical switches ITSO_7840_SiteA_FID2 and ITSO_7840_SiteA_FID2 was created.

Example 6-8 fabricshow

```
ITSO_7840_SiteA_FID2:FID2:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-------------------------------------------------------------------------
 11: fffc0b 10:00:50:eb:1a:d7:83:82 10.18.228.217   10.18.228.97
>"ITSO_7840_SiteA_FID2"
 21: fffc15 10:00:50:eb:1a:36:1d:3a 10.18.228.216   10.18.228.98
"ITSO_7840_SiteB_FID2"

The Fabric has 2 switches

ITSO_7840_SiteA_FID2:FID2:admin>
```

IBM Network Advisor GUI

The following procedure is how to enable the Virtual Fabric feature and configure logical switch for FCIP connection by using IBM Network Advisor GUI:

1. Enable Virtual Fabrics in each physical chassis. Before logical switch creation, enable Virtual Fabrics on the physical chassis.

   To enable this feature or check the status, select the physical chassis that you want to enable this feature on, and click **Configure** → **Virtual Fabric** → **Enable** (Figure 6-2). Here this feature has been enabled by default.

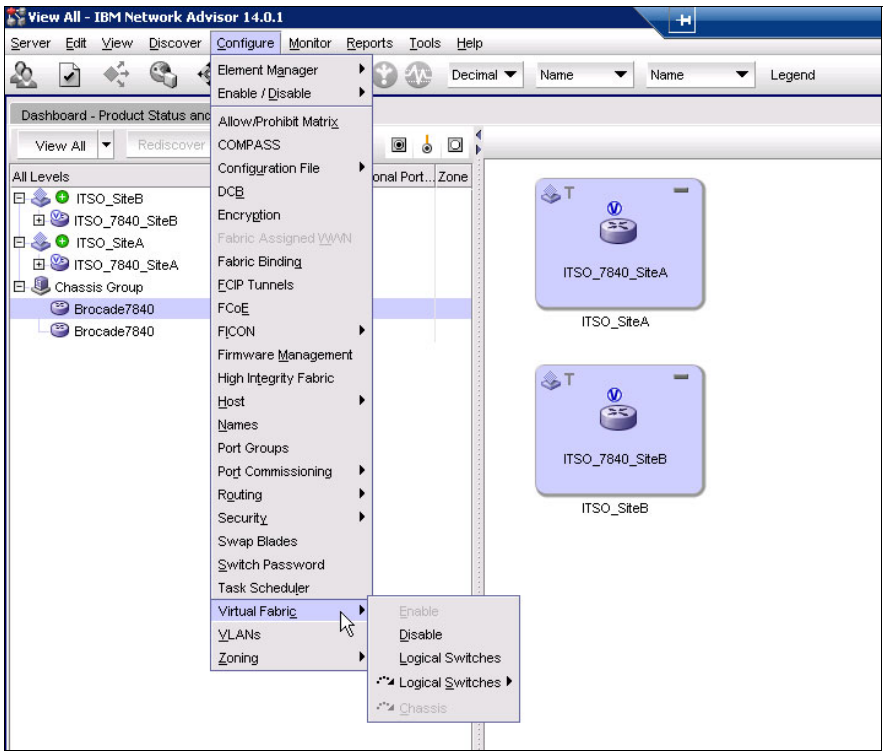


Figure 6-2 Enable Logical Switch feature

2. To create a logical switch, select **Configure → Virtual Fabric** (Figure 6-3).

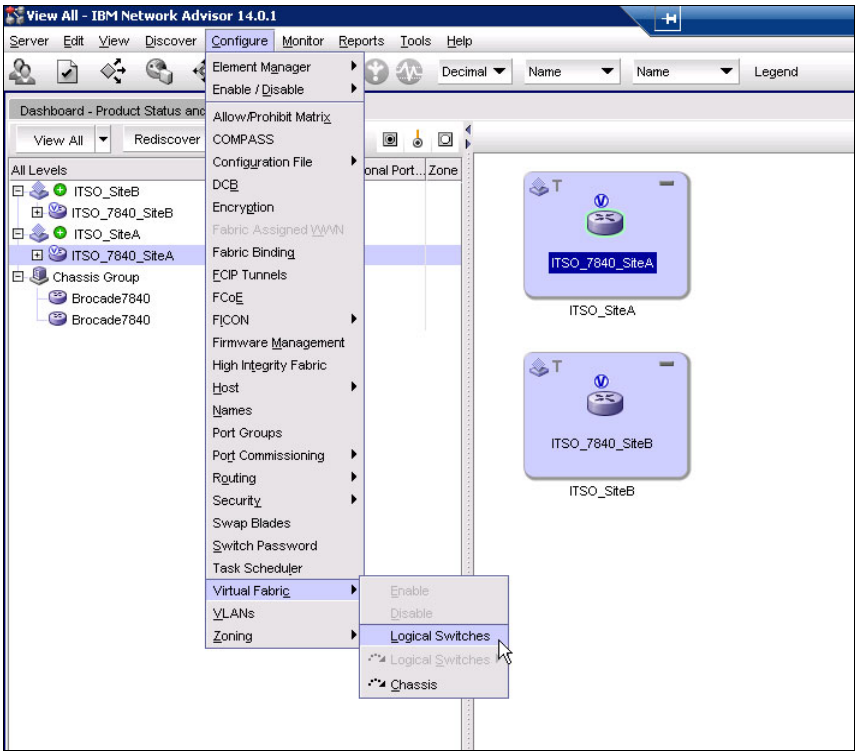


Figure 6-3 Menu for logical switch configuration

3. The Logical Switches window is displayed (Figure 6-4). Select the physical chassis **IBM42B-R_SiteA** in the **Chassis** list box.

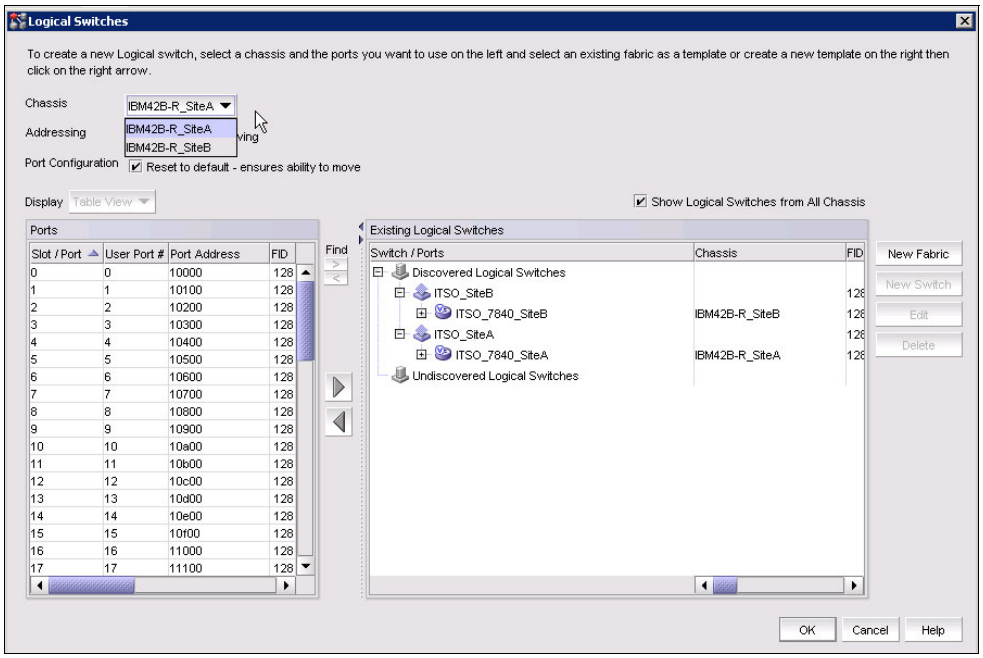


Figure 6-4 Select chassis for logical switch configuration

4. Select the physical chassis for which to create the logical switch in the **Existing logical switches** list and click **New Switch**.
5. The New Logical Switch window is displayed (Figure 6-5). Enter a logical fabric ID in the **Logic Fabric ID** field. Here we assign fabric ID 1 for the first logical switch, which is the base switch for the XISL connection.

   To make the logical switch the base switch for the XISL connection, clear the **Base fabric for Transport (XISL)** check box and select **Base switch**.

   Leave other fabric parameters unchanged and accept the default values. For more information about these fabric parameters, see *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331.

Notes:

- The fabric ID uniquely identifies the logical switch within a chassis. You cannot define multiple logical switches with the same FID within the chassis. A logical switch in one chassis can communicate with a logical switch in another chassis only if the two logical switches have the same FID.
- The default logical switch is assigned the FID 128, which cannot be changed.

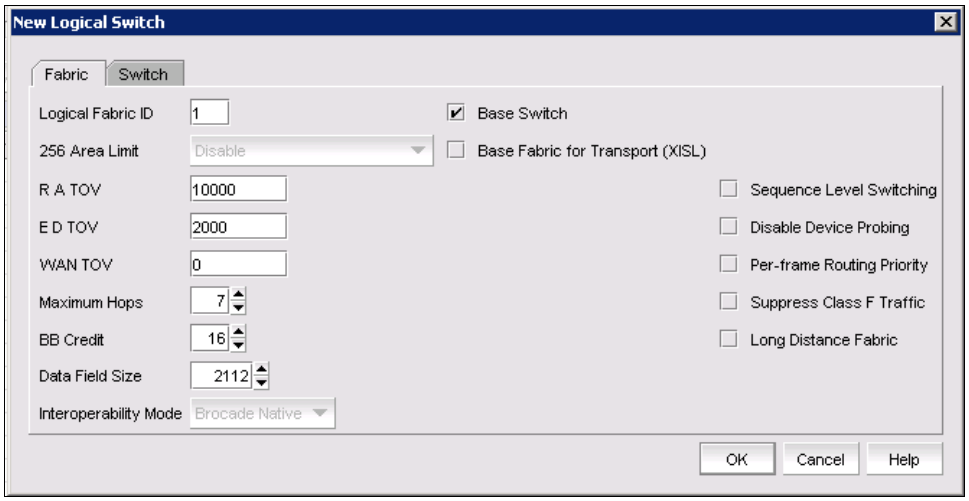


Figure 6-5 Create the base logical switch

6. Click the **Switch** tab and enter a name and domain ID for the base logical switch. Here we assign domain ID 10 for the domain ID and enter the name `ITSO_7840_SiteA_FID1` for the base logical switch (Figure 6-6). The domain ID should be unique to other logical switches within the same fabric.

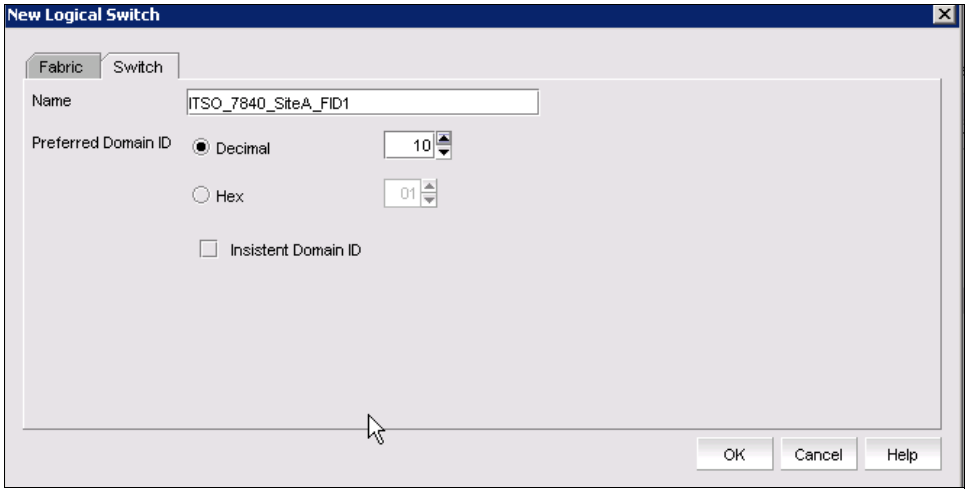


Figure 6-6 Assign a name and Domain ID for the base logical switch

7. Click **OK**. The new logical switch is displayed in the **Existing Logical Switches** list (Figure 6-7).

   The new logical switch has been created without any ports. Ports must be assigned to it. To assign ports, follow these steps:

   a. Select the ports to include in this logical switch from the **Ports** list in the left pane.

   b. Select the base logical switch that was created before in the **Existing logical switches** list.

   c. Click the right arrow button to move the selected ports to the base logical switch.

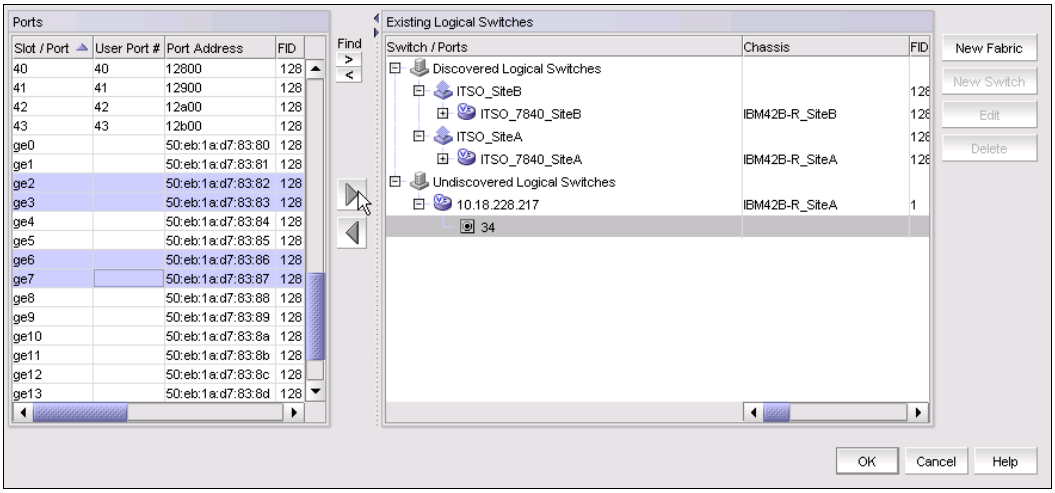


Figure 6-7 Select ports for the base logical switch

8. As shown in Figure 6-8, the following ports are assigned to the base logical switch: ge2, ge3, ge6, ge7, and 34 (VE port). Click **OK** to confirm the setting and complete the task of logical switch creation.

Note: Ports are disabled while they are moved between multiple logical switches.

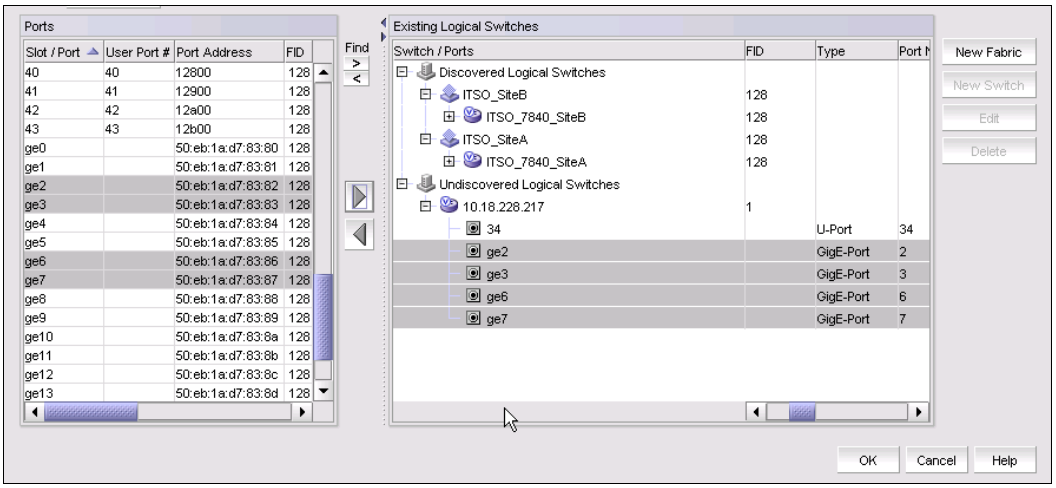


Figure 6-8 Move ports to base logical switch

9. Repeat steps 2 - 8 to create the second logical switch, which is used for the FC connection of the IBM Storwize V7000 storage system.

   Assign the domain ID 11 and enter the name `ITSO_7840_SiteA_FID2` for the second logical switch (Figure 6-9).

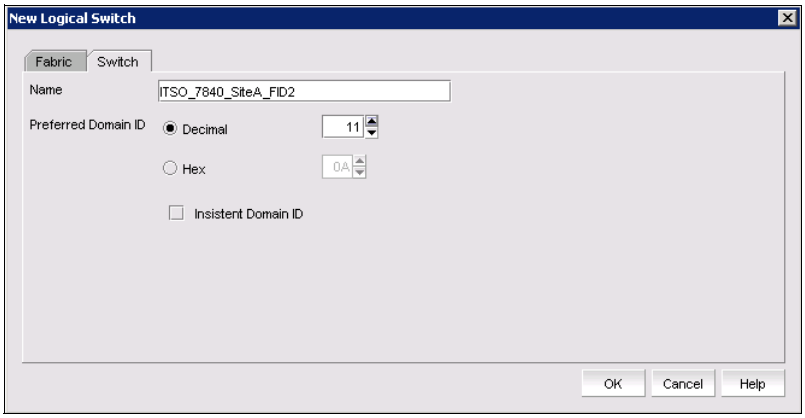


Figure 6-9 Assign name and Domain ID for logical switch ITSO_7840_SiteA_FID2

10. Enter the **Logic Fabric ID** for the second logical switch (Figure 6-10). Here we assign fabric ID 2 for it.

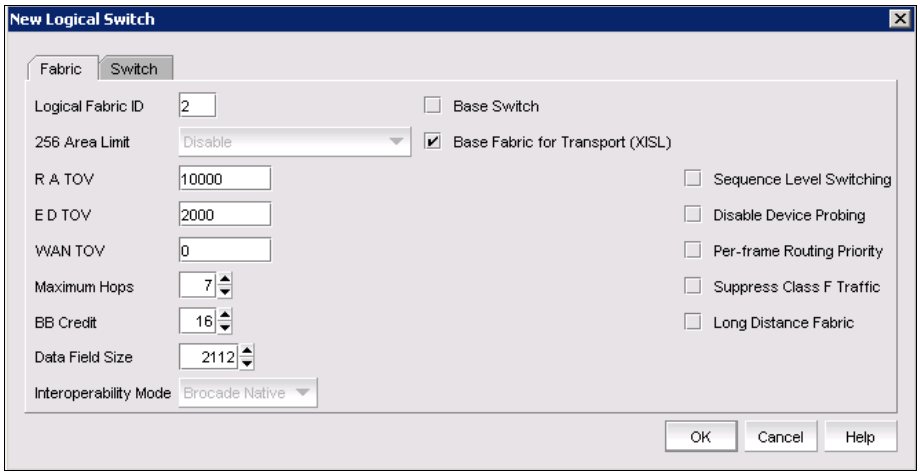


Figure 6-10 Setting FID for logical switch ITSO_7840_SiteA_FID2

11. We select the **Base Fabric for Transport (XISL)** check box because we want to use the XISL from the base switch for the Site B connection. For more information about these fabric parameters, see *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331.

12. Assign ports for the logical switch `ITSO_7840_SiteA_FID2` (Figure 6-11). The FC ports that are connected to V7000 storage system should be assigned to the logical switch. Click **OK** to confirm these configurations. The creation of the logical switch ITSO_7840_SiteA_FID2 is completed.

Note: Ports are disabled while they are moved between multiple logical switches.

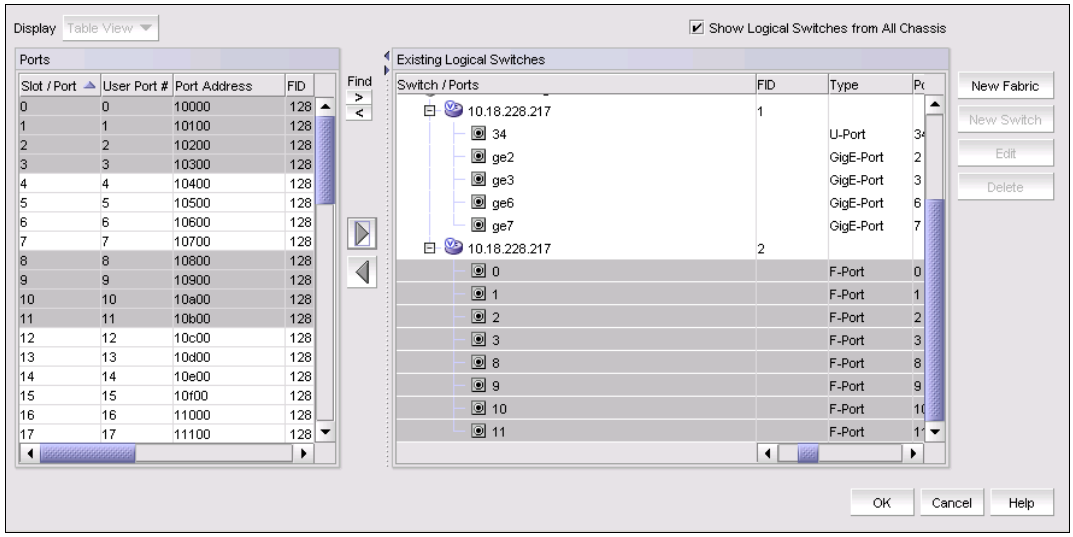


Figure 6-11 Assign ports for logical switch ITSO_7840_SiteA_FID2

13. Repeat steps 2 - 11 to create two logical switches on the chassis of Site B. Make sure that the two logical switches which are created on Site B have the same Fabric ID as Site A.

14. For FCIP configuration on the logical switch with the GUI, discover the fabrics to find the new logical switches. To discover fabrics, select **Discover** → **Fabrics** (Figure 6-12).

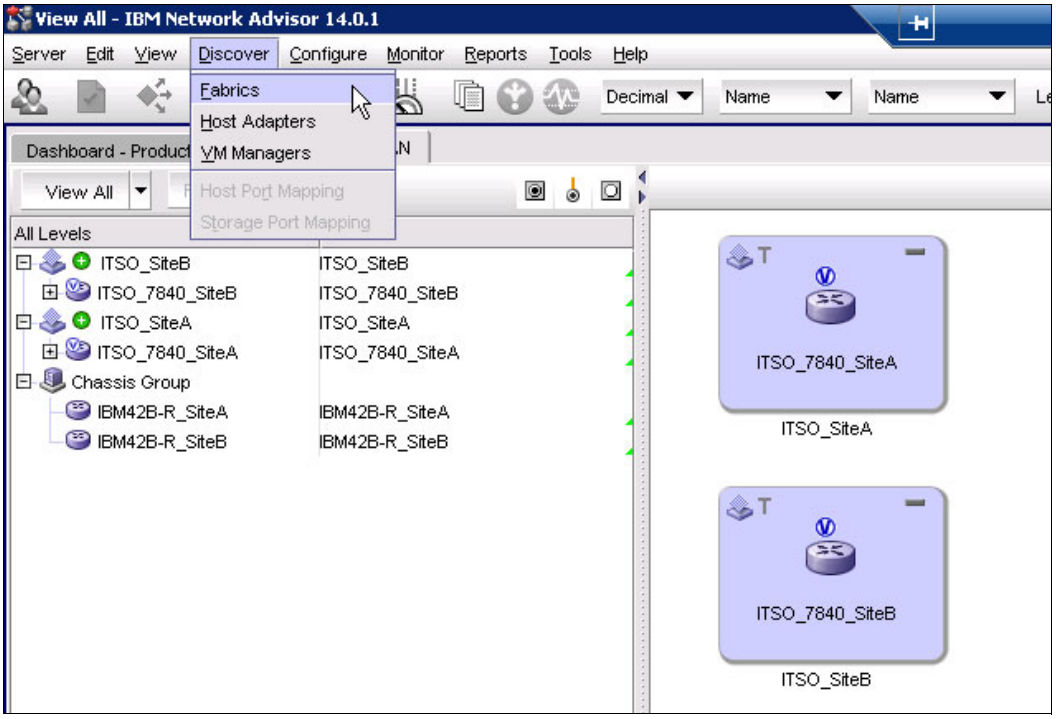


Figure 6-12 Menu for fabric discovery

15. Click **Add** to discover new fabrics (Figure 6-13).

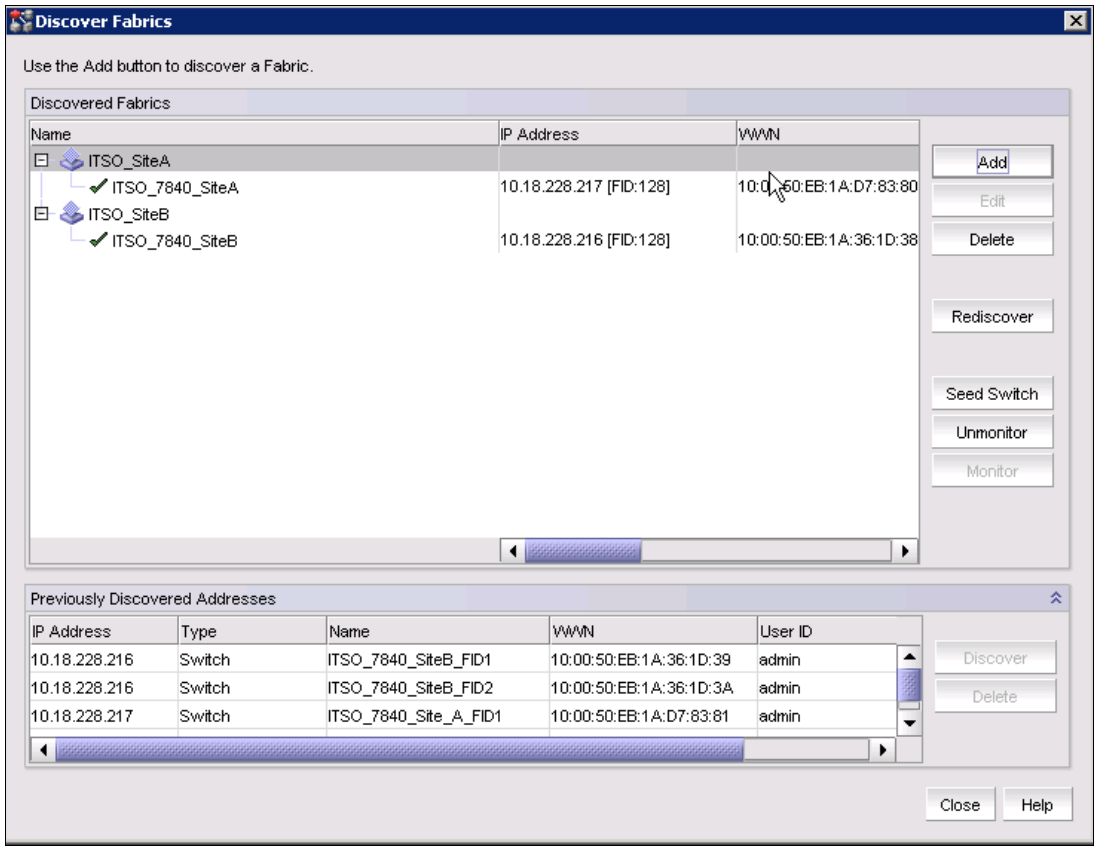


Figure 6-13 Add fabric for new logical switch

16. The Add Fabric Discovery window is displayed. Enter the management IP address and user authentication of the physical chassis (Figure 6-14).

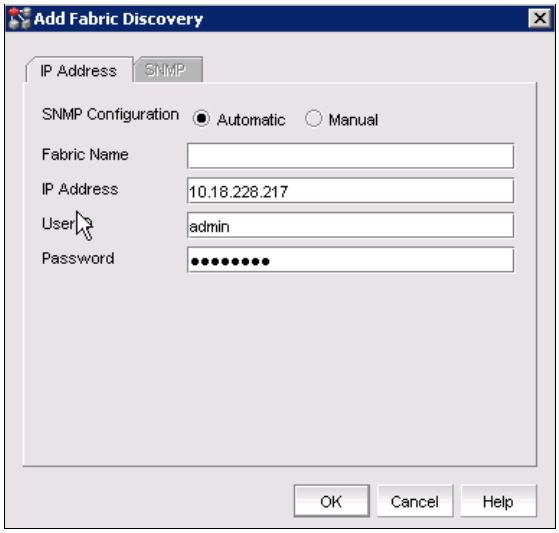


Figure 6-14 Input for fabric discovery

17. The result of fabric discovery is shown in Figure 6-15.

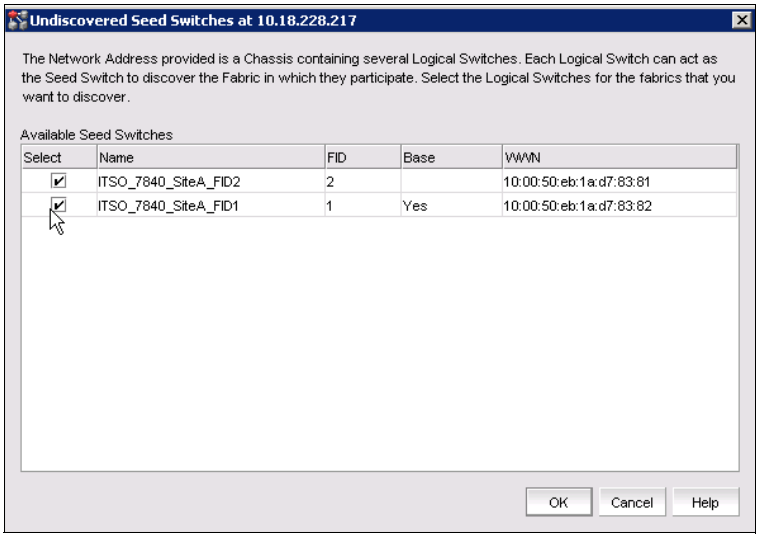


Figure 6-15 Results of fabric discovery

18. Click **OK** to add these fabrics and see the results (Figure 6-16).

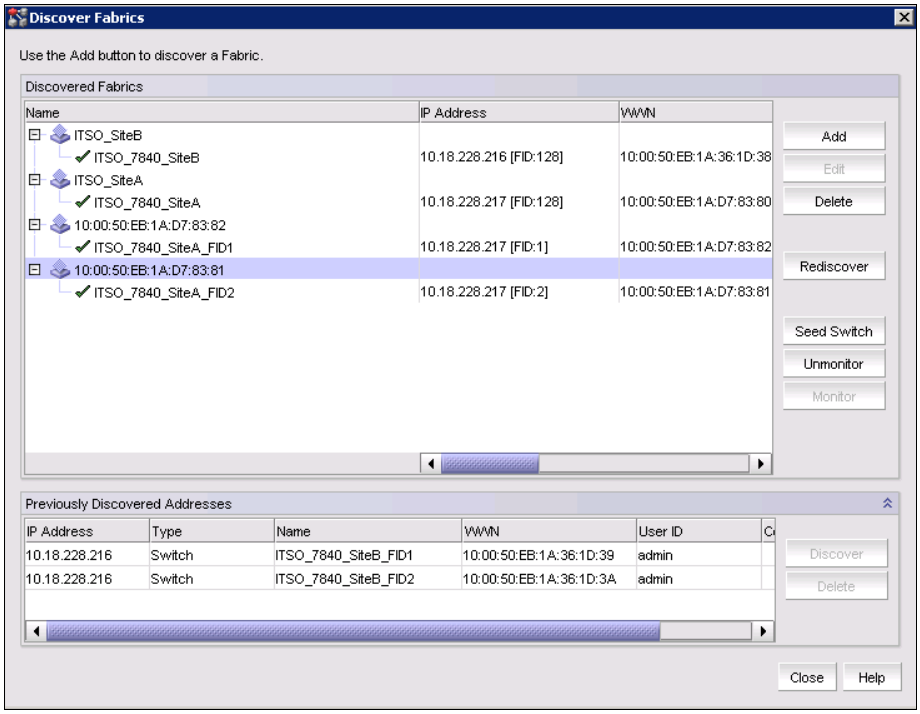


Figure 6-16 Result of adding fabric discovery

19. It is optional to rename the new fabric to a more meaningful name. Figure 6-17 shows how to display information.

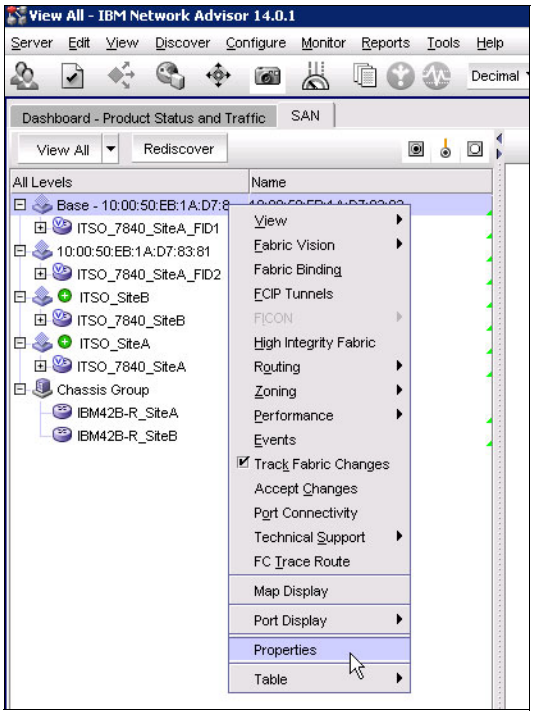


Figure 6-17 Check the properties of the fabric

We rename it to ITSO_SiteA_FID1/2 and ITSO_SiteB_FID1/2. See Figure 6-18.

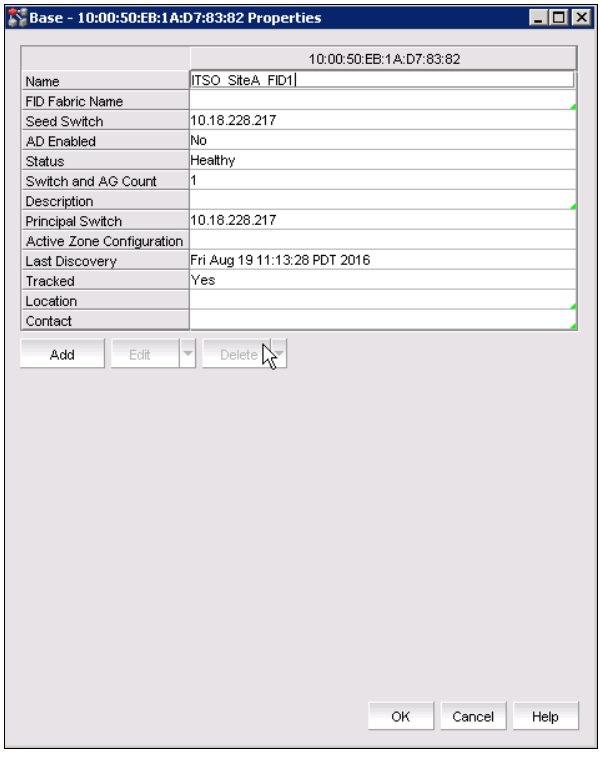


Figure 6-18 Change the name of the fabric

20. Create the FCIP Tunnel for the base logical switches of SiteA and SiteB. See Chapter 4, “FCIP replication” on page 93.

    Figure 6-19 shows the first circuit of the FCIP tunnel between the two base logical switches.

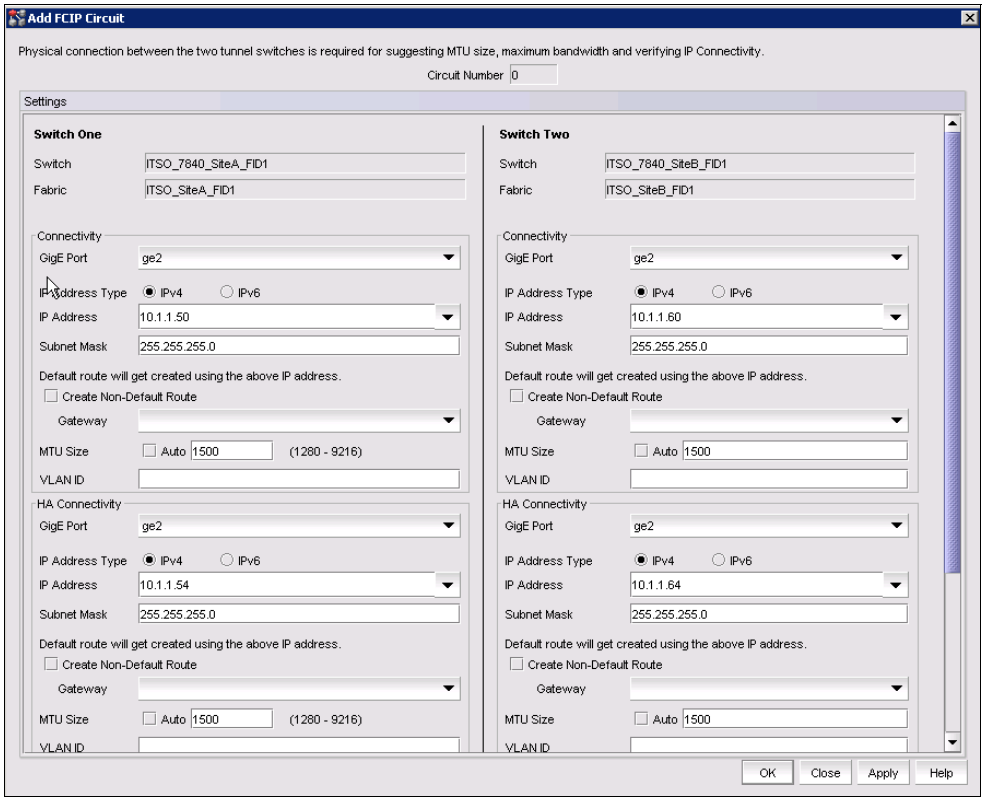


Figure 6-19 Add an FCIP circuit for the logical switch

Figure 6-20 shows the full configuration of the FCIP tunnel between the two base logical switches.

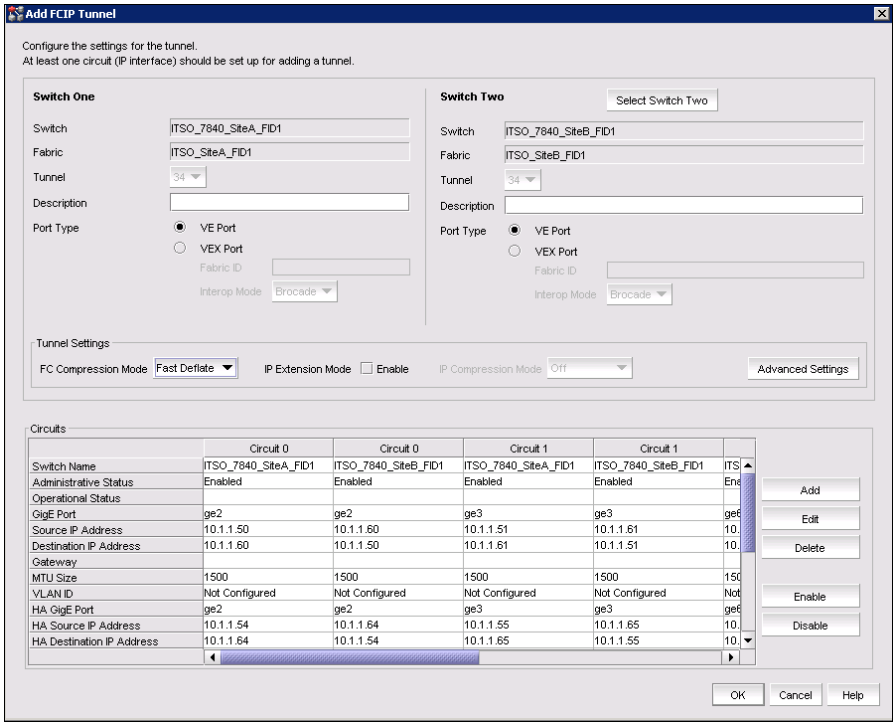


Figure 6-20 Add FCIP tunnel for logical switch

21. After FCIP creation, the status of the XISL connection between logical switches is shown in Figure 6-21.

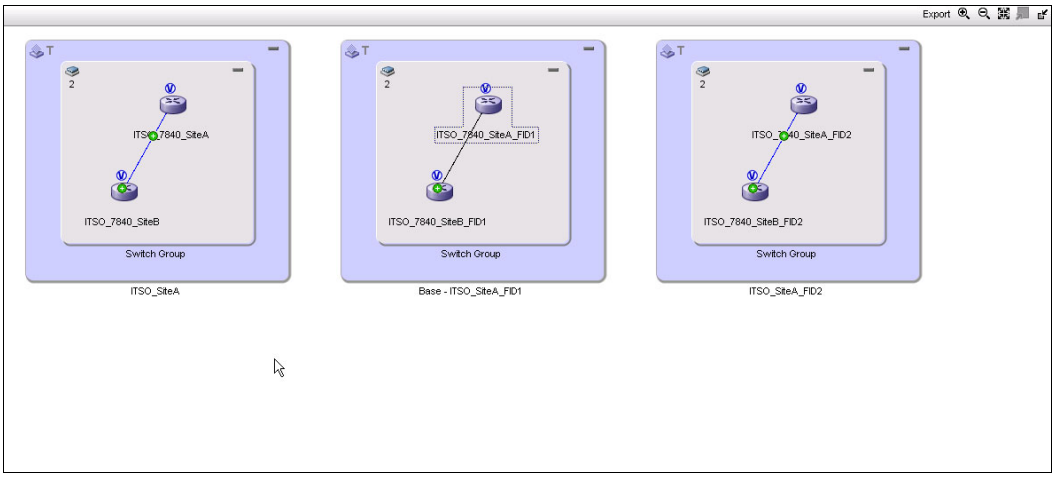


Figure 6-21 Status of the XISL connection

We can also double-click the connection line between two logical switches to view the properties of XISL connections (Figure 6-22).

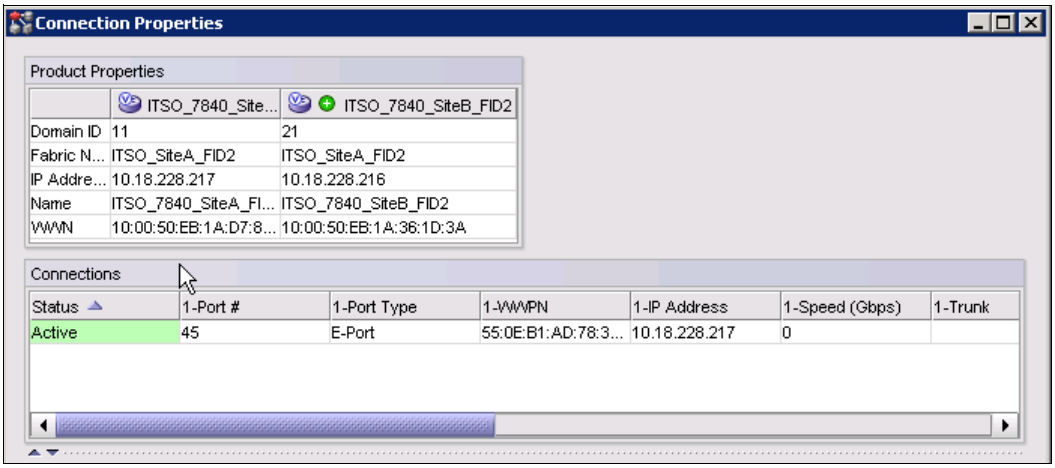


Figure 6-22 Properties of XISL connection

6.2.2 Start storage replication

After the fabric of the virtual switches with the fabric 2 is built, the status of the consistency group displays as “Consistent Synchronized,” which means the data from Site B is consistent and synchronized with the data from Site A (Figure 6-23). For more information about the implementation of IBM Storwize V7000 remote mirroring, see *IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services*, SG24-7574.

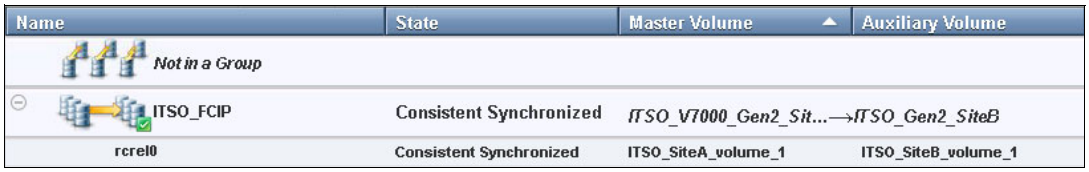


Figure 6-23 Status of V7000 remote replication

6.3 Implementing the Integrated Routing concept

This section describes an implementation of the integrated routing concept. For more information about Fibre Channel routing, see *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331.

6.3.1 Lab configuration: Overview

This section describes the lab example that we use in this chapter for the implementation of an integrated routing solution. In our configuration, there is an IBM SAN Volume Controller stretched cluster within site A and an IBM Storwize V7000 storage system in site B. The main goal is to provide the quorum volume for the IBM SAN Volume Controller stretched cluster system by IBM Storwize V7000 storage system in site B.

Note: For SAN Volume Controller version 7.6 and later, IP quorum applications can be used as a third site to house quorum devices. To use an IP-based quorum application as the quorum device for the third site, no Fibre Channel connectivity is used. For more information about IP quorum, see *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V7.6*, SG24-7933.

In our lab configuration example, the following points must be noted:

- We describe the implementation of an integrated routing solution in one fabric.
- All IBM SAN Volume Controller ports that are dedicated for host and storage traffic are connected to a single physical 8501 SAN switch that is located in site A.
- All IBM Storwize V7000 storage system ports are connected to a single physical 8501 SAN switch located in site B.
- We do not show the implementation of the private SAN that is used for the IBM SAN Volume Controller stretched cluster system.
- We do not discuss IBM best practices for implementing IBM SAN Volume stretched cluster systems.

For more information about IBM SAN Volume Controller stretched cluster, see *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V7.6*, SG24-7933, and *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

Figure 6-24 shows our lab configuration that is used for all implementation steps in this chapter.

Note: The tunnel that is defined on VE port 34 is based on four circuits with two failover groups. Circuits are defined on IP interfaces ge2, ge3, ge6, and ge7.

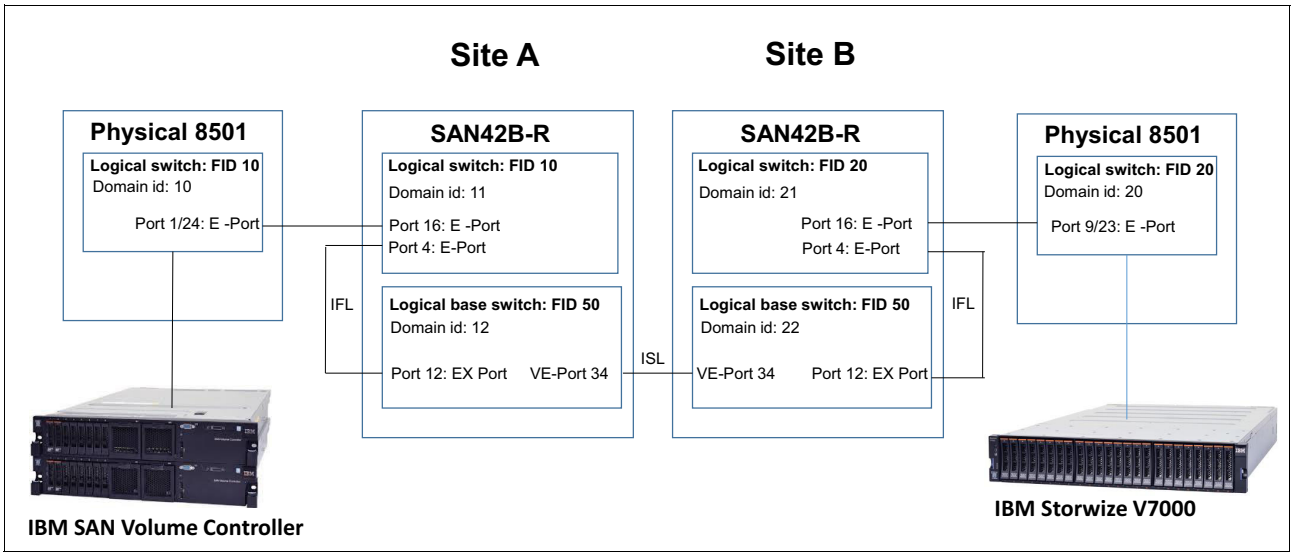


Figure 6-24 Lab configuration overview

6.3.2 Preferred practices

This section gives you a brief overview of an edge-backbone-edge architecture. For more information about FCIP architectures, see Chapter 3, “Extension architectures” on page 45.

Figure 6-25 shows an Edge-Backbone architecture.

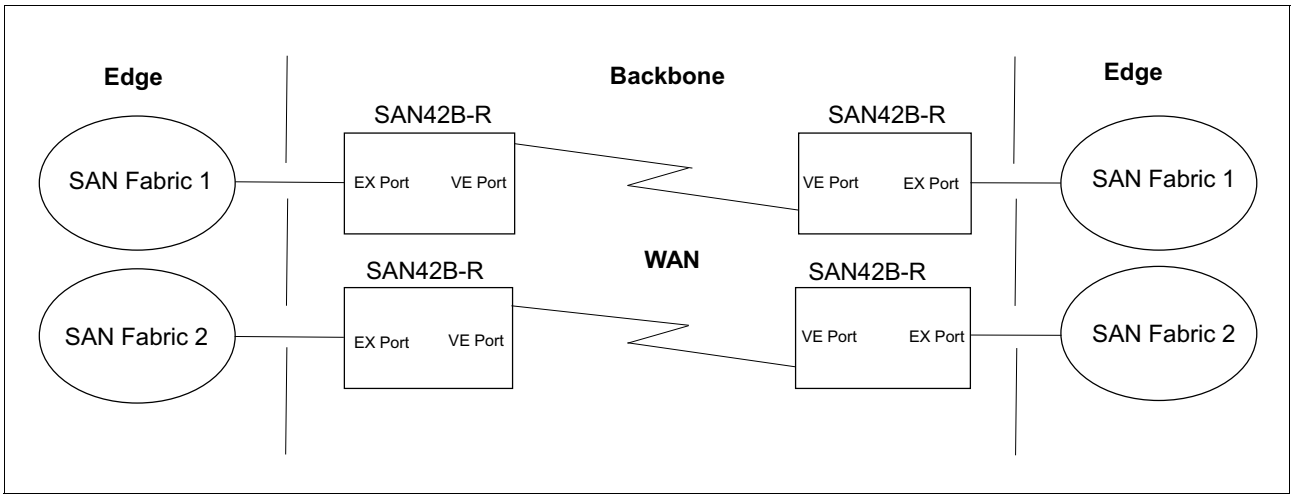


Figure 6-25 Preferred practice for an integrated routing concept is to apply a Edge-Backbone architecture.

6.3.3 Implementation overview

The following configuration tasks are required for implementing an edge-backbone-edge architecture by using EX Ports on FCIP routers:

- Preparing physical connections
- Site A: Configuring the edge fabric
- Site B: Configuring the edge fabric
- Site A: Configuring logical switches on the FCIP router
- Site B: Configuring logical switches on the FCIP router
- Configuring the inter fabric link
- Configuring the tunnel on VE Port 34
- Creating LSAN zones
- Verification
- Providing a quorum from V7000 Site B to the IBM SAN Volume Controller stretched cluster in Site A

6.3.4 Preparing physical connections

An inter fabric link requires an additional physical Fibre Channel cable, which connects the logical base switch with the logical switch that is connected to the edge fabric, as shown in the lab configuration overview in 6.3.1, “Lab configuration: Overview” on page 171.

6.3.5 Site A: Configuring the edge fabric

This section describes the implementation of a logical switch on the edge fabric on site A used for IBM SAN Volume Controller stretched cluster systems public fabric.

Implementation steps with the CLI

Use the following steps to configure the edge fabric with the CLI:

1. Create a logical switch.

   A new logical switch with fabric ID 10 must be created on physical switch 8510 as shown in Example 6-9.

   Example 6-9 Create a new logical switch with FID 10

```
ITSO_8510_SiteA:FID128:admin> lscfg --create 10
A Logical switch with FID 10 will be created with default configuration.
Would you like to continue [y/n]?: y
About to create switch with fid=10. Please wait...
Logical Switch with FID (10) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
ITSO_8510_SiteA:FID128:admin>
```

2. Configure the logical switch.

   Example 6-10 shows the configuration of domain ID, switchname, and XISL settings. EX_Port configuration on FCIP routers base fabric requires XISL to be disabled in edge fabrics.

   Example 6-10 Configure the logical switch

```
ITSO_8510_SiteA:FID128:admin> setcontext 10
switch_10:FID10:admin>
switch_10:FID10:admin> switchdisable
switch_10:FID10:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 10
Allow XISL Use (yes, y, no, n): [yes] n
WARNING!! Disabling this parameter will cause removal of LISLs to
 other logical switches. Do you want to continue? (yes, y, no, n): [no] y
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_10:FID10:admin>
switch_10:FID10:admin> switchname ITSO_8510_SiteA_FID10_pub
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_10:FID10:admin>
switch_10:FID10:admin> switchenable
switch_10:FID10:admin>
```

   The configuration of the logical switch with fabric ID 10 can be verified with the **switchshow** command, as shown in Example 6-11.

   Example 6-11 Verify the logical switch

```
ITSO_8510_SiteA_FID10_pub:FID10:admin> setcontext 10
ITSO_8510_SiteA_FID10_pub:FID10:admin> switchshow
switchName:     ITSO_8510_SiteA_FID10_pub
switchType:     121.0
switchState:    Online
switchMode:     Native
```

```
switchRole:     Principal
switchDomain:   10
switchId:       fffc0a
switchWwn:      10:00:00:05:33:96:f4:02
zoning:         OFF
switchBeacon:   OFF
FC Router:      OFF
HIF Mode:       OFF
Allow XISL Use: OFF
LS Attributes:  [FID: 10, Base Switch: No, Default Switch: No, Address Mode 0]

Index Slot Port Address Media  Speed        State    Proto
==========================================================
==============================
No ports found in the system!!!
ITSO_8510_SiteA_FID10_pub:FID10:admin>
```

3. Assign physical ports to the logical switch.

   You must assign physical ports connected to dedicated host and storage ports of the IBM SAN Volume Controller stretched cluster system to the logical switch with fabric ID 10 with the **lscfg** command, as shown in Example 6-12.

   Example 6-12 Assign physical port to logical switch with fabric ID 10 to form with edge fabric

```
ITSO_8510_SiteA_FID10_pub:FID10:admin> lscfg --config 10 -slot 1 -port 24
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
```

   The port is disabled, and must be enabled on the logical switch with fabric ID 10 by using the **portcfgpersistentenable** command, as shown in Example 6-13.

   Example 6-13 Enable Ports by portcfgpersistentenable

```
ITSO_8510_SiteA_FID10_pub:FID10:admin> portcfgpersistentenable 1/7
```

Note: E_Port 1/24 on `ITSO_8510_SiteA_FID10_pub` remains disabled until the logical switch with fabric ID 10 on `ITSO_7840_SiteA` is configured and ready for fabric build.

This section does not describe best practices for IBM SAN Volume Controller stretched cluster system configuration. See *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

Implementation steps with the IBM Network Advisor GUI

This section shows the implementation steps on IBM Network Advisor GUI:

1. To start the configuration steps with the GUI, log in to the IBM Network Advisor GUI. For more details about how to log in to the GUI, see Chapter 4, “FCIP replication” on page 93.
2. Review the SAN tab of the overview page to make sure the four SAN fabric (2 SAN42B-R and Brocade 8510) switches are already added into fabric discovery (Figure 6-26). If not, follow the steps to add fabric discovery which are described from the steps 14 on page 165 to 18 on page 167.

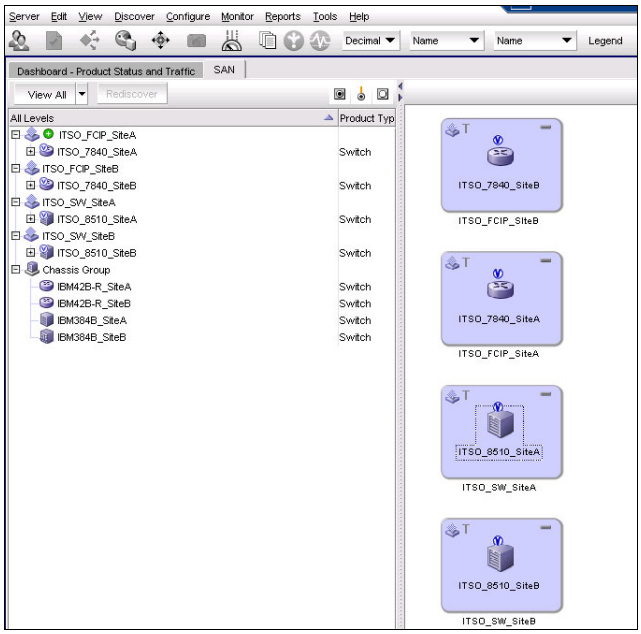


Figure 6-26 Fabric Overview

3. Create a private logical switch in Site A for SVC Stretched Cluster internal SAN communication. Assign half of the FC ports from SVC into this private logical switch. This section does not cover SVC Stretched Cluster implementation. For more information, see *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V7.6*, SG24-7933

4. Create a logical switch on physical chassis ITSO_8510_SiteA for SVC public SAN. To create a logical switch, select **Configure** → **Virtual Fabric**. The logical switches window is displayed (Figure 6-27). Select the physical chassis **IBM384B_SiteA** in the **Chassis** list box.

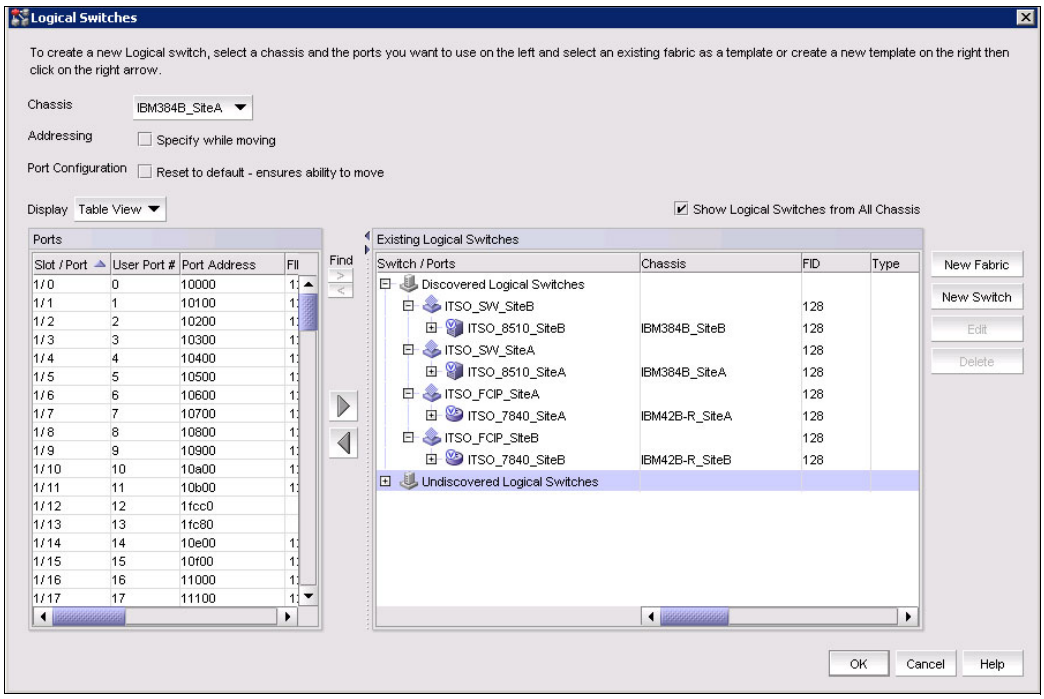


Figure 6-27 Adding logical switch for the SVC public SAN

5. Select **Undiscovered Logical Switches** in the **Existing logical switches** list, and click **New Switch.**

6. Assign a fabric ID to this logical switch (Figure 6-28). Clear the **Base Fabric for Transport (XISL)** and **Base Switch** check boxes. Here we use 10 for the fabric ID of this logical switch.

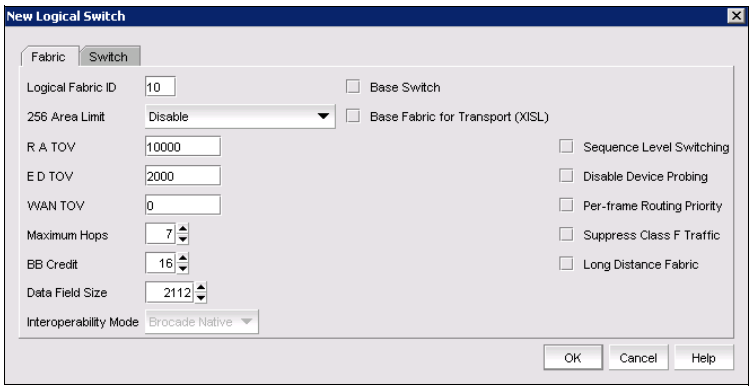


Figure 6-28 Assign fabric ID

7. Assign the switch name and domain ID to this logical switch (Figure 6-29). Here we use 10 for the domain ID of this logical switch.

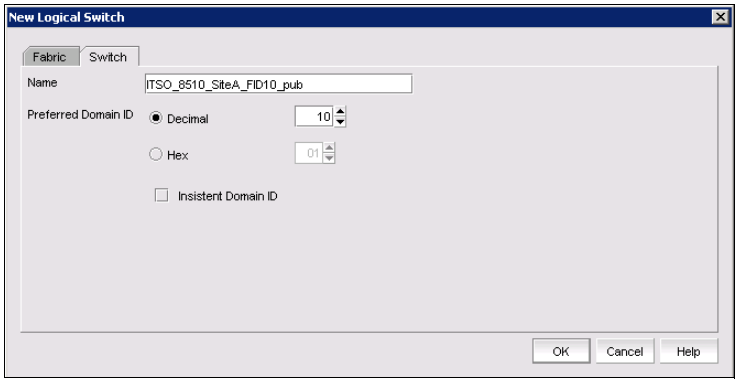


Figure 6-29 Assign switch name and domain ID

8. Assign physical ports to this logical switch (Figure 6-30). Here we assign the following ports to it: Four ports for SVC connection, four ports for backend storage connection, and one port (Port 1/24) for ISL connection to `ITSO_7840_SiteA`.

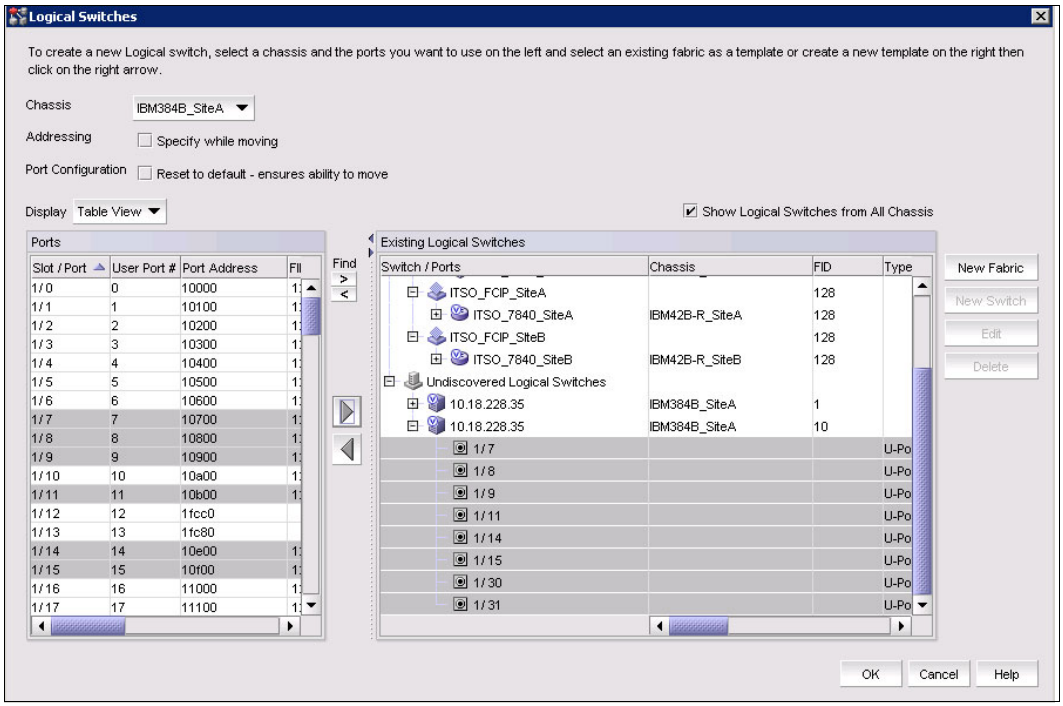


Figure 6-30 Assign ports to logical switch

9. Click **OK** to confirm these settings and complete the creation of logical switch `ITSO_8510_SiteA_FID10_pub`.

6.3.6 Site B: Configuring the edge fabric

This section describes the implementation of the logical switches on site B that are used for the IBM Storwize V7000 storage system that is a quorum source for the IBM SAN Volume Controller stretched cluster configuration on site A. See *Implementing the IBM Storwize V7000 and IBM Spectrum Virtualize V7.6*, SG24-7938 for more information.

Implementation steps with the CLI

This section shows the implementation with the CLI:

1. Create a logical switch on physical switch 8510.

   Example 6-14 shows how to create a new logical switch with fabric ID 20 on the ITSO_8510_SiteB switch.

   Example 6-14 Create a new logical switch with FID 20

```
ITSO_8510_SiteB:FID128:admin> lscfg --create 20
A Logical switch with FID 20 will be created with default configuration.
Would you like to continue [y/n]?: y
About to create switch with fid=20. Please wait...
Logical Switch with FID (20) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
```

   Example 6-15 shows how to create a new base switch with fabric ID 50 on the switch.

   Example 6-15 Create a new logical switch with FID 50

```
ITSO_7840_SiteB_FID20:FID20:admin> lscfg --create 50 -base
Creation of a base switch requires that the proposed new base switch on this
system be disabled.
Would you like to continue [y/n]?: y
About to create switch with fid=50. Please wait...
Logical Switch with FID (50) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
ITSO_7840_SiteB_FID20:FID20:admin>
```

2. Configure the logical switch.

   Example 6-16 and Example 6-17 on page 180 show the configuration of the domain ID, switch name, and XISL setting. Note that XISL must be disabled.

   Example 6-16 Configure the logical switch with fabric ID 20

```
ITSO_8510_SiteB:FID128:admin> setcontext 20
Warning: Default password not changed for 'root' and 'factory'. Please login as
'root' to change them.
Warning: Default password not changed for 'root' and 'factory'. Please login as
'root' to change them.
switch_20:FID20:admin> switchdisable
switch_20:FID20:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 20
    Allow XISL Use (yes, y, no, n): [yes] n
WARNING!! Disabling this parameter will cause removal of LISLs to
 other logical switches. Do you want to continue? (yes, y, no, n): [no] y
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_20:FID20:admin> switchname ITSO_8510_SiteB_FID20
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
```

```
switch_20:FID20:admin> switchenable
switch_20:FID20:admin>
switch_20:FID20:admin> setcontext 20
Warning: Default password not changed for 'root' and 'factory'. Please login as
'root' to change them.
Warning: Default password not changed for 'root' and 'factory'. Please login as
'root' to change them.
```

Example 6-17 shows configuring the logical switch with FID 50.

Example 6-17 Configure the logical switch with FID 50

```
ITSO_7840_SiteB_FID20:FID20:admin> setcontext 50
switch_50:FID50:admin>
switch_50:FID50:admin> switchdisable
switch_50:FID50:admin>
switch_50:FID50:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 22
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_50:FID50:admin> switchenable
switch_50:FID50:admin> switchname ITSO_7840_SiteB_base_FID50
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_50:FID50:admin> setcontext 50
ITSO_7840_SiteB_base_FID50:FID50:admin>
```

3. Assign physical ports to virtual switches.

   Physical ports connected to IBM Storwize V7000 storage system located in site B must be assigned to a logical switch with fabric ID 20 with the **lscfg** command, as shown in Example 6-18.

   Example 6-18 Assign physical port to logical switch with fabric ID 20 to form with edge fabric

```
ITSO_8510_SiteB_FID20:FID20:admin> lscfg --config 20 -slot 9 -port 25
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_8510_SiteB_FID20:FID20:admin>
```

4. Enable ports with the **portcfgpersistentenable** command.

   All ports connected to IBM Storwize V7000 storage system in site B must be enabled with the **portcfgpersistentenable** command, as shown in Example 6-19.

   Example 6-19 Enable Ports by portcfgpersistentenable command

```
ITSO_8510_SiteB_FID20:FID20:admin> portcfgpersistentenable 9/25
```

Note: E_Port 9/25 on `ITSO_8510_SiteB_FID20` remains disabled until logical switch with fabric ID 20 on `ITSO_7840_SiteB` is configured and ready for fabric build.

Implementation steps with the IBM Network Advisor GUI

To create the logical switch ITSO_8510_SiteB_FID10 on the physical chassis *IBM384B_SiteB,* repeat the steps from step 4 on page 177 to 9 on page 178. This logical switch is for quorum storage remote connection on Site B.

Here we assign fabric ID 20 and domain ID 20 to this logical switch. We also clear the check boxes **Base Fabric for Transport (XISL)** and **Base Switch.** We assign four physical ports for quorum storage FC connection and one port (Port 9/12) for ISL connection to ITSO_7840_SiteB.

6.3.7 Site A: Configuring logical switches on the FCIP router

This section describes the implementation steps of the logical switch with fabric ID 10 connected to edge fabric and the logical base switch with the fabric ID 50.

Implementation steps with the CLI

The following commands show the implementation of a virtual fabric with a logical switch connected to edge fabric and a virtual fabric with a logical base switch.

1. Create a logical switch.

   Example 6-20 shows the creation of a logical switch with fabric ID 10.

   Example 6-20 Create a new virtual fabric with fid10

```
ITSO_7840_SiteA:FID128:admin> lscfg --create 10
A Logical switch with FID 10 will be created with default configuration.
Would you like to continue [y/n]?: y
About to create switch with fid=10. Please wait...
Logical Switch with FID (10) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
```

   Example 6-21 shows the creation of a base switch with fabric ID 50.

   Example 6-21 Create a new virtual fabric with fabric ID 50 as a base switch

```
ITSO_7840_SiteA_FID10:FID10:admin> lscfg --create 1 50 -base
Creation of a base switch requires that the proposed new base switch on this
system be disabled.
Would you like to continue [y/n]?: y
About to create switch with fid=50. Please wait...
Logical Switch with FID (50) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
```

2. Configure the logical switch.

   Example 6-22 and Example 6-23 show the configuration of domain ID, XISL, and switch name. XISL must be disabled.

   Example 6-22 Configure logical switch with fabric ID 10

```
ITSO_7840_SiteA:FID128:admin> setcontext 10
switch_10:FID10:admin>
switch_10:FID10:admin> switchdisable
```

```
switch_10:FID10:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 11
    Allow XISL Use (yes, y, no, n): [yes] no
WARNING!! Disabling this parameter will cause removal of LISLs to
 other logical switches. Do you want to continue? (yes, y, no, n): [no] yes
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_10:FID10:admin> switchname ITSO_7840_SiteA_FID10
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_10:FID10:admin> setcontext 10
ITSO_7840_SiteA_FID10:FID10:admin> switchenable
ITSO_7840_SiteA_FID10:FID10:admin>
```

Example 6-23 shows how to configure the logical switch base switch with fabric ID 50.

Example 6-23 Configure logical switch base switch with fabric ID 50

```
ITSO_7840_SiteA_FID10:FID10:admin> setcontext 50
switch_50:FID50:admin> switchdisable
switch_50:FID50:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 12
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_50:FID50:admin>
switch_50:FID50:admin> switchname ITSO_7840_SiteA_base_FID50
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_50:FID50:admin> setcontext 50
ITSO_7840_SiteA_base_FID50:FID50:admin> switchenable
ITSO_7840_SiteA_base_FID50:FID50:admin>
```

3. Assign physical ports to logical switches.

   Port 4 and 16 must be assigned to logical switch `ITSO_8501O_SiteA_FID10_pub` and port 12 must be assigned to logical base switch `ITSO_7840_SiteA_base_FID50` with the **`lscfg`** command, as shown in Example 6-24.

Example 6-24 Assign physical port to logical switch with fabric ID 10

```
ITSO_7840_SiteA_FID10:FID10:admin> lscfg --config 10 -port 4
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_FID10:FID10:admin>
```

IP interfaces ge2, ge3, ge6, ge7, and VE Port 34 must be assigned to logical base switch with fabric ID 50.

Example 6-25 shows the assignment of the physical ports to logical switch with fabric ID 50.

Example 6-25 Assign physical port to logical switch with fabric ID 10

```
ITSO_7840_SiteA_base_FID50:FID50:admin> lscfg --config 50 -port ge2-3
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Note: All ports assigned to logical switches with fabric ID 10 and 50 remain disabled until logical switch configuration on ITSO_7840_SiteB is completed.

Implementation steps with the IBM Network Advisor GUI

Complete the following steps with the IBM Network Advisor GUI:

1. Create a base logical switch on the physical chassis ITSO_7840_SiteA for FCIP tunnel connection. To create the logical switch, click **Configure** → **Virtual Fabric**, and select the physical chassis **IBM42B-R_SiteA** in the **Chassis** list box.
2. Select **Undiscovered Logical Switches** in the **Existing logical switches** list, and click **New Switch**.
3. Assign a fabric ID to this logical switch (Figure 6-31). Clear the check box **Base Fabric for Transport (XISL)**, and select **Base Switch**. Here we use 50 for the fabric ID of the base logical switch.

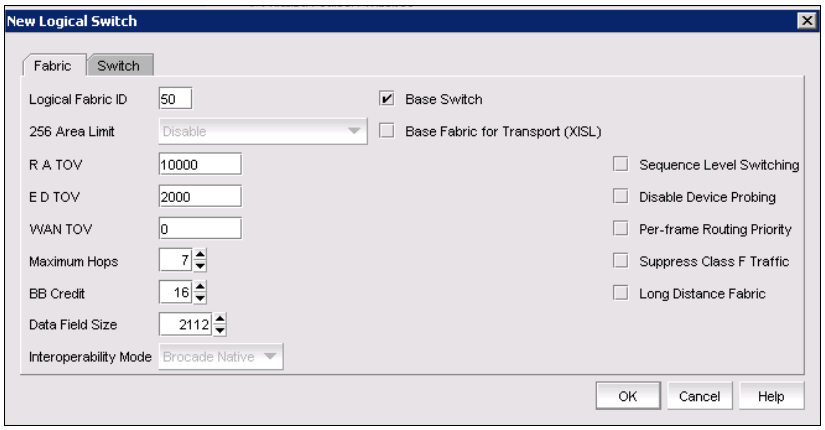


Figure 6-31 Assign the fabric ID for the base logical switch

4. Assign the switch name and domain ID to this logical switch (Figure 6-32). Here we use 12 for the domain ID of this logical switch.

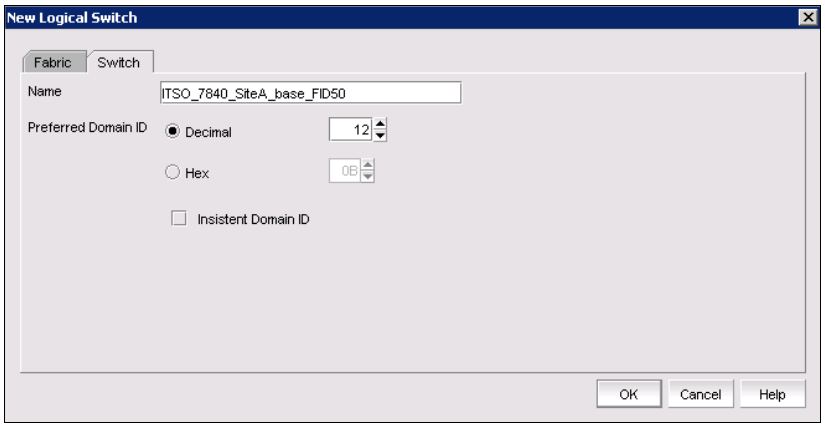


Figure 6-32 Assign domain ID for base logical switch

5. Assign physical ports to this logical switch (Figure 6-33). Here we assign the following ports: One port (Port 12) for IFL connection to the other logical switch that will be created later, four GE ports (ge2, ge3, ge6, ge7), and one VE port (34) for FCIP tunnel connection.

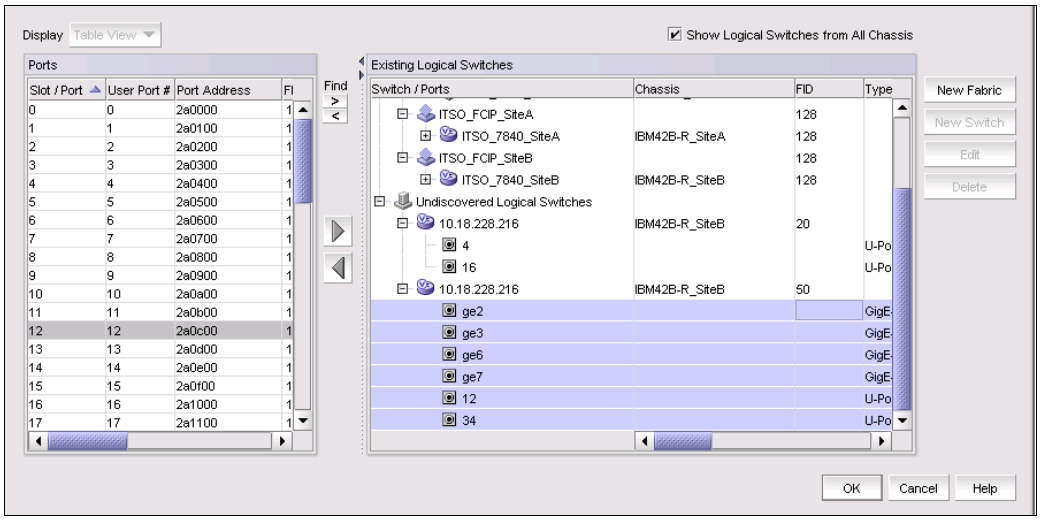


Figure 6-33 Assign ports to base logical switch

6. Click **OK** to confirm these setting and complete the creation of logical switch `ITSO_7840_SiteA_base_FID50`.
7. Create another logical switch on the same physical chassis `ITSO_7840_SiteA` for ISL connection to `ITSO_8510_SiteA` switch. To create the logical switch, click **Configure** → **Virtual Fabric**, and select the physical chassis **IBM42B-R_SiteA** in the **Chassis** list box.
8. Select **Undiscovered Logical Switches** in the **Existing logical switches** list, and click **New Switch**.

9. Use the same fabric ID (10) of `ITSO_8510_SiteA_FID10_pub` for this logical switch (Figure 6-34). Clear the check boxes **Base Fabric for Transport (XISL)** and **Base Switch**.

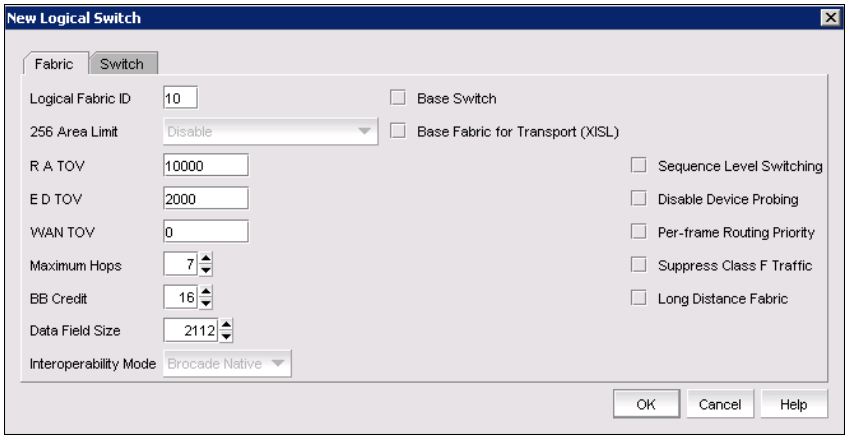


Figure 6-34 Assign fabric ID to logical switch

10. Assign the switch name and domain ID to this logical switch (Figure 6-35). The domain ID should be different from the one of `ITSO_8510_SiteA_FID10_pub`. Here we use 11 for the domain ID of this logical switch.

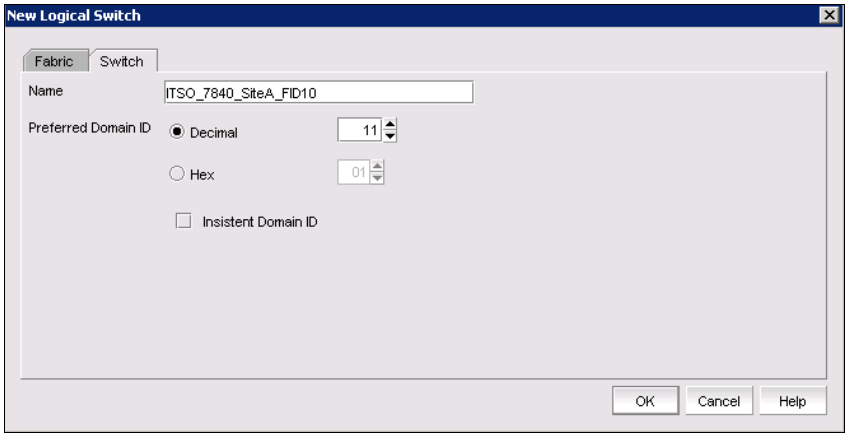


Figure 6-35 Assign domain ID for logical switch

11. Assign physical ports to this logical switch (Figure 6-36). Here we assign the following ports to it: One port (Port 4) for IFL connection to the base logical switch on the same chassis, and one port (Port 16) for ISL connection to `ITSO_8510_SiteA`.

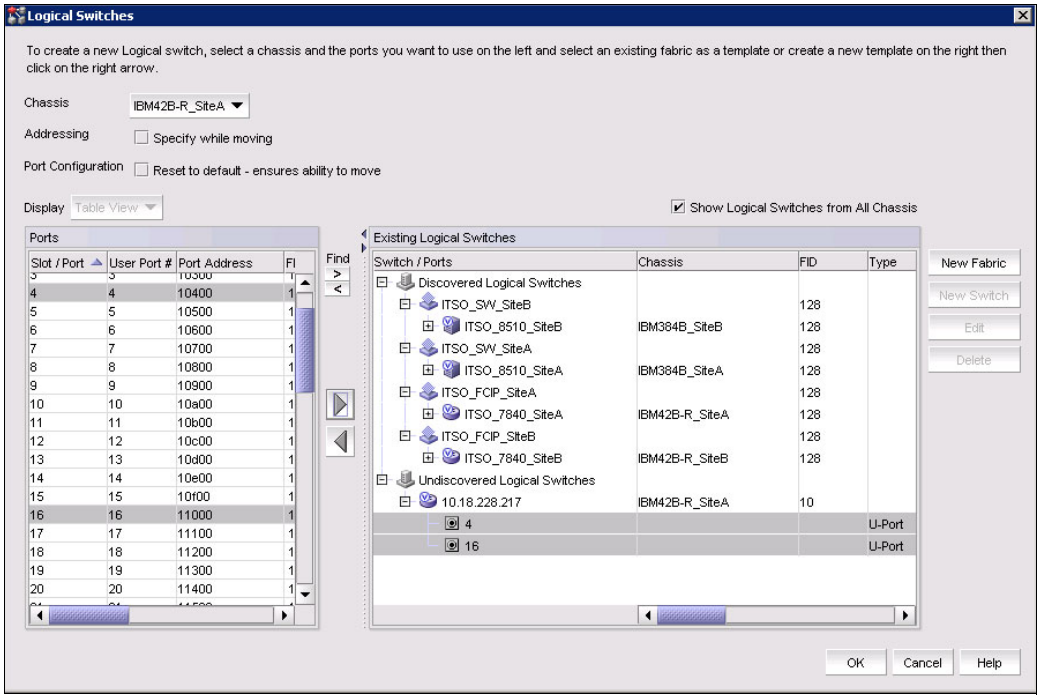


Figure 6-36 Assign ports to logical switch

6.3.8 Site B: Configuring logical switches on the FCIP router

This section describes the implementation of the logical switch connected to the edge fabric in site B and the logical base switch.

Implementation steps with the CLI

The following commands show the implementation of a virtual fabric with a logical switch connected to edge fabric and a virtual fabric with a logical base switch:

1. Create a virtual fabric with fabric ID 20.

   Example 6-26 shows the creation of a logical switch with fabric ID 10.

Example 6-26 Create a new virtual fabric with 20

```
ITSO_7840_SiteB:FID128:admin> lscfg --create 20
A Logical switch with FID 20 will be created with default configuration.
Would you like to continue [y/n]?: y
About to create switch with fid=20. Please wait...
Logical Switch with FID (20) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
ITSO_7840_SiteB:FID128:admin>
```

Example 6-27 shows the creation of a new base switch fabric ID 50.

Example 6-27 Create a new virtual fabric with fabric ID 50 as a base switch

```
ITSO_7840_SiteB_FID20:FID20:admin> lscfg --create 50 -base
Creation of a base switch requires that the proposed new base switch on this
system be disabled.
Would you like to continue [y/n]?: y
About to create switch with fid=50. Please wait...
Logical Switch with FID (50) has been successfully created.
Logical Switch has been created with default configurations.
Please configure the Logical Switch with appropriate switch
and protocol settings before activating the Logical Switch.
ITSO_7840_SiteB_FID20:FID20:admin>
```

2. Configure the logical switch.

   Example 6-28 and Example 6-29 show the configuration of the domain id, switchname, and XISL setting. XISL must be disabled.

Example 6-28 Configure logical switch with fabric ID 20

```
TSO_7840_SiteB:FID128:admin> setconn text 20
switch_20:FID20:admin>
switch_20:FID20:admin> switchdisable
switch_20:FID20:admin>
switch_20:FID20:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 21
    WWN Based persistent PID (yes, y, no, n): [no]
    F-Port Device Update Mode: (on, off): [off]
    Allow XISL Use (yes, y, no, n): [yes] no
WARNING!! Disabling this parameter will cause removal of LISLs to
 other logical switches. Do you want to continue? (yes, y, no, n): [no] yes
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_20:FID20:admin> switchname ITSO_7840_SiteB_FID20
Done.
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_20:FID20:admin> setcontext 20
ITSO_7840_SiteB_FID20:FID20:admin>
ITSO_7840_SiteB_FID20:FID20:admin>switchenable
ITSO_7840_SiteB_FID20:FID20:admin>
```

Example 6-29 shows how to configure the logical base switch with fabric ID 50.

Example 6-29 Configure logical base switch with fabric ID 50

```
switch_50:FID50:admin> configure
Configure...
  Fabric parameters (yes, y, no, n): [no] y
    Domain: (1..239) [1] 22
WARNING: The domain ID will be changed. The port level zoning may be affected
switch_50:FID50:admin> switchenable
switch_50:FID50:admin>
switch_50:FID50:admin> switchname ITSO_7840_SiteB_FID50
Done.
```

```
Switch name has been changed.Please re-login into the switch for the change to
be applied.
switch_50:FID50:admin> setcontext 50
ITSO_7840_SiteB_FID50:FID50:admin>
```

3. Assign physical ports to logical switches with fid 20 and fid 50.

   Port 4 and 16 must be assigned to logical switch `ITSO_8501O_SiteA_FID1O_pub`, port 12 must be assigned to logical base switch `ITSO_7840_SiteA_base_FID50` with the **`lscfg`** command, as shown in Example 6-30.

   Example 6-30 Assign physical port to logical switch with fabric ID 20

```
TSO_7840_SiteB_FID2O:FID2O:admin> lscfg --config 20 -port 16
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteB_FID2O:FID2O:admin>
```

   IP interfaces ge2, ge3, ge6, and ge7, and VE Port 34 must be assigned to logical switch with fabric ID 50, as shown in Example 6-27 on page 187.

   Example 6-31 shows the assignment of the physical ports to logical switch with fabric ID 50.

   Example 6-31 Assign physical port to logical switch with fabric ID 20

```
ITSO_7840_SiteB_base_FID5O:FID5O:admin> lscfg --config 50 -port ge2-3
This operation requires that the affected ports be disabled.
Would you like to continue [y/n]?: y
Making this configuration change.  Please wait...
Configuration change successful.
Please enable your ports/switch when you are ready to continue.
ITSO_7840_SiteB_base_FID5O:FID5O:admin>
```

   After the creation of logical switches in both edge fabrics and FCIP routers, the configuration must be completed by enabling the E_Ports between the following logical switches:

   – ITSO_8510_SiteA_FID10 --> ITSO_7840_SiteA_FID10
   – ITSO_8510_SiteB_FID20 --> ITSO_7840_SiteB_FID20

   E_Ports must be enabled by the **`portcfgpersistentenable`** command. After enabling E_Ports, all logical switches within a fabric can be shown by the **`fabricshow`** command, as shown in Example 6-32 and Example 6-33 on page 189.

   Example 6-32 SiteA: Show all logical switches of the fabric by fabricshow command.

```
ITSO_851O_SiteA_FID1O_pub:FID1O:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-----------------------------------------------------------------------
 1O: fffc0a 1O:OO:OO:05:33:96:f4:02 10.18.228.35    0.0.0.0
>"ITSO_851O_SiteA_FID1O_pub"
 11: fffc0b 1O:OO:50:eb:1a:d7:83:81 10.18.228.217   0.0.0.0
"ITSO_7840_SiteA_FID1O"
The Fabric has 2 switches
```

Example 6-33 shows all logical switches at site B.

Example 6-33 SiteB: Show all logical switches of the fabric with the fabricshow command

```
ITSO_8510_SiteB_FID20:FID20:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-------------------------------------------------------------------------
 20: fffc14 10:00:00:05:1e:46:8a:01 10.18.228.106   0.0.0.0
>"ITSO_8510_SiteB_FID20"
 21: fffc15 10:00:50:eb:1a:36:1d:39 10.18.228.216   0.0.0.0
"ITSO_7840_SiteB_FID20"
The Fabric has 2 switches
```

Implementation steps with the IBM Network Advisor GUI

Repeat the steps from 1 on page 183 to 11 on page 186 to create two logical switches on the physical chassis `ITSO_7840_SiteB`: `ITSO_7840_SiteB_base_FID50` and `ITSO_7840_SiteB_FID20`.

For the base logical switch `ITSO_7840_SiteB_base_FID50`, we assign the same fabric ID (50) but different domain ID (22) than the logical switch `ITSO_7840_SiteA_base_FID50`. We also assign the following ports to it: One port (Port 12) for IFL connection to the other logical switch that will be created later, four GE ports (ge2, ge3, ge6, ge7), and one VE port (34) for FCIP tunnel connection.

For the logical switch `ITSO_7840_SiteB_FID20`, we assign the same fabric ID (20), but different domain ID (21) to the logical switch `ITSO_8510_SiteB_FID20`. We also assign the following ports to it: One port (Port 4) for IFL connection to the base logical switch on the same chassis, and one port (Port 16) for ISL connection to ITSO_8510_SiteB.

After the creation of all the logical switches, follow the steps from 14 on page 165 to 18 on page 167 to add fabric discovery for the new logical switches. It is necessary to add these logical switches into fabric discovery before the next steps. Figure 6-37 and Figure 6-38 on page 190 show the result of adding discovery.

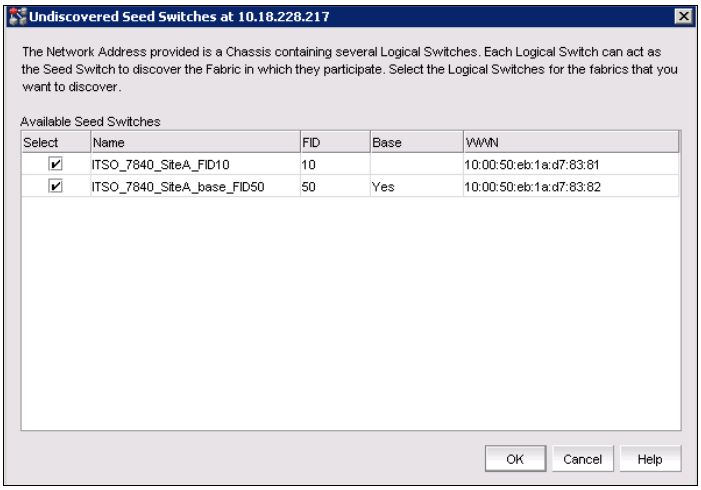


Figure 6-37 Add fabric discovery for Site A

Figure 6-38 shows the result of adding discovery for Site B.

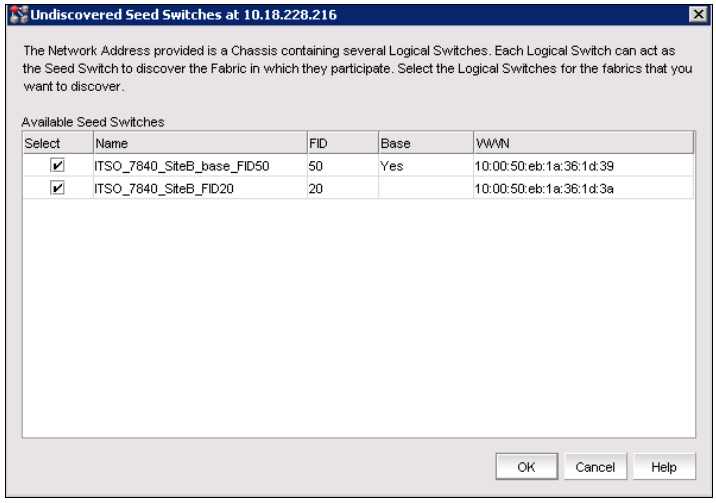


Figure 6-38 Add fabric discovery for Site B

6.3.9 Configuring the inter fabric link

This section describes the configuration of EX_Ports on `ITSO_7840_SiteA` and `ITSO_7840_SiteB`. The prerequisite for this configuration is an inter-fabric link, connecting logical switch with the fabric ID 10 and logical base switch with fabric ID 50 by using a physical Fibre Channel cable.

Constraints of virtual fabrics and EX_Ports

The following points must be considered before implementing EX_Ports and virtual fabrics:

- The base switch does not have any devices connected. The base fabric can have devices in remote layer 2 switches. Traffic between those devices is supported.
- EX_Ports and VEX_Ports are supported in the base switch. EX_Ports cannot be part of any switch other than the base switch.
- EX_Ports cannot connect to a fabric that has a logical switch with the Allow XISL use mode on. The port is disabled with the reason “Conflict: XISL capability domain.”
- A logical fabric cannot have EX_Ports using XISLs and cannot serve as a backbone to any EX_Port traffic.
- Traffic Isolation zones with no failover option are not supported in logical fabrics. TI zones defined in the base fabric for logical fabric traffic must allow failover.

Configuring EX ports with the CLI

This section describes the implementation steps for configuring EX_Ports on a base switch:

1. Disable Port 12 on logical switches ITSO_7840_SiteA_base_FID50 and ITSO_7840_SiteB_base_FID50 with the **portdisable** command, as shown in Example 6-34.

   Note: Prior EX_Port configuration, port 4 on ITSO_7840_SiteA_FID10 and ITSO_7840_SiteB_FID20 were disabled.

   Example 6-34 Disable port 12 for EX_Port configuration

   ```
   ITSO_7840_SiteA_base_FID50:FID50:admin> portdisable 12
   ```

2. Configure the EX_Port with the **portcfgexport** command.

   Example 6-35 and Example 6-36 show the configuration of EX_Port on ITSO_7840_SiteA_base_FID50 and ITSO_7840_SiteB_base_FID50.

   During EX_Port configuration, the port number is required, whereas the edge fabric ID and front domain id parameters are optional. The parameter **-f** specifies the edge fabric ID, which is the number you must assign to the EX_Port. If no parameter is specified for fabric ID and domain id, a number is assigned by the system.

   It is preferred to specify a unique fabric ID parameter for EX_Ports connected to same edge fabric. In our example, we use fabric ID 100 for the EX_Port on ITSO_7840_SiteA_base_FID50 and fabric ID 101 for EX_Port on ITSO_7840_SiteB_base_FID50.

   Example 6-35 shows an example for Site A.

   Example 6-35 Site A: Configure port 12 as an EX_Port

   ```
   ITSO_7840_SiteA_base_FID50:FID50:admin> portcfgexport 12 -a 1 -f 100
   ```

   Example 6-36 shows an example for Site B.

   Example 6-36 Site B: Configure port 12 as an EX_Port

   ```
   ITSO_7840_SiteB_base_FID50:FID50:admin> portcfgexport 12 -a 1 -f 101
   ```

   Note: Not specifying the **-f** parameter in our example would cause a fabric ID conflict because the fid used for EX_Port on ITSO_7840_SiteA and the fabric ID used for EX_Port on ITSO_7840_SiteB would be set to same value by the system. It is not supported to specify the same fabric ID for EX_Ports connecting to different edge fabrics.

   For more information about the **portcfgexport** command, see the *Brocade Fabric OS Command Reference*:

   https://www.brocade.com/content/dam/common/documents/content-types/command-reference-guide/fos-800-commandref.pdf

3. Enable EX_Ports on the logical base switches.

   By **portcfgpersistentenable** command, the port 12 on base switches on both sites must be enabled. Example 6-37 shows Site A.

   Example 6-37 Site A: Configure port 12 as an EX_Port

   ```
   ITSO_7840_SiteB_base_FID50:FID50:admin> portcfgexport 12 -a 1 -f 100
   ```

Example 6-38 shows an example for Site B.

Example 6-38 Site B: Configure port 12 as an EX_Port

```
ITSO_7840_SiteB_base_FID50:FID50:admin> portcfgexport 12 -a 1 -f 101
```

Note: Port 4 on logical switches ITSO_7840_SiteA_FID10 and ITSO_7840_SiteA_FID20 must be enabled for EX_port verification.

4. Show the EX_Port configuration details.

   The EX_Port configuration can be verified with the **portcfgexport** command.

   Example 6-39 shows Site A.

Example 6-39 Site A: Show current EX_Port configuration settings

```
ITSO_7840_SiteA_base_FID50:FID50:admin> portcfgexport 12
        Port   12   info
Admin:                  enabled
State:                  OK
Pid format:             core(N)
Operate mode:           Brocade Native
Edge Fabric ID:         100
Front Domain ID:        160
Front WWN:              55:0e:b1:ad:78:38:2e:64
Principal Switch:       10
Principal WWN:          10:00:00:05:33:96:f4:02
Fabric Parameters:      Auto Negotiate
R_A_TOV:                10000(N)
E_D_TOV:                2000(N)
Authentication Type: None
DH Group: N/A
Hash Algorithm: N/A
Encryption:     OFF
Compression:    OFF
Forward Error Correction: ON
Edge fabric's primary wwn: N/A
Edge fabric's version stamp: N/A

ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Example 6-40 shows Site B.

Example 6-40 Site B: Show current EX_Port configuration settings

```
ITSO_7840_SiteB_base_FID50:FID50:admin> portcfgexport 12
        Port   12   info
Admin:                  enabled
State:                  OK
Pid format:             core(N)
Operate mode:           Brocade Native
Edge Fabric ID:         101
Front Domain ID:        160
Front WWN:              55:0e:b1:a3:61:d3:ae:65
Principal Switch:       20
Principal WWN:          10:00:00:05:1e:46:8a:01
Fabric Parameters:      Auto Negotiate
```

```
R_A_TOV:                10000(N)
E_D_TOV:                2000(N)
Authentication Type: None
DH Group: N/A
Hash Algorithm: N/A
Encryption:     OFF
Compression:    OFF
Forward Error Correction: ON
Edge fabric's primary wwn: N/A
Edge fabric's version stamp: N/A

ITSO_7840_SiteB__base_FID50:FID50:admin>
```

Note: If you do not specify the **-d** parameter for the **portcfgexport** command, the front domain ID 160 is chosen by the system.

5. Verify the EX_Port status.

   You can use the **switchshow** command to verify the status of the EX_Ports on the logical base switches.

   Example 6-41 shows Site A.

Example 6-41 Site A: Switchshow

```
ITSO_7840_SiteA_base_FID50:FID50:admin> switchshow
switchName:     ITSO_7840_SiteA_base_FID50
switchType:     148.0
switchState:    Online
switchMode:     Native
switchRole:     Subordinate
switchDomain:   12
switchId:       fffc0c
switchWwn:      10:00:50:eb:1a:d7:83:82
zoning:         OFF
switchBeacon:   OFF
FC Router:      ON
HIF Mode:       OFF
LS Attributes:  [FID: 50, Base Switch: Yes, Default Switch: No, Address Mode 0]

Index Port Address  Media Speed   State       Proto
==================================================
  12  12   0c0100   id    N16     Online      FC  EX-Port
10:00:50:eb:1a:d7:83:81 "" (fabric id = 100 )(Trunk master)
  34  34   0c0000   --    --       Online      VE  VE-Port
10:00:50:eb:1a:36:1d:3a "ITSO_7840_SiteB_FID50" (upstream)
      ge2           id    10G     Online      FCIP
      ge3           id    10G     Online      FCIP
      ge6           id    10G     Online      FCIP
      ge7           id    10G     Online      FCIP
ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Example 6-42 shows Site B.

Example 6-42 Site B: Switchshow

```
ITSO_7840_SiteB_FID50:FID50:admin> switchshow
switchName:     ITSO_7840_SiteB_FID50
switchType:     148.0
switchState:    Online
switchMode:     Native
switchRole:     Principal
switchDomain:   22
switchId:       fffc16
switchWwn:      10:00:50:eb:1a:36:1d:3a
zoning:         OFF
switchBeacon:   OFF
FC Router:      ON
HIF Mode:       OFF
LS Attributes:  [FID: 50, Base Switch: Yes, Default Switch: No, Address Mode 0]

Index Port Address  Media Speed   State       Proto
==================================================
  12  12   160000   id    N16     Online      FC  EX-Port
10:00:50:eb:1a:36:1d:39"ITSO_7840_SiteB_FID20"(fabricid=101)(Trunkma
ster)
  34  34   160100   --    --       Online      VE  VE-Port
10:00:50:eb:1a:d7:83:82 "ITSO_7840_SiteA_base_FID50" (downstream)
      ge2           id    10G     Online      FCIP
      ge3           id    10G     Online      FCIP
      ge6           id    10G     Online      FCIP
      ge7           id    10G     Online      FCIP
ITSO_7840_SiteB_FID50:FID50:admin>
```

You can run the **fcrfabricshow** command to display the logical switches within the backbone as shown in Example 6-43.

Example 6-43 Enable EX_Port

```
ITSO_7840_SiteA_base_FID50:FID50:admin> fcrfabricshow
FC Router WWN: 10:00:50:eb:1a:d7:83:82, Dom ID:  12,
Info: 10.18.228.217, "ITSO_7840_SiteA_base_FID50"
   EX_Port       FID    Neighbor Switch Info (enet IP, WWN, name)
   ---------------------------------------------------------------------------
      12        100     10.18.228.217    10:00:50:eb:1a:d7:83:81
"ITSO_7840_SiteA_FID10"
ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Configuring the EX Port with the IBM Network Advisor GUI

Complete these steps to configure the EX port with the GUI:

1. To configure the EX port, select **Configure → Routing → Configuration** from the main menu (Figure 6-39).

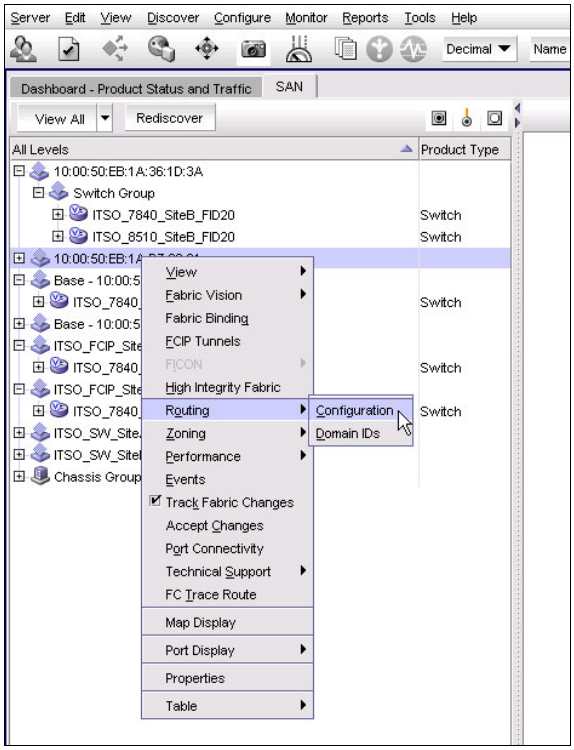


Figure 6-39 The routing configuration menu

The **Router Configuration** window is displayed (Figure 6-40). To configure routing, complete these steps:

a. Select the base logical switch that we want to configure integrate routing from the list of **Available Routers**.

b. Select the Fabric ID (100) to be assigned to backbone switch.

c. Click the right arrow button to move the logical switch to the list of **Selected Router**.

d. Click **OK** on the Router Configuration window (Figure 6-40).

The Element Manager starts automatically and opens the FC Router window and Port Configuration wizard. For more details, see the *Brocade Web Tools Administrator's Guide*, 53-1003989-01:

https://www.brocade.com/content/html/en/administration-guide/fos-800-webtools/index.html

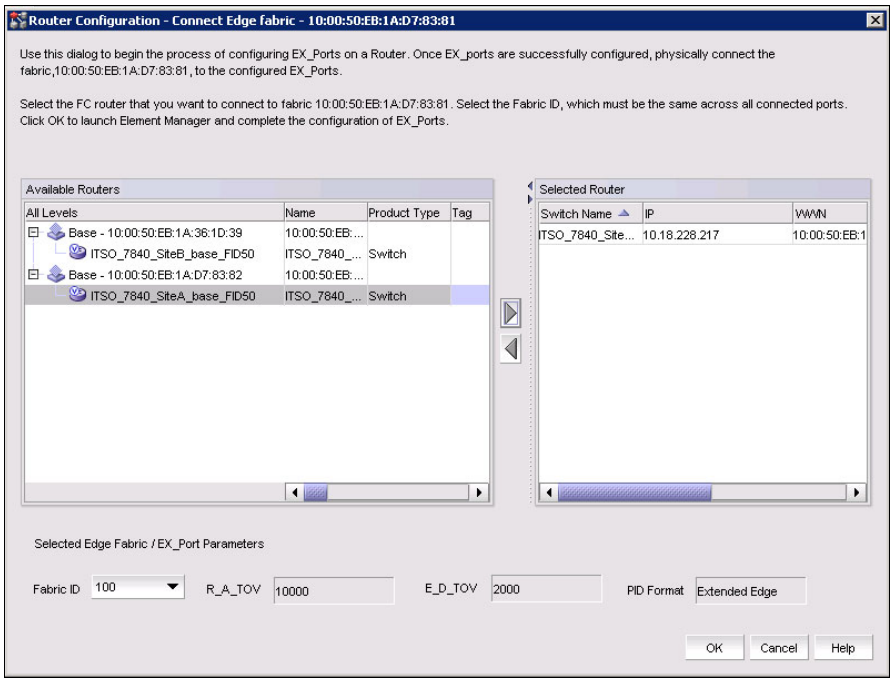


Figure 6-40 Select the logical switch for the routing configuration

2. Follow the instructions in the port configuration wizard to configure the EX port:
   a. Select the port to be configured as an EX_Port (Figure 6-41). Here we select port 12.

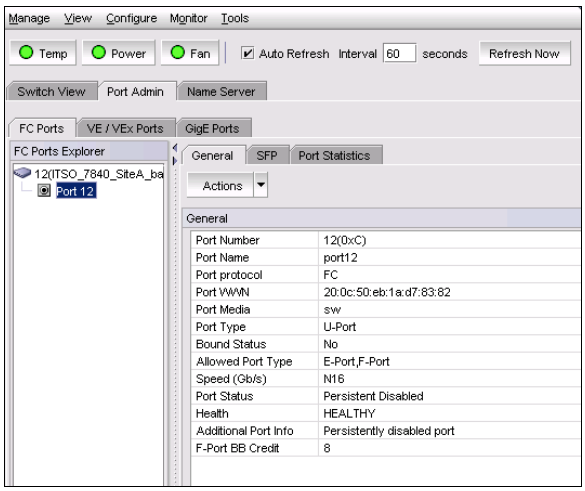


Figure 6-41 Select ports for routing configuration

   b. Click **Actions** from the General tab and select **Edit** (Figure 6-42).

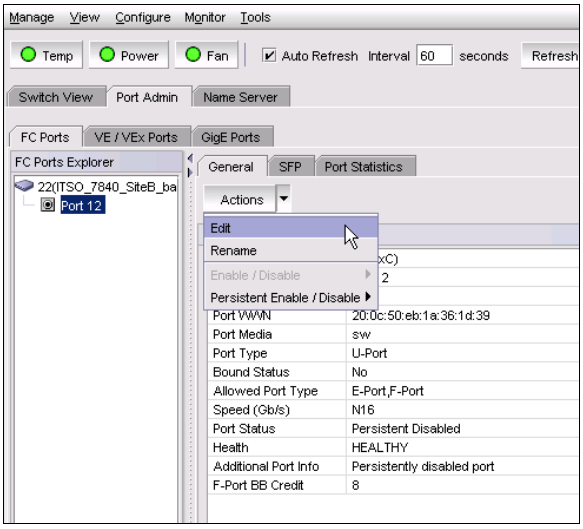


Figure 6-42 Edit port for routing configuration

c. To specify port parameters, select EX-Port and assign fabric ID for the connection of the edge fabric. Here we assign fabric ID 100 for it (Figure 6-43).

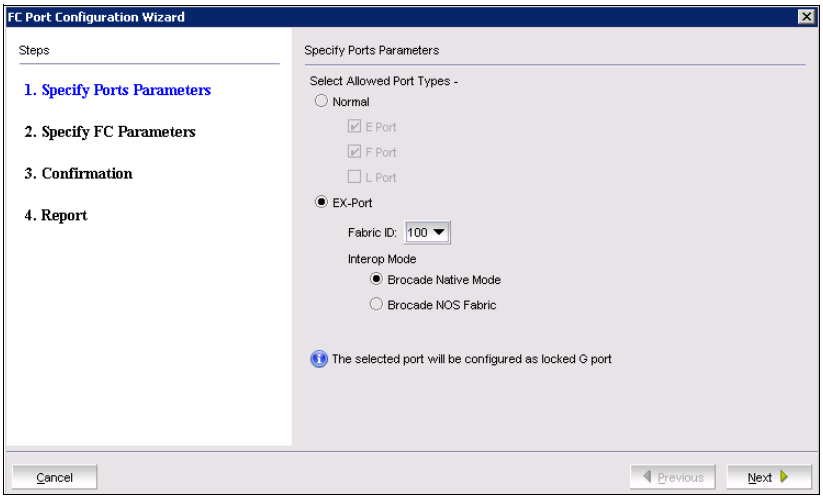


Figure 6-43 Select EX port and fabric ID

d. Leave other parameters unchanged to accept the default values and click **Save** to confirm the setting.

3. Select this port and click **Actions** → **Persistent Enable/Disable** → **Enable** to enable this port (Figure 6-44).

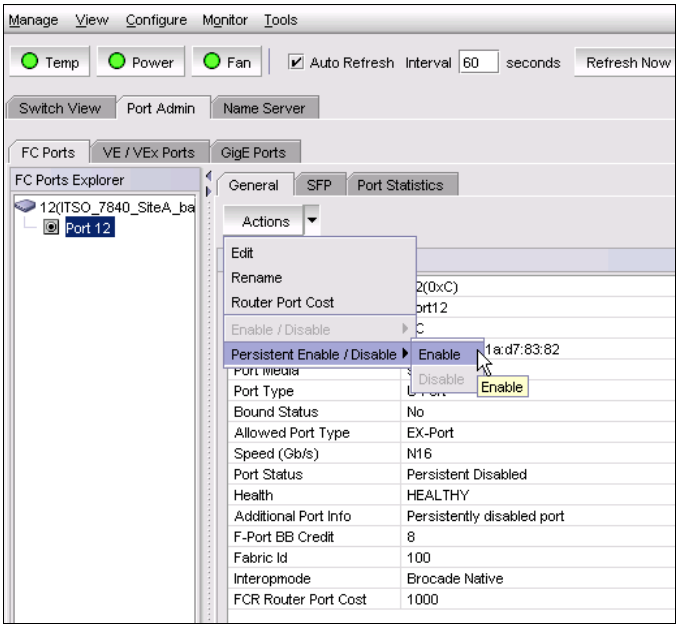


Figure 6-44 Enable the port

After these steps, refresh the overview page of IBM Network Advisor GUI and get the fabric graph shown in Figure 6-45.

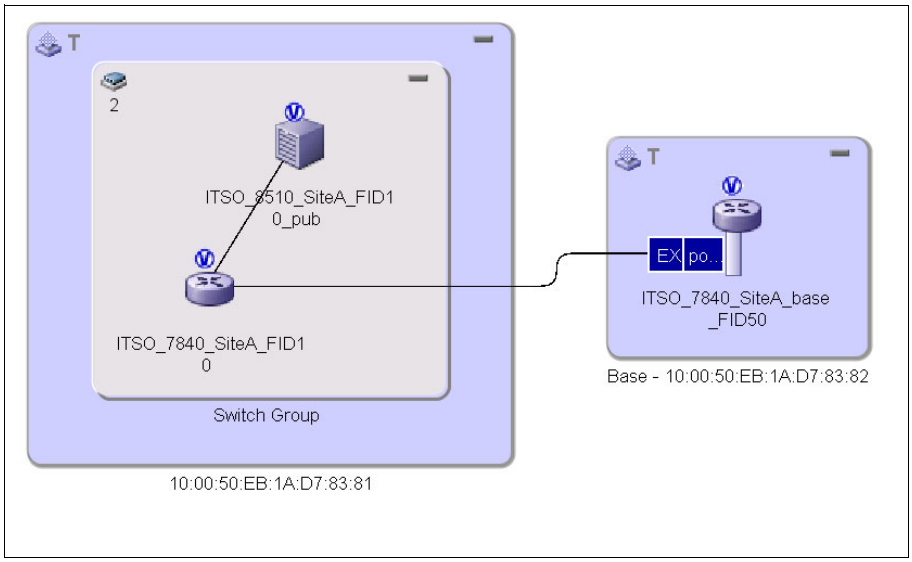


Figure 6-45 Fabric overview after EX port configuration on Site A

4. Repeat steps 1 - 3 to configure EX ports on the logic switch `ITSO_7840_SiteB_base_FID50`. Make sure that the fabric ID that you configure on the EX port should be different from the fabric ID (100) of the EX port on the base logical switch `ITSO_7840_SiteA_base_FID50`. Here we assign 101 to it (Figure 6-46).

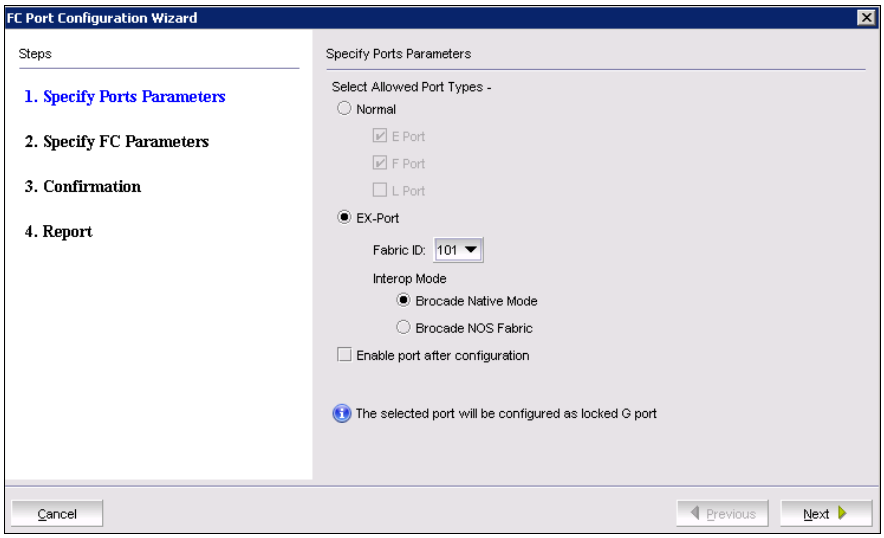


Figure 6-46 Assign fabric ID for EX port on site B

5. After these steps, the new fabric graph is shown (Figure 6-47).

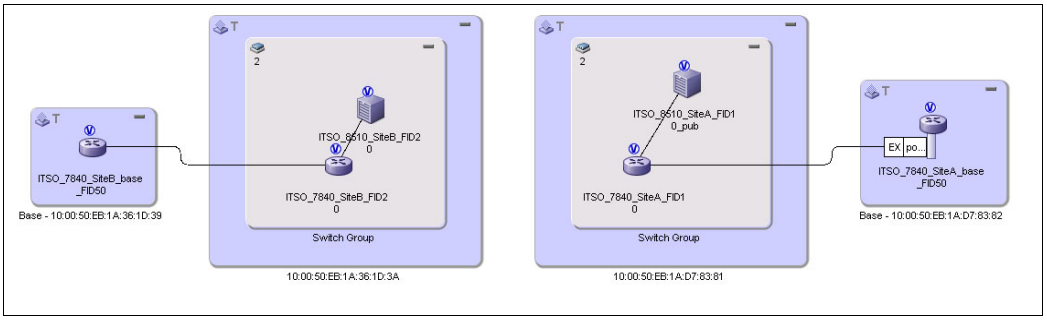


Figure 6-47 Fabric overview after routing configuration

6.3.10 Configuring the tunnel on VE Port 34

In our example, VE Port 34 has four circuits with two failover groups. For tunnel creation, see Chapter 4, “FCIP replication” on page 93.

Enabling the FCIP tunnel with the CLI

In our example, to enable the FCIP tunnel on logical switch `ITSO_7840_SiteA_base_FID50` and `ITSO_7840_SiteB_base_FID50`, VE Port 34 must be enabled by using the **`portcfgpersistentenable`** command, as shown in Example 6-44.

Example 6-44 Enable EX_Port

```
ITSO_7840_SiteA_base_FID50:FID50:admin> portcfgpersistentenable 34
```

After the port is enabled, the status of the tunnel can be verified with the **`portshow fciptunnel`** command as shown in Example 6-45 and Example 6-46.

Example 6-45 Show tunnel status

```
ITSO_7840_SiteA_base_FID50:FID50:admin> portshow fciptunnel

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 34    -         Up       -------a-      49s    0.01    0.01   6      -       -
--------------------------------------------------------------------------------
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext

ITSO_7840_SiteA_base_FID50:FID50:admin>
```

After enabling FCIP tunnel, the base switches `ITSO_7840_SiteA_base_FID50` and `ITSO_7840_SiteA_base_FID50` become a fabric, as shown in Example 6-46.

Example 6-46 fabricshow command line output

```
ITSO_7840_SiteA_base_FID50:FID50:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-------------------------------------------------------------------------
 12: fffc0c 10:00:50:eb:1a:d7:83:82 10.18.228.217   0.0.0.0
"ITSO_7840_SiteA_base_FID50"
```

```
 22: fffc16 10:00:50:eb:1a:36:1d:39 10.18.228.216   0.0.0.0
>"ITSO_7840_SiteB_base_FID50"

The Fabric has 2 switches

ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Example 6-47 shows information about FC routers that exist in our backbone fabric.

Example 6-47 fcrfabricshow command line output

```
ITSO_7840_SiteA_base_FID50:FID50:admin> fcrfabricshow
FC Router WWN: 10:00:50:eb:1a:d7:83:82, Dom ID:  12,
Info: 10.18.228.217, "ITSO_7840_SiteA_base_FID50"
   EX_Port       FID    Neighbor Switch Info (enet IP, WWN, name)
   ------------------------------------------------------------------------
      12        100     10.18.228.217    10:00:50:eb:1a:d7:83:81
"ITSO_7840_SiteA_FID10"

FC Router WWN: 10:00:50:eb:1a:36:1d:39, Dom ID:  22,
Info: 10.18.228.216, "ITSO_7840_SiteB_base_FID50"
   EX_Port       FID    Neighbor Switch Info (enet IP, WWN, name)
   ------------------------------------------------------------------------
      12        101     10.18.228.216    10:00:50:eb:1a:36:1d:3a
"ITSO_7840_SiteB_FID20"

ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Enabling the FCIP tunnel with the IBM Network Advisor GUI

To create the FCIP tunnel between the two base logical switches (`ITSO_7840_SiteA_Base_FID50` and `ITSO_7840_SiteB_Base_FID50`), repeat step 20 on page 169 for FCIP tunnel creation (Figure 6-48). For more information about FCIP tunnel implementation, see Chapter 4, “FCIP replication” on page 93.

In our example, VE Port 34 has four circuits with two failover groups (Figure 6-48).

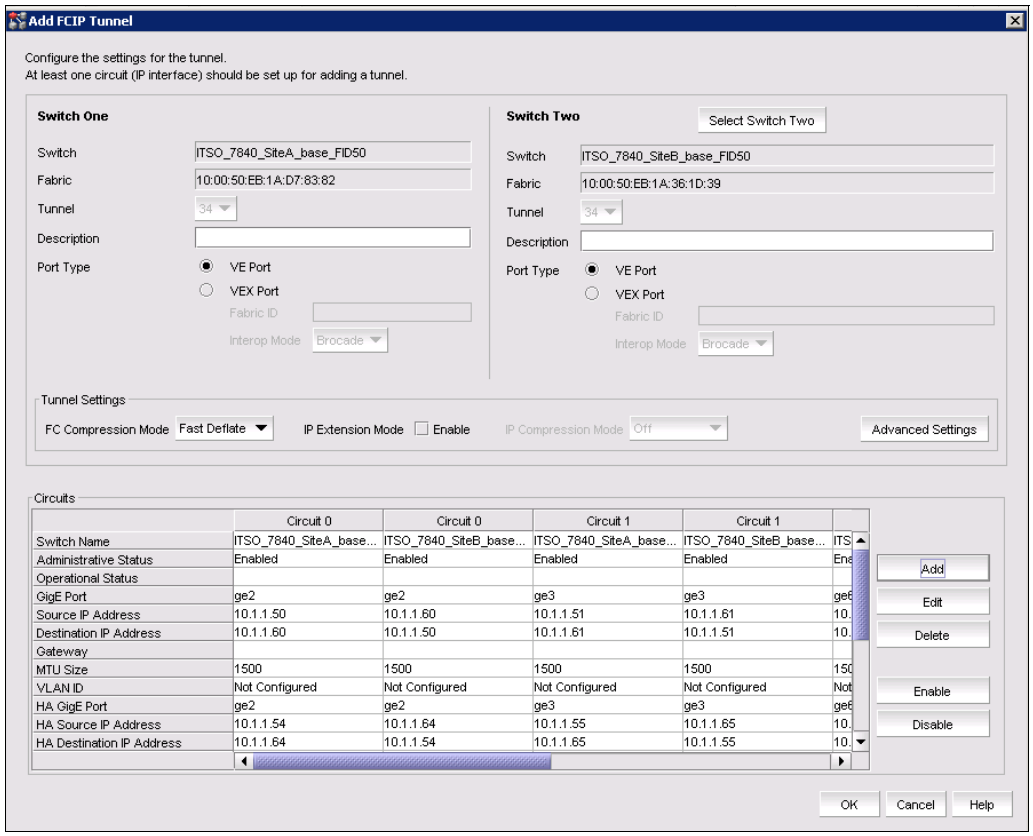


Figure 6-48 Create FCIP tunnel on base logical switches

After creating the FCIP tunnel, we complete all of the configuration for this scenario. Figure 6-49 shows the final graph of fabrics.

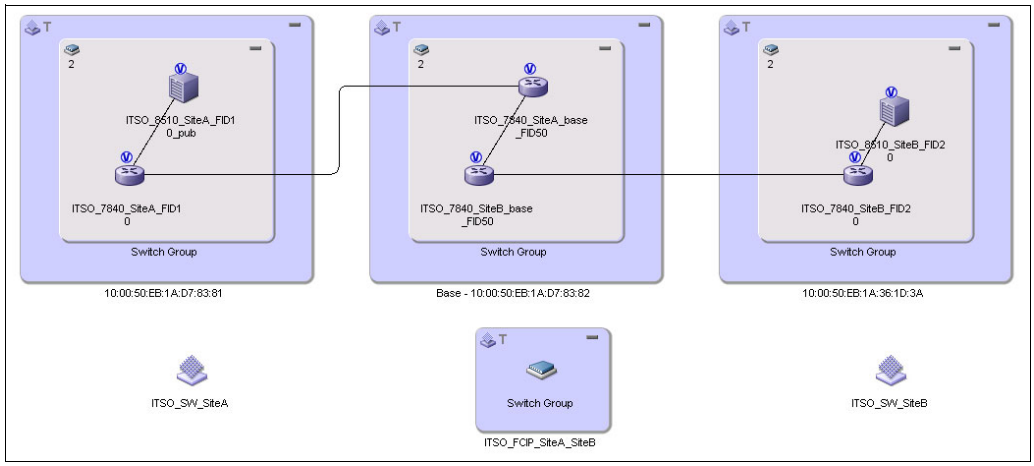


Figure 6-49 Final status of fabrics

6.3.11 Creating LSAN zones

The following requirements must be noted before you implement LSAN zoning:

- LSAN zones begin with the characters LSAN_ (LSAN_ZoneName), lsan_(lsan_ZoneName), or Lsan_ (Lsan_ZoneName).
- LSAN zone names do not have to match in edge and backbone fabric.
- The LSAN zone must be defined and enabled in each fabric that shares a particular device. In our case, we need to create a separate LSAN zone in each switch.
- LSAN zone members must be identified by their PWWN.

You must create an LSAN zone in edge fabric site A and site B. LSAN zone creation in the backbone fabric is optional.

Creating LSAN zones with the CLI

In our example, an LSAN zone must be created, containing IBM SAN Volume Controller ports connected to edge fabric in site A and IBM Storwize V7000 storage system ports connected to the edge fabric in site B.

Example 6-48 shows the creation of LSAN zones in edge fabrics within site A.

Example 6-48 ITSO_SiteA_FID10_pub edge fabric: Create and activate LSAN zone

```
TSO_7840_SiteA_FID10:FID10:admin> zonecreate
"LSAN_zone_SVC_SiteA_V7000_SiteB","50:05:07:68:02:10:a7:fe;50:05:07:68:02:20:a7:be
;50:05:07:68:02:20:a7:fe;50:05:07:68:02:10:a7:be;50:05:07:68:0c:23:00:00;50:05:07:
68:0c:24:00:00;50:05:07:68:0c:23:05:08;50:05:07:68:0c:24:05:08"
ITSO_7840_SiteA_FID10:FID10:admin> cfgadd
"ITSO_SiteA_PUB","LSAN_zone_SVC_SiteA_V7000_SiteB"
ITSO_7840_SiteA_FID10:FID10:admin> cfgsave
WARNING!!!
The changes you are attempting to save will render the
Effective configuration and the Defined configuration
inconsistent. The inconsistency will result in different
Effective Zoning configurations for switches in the fabric if
a zone merge or HA failover happens. To avoid inconsistency
it is recommended to commit the configurations using the
'cfgenable' command.

Do you want to proceed with saving the Defined
zoning configuration only?  (yes, y, no, n): [no] y
Updating flash ...
ITSO_7840_SiteA_FID10:FID10:admin> cfgenable "ITSO_SiteA_PUB"
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected. If the update includes changes
to one or more traffic isolation zones, the update may result in
localized disruption to traffic on ports associated with
the traffic isolation zone changes
Do you want to enable 'ITSO_SiteA_PUB' configuration  (yes, y, no, n): [no] y
zone config "ITSO_SiteA_PUB" is in effect
Updating flash ...
ITSO_7840_SiteA_FID10:FID10:admin>
```

Example 6-49 shows the creation of a new LSAN zone and of a new configuration set within the edge fabric in site B.

Example 6-49 ITSO_SiteB_FID20_pub edge fabric: Create and activate of LSAN zone

```
ITSO_7840_SiteB_FID20:FID20:admin> zonecreate
"LSAN_zone_SVC_SiteA_V7000_SiteB","50:05:07:68:02:10:a7:fe;50:05:07:68:02:20:a7:be
;50:05:07:68:02:20:a7:fe;50:05:07:68:02:10:a7:be;50:05:07:68:0c:23:00:00;50:05:07:
68:0c:24:00:00;50:05:07:68:0c:23:05:08;50:05:07:68:0c:24:05:08"
ITSO_7840_SiteB_FID20:FID20:admin> cfgcreate
"ITSO_SiteB_PUB","LSAN_zone_SVC_SiteA_V7000_SiteB"
ITSO_7840_SiteB_FID20:FID20:admin>
ITSO_7840_SiteB_FID20:FID20:admin>
ITSO_7840_SiteB_FID20:FID20:admin> cfgsave
You are about to save the Defined zoning configuration. This
action will only save the changes on Defined configuration.
If the update includes changes to one or more traffic isolation
zones, you must issue the 'cfgenable' command for the changes
to take effect.
Do you want to save the Defined zoning configuration only?  (yes, y, no, n): [no]
y
Updating flash ...
ITSO_7840_SiteB_FID20:FID20:admin> cfgenable "ITSO_SiteB_PUB"
You are about to enable a new zoning configuration.
This action will replace the old zoning configuration with the
current configuration selected. If the update includes changes
to one or more traffic isolation zones, the update may result in
localized disruption to traffic on ports associated with
the traffic isolation zone changes
Do you want to enable 'ITSO_SiteB_PUB' configuration  (yes, y, no, n): [no] y
zone config "ITSO_SiteB_PUB" is in effect
Updating flash ...
ITSO_7840_SiteB_FID20:FID20:admin>
```

The LSAN zoning for edge fabrics is completed.

Creating LSAN zones with the IBM Network Advisor GUI

To create LSAN zones with the GUI, follow these steps:

1. Select a backbone fabric from the product list, and then click **Configure** → **Zoning** → **LSAN Zoning** (**Device Sharing**) as shown in Figure 6-50.

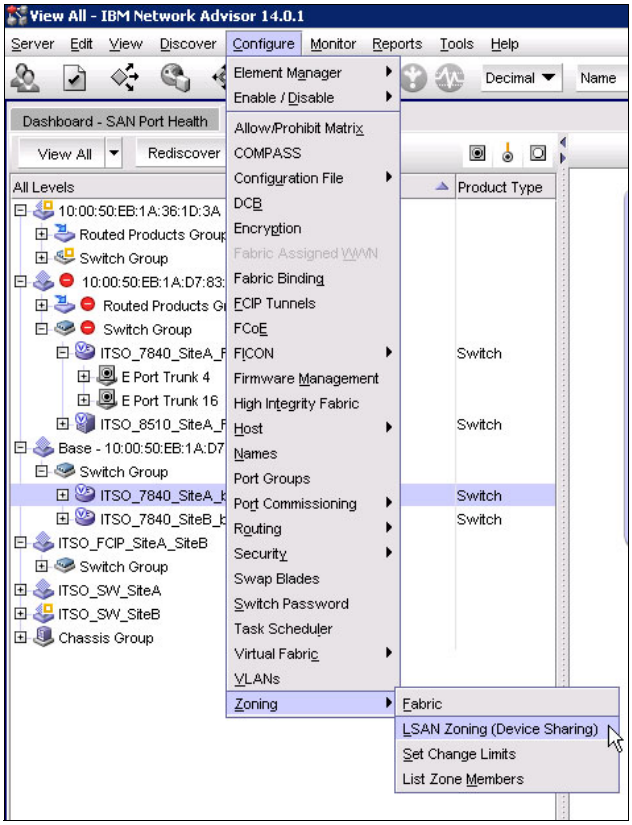


Figure 6-50 Menu for LSAN Zoning

2. The Zoning window is displayed (Figure 6-51).

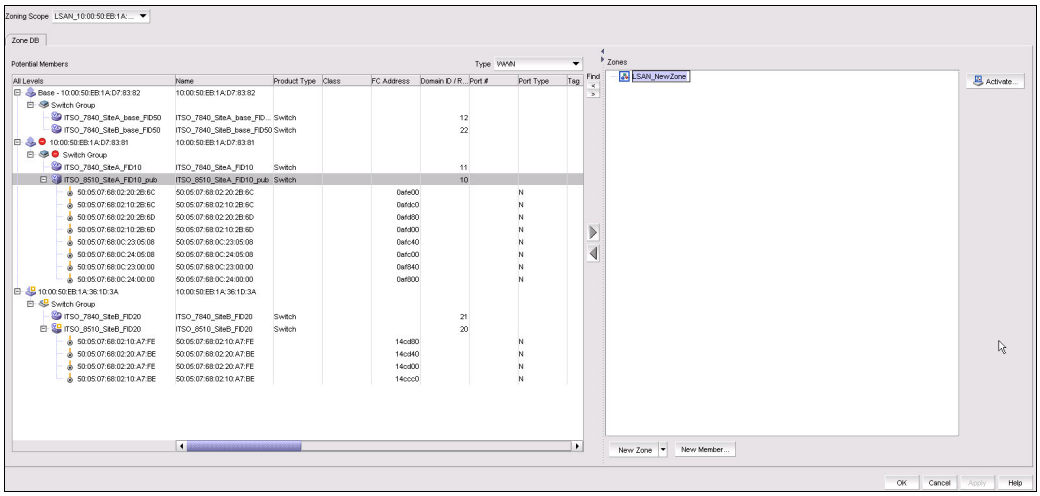


Figure 6-51 LSAN Zoning window

3. Click **New Zone** and the prefix **LSAN_** is automatically added in the text field of the new LSAN zone.
4. Enter a name for the zone and press **Enter**.
5. Add members to the new LSAN Zone:
   a. Select members that are required to add to this LSAN zone from the **Potential Members** list.
   b. Click the right arrow between the **Potential Members** list and the **Zones** list to add them to the zone (Figure 6-52).

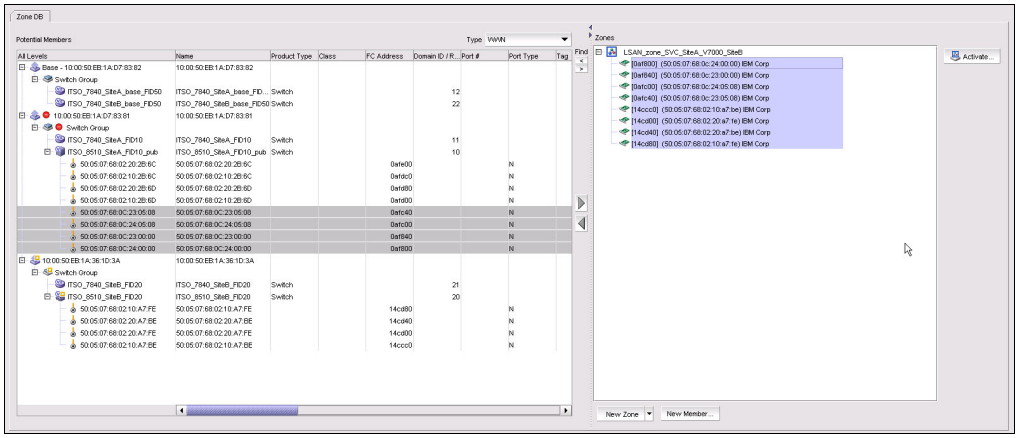


Figure 6-52 Create a LSAN zone

6. Click **Active** to activate the LSAN Zones that we just added.
7. Review the information in the Activate LSAN Zones window and click **OK** to confirm the activation (Figure 6-53).

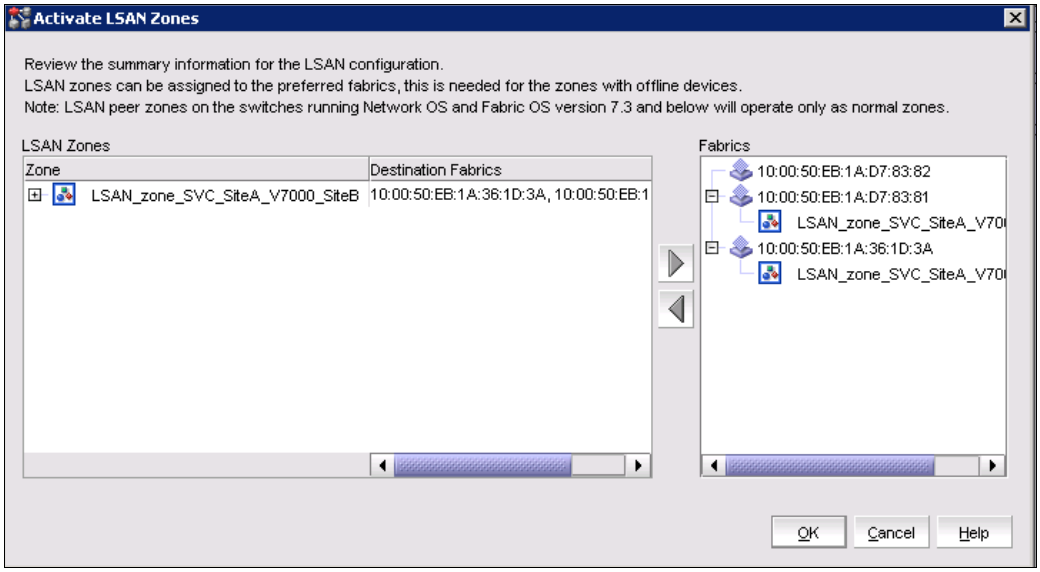


Figure 6-53 Active LSAN Zoning

6.3.12 Verification

This section describes the verification of configuration steps on edge fabrics and backbone fabrics.

Edge fabric verification

The following example shows the **fabricshow** command line output on edge fabric switches ITSO_8510_SiteA_FID10_pub and ITSO_8510_SiteB_FID20.

Example 6-50 shows all logical switches that belong to the edge fabric in site A. fcr_fd_160 is the backbone switch in site A, and fcr_xd_200_101 is the backbone switch in site B.

Example 6-50 ITSO_SiteA_FID10: edge fabric: fabricshow command output

```
ITSO_8510_SiteA_FID10_pub:FID10:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-------------------------------------------------------------------------
 10: fffc0a 10:00:00:05:33:96:f4:02 10.18.228.35    0.0.0.0
>"ITSO_8510_SiteA_FID10_pub"
 11: fffc0b 10:00:50:eb:1a:d7:83:81 10.18.228.217   0.0.0.0
"ITSO_7840_SiteA_FID10"
160: fffca0 55:0e:b1:ad:78:38:2e:64 0.0.0.0         0.0.0.0 "fcr_fd_160"
200: fffcc8 55:0e:b1:ad:78:3c:0f:09 0.0.0.0         0.0.0.0 "fcr_xd_200_101"
The Fabric has 4 switches
ITSO_8510_SiteA_FID10_pub:FID10:admin>
```

Example 6-51 shows all logical switches that belong to the edge fabric in site B. fcr_fd_160 is the backbone switch in site A, and fcr_xd_200_101 is the backbone switch in site B.

Example 6-51 ITSO_SiteB_FID20: edge fabric: fabricshow command output

```
ITSO_8510_SiteB_FID20:FID20:admin> fabricshow
Switch ID   Worldwide Name          Enet IP Addr    FC IP Addr      Name
-------------------------------------------------------------------------
 20: fffc14 10:00:00:05:1e:46:8a:01 10.18.228.106   0.0.0.0
>"ITSO_8510_SiteB_FID20"
 21: fffc15 10:00:50:eb:1a:36:1d:39 10.18.228.216   0.0.0.0
"ITSO_7840_SiteB_FID20"
160: fffca0 55:0e:b1:a3:61:d3:ae:65 0.0.0.0         0.0.0.0 "fcr_fd_160"
200: fffcc8 55:0e:b1:a3:61:d7:8f:06 0.0.0.0         0.0.0.0 "fcr_xd_200_100"
The Fabric has 4 switches
ITSO_8510_SiteB_FID20:FID20:admin>
```

Example 6-52 shows the **nsshow** command line output on ITSO_8510_SiteA_FID10. You can use the **nsshow** command to display local name server information for devices connected to the switch.

Example 6-52 ITSO_8510_SiteA_FID10_pub edge fabric: nsshow

```
ITSO_8510_SiteA_FID10_pub:FID10:admin> nsshow
{
 Type Pid    COS     PortName                NodeName                 TTL(sec)
 N    0af800;      3;50:05:07:68:0c:24:00:00;50:05:07:68:0c:00:00:00; na
    FC4s: FCP [IBM     2145            0000]
    Fabric Port Name: 20:1f:00:05:33:96:f4:02
    Permanent Port Name: 50:05:07:68:0c:24:00:00
```

```
   Port Index: 31
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
   LSAN: Yes
N    0af840;      3;50:05:07:68:0c:23:00:00;50:05:07:68:0c:00:00:00; na
   FC4s: FCP
   Fabric Port Name: 20:1e:00:05:33:96:f4:02
   Permanent Port Name: 50:05:07:68:0c:23:00:00
   Port Index: 30
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
   LSAN: Yes
N    0afc00;      3;50:05:07:68:0c:24:05:08;50:05:07:68:0c:00:05:08; na
   FC4s: FCP [IBM     2145            0000]
   Fabric Port Name: 20:0f:00:05:33:96:f4:02
   Permanent Port Name: 50:05:07:68:0c:24:05:08
   Port Index: 15
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
   LSAN: Yes
N    0afc40;      3;50:05:07:68:0c:23:05:08;50:05:07:68:0c:00:05:08; na
   FC4s: FCP
   Fabric Port Name: 20:0e:00:05:33:96:f4:02
   Permanent Port Name: 50:05:07:68:0c:23:05:08
   Port Index: 14
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
   LSAN: Yes
N    0afd00;      3;50:05:07:68:02:10:2b:6d;50:05:07:68:02:00:2b:6d; na
   FC4s: FCP [IBM     2145            0000]
   Fabric Port Name: 20:0b:00:05:33:96:f4:02
   Permanent Port Name: 50:05:07:68:02:10:2b:6d
   Port Index: 11
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
   LSAN: No
N    0afd80;      3;50:05:07:68:02:20:2b:6d;50:05:07:68:02:00:2b:6d; na
   FC4s: FCP [IBM     2145            0000]
   Fabric Port Name: 20:09:00:05:33:96:f4:02
   Permanent Port Name: 50:05:07:68:02:20:2b:6d
   Port Index: 9
   Share Area: Yes
   Device Shared in Other AD: No
   Redirect: No
   Partial: No
```

```
    LSAN: No
 N    0afdc0;      3;50:05:07:68:02:10:2b:6c;50:05:07:68:02:00:2b:6c; na
    FC4s: FCP
    Fabric Port Name: 20:08:00:05:33:96:f4:02
    Permanent Port Name: 50:05:07:68:02:10:2b:6c
    Port Index: 8
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: No
 N    0afe00;      3;50:05:07:68:02:20:2b:6c;50:05:07:68:02:00:2b:6c; na
    FC4s: FCP [IBM     2145            0000]
    Fabric Port Name: 20:07:00:05:33:96:f4:02
    Permanent Port Name: 50:05:07:68:02:20:2b:6c
    Port Index: 7
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: No
The Local Name Server has 8 entries }
ITSO_8510_SiteA_FID10_pub:FID10:admin>
```

Example 6-53 shows the **nsshow** command line output on ITSO_8510_SiteB_FID20.

Example 6-53 ITSO_S8510_SiteB_FID20 edge fabric: nsshow command output

```
ITSO_8510_SiteB_FID20:FID20:admin> nsshow
{
 Type Pid    COS     PortName                NodeName                  TTL(sec)
 N    14ccc0;      3;50:05:07:68:02:10:a7:be;50:05:07:68:02:00:a7:be; na
    FC4s: FCP
    Fabric Port Name: 20:cc:00:05:1e:46:8a:01
    Permanent Port Name: 50:05:07:68:02:10:a7:be
    Port Index: 204
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: Yes
 N    14cd00;      3;50:05:07:68:02:20:a7:fe;50:05:07:68:02:00:a7:fe; na
    FC4s: FCP [IBM     2145            0000]
    Fabric Port Name: 20:cb:00:05:1e:46:8a:01
    Permanent Port Name: 50:05:07:68:02:20:a7:fe
    Port Index: 203
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: Yes
 N    14cd40;      3;50:05:07:68:02:20:a7:be;50:05:07:68:02:00:a7:be; na
    FC4s: FCP
    Fabric Port Name: 20:ca:00:05:1e:46:8a:01
    Permanent Port Name: 50:05:07:68:02:20:a7:be
    Port Index: 202
```

```
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: Yes
 N    14cd80;      3;50:05:07:68:02:10:a7:fe;50:05:07:68:02:00:a7:fe; na
    FC4s: FCP [IBM     2145            0000]
    Fabric Port Name: 20:c9:00:05:1e:46:8a:01
    Permanent Port Name: 50:05:07:68:02:10:a7:fe
    Port Index: 201
    Share Area: Yes
    Device Shared in Other AD: No
    Redirect: No
    Partial: No
    LSAN: Yes
The Local Name Server has 4 entries }
```

Backbone fabric

Example 6-54 shows the dedicated host and storage ports of our IBM SAN Volume Controller within site A as *EXIST* and the IBM Storwize V7000 storage system ports within site B as *imported* on the `ITSO_7840_SiteA_base_FID50` switch.

Example 6-54 Show lsan configuration on backbone fabric

```
ITSO_7840_SiteA_base_FID50:FID50:admin> lsanzoneshow -s
Fabric ID: 100 Zone Name: LSAN_zone_SVC_SiteA_V7000_SiteB
        50:05:07:68:02:10:a7:fe  Imported
        50:05:07:68:02:20:a7:be  Imported
        50:05:07:68:02:20:a7:fe  Imported
        50:05:07:68:02:10:a7:be  Imported
        50:05:07:68:0c:23:00:00  EXIST
        50:05:07:68:0c:24:00:00  EXIST
        50:05:07:68:0c:23:05:08  EXIST
        50:05:07:68:0c:24:05:08  EXIST
Fabric ID: 101 Zone Name: LSAN_zone_SVC_SiteA_V7000_SiteB
        50:05:07:68:02:10:a7:fe  EXIST
        50:05:07:68:02:20:a7:be  EXIST
        50:05:07:68:02:20:a7:fe  EXIST
        50:05:07:68:02:10:a7:be  EXIST
        50:05:07:68:0c:23:00:00  Imported
        50:05:07:68:0c:24:00:00  Imported
        50:05:07:68:0c:23:05:08  Imported
        50:05:07:68:0c:24:05:08  Imported
ITSO_7840_SiteA_base_FID50:FID50:admin>
```

Example 6-55 shows the dedicated host/storage ports of our IBM SAN Volume Controller within site A as *Imported* and the IBM Storwize V7000 storage system ports within site B as *EXIST* on `ITSO_7840_SiteB_base_FID50` switch.

Example 6-55 ITSO_SiteB_FID20_pub edge fabric: Create and activate of LSAN zone

```
ITSO_7840_SiteB_FID50:FID50:admin> lsanzoneshow -s
Fabric ID: 100 Zone Name: LSAN_zone_SVC_SiteA_V7000_SiteB
        50:05:07:68:02:10:a7:fe  Imported
        50:05:07:68:02:20:a7:be  Imported
```

```
        50:05:07:68:02:20:a7:fe  Imported
        50:05:07:68:02:10:a7:be  Imported
        50:05:07:68:0c:23:00:00  EXIST
        50:05:07:68:0c:24:00:00  EXIST
        50:05:07:68:0c:23:05:08  EXIST
        50:05:07:68:0c:24:05:08  EXIST
Fabric ID: 101 Zone Name: LSAN_zone_SVC_SiteA_V7000_SiteB
        50:05:07:68:02:10:a7:fe  EXIST
        50:05:07:68:02:20:a7:be  EXIST
        50:05:07:68:02:20:a7:fe  EXIST
        50:05:07:68:02:10:a7:be  EXIST
        50:05:07:68:0c:23:00:00  Imported
        50:05:07:68:0c:24:00:00  Imported
        50:05:07:68:0c:23:05:08  Imported
        50:05:07:68:0c:24:05:08  Imported
ITSO_7840_SiteB_FID50:FID50:admin>
```

6.3.13 Providing a quorum from V7000 Site B to the IBM SAN Volume Controller stretched cluster in Site A

Figure 6-54 shows that the IBM Storwize V7000 storage system located on site B can be configured as a back end storage on our IBM SAN Volume Controller stretched cluster system within Site A.

The IBM Storwize V7000 storage system located in site B can be used for IBM SAN Volume Controller stretched cluster quorum disk configuration (Figure 6-54).

Figure 6-54 Controllers available on IBM San Volume Controller Stretched cluster on site a.

For more comprehensive information about IBM SAN Volume Controller stretched cluster configuration, see IBM Knowledge Center and *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

7

Troubleshooting and monitoring

This chapter gives you an overview of available troubleshooting tools.

It provides the following information:

- General problem determination
- IBM Network Advisor
- Using the portshow command
- WAN tool analysis

7.1 General problem determination

This section introduces available troubleshooting tools and provides a set of methods.

7.2 IBM Network Advisor

IBM Network Advisor is a management application that provides easy and centralized management of the network, and quick access to all product configuration applications. You can configure, manage, and monitor a network with ease by using this application.

Consult your IBM Network Documentation. You can also refer to the *Brocade Network Advisor SAN User Manual*, 53-1004147-01, at the following website:

`http://my.brocade.com`

7.2.1 The IBM Network Advisor dashboard

The IBM Network Advisor dashboard can be accessed by clicking the Dashboard - Product Status and Traffic tab on the main window. It provides a high-level overview of the network and the current state of the management devices. You can use the dashboards to easily check the status of the devices in the network. The dashboards also provide several features to help you quickly access reports, device configurations, and system logs.

Figure 7-1 shows the IBM Network Advisor Dashboard main window.

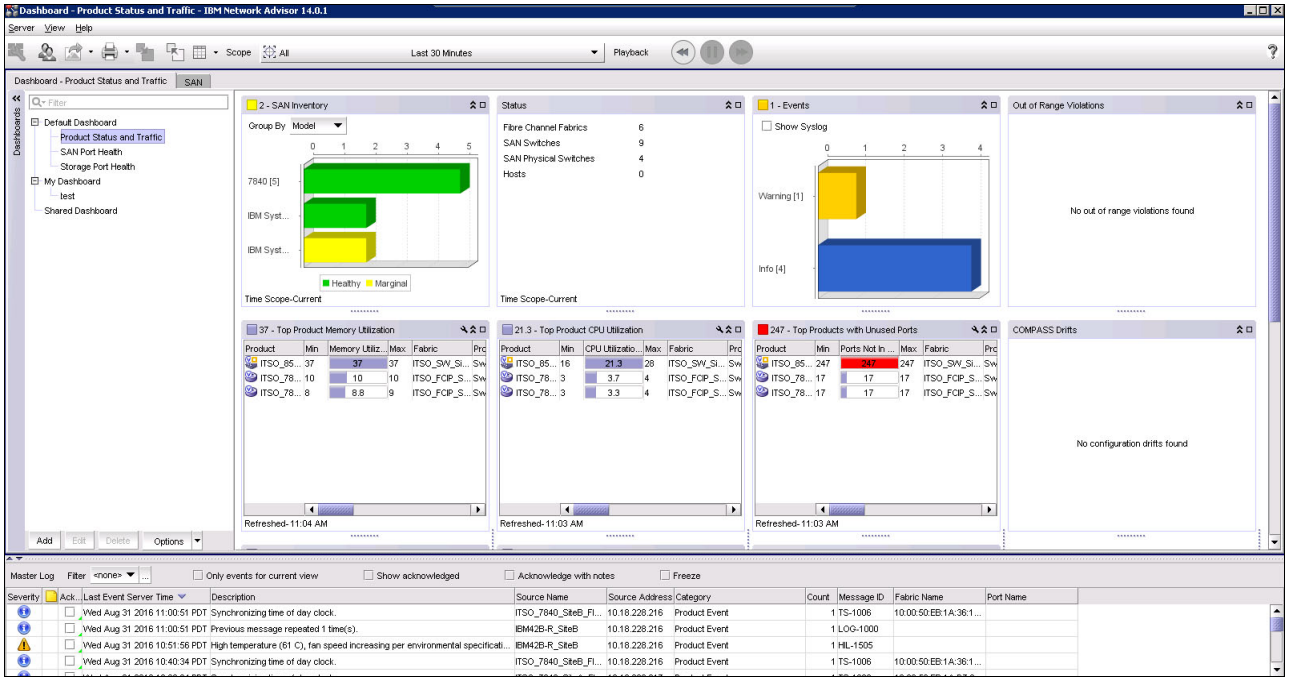


Figure 7-1 Dashboard of IBM Network Advisor

The dashboard includes the following components:

- Menu bar: Lists commands that you can run on the dashboard. The dashboard also provides a menu to reset the dashboard back to the defaults. You can reset the dashboard back to the default settings by right-clicking and selecting **Reset to Default**.
- Toolbar: Provides buttons that enable quick access to windows and functions.
- Dashboard tab: Provides a high-level overview of the network that is managed by the management application server.
- SAN: Displays the master log, Minimap, Connectivity map (topology), and product list.
- Widgets: Displays the operational status, inventory status, event summary, performance monitors, and overall network or fabric status.
- Master log: Displays all events that have occurred on the management application.
- Status bar: Displays the connection, port, product, fabric, special event, Call Home, backup status, and server and user data.

The Dashboard can also be customized according to specific requirements, such as network scope and time interval.

For more information about these IBM Network Advisor dashboard features and the multiple dashboard widgets, see *IBM Network Advisor SAN User Manual*, SC27-5423-01.

7.2.2 MAPS

The Monitoring and Alerting Policy Suite (MAPS) is an optional network health monitor that allows you to enable each switch to constantly monitor its Ethernet fabric for potential faults. MAPS tracks various Ethernet fabric measures and events. Monitoring Ethernet fabric-wide events, ports, and environmental parameters enables early fault detection and isolation, as well as performance measurement. For more information about MAPS, see Chapter 2, “The IBM System Storage SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade” on page 13.

MAPS license requirements

MAPS is supported on Fabric OS devices running that are running version 7.1 or earlier with the Fabric Watch and Performance Monitor license.

MAPS is supported on Fabric OS devices running 7.2 or later with the Fabric Vision license. MAPS must be enabled on the device. For how to enable MAPS, see “Enabling MAPS on a device” on page 216.

Enabling MAPS on a device

Follow these steps to enable MAPS on switches:

1. Click **Monitor → Fabric Vision → MAPS → Enable**. The Enable MAPS window is displayed (Figure 7-2).

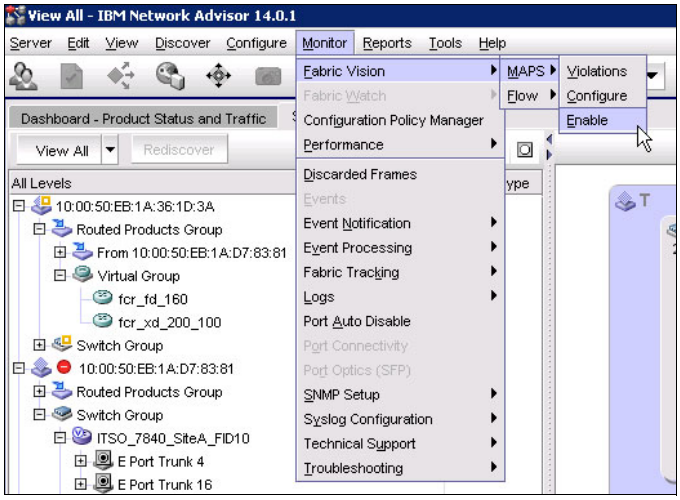


Figure 7-2 Menu of enabling MAPS

2. Select one or more switches on which you want to enable MAPS in the **Available Switches** list.
3. Click the right arrow button to move the selected switches to the correct list.
4. Click **OK** to confirm the setting and complete the task.

Enable or disable policy actions for policies

A MAPS policy is a set of rules that define thresholds for measures and action to take when a threshold is triggered. When you enable a policy, all of the rules in the policy are in effect. A rule associates a condition with actions that need to be triggered when the specified condition is evaluated to be true.

Note: At any time, one policy must always be active on the switch. You can have an active policy with no rules, but you must have an active policy. You cannot disable the active policy. You can only change the active policy by enabling a different policy.

MAPS provides actions (event notifications) in several different formats to ensure that event details are accessible from all platforms and operating systems. In response to an event, MAPS can record event data as any of the alarm options.

You can define what actions are allowable on the device, regardless of the actions specified in the individual rules in a policy. To enable or disable policy actions, follow these steps:

1. Right-click a device in the Product List or Connectivity Map and select **Fabric Vision** → **MAPS** → **Configure** (Figure 7-3).

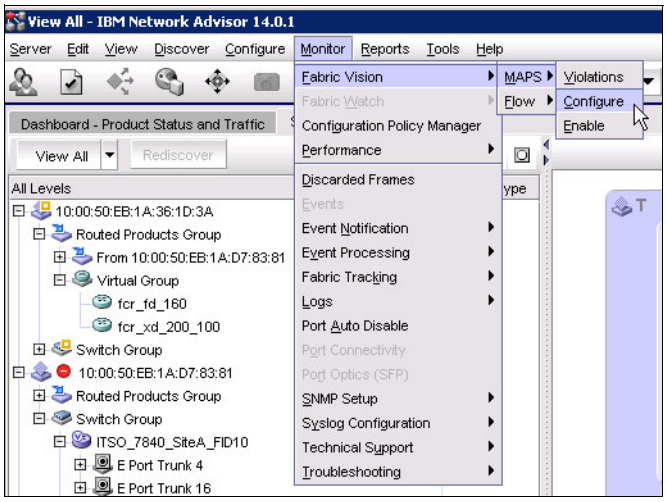


Figure 7-3 Menu of MAPS configuration

2. The MAPS Configuration dialog box displays (Figure 7-4 on page 218).

   To configure MAPS actions for all policy rules on all fabrics, select **All fabrics** and click **Actions**.

   To configure MAPS actions for all policy rules on the selected fabrics, select one or more fabrics, and click **Actions**.

   To configure MAPS actions for all policy rules on the selected switches, select one or more switches and click **Actions**.

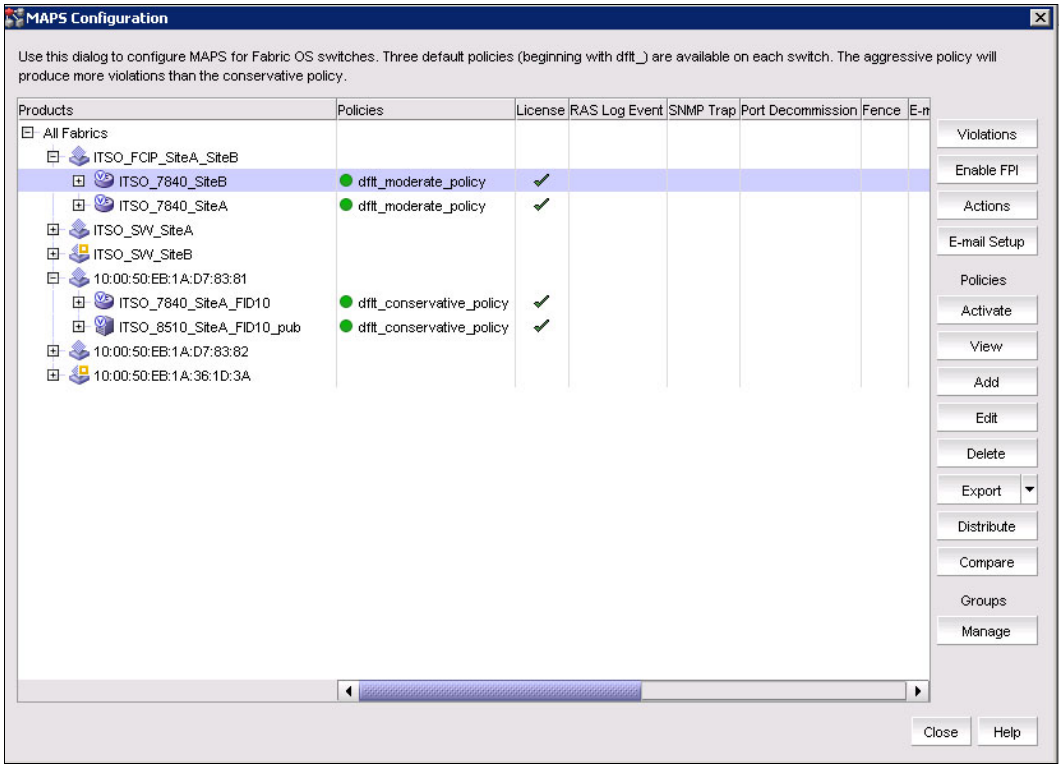


Figure 7-4 Dialog box for MAPS configuration

3. Select the associated check box for each action you want to enable.

   Several actions can be enabled or disabled:

   – **RAS Log Event**: Following an event, MAPS adds an entry to the internal event log for an individual switch. The RAS log stores event information, but does not actively send alerts.

   – **SNMP traps**: SNMP performs an operation called a trap that notifies a management station using SNMP when events occur. For more details about SNMP configuration, see the *Monitoring and Alerting Policy Suite Administrator's Guide* that is available at the following website:

     `http://my.brocade.com`

   – **Fence**: Port fencing takes the ports offline if the user-defined thresholds are exceeded.

   – **Email**: An email alert sends information about a switch event to a specified email address.

   – **SFP Marginal**: When a threshold is triggered, the SFP transceiver status changes to a marginal status icon on the Dashboard and SAN tabs.

   – **Switch Status Critical**: Set the switch status to critical.

   – **Switch Status Marginal**: Set the switch status to marginal.

Enable all actions by clicking **Enable All**. Disable all actions by clicking **Disable All** (Figure 7-5).

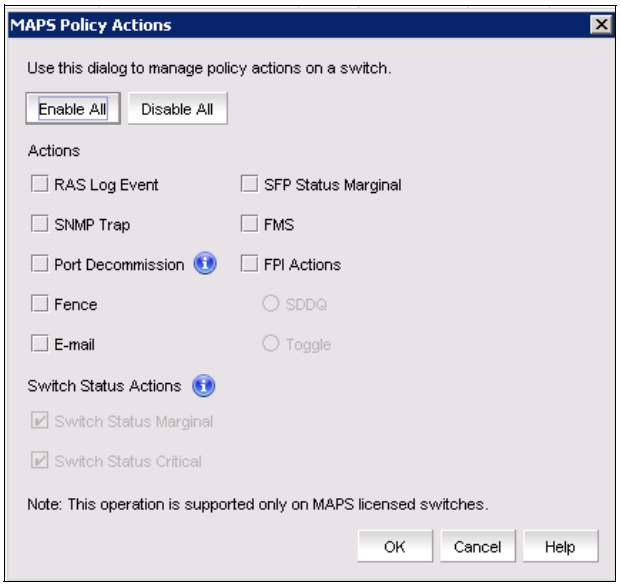


Figure 7-5 Enable or disable MAPS policy actions

4. Click **OK** in the **MAPS Policy Actions** window to complete this task.

Configuring a MAPS policy

Follow these steps to configure a MAPS policy:

1. Right-click a device in the Product List or Connectivity Map and select **Fabric Vision** → **MAPS** → **Configure**. The **MAPS configuration** window is displayed (Figure 7-4 on page 218).

2. Click **Add** to add a policy (Figure 7-6).

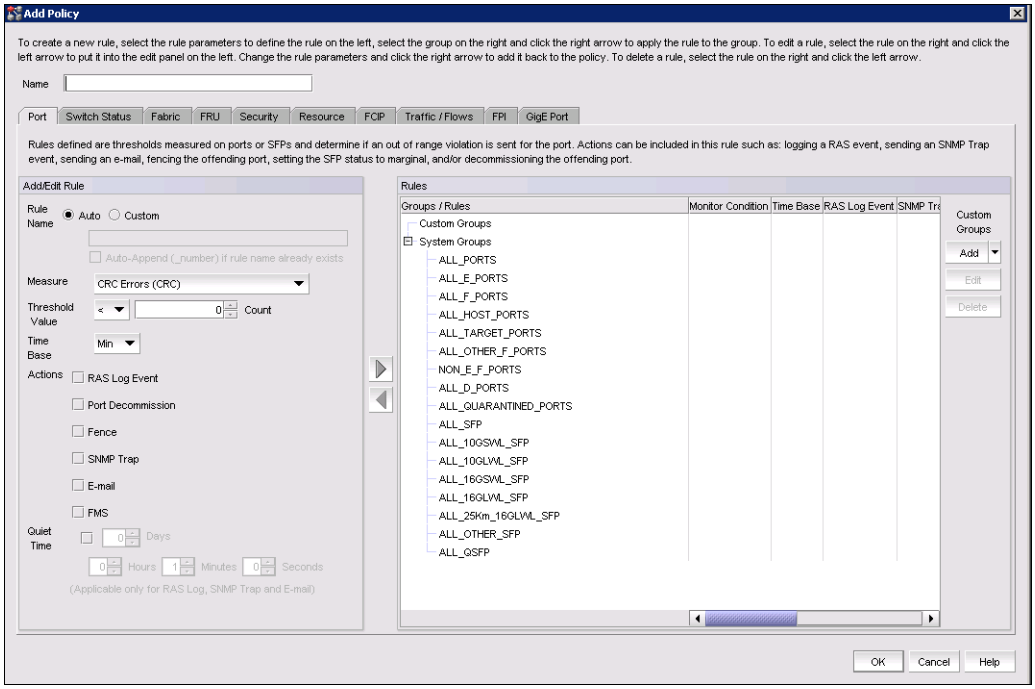


Figure 7-6 Add a policy

3. Enter a name of the policy in the **Name** field:

> **Note:** The policy name can be up to 32 characters and can only contain of alphanumeric and underscore characters.

– Select one of the category tabs to configure the policy measures. Here we select FCIP. For a complete list of categories and the associated measures and actions, see *Monitoring and Alerting Policy Suite Administrator's Guide* that is available at the following website:

`http://my.brocade.com`

– Select the **Auto** option (default) to generate the rule name automatically in the **Rule Name** area or select the **Customer** option to provide a user-defined name.

– Select a measure from the **Measure** list. For a complete list of categories and the associated measures and actions, see *Monitoring and Alerting Policy Suite Administrator's Guide* that is available at the following website:

`http://my.brocade.com`

– Select a logical operator from the Threshold list and enter a threshold value in the **Threshold** field.

– Select one of the following durations to monitor the counter from the **Time Base** list.

– From the **Actions** check boxes, select the check box for each action you want to occur when a threshold is crossed.

– Add the rule to a group by selecting the group in the **Rules** area and clicking the right arrow button to move the new rule to the selected group.

4. Click **OK** to add the policy to the MAPS Configuration window (Figure 7-7).

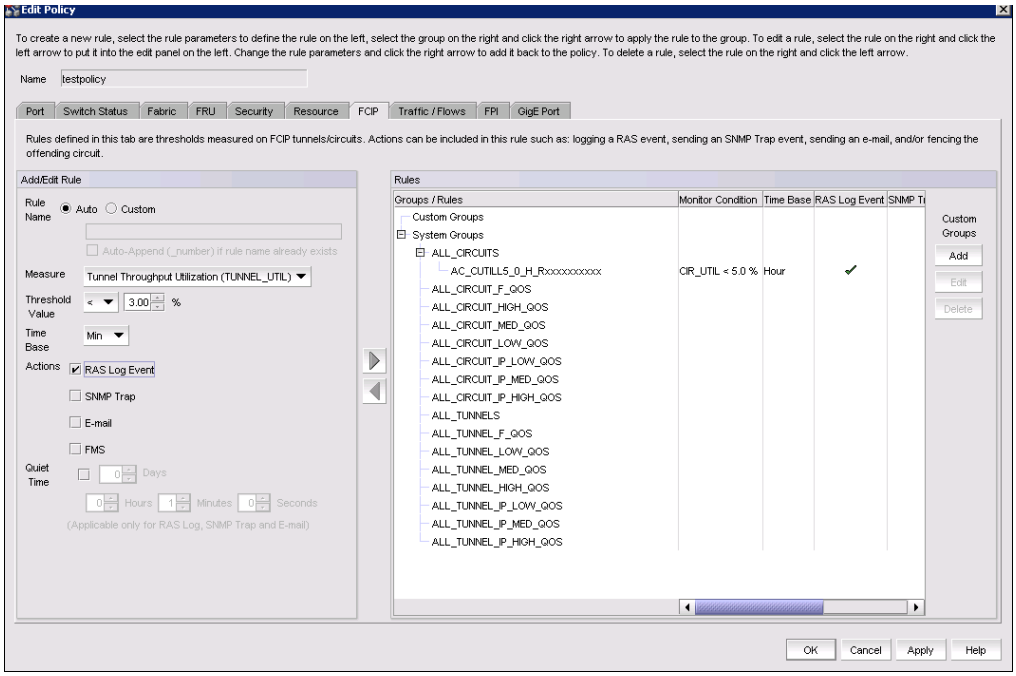


Figure 7-7 Details of MAPS configuration

Activating a MAPS policy

The new MAPS policy cannot be active automatically after creation. To active a MAPS policy, complete these steps:

1. Right-click a device in the Product List or Connectivity Map and select **Fabric Vision** → **MAPS** → **Configure**. The MAPS configuration window is displayed (Figure 7-4 on page 218).

2. Select an inactive policy in the list and click **Activate** (Figure 7-8 on page 222). The inactive policies have no green indicator on the left.

Note: Only one policy can be active on a switch at a time. You can activate policies for multiple switches at once by selecting the policy you want to activate for each switch and clicking **Activate**.

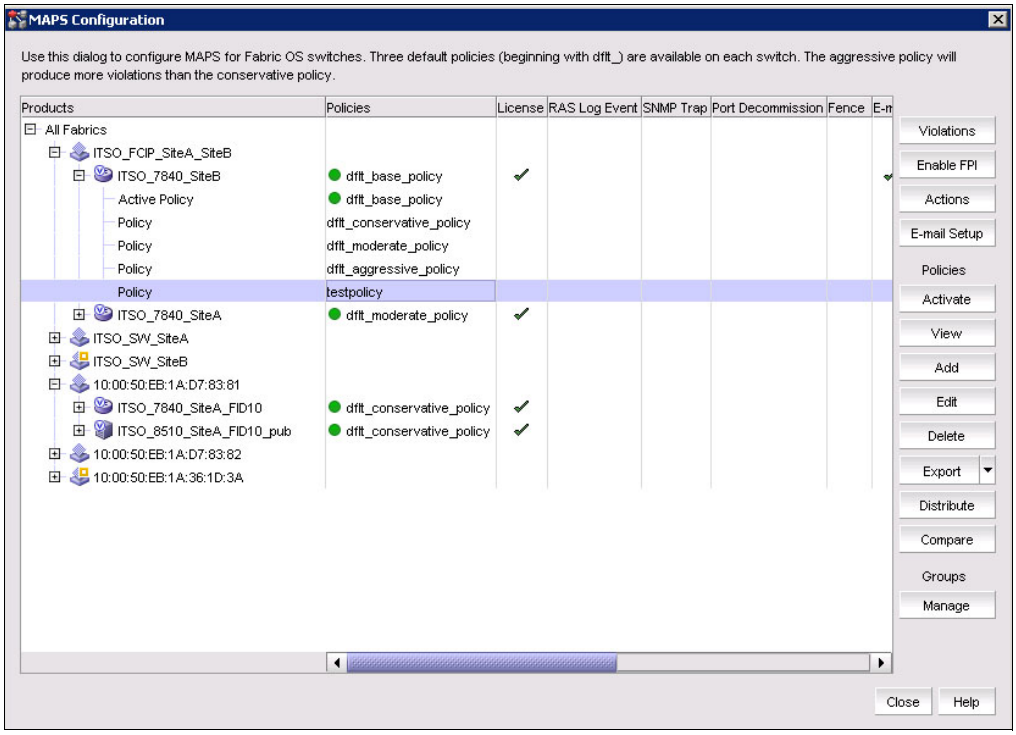


Figure 7-8 Active MAPS policy

Deleting a MAPS policy

Follow these steps to delete a MAPS policy:

1. Right-click a device in the Product List or Connectivity Map and select **Fabric Vision** → **MAPS** → **Configure**. The MAPS configuration window is displayed (Figure 7-4 on page 218).
2. Select the policies that you want to delete in the list and click **Delete**. You can delete one or more policies from switches.

Note: The default or active policies cannot be deleted.

3. Click **Yes** on the confirmation message (Figure 7-9).

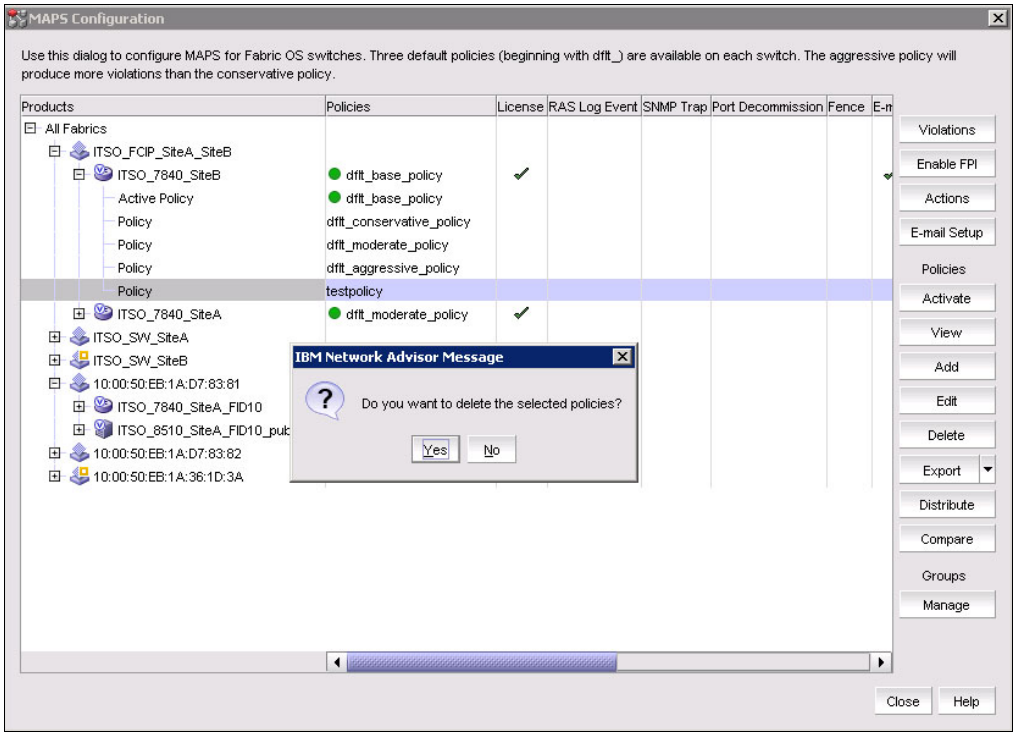


Figure 7-9 Delete MAPS policy

Viewing MAPS violations and events

MAPS violation data is stored in the database for 30 days. The system purges old data (older than 30 days) every night at 12 midnight.

To view MAPS violations, follow these steps:

1. Right-click a device in the Product List or Connectivity Map and select **Fabric Vision → MAPS → Violations**. The MAPS Violations window is displayed (Figure 7-10).

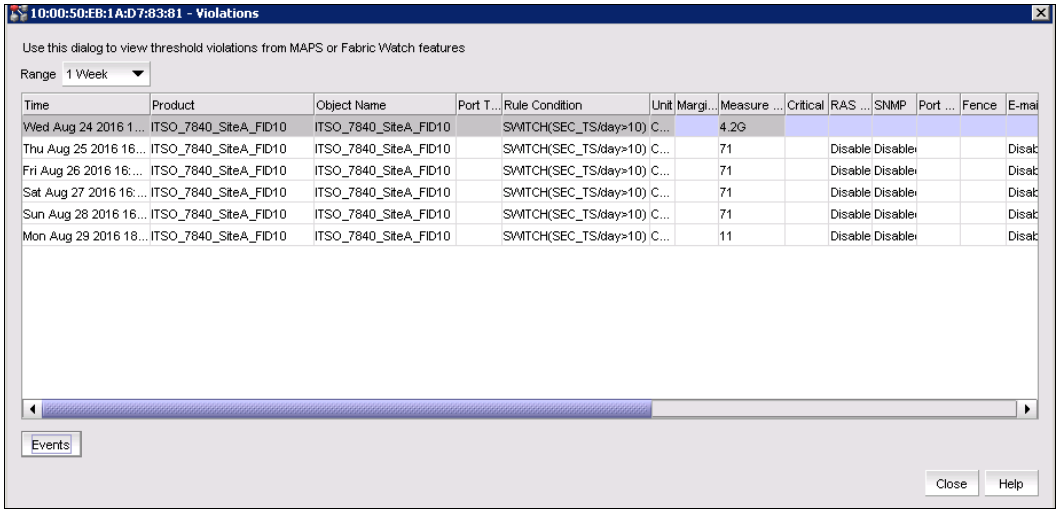


Figure 7-10 MAPS violations

2. Display data for a specific duration by selecting options from the **Range** list. The default option is 30 minutes.
3. Review the detailed data.
4. Select one or more violations and click **Events** to open the MAPS Violation Master Log Events window. The MAPS Violation Master Log Events window is displayed (Figure 7-11).

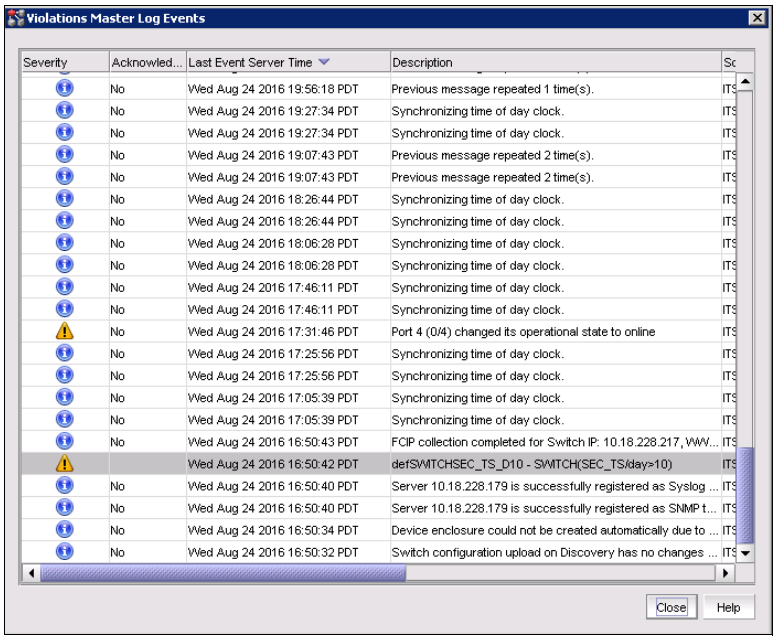


Figure 7-11 MAPS log events

Note: The events display for the selected time range (50% of the events before the selected violations and 50% after) up to a maximum of 200 event rows. For example, if you select one MAPS violation and set the time range to 1 hour, events display for 30 minutes before and after the selected violations. If the number of events within the selected the time range exceeds the maximum number of events (200), the time range changes for the maximum number of events.

5. Review the detail data of the log events.

7.2.3 Flow Vision

Flow Vision is a network diagnostic tool that provides a unified platform to manage traffic-related applications on Fabric OS devices. You can use this platform to simulate, monitor, and capture the network's traffic patterns, and to make decisions for troubleshooting based on the collected statistical data. For more details about the introduction of this feature, see Chapter 2, "The IBM System Storage SAN42B-R extension switch and the IBM b-type Gen 6 Extension Blade" on page 13.

Before monitoring a flow, you must add flow definition to define criteria that uniquely identify the flow. A flow definition includes basic criteria such as a flow name, source identifier (SID), destination identifier (DID), ingress port, egress port, and flow direction.

There are three traffic-related features in the flow definition:

- Flow Monitor: This feature monitors the network's traffic pattern and provides statistics for the defined flow. It is supported by IBM SAN42B-R and SAN06B-R.
- Flow Mirror: This feature allows you to select a traffic pattern and mirror this traffic to the CPU, enabling debugging that does not disturb the existing connections. It is supported by IBM SAN42B-R and SAN06B-R.
- Flow Generator: This feature simulates and generates traffic for the specified flows. It can create and activate multiple custom flows in the fabric. It is supported by IBM SAN42B-R, but not SAN06B-R.

Adding a flow definition

Complete the following steps to add flow generator for data traffic generation:

1. Before adding a flow generator, enable source simulator port (SIM) port mode on the switch ports that are connected to the source and destination devices for the new flow.

   To configure SIM mode for source device, select the port on the local switch for the source device in the Product List, and then select **Monitor → Fabric Vision → Flow → SIM Mode → Enable**.

   To configure SIM mode for destination device, select the port on the local switch for the destination device in the **Product List**, and then select **Monitor → Fabric Vision → Flow → SIM Mode → Enable** (Figure 7-12)**.**

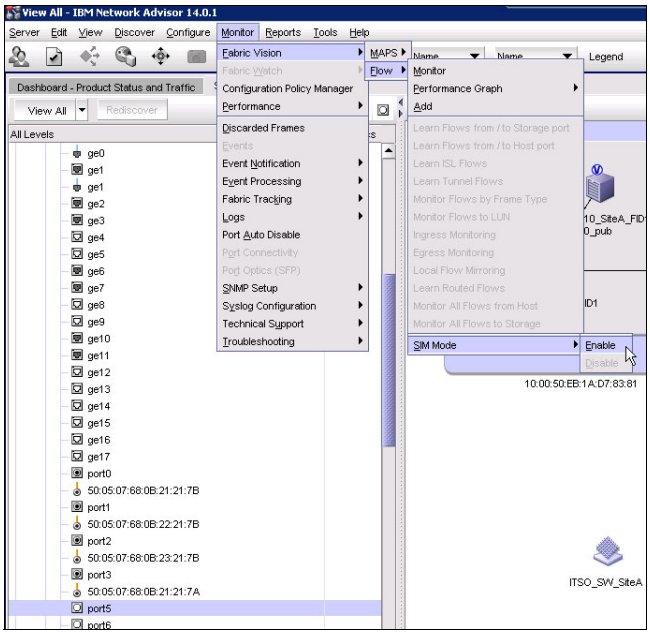


Figure 7-12 Menu for SIM port

Note: There are several limitations and prerequisites that apply specifically to the Flow Generator feature. For more information, see *IBM Network Advisor SAN User Manual*, SC27-5423-01.

2. Select the switch or switch port from the **Connectivity Map** or **Product List**, and then select **Monitor** → **Fabric Vision** → **Flow** → **Add** (Figure 7-13).

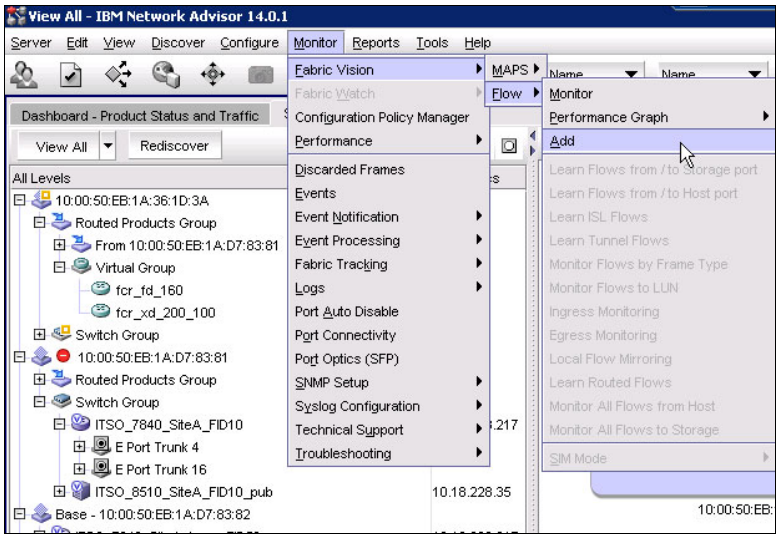


Figure 7-13 Menu for adding a flow

3. The Add Flow Definition window is displayed. Enter a name for the new flow (Figure 7-14).

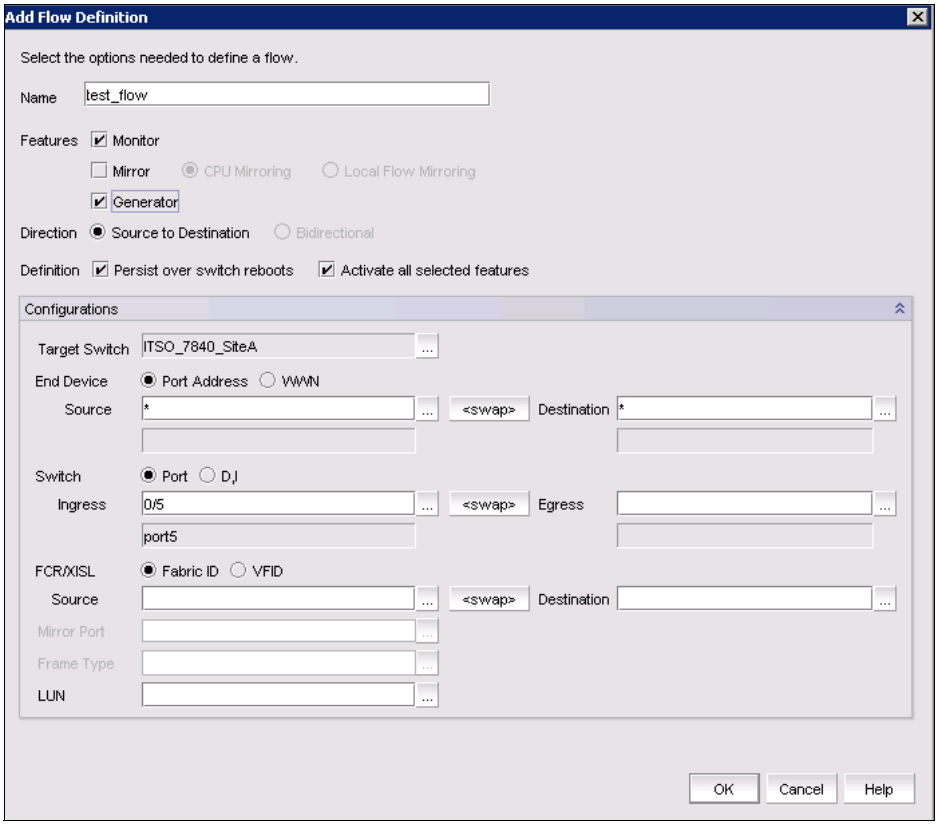


Figure 7-14 Add flow definition

4. Select the feature of the flow we create. Here we select **Generator**.
5. Select **Source to Destination** or **Bidirectional** flow.

6. Select **Persist over switch reboots** to persist flow definitions over reboots.
7. Select **Target Switch** which the flow is created on (Figure 7-15).

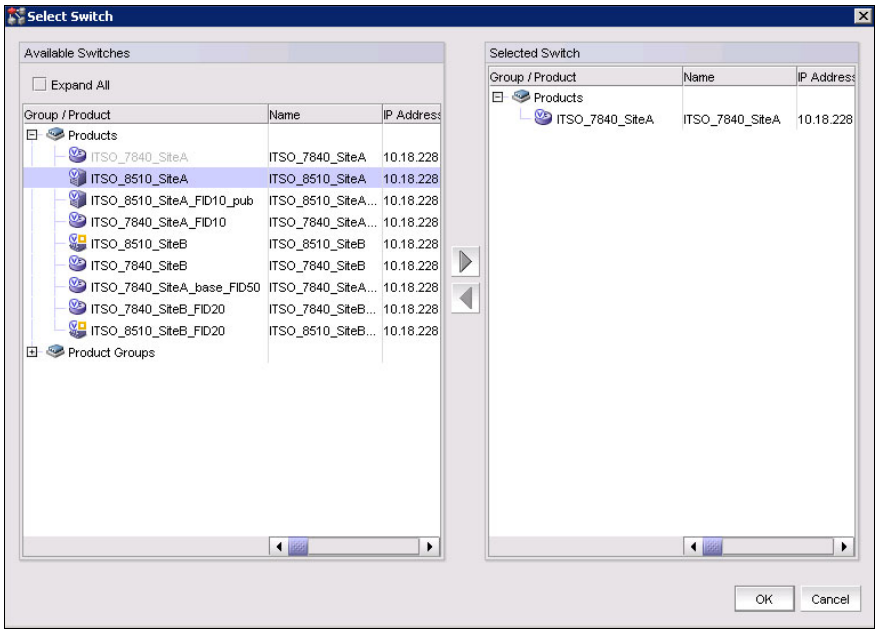


Figure 7-15 Select the switch for flow creation

8. Click the ellipsis button to the right of **Source** and **Destination** fields to display the Select Device port window. Select the port ID or WWN from the list of **Available Device Ports** (Figure 7-16). Click the right arrow to move the selected port to the Selected Device Port pane and click **OK** to confirm the selection.

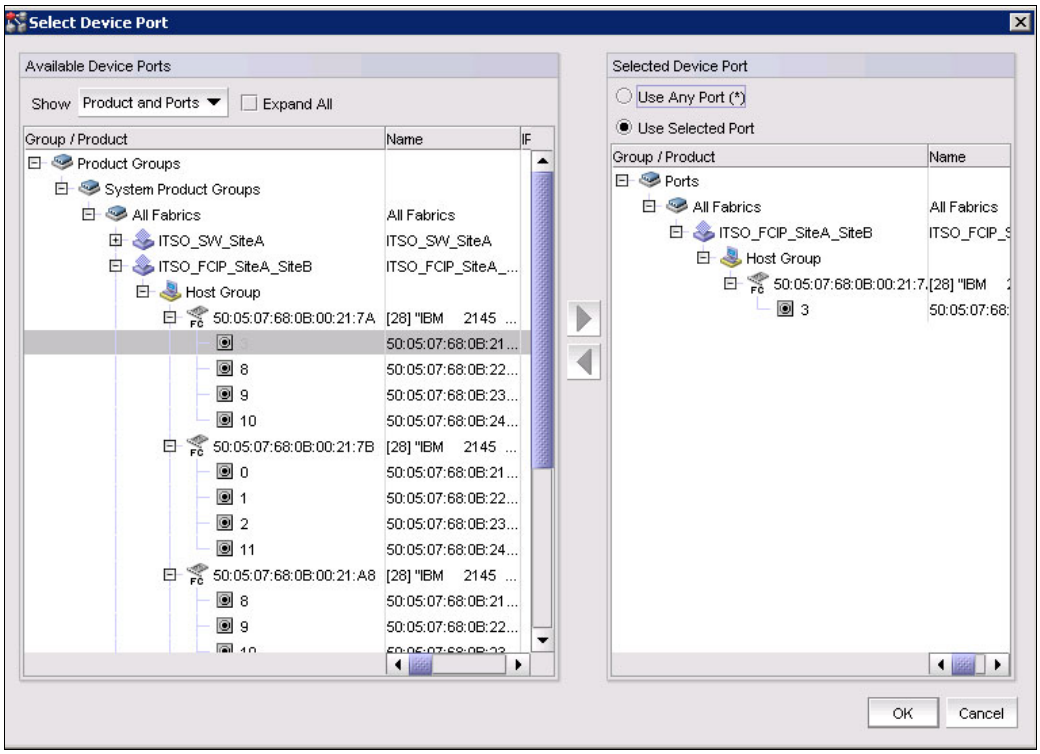


Figure 7-16 Select port for flow creation

9. Click the ellipsis button to the right of the **Ingress** and **Egress** fields to display the Select Switch Ports window. Select the Port ID or WWN from the list of **Available Device Ports** (Figure 7-16 on page 227). Click the right arrow to move the selected port to the **Selected Device Port** pane and click **OK** to confirm the selection (Figure 7-17).

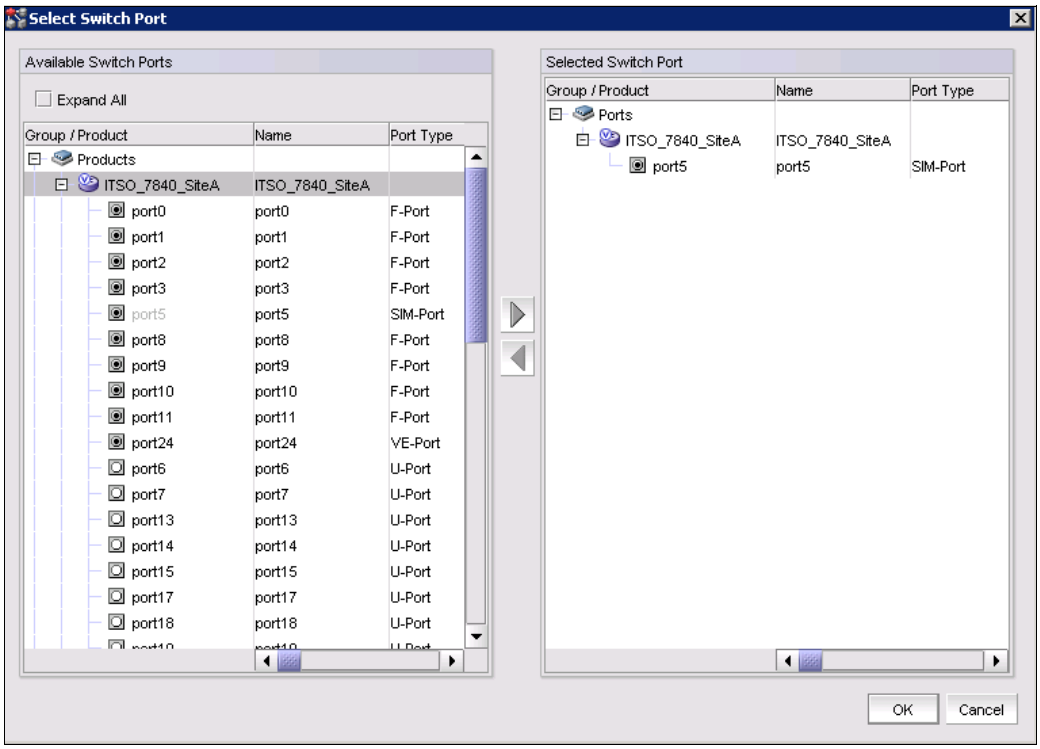


Figure 7-17 Select Ingress and Egress ports

10. Select **OK** to save the definition.

Monitoring flows

To monitor flows in the Flow Vision window, complete the following steps:

1. Select the switch or switch port from the Connectivity Map or Product List, and then select **Monitor** → **Fabric Vision** → **Flow** → **Monitor**. The Flow Vision window is displayed (Figure 7-18).

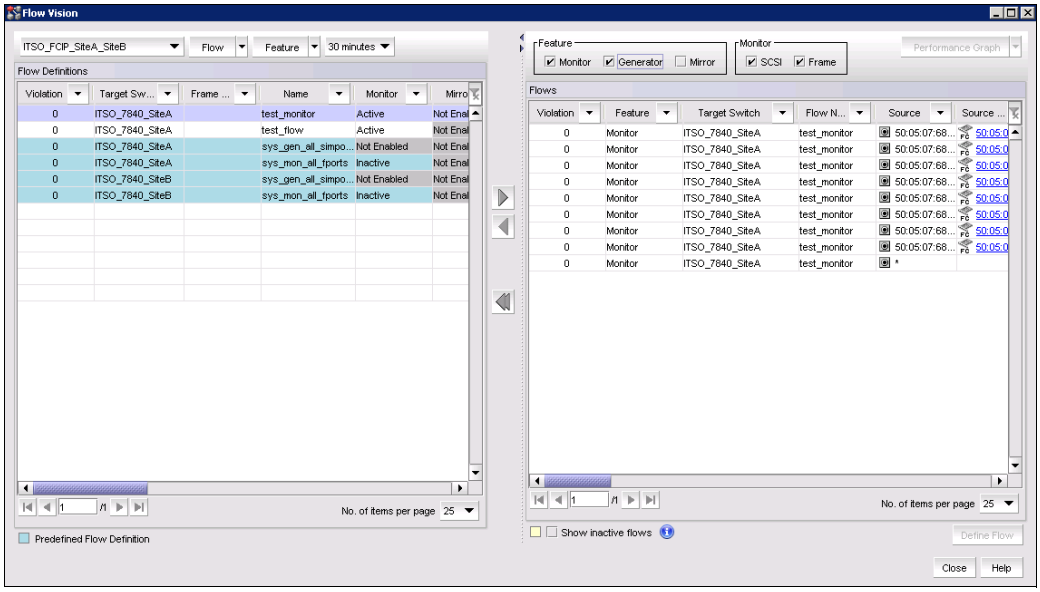


Figure 7-18 Configure Flow Vision

2. Select the flows that were created from Flow Definitions and click the right arrow to move them to the Flows pane.
3. Review the detail data of the flows. For more information about displaying data and launching more tools, see the *Brocade Network Advisor SAN User Manual*, 53-1004147-01, which can be downloaded from the following website:

   http://my.brocade.com

7.3 Using the portshow command

This section provides a brief overview of available commands to display information about IP interfaces, IP routes, switch mode, link aggregation groups, and IP tunnels.

7.3.1 Displaying IP interfaces

Example 7-1 displays information about IP interface ge0.dp0.

Example 7-1 Display information about IP interface

```
ITSO_7840_SiteA:FID128:admin> portshow ipif ge0.dp0

 Port         IP Address                    / Pfx  MTU   VLAN  Flags
-------------------------------------------------------------------------
 ge0.dp0      10.1.1.10                     / 24   1500  0     U R M
```

```
--------------------------------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
       N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport
```

7.3.2 Displaying IP routes

This section shows how to use the **portshow iproute** command to display information about IP routes.

7.3.3 Displaying switch mode information

Example 7-2 shows the switch mode.

Example 7-2 Display switch mode

```
ITSO_7840_SiteA:FID128:admin> extncfg --show
APP Mode is HYBRID (FCIP with IPEXT)
VE-Mode: configured for 10VE mode.
ITSO_7840_SiteA:FID128:admin>
```

To use IP extension features, you must operate the switch in hybrid mode.

7.3.4 Displaying LAG information

Link aggregation group information can be displayed with the **portshow lag --detail** command. Example 7-3 shows information about link aggregation groups.

Example 7-3 Display link aggregation group information

```
hydswitchb1:FID128:admin> portshow lag --detail

LAG : lag1
------------------------------------------------------------
Oper State : Online
Port Count : 2
port      AdminSt   Oper state     Speed     AutoNeg
ge10      Enabled   ONLINE         10G       Disabled
ge11      Enabled   ONLINE         10G       Disabled

hydswitchb1:FID128:admin>
```

7.3.5 Displaying tunnel information

Example 7-4 shows how to display FCIP tunnel information with the **portshow fciptunnel** command.

Example 7-4 Display FCIP tunnel information

```
ITSO_7840_SiteA_base_FID50:FID50:admin> portshow fciptunnel all -c

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 34    -         Up       -------a- 7h14m29s    0.81    0.81   1      -       -
```

```
 34    0 ge2     Up      ----ah--4 7h14m29s    0.41    0.41   1  5000/10000 0/-
 34    1 ge3     Up      ----ah--4 7h14m18s    0.41    0.41   1  5000/10000 0/1
 34    2 ge6     Up      ----ah--4  7h14m6s    0.00    0.00   1  5000/10000 1/-
 34    3 ge7     Up      ----ah--4 7h13m47s    0.00    0.00   1  5000/10000 1/1
--------------------------------------------------------------------------------
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA

ITSO_7840_SiteA_base_FID50:FID50:admin>
```

7.4 WAN tool analysis

WAN tool analysis provides a set of troubleshooting methods that are designed for end-to-end connection testing. The tools are based on the **portcmd** command. The following options are available:

- **-- Tperf** sends test data between a source Ethernet port and a destination port.
- **-- ping** tests the connection between a source Ethernet port and a destination port.
- **-- traceroute** traces routes from a source Ethernet port to a destination port.
- **-- wtool** generates traffic between a pair of IP addresses to test connection for maximum throughput, link issues, congestion, and out of order delivery.
- **fciptunnel --perf** shows performance data that is generated by WAN analysis.

7.4.1 Using ping

The **portcmd --ping** command tests a connection between a source Ethernet port and a destination Ethernet port. The general syntax of the **portCmd --ping** command is built as follows:

```
portCmd --ping slot/ge-port -s source_ip -d destination_ip -n num_request -q
diffserv -t -ttl -w wait_time -z size -v vlan_id -c L2_Cos
```

If you want to test a connection between VLAN, the **portcfg vlantag** command must be used. If there is no active tunnel and you want to test a VLAN connection, VLAN entries must be manually added to the VLAN table on the source and remote site.

7.4.2 Using traceroute

The **portcmd traceroute** command can be used for tracing routes from a local Ethernet port to a destination Ethernet port. The general syntax of the **portCmd --traceroute** command is built as follows:

```
portCmd --traceroute slot/ge-port -s source_ip -d destination_ip -h max_hops -f
first_ttl -q diffserv -w wait -time -z size -v vlan_id -c L2_Cos
```

This command can also be used for tracing routes across VLANs. Note that if there is no active tunnel and you want to trace routes across VLAN connection, VLAN entries must be added manually to the VLAN table on the source and remote side with the **portcfg vlantag** command.

7.4.3 Using the WAN tool

The WAN tool allows the administrator to generate traffic on specific circuits to test the connectivity for network issues like congestion, throughput, and other network issues.

Prerequisites

Consider the following requirements before you use the WAN tool:

- The WAN tool is supported only by the IBM SAN42B-R Extension Switch and IBM b-type Gen 6 Extension Blade.
- A maximum of eight user-configured WAN Tool sessions are supported per DP complex.
- Each session can support a 10 Gbps connection as a maximum.
- A test session can run over an IP path being used by an existing circuit between two switches. However, you must disable the circuit at each end before you configure the session.
- You must configure the WAN Tool session on the switch at each end of the circuit.
- After configuration, you can start a test from one switch only to test unidirectional traffic to the opposite switch, or you can test bidirectional traffic between both switches using the bidirectional option. If bidirectional is specified for the test session, you can start the session at either switch.
- You can configure multiple test sessions (one per circuit) for a single port, but the total rate configured for all sessions must be equal to or less then the physical speed of the port (40 Gbps, 10 Gbps, or 1 Gbps). For example, on a 10 Gbps port, you can configure four 2.5 Gbps sessions. As another example, on a 40 Gbps interface, you can configure four 10 Gbps sessions.
- The default MTU size used in the test session is 1500, but jumbo frames are supported.

The following misconfigurations can cause issues when you use the WAN:

- The test rate defined for WAN tool session on the source switch does not match the rate defined on the remote switch.
- The test rate defined on a circuit exceeds the supported amount for the port.
- Specified source and destination IP addresses is not correct on switches.

Example

In our lab configuration, we implemented a WAN tool session. The following steps are required for implementing a WAN test tool session:

1. Connect to Switch as an admin user.
2. The connection between the IP pair which wanted to be tested needs to be disabled. Example 7-5 shows how to disable circuit 0 on VE Port 24 on site A.

Example 7-5 Site A: Disable circuit 0 on VE Port 24

```
ITSO_7840_SiteA:FID128:admin> portCfg fcipcircuit 24 modify 0 --admin-status
disable
```

```
!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a
brief period of time. This operation will bring the existing tunnel down (if
tunnel is up) before applying new configuration.

Continue with Modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteA:FID128:mmet>
```

Example 7-6 shows how to disable circuit 0 on VE Port 24 on site B.

Example 7-6 Disable circuit 0 on VE Port 24

```
ITSO_7840_SiteB:FID128:admin> portCfg fcipcircuit 24 modify 0 --admin-status
disable

!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a
brief period of time. This operation will bring the existing tunnel down (if
tunnel is up) before applying new configuration.

Continue with Modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

3. Verify that the circuit is down.

   Example 7-7 verifies whether the circuit with index 0 is down.

Example 7-7 Verify if port is disabled

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel -c

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
-
 24    -         Up      --i----a-  2h57m2s    0.66    0.66   1      -       -
 24    0 ge0     Disable ----ah--4 2h56m45s    0.00    0.00   1 10000/10000 0/-
 24    1 ge0     Up      ----ah--4 2h27m47s    0.33    0.33   1 10000/10000 0/1
 24    2 ge1     Up      ----ah--4 2h52m58s    0.20    0.20   1 10000/10000 1/-
 24    3 ge1     Up      ----ah--4  2h31m6s    0.00    0.00   1 10000/10000 1/1
--------------------------------------------------------------------------------
-
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
       (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                  ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

4. Create a new wtool session.

   Example 7-8 shows the creation of a new wtool session with source and destination address on site A.

Example 7-8 Site A: Create a wtool session between two interfaces

```
ITSO_7840_SiteB:FID128:admin> portcmd --wtool 0 create --src 10.1.1.30 --dst
10.1.1.10 --rate 10000000 -time 1
```

```
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

Example 7-9 shows the creation of a new wtool session with source and destination addresses on site B.

Example 7-9 Site B: Create a wtool session between two interfaces

```
ITSO_7840_SiteB:FID128:admin> portcmd --wtool 0 create --src 10.1.1.30 --dst
10.1.1.10 --rate 10000000 -time 1
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

5. Set the status to enabled.

   Example 7-10 shows how to set the status of wtool session with index 0 to enabled.

Example 7-10 Site A: Set status of wtool session to enabled

```
ITSO_7840_SiteA:FID128:admin> portcmd --wtool 0 modify -a enable
!!!! WARNING !!!!
Modify operation will disrupt traffic on the WAN Tool session specified. This
operation will bring the existing WAN Tool session down before applying the new
configuration.
Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

6. Verify the session.

   Example 7-11 shows how to verify the status of a wtool session. The status must be `up`.

Example 7-11 Verify status of session

```
ITSO_7840_SiteA:FID128:mmet> portcmd --wtool 0 show
 Session OperSt GE-Pt   LocalIP         RemoteIp         TxMBps  RxMBps  Drop%
--------------------------------------------------------------------------------
-
 0       Up     ge0     10.1.1.10       10.1.1.30          0.65    0.65   0.00
--------------------------------------------------------------------------------
-
ITSO_7840_SiteA:FID128:admin>
```

7. Start the wtool session.

   Example 7-12 shows how to start a wtool session.

Example 7-12 Start wtool

```
ITSO_7840_SiteA:FID128:mmet> portcmd --wtool 0 start
Operation Succeeded
ITSO_7840_SiteB:FID128:admin> portcmd --wtool 0 start
Operation Succeeded
```

Example 7-13 shows how to verify the wtool session on site A.

Example 7-13 Verify that the session from Site A has started

```
ITSO_7840_SiteA:ITSO_7840_SiteA:FID128:admin> portcmd --wtool show --detail

 WTool Session: 0 (DP0)
```

```
================================================
 Admin / Oper State     : Enabled / Running
 Up Time                : 42s
 Run Time               : 12s
 Time Remaining         : 48s
 IP Addr (L/R)          : 10.1.1.10 ge0 <-> 10.1.1.30
 IP-Sec Policy          : (none)
 PMTU Discovery (MTU)   : disabled (1500)
 Bi-Directional         : disabled
 L2CoS / DSCP           : (none) / (none)
 Configured Comm Rate   : 10000000 kbps
 Peer Comm Rate         : 10000000 kbps
 Actual Comm Rate       : 10000000 kbps
 Tx rate                : 9955022.02 Kbps (1244.38 MB/s)
 Rx rate                : 9960078.02 Kbps (1245.01 MB/s)
 Tx Utilization         : 99.55%
 Rx Utilization         : 99.60%
 RTT (Min/Max)          : 1 ms/1 ms
 RTT VAR (Min/Max)      : 1 ms/1 ms
 Local Session Statistics
  Tx pkts               : 8932008
 Peer Session Statistics
  Rx pkts               : 8929500
  Ooo pkts              : 0
  Drop pkts             : 0 (0.00%)

ITSO_7840_SiteA:FID128:admin>
```

Example 7-14 shows how to verify the wtool session on site B.

Example 7-14 Verify that the session from Site B has started

```
ITSO_7840_SiteB:FID128:admin>  portcmd --wtool show --detail

 WTool Session: 0 (DP0)
================================================
 Admin / Oper State     : Enabled / Running
 Up Time                : 1m16s
 Run Time               : 48s
 Time Remaining         : 12s
 IP Addr (L/R)          : 10.1.1.30 ge0 <-> 10.1.1.10
 IP-Sec Policy          : (none)
 PMTU Discovery (MTU)   : disabled (1500)
 Bi-Directional         : disabled
 L2CoS / DSCP           : (none) / (none)
 Configured Comm Rate   : 10000000 kbps
 Peer Comm Rate         : 10000000 kbps
 Actual Comm Rate       : 10000000 kbps
 Tx rate                : 9951435.90 Kbps (1243.93 MB/s)
 Rx rate                : 9956222.62 Kbps (1244.53 MB/s)
 Tx Utilization         : 99.51%
 Rx Utilization         : 99.56%
 RTT (Min/Max)          : 1 ms/2 ms
 RTT VAR (Min/Max)      : 1 ms/2 ms
 Local Session Statistics
  Tx pkts               : 38995848
```

```
 Peer Session Statistics
  Rx pkts               : 36790930
  Ooo pkts              : 0
  Drop pkts             : 0 (0.00%)

ITSO_7840_SiteB:FID128:admin>
```

After the test, the WAN tool session must be removed and the status of both interfaces must be updated to `enabled`, as shown in Example 7-15 and Example 7-16.

Example 7-15 shows how to delete a session.

Example 7-15 Delete a WAN tool session

```
TSO_7840_SiteA:FID128:admin> portcmd --wtool 0 delete
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 7-16 shows how to enable ports.

Example 7-16 Enable ports

```
ITSO_7840_SiteA:FID128:admin> portCfg fcipcircuit 24 modify 0 --admin-status
enable

!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a brief
period of time. This operation will bring the existing tunnel down (if tunnel is
up) before applying new configuration.

Continue with Modification (Y,y,N,n): [ n]      y
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 7-17 shows how to verify the circuit status after the wtool test.

Example 7-17 Verify the status after test

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel --circuit

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
 24    -         Up      --i----a-   3h5m9s    0.67    0.67   1      -       -
 24    0 ge0     Up      ----ah--4       1s    0.00    0.00   2 10000/10000 0/-
 24    1 ge0     Up      ----ah--4 2h35m53s    0.33    0.33   1 10000/10000 0/1
 24    2 ge1     Up      ----ah--4   3h1m4s    0.33    0.33   1 10000/10000 1/-
 24    3 ge1     Up      ----ah--4 2h39m12s    0.00    0.00   1 10000/10000 1/1
--------------------------------------------------------------------------------
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

7.4.4 Service level agreement

Service level agreement (SLA) works with the existing WAN Tool features to provide automatic testing of a circuit before the circuit is placed in service. The primary purpose of SLA is to provide automated testing of a circuit. The SLA checks the circuit for packet loss percentage. If you need to verify the circuit for additional network performance, such as throughput, congestion, and out of order delivery, use the WAN Tool to run tests manually.

You must configure an SLA session at each end of the circuit being tested. The SLA session uses information from the circuit configuration to configure and establish the SLA connections. If the circuit configurations specify different transmission rates, SLA negotiates and uses the lower configured rate. This process allows SLA to start even when circuit configurations have a minor mismatch.

In addition to packet loss, the SLA can also test for timeout duration. If the timeout value is reached during the SLA session, the session is terminated and the circuit put into service.

Configured SLA sessions are persistent across reboots. This configuration is because circuit configurations are persistent across reboots and the SLA is part of the circuit configuration. However, user-configured WAN Tool sessions are not persistent.

Any attempt to modify a session while it is active is blocked, which means the WAN Tool commands cannot be used while an SLA session is running. Up to 16 SLA sessions can be defined per DP.

Note: During a HCL reboot, SLA is disabled and no new SLA sessions can be created until all HCL operations are complete. After all HCL operations are complete, SLA is reenabled.

To configure or abort SLA sessions, complete these steps:

1. Use the **portcfg sla** command to create an SLA session. You must create an SLA session at each end of the circuit, but the session names need not match (Example 7-18 and Example 7-19).

Example 7-18 Create SLA sessions of Site A

```
ITSO_7840_SiteA:FID128:admin> portcfg sla sla_A create --loss 3
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
ITSO_7840_SiteA:FID128:admin> portcfg sla sla_B create --loss 0.5 --runtime 2
--timeout 3
Operation Succeeded
```

Note: The valid range of SLA packet loss percentage is 0.05 - 5.00.

Example 7-19 Create SLA sessions for Site B

```
ITSO_7840_SiteB:FID128:admin> portcfg sla sla_A create --loss 3
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
ITSO_7840_SiteB:FID128:admin> portcfg sla sla_B create --loss 0.5 --runtime 2
--timeout 3
Operation Succeeded
```

Here we create two SLA sessions. The first one named sla_A is created with a package-loss percentage of 3%. It runs for the default 5 minutes. The second one named sla_B is created with a package-loss percentage of 0.5%. It runs for 2 minutes and times out after 3 minutes.

2. Use the **portcfg fcipcircuit** command to assign an SLA to a circuit. Here we modify a circuit and assign the SLA *sla_A* to the circuit. Both of the two ends should be configured with a matching SLA. Example 7-20 shows Site A.

Example 7-20 Configure SLA for circuits for Site A

```
ITSO_7840_SiteA:FID128:admin> portcfg fcipcircuit 24 modify 0 --sla sla_A
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
```

Example 7-21 shows Site B.

Example 7-21 Configure SLA for circuits of Site B

```
ITSO_7840_SiteB:FID128:admin> portcfg fcipcircuit 24 modify 0 --sla sla_A
Operation Succeeded
ITSO_7840_SiteB:FID128:admin>
```

3. Use the **portshow fciptunnel** command to display the status of a circuit and see whether the SLA session is actively testing the circuit (Example 7-22).

Example 7-22 Display the status of SLA sessions

```
ITSO_7840_SiteA:FID128:admin> portshow fciptunnel -c

 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
--------------------------------------------------------------------------------
-
 24    -         Up       --i----a-      13s    0.00    0.00   6      -       -
 24    0 ge0     Test     --S-a---4   11m47s    0.00    0.00   3  5000/10000 0/-
 24    1 ge0     Up       ----a---4      13s    0.00    0.00   7 10000/10000 0/1
 24    2 ge1     Up       ----a---4      13s    0.00    0.00   6 10000/10000 1/-
 24    3 ge1     Up       ----a---4      13s    0.00    0.00   6 10000/10000 1/1
--------------------------------------------------------------------------------
-
 Flags (tunnel): i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA

ITSO_7840_SiteA:FID128:admin>
```

The output shows that tunnel 24, circuit 0 is under active test and has an SLA configured.

4. Use the **portcmd --wtool show --detail** command to display details about active WAN Tool sessions (Example 7-23).

Example 7-23 Display details of SLA sessions

```
ITSO_7840_SiteA:FID128:admin> portcmd --wtool show --detail

 WTool Session: 24.0 (DP0)
==============================================
```

```
 Admin / Oper State     : Enabled / Running
 Up Time                : 1m0s
 Run Time               : 0s
 Time Out               : -
 Time Remaining         : 4m1s
 IP Addr (L/R)          : 10.1.1.10 ge0 <-> 10.1.1.30
 IP-Sec Policy          : myPolicy1
 PMTU Discovery (MTU)   : disabled (1500)
 Bi-Directional         : disabled
 L2CoS / DSCP           : (none) / (none)
 Configured Comm Rate   : 5000000 kbps
 Peer Comm Rate         : 5000000 kbps
 Actual Comm Rate       : 5000000 kbps
 Tx rate                : 4998706.53 Kbps ( 624.84 MB/s)
 Rx rate                : 4999496.13 Kbps ( 624.94 MB/s)
 Tx Utilization         : 99.97%
 Rx Utilization         : 99.99%
 RTT (Min/Max)          : 1 ms/3 ms
 RTT VAR (Min/Max)      : 1 ms/5 ms
 Local Session Statistics
  Tx pkts               : 25121707
 Peer Session Statistics
  Rx pkts               : 25117127
  Ooo pkts              : 0
  Drop pkts             : 0 (0.00%)
```

5. Use the **portcmd --wtool <*session*> stop** command to abort an SLA session. The session ID information is obtained from the **portcmd --wtool show** command (Example 7-24).

Example 7-24 Abort SLA session

```
ITSO_7840_SiteA:FID128:admin> portcmd --wtool show

 Session OperSt GE-Pt   LocalIP         RemoteIp         TxMBps  RxMBps  Drop%
-------------------------------------------------------------------------------
 24.0    Running  ge0   10.1.1.10       10.1.1.30        624.92  624.82   0.00
-------------------------------------------------------------------------------
ITSO_7840_SiteA:FID128:admin> portcmd --wtool 24.0 stop
Operation Succeeded
ITSO_7840_SiteA:FID128:admin>
ITSO_7840_SiteA:FID128:admin> portcmd --wtool show

 Session OperSt GE-Pt   LocalIP         RemoteIp         TxMBps  RxMBps  Drop%
-------------------------------------------------------------------------------
 24.0    Disable  ge0   10.1.1.10       10.1.1.30          0.00    0.00   0.00
-------------------------------------------------------------------------------
```

The SLA session 24.0 is aborted.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186
- *IBM Storage Networking SAN512B-6 and SAN256B-6 Directors*, REDP-5398
- *IBM System Storage b-type Multiprotocol Routing: An Introduction and Implementation*, SG24-7544
- *IBM System Storage SAN06B-R Extension Switch*, TIPS1126
- *IBM System Storage SAN42B-R Extension Switch*, TIPS1209
- *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116
- *Implementing or Migrating to an IBM Gen 5 b-type SAN*, SG24-8331

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- *Brocade Network Advisor SAN User Manual*, 53-1004147-01

  http://my.brocade.com

- *Brocade Web Tools Administrator's Guide*, 53-1003989-01

  https://www.brocade.com/content/html/en/administration-guide/fos-800-webtools/index.html

- *Define FC-IP Routing with IBM SAN42B-R* white paper

  https://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102669

Online resources

These websites are also relevant as further information sources:

- *Brocade Fabric OS Extension Configuration Guide 8.0.1* for more comprehensive information:

  http://www.brocade.com/content/html/en/configuration-guide/fos-801-extension/GUID-647FE9BE-0AA2-4774-B8E2-3BB2986D74F1-homepage.html

- Brocade 7840 Extension Switch Technical Specifications

  http://bit.ly/2cTiPEc

- IBM V7000 Knowledge center

  http://www.ibm.com/support/knowledgecenter/ST3FR7/

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5404-00

ISBN 0738455989

Printed in U.S.A.