WHITE PAPER

CONNECTING SANS OVER METROPOLITAN AND WIDE AREA NETWORKS

Internetworking technologies help link Fibre Channel SANs over existing long-distance connections

BROCADE
The intelligent platform for networking storage

As Storage Area Network (SAN) infrastructures continue to expand, so does the need to connect storage devices over longer distances in heterogeneous environments. In fact, many organizations are beginning to connect local SAN islands over existing high-speed public and private networks—an approach that enables new types of applications that leverage a geographically dispersed, yet interconnected SAN infrastructure. Typical applications include wide area data replication, high-speed remote centralized backup, cost-effective remote storage centralization, business continuity, and storage outsourcing.

This paper describes the technologies available for connecting Fibre Channel SANs over Metropolitan Area Networks (MANs) and Wide Area Networks (WANs)—a concept known as "SAN internetworking." It also highlights the excellent performance of the Brocade® Extended Fabrics™ feature at extremely long distances and includes interconnectivity performance results and examples of tested storage solutions.

The Expansion of SANs

As SANs become a more strategic component of the IT infrastructure, organizations are seeking new ways to connect existing local SAN islands into larger SAN fabrics that span the enterprise and beyond. To accomplish this goal, organizations are leveraging existing high-speed public and private networks to connect Fibre Channel SANs over MANs and WANs. Many of these organizations have discovered that one of the best ways to achieve excellent performance at longer connectivity distances is to use the Brocade Extended Fabrics feature, which maintains high throughput at extended distances.

In general, MANs span up to 120 km and transfer data using the native Fibre Channel protocol, while WANs span the globe and transfer data using protocol translation (such as Fibre Channel over ATM). This extended connection strategy enables new types of applications specially designed to take advantage of this geographically dispersed yet interconnected SAN infrastructure. The following sections describe the technologies available today for extending SANs over MANs and WANs in heterogeneous environments. Also included are performance results and possible applications for these types of configurations.

SAN Internetworking Overview

A Fibre Channel fabric is a network configuration of one or more Fibre Channel switches connected by the switches' expansion ports (E_Ports) in various topologies. Organizations that interconnect SAN fabrics across the enterprise over a private or common-carrier infrastructure can deploy a variety of solutions, depending on their particular application or business requirements. A Brocade SilkWorm® Fibre Channel fabric infrastructure supports multiple configurations for extending SANs over longer distances, including:

- An Extended Fabrics configuration for fabric connectivity over MANs enhanced by the Brocade Extended Fabrics feature.
- A Remote Fabric configuration for fabric connectivity over WANs using techniques such as the Brocade Remote Switch™ feature.
- A Remote Device Mapping configuration for remote device connectivity to a local SAN over a WAN. This method involves importing devices from remote SANs into a local SAN and making them appear as local LUNs.

These configurations are described in more detail in the following sections.

Extended Fabric Configuration

As shown in Figure 1, a fabric can extend over long distances without protocol translation. The Brocade fabric operating system, Fabric OS,™ provides a special Extended Fabrics feature that provides additional E_Port buffers to enable higher throughput at longer distances.

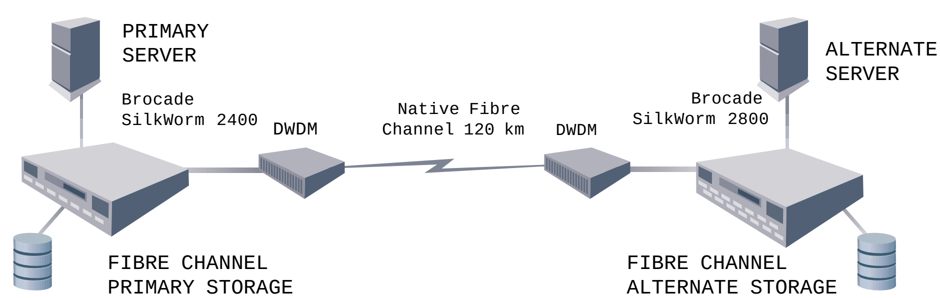


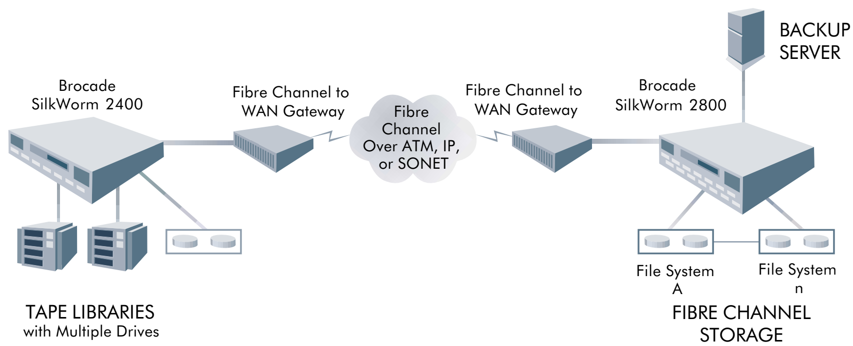
Figure 1.
A typical metropolitan area SAN that uses the Brocade Extended Fabrics feature

High-bandwidth links enable the deployment of innovative applications—such as synchronous data replication, which facilitates business continuity and faster recovery from disasters while maintaining data integrity.

Remote Fabric and Remote Device Mapping Configurations

Figure 2 shows connectivity over global distances for Remote Fabric and Remote Device Mapping configurations, which use existing WAN infrastructures and Fibre Channel encapsulation over WAN protocols such as ATM, IP, and SONET. Organizations can use either type of configuration for applications such as remote tape vaulting (backup) or storage replication.

Figure 2.
Wide area SANs connected over existing WAN infrastructures and protocols



A Remote Fabric configuration is a fabric that spans across WANs by using protocol translation (a process also known as "tunneling") such as Fibre Channel over ATM or Fibre Channel over IP. Brocade supports Remote Switch configurations (a subset of Remote Fabric configurations) that enable fabric connectivity of two switches over long distances by encapsulating Fibre Channel over ATM. This type of configuration is currently supported over OC-3.

A Remote Device Mapping configuration attaches (imports) remote storage devices to a local SAN over WANs without extending the fabric. Instead, these configurations employ a gateway that makes the remote devices appear as local SCSI (LUN) devices. Brocade has tested and certified gateways that support Fibre Channel over ATM at SONET rates of OC-3 and OC-12 in multipoint-to-multipoint configurations.

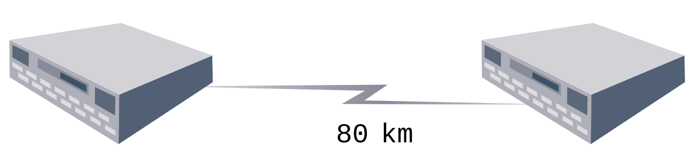
Connecting SANs Over Metropolitan Area Networks (MANs)

Because fiber optic communication is handled in a point-to-point manner for MANs, organizations can use various methods to extend SAN fabrics over a metropolitan area:

- Single-mode Long Wavelength (LWL) Gigabit Interface Converters (GBICs)
- Extended Long Wavelength (ELWL) GBICs
- Link extenders
- Dense Wave Division Multiplexers (DWDM)

Single-mode LWL GBICs extend the light wavelength to distances of 10 km, while ELWL GBICs can extend the light to distances up to 80 km. Brocade has tested their operation at 75 km. Figure 3 shows how two switches can connect to a fabric by using ELWL GBICs.

Figure 3.
Interconnectivity through ELWL GBICs



Link extenders are small external devices that can increase link distance to 120 km. Brocade has tested their operation at 100 km. Link Extenders can be used in conjunction with Short WaveLength (SWL) GBICs, as shown in Figure 4.

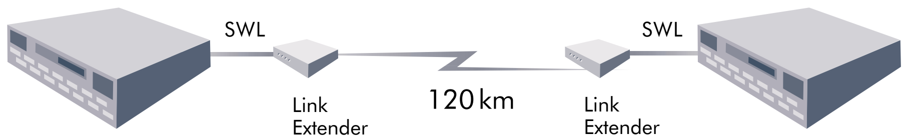


Figure 4.
Interconnectivity through link extenders

DWDM devices are used for multiplexing multiple 1 Gbit/sec (or higher) channels on a single fiber. These optical multiplexers are transparent to the underlying protocols, which means that organizations can use a single DWDM device to transfer Gigabit Ethernet, Gigabit Fibre Channel, ESCON, and SONET on a single fiber—each with its own wavelength.

Organizations can configure DWDM devices as point-to-point configurations or cumulative point-to-point configurations to form a ring. Most DWDM devices support immediate automatic failover to a redundant physical link if the main link is inaccessible. In a ring topology, only a single link is needed between nodes—if a link fails, the light is switched to the reverse direction to reach its target. Certain types of DWDM equipment can add and drop wavelengths—enabling wavelength routing in or out of a ring at 70 km to more than 160 km.

DWDM equipment is available in two basic classes—edge class (for enterprise access) and core class (for carriers). For the edge class, DWDM devices are usually smaller and less expensive, and provide fewer channels. Figure 5 shows how an organization can connect two sites over 50 km. The dual Inter-Switch Links (ISLs) between the switch and DWDM devices provide greater bandwidth (2 Gbits/sec instead of 1 Gbit/sec) but are not required. The DWDM devices can have a hot standby-protected link (the dashed line) that is automatically invoked if the main link fails. The protected link should reside on a separate physical path.

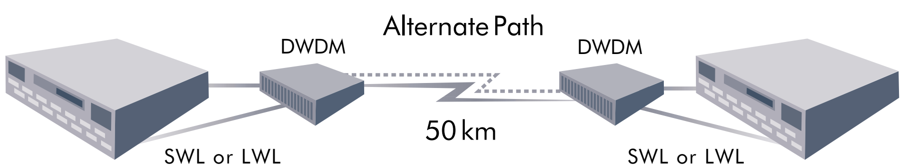
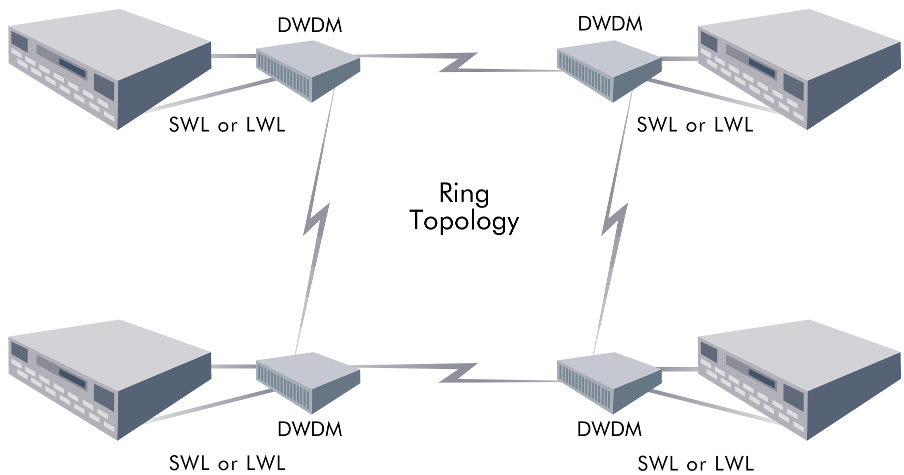


Figure 5.
Switch interconnectivity with DWDM

WHITE PAPER

For the core class, the DWDM equipment is larger and more expensive, and provides more channels. As shown in Figure 6, this DWDM equipment enables ring configurations and provides add and drop capabilities. (Refer to Appendix A for an example of service provisioning across four sites.)

Figure 6.
Switch interconnectivity through DWDM in a ring configuration



Performance Test Results

The following sections describe a test of three SAN fabrics that span 0 km (switches right next to each other), 100 km, and 100 km with the Brocade Extended Fabrics feature providing additional E_Port buffers. Finisar FLX2000 link extenders enabled the longer distance connectivity. For I/O performance testing, the configuration included a PC with an Emulex Host Bus Adapter (HBA) on one end and a JBOD storage device on the remote end. For I/O traffic generation and performance monitoring, Intel IOMeter was used. All eight drives in the JBOD storage device were used, with IOMeter creating a RAID-like striping effect to produce maximum performance. For traffic generation and raw Fibre Channel frame performance monitoring, a Finisar Frame Shooter was used.

I/O Data Rate (Throughput in MB Per Second)

The following figures show the data rate for 100 percent sequential read and write I/Os. On the read side, throughput greatly improved with the use of the Extended Fabrics feature, which provides enhanced buffering capabilities even at small size I/Os. When I/Os reached 128 KB and larger sizes, the performance of the 100 km fabric with the Extended Fabrics feature was almost identical to the 0 km fabric.

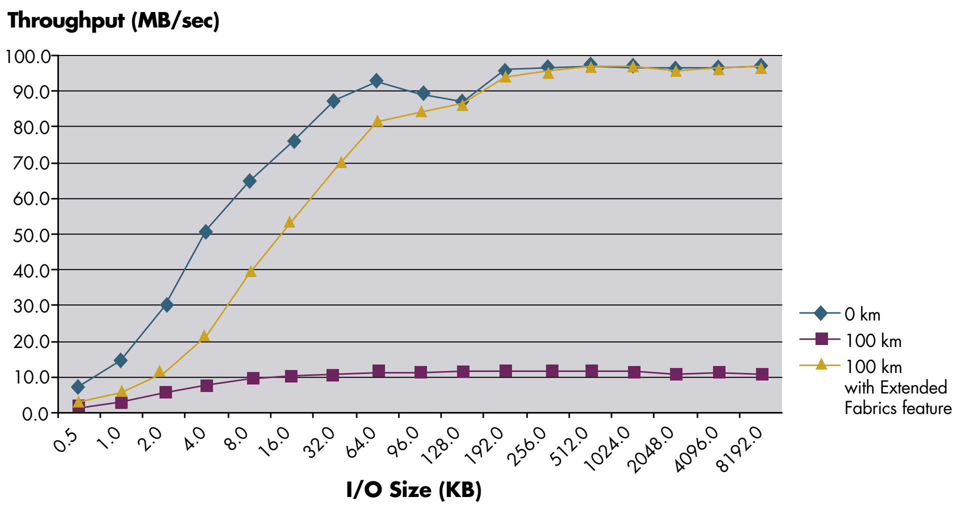


Figure 7.
I/O throughput with 100 percent sequential reads

On the write side, there was almost no difference among the three configurations at small size I/Os (up to 8 KB). However, when I/O reached 16 KB and larger sizes, the performance differences were significant—with the performance of the 100 km fabric with the Extended Fabrics feature only slightly below that of the 0 km fabric. These tests demonstrate that the Extended Fabrics feature provides near-full Fibre Channel performance at extended distances, especially for larger workloads.

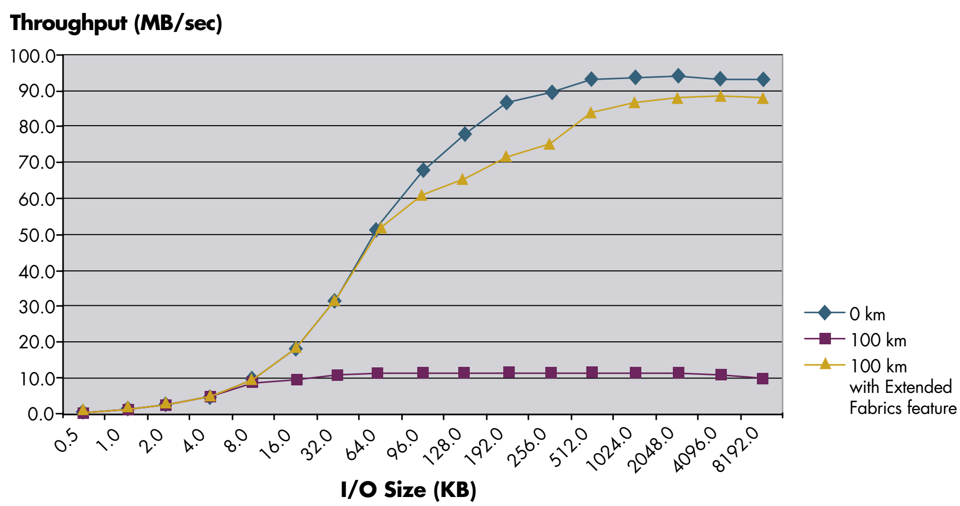


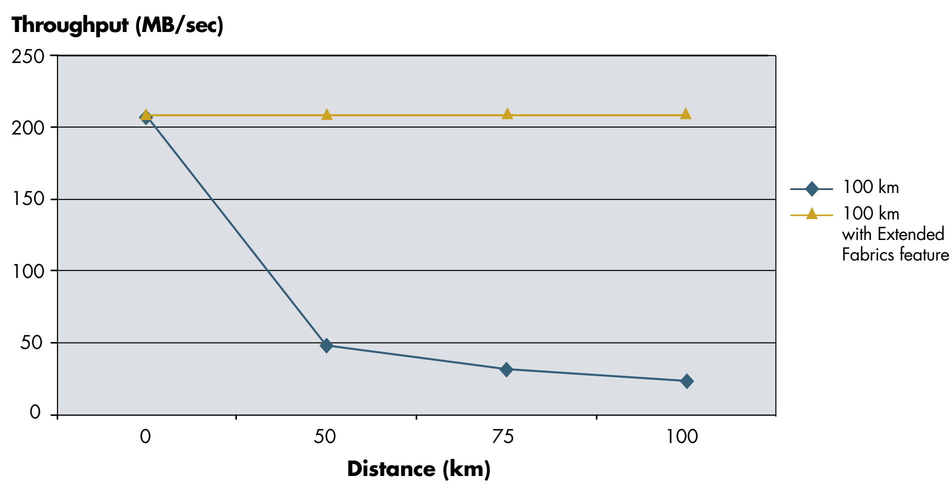
Figure 8.
I/O throughput with 100 percent sequential writes

Raw Frames Data Rate (Throughput in MB Per Second)

Another test measured duplex raw Fibre Channel frame throughput over various distances with and without the Extended Fabrics feature. In theory, providing enough buffering for long distances at 1 Gbit/sec requires approximately five buffers for every 10 km.

In this test, two Finisar Frame Shooter cards generated the duplex traffic of raw frames. Figure 9 compares fabrics connected over various distances—with and without the Extended Fabrics feature. The test results confirmed that Extended Fabrics E_Port buffering provides an adequate number of buffers to enable operation at full-duplex bandwidth (200 MB/sec) at distances up to 120 km. The reason the throughput line in the graphic actually exceeds 200 MB/sec is because it contains the primary data traffic (200 MB/sec) along with overhead traffic that travels through the port.

Figure 9.
Duplex raw frame throughput



Extending Distance Connectivity with E_Port Repeaters

Another test compared the difference in throughput at a distance of 240 km for a configuration using the ONI Systems Online 9000 between directly connected E_Port switches and a configuration using E_Port Repeaters. Figure 10 shows the configuration without the E_Port repeaters while Figure 11 shows the configuration with the repeaters.

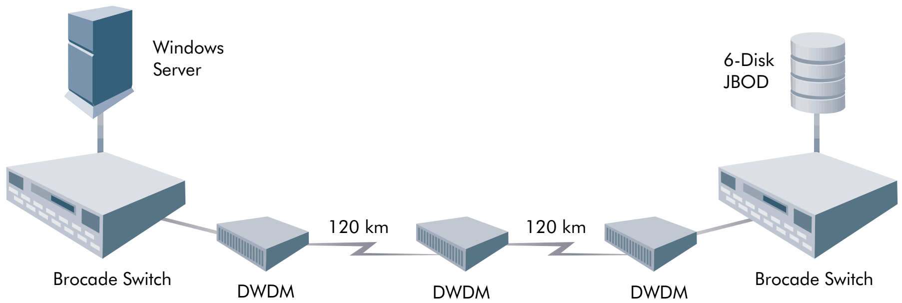


Figure 10.
Standard long-distance connectivity over DWDM

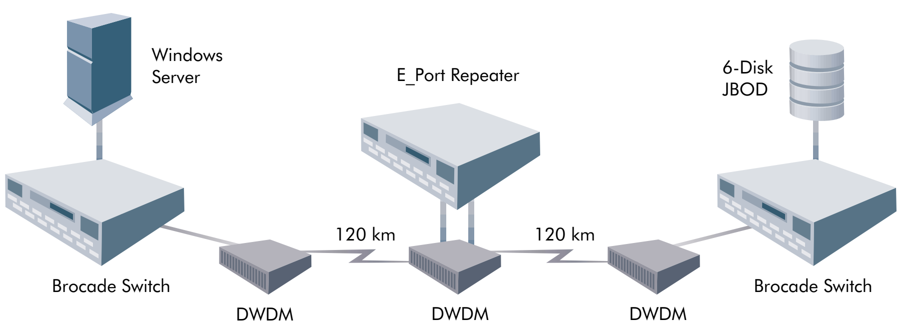


Figure 11.
Long-distance connectivity over DWDM with E_Port Repeaters

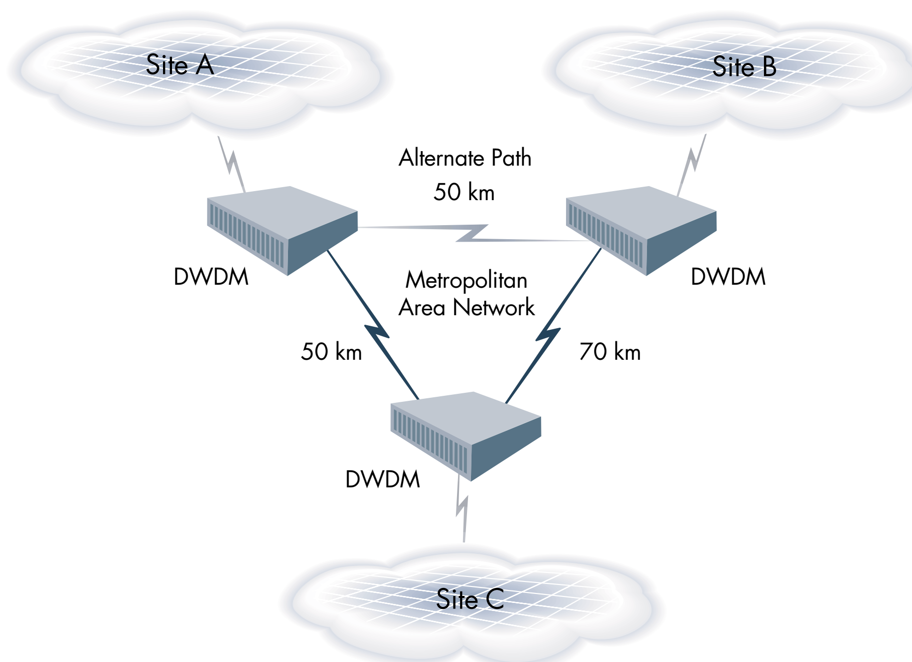
This particular test consisted of a Windows server, a six-disk JBOD, three Brocade SilkWorm switches (with the Extended Fabrics feature enabled), and ONI Systems Online 9000 metropolitan DWDM. IOMeter generated the data traffic and captured the performance. At 1 MB block sizes, the throughput with E_Port Repeaters was an impressive 93 MB/sec over 240 km. However, when the intermediate switch was not used, throughput was only 48 MB/sec. These results clearly show the higher performance of E_Port Repeaters, which can be extended over longer distances.

MAN-Based Applications

Many types of applications can benefit from a MAN-based SAN configuration. The most common applications include those for remote storage centralization (such as a service provider model), centralized remote backup, and business continuity.

Figure 12 shows an optical DWDM ring topology, which provides redundant paths and has the ability to fail over from a disconnected path to an alternate path. Site B has a 70 km connection (primary path) to Site C. When that connection goes offline, Site B uses the alternate path (other direction) over DWDM to restore Site B's connection to Site C. This path (B to A to C) spanned 100 km (50 km plus 50 km). Because of the extended buffering at the Fibre Channel switch E_Ports, the primary and alternate paths provided nearly the same level of data access performance during testing.

Figure 12.
Redundant paths in a DWDM-based MAN



Storage Centralization Over a SAN/DWDM Infrastructure

Organizations can centralize storage across a campus or geographically dispersed environment, or even remotely outsource the work to a Storage Services Provider (SSP). Figure 13 shows an SSP configuration where a designated site (Site C, the SSP) provides storage to multiple sites over MAN-based SANs in heterogeneous environments.

In this example Sites A and B subscribe to Site C, the SSP. Brocade Zoning™ is used to isolate heterogeneous fabrics, thereby controlling the amount of storage each customer site can access. Two fabric zones—one for Site A and the other for Site B—isolate storage for the two sites.

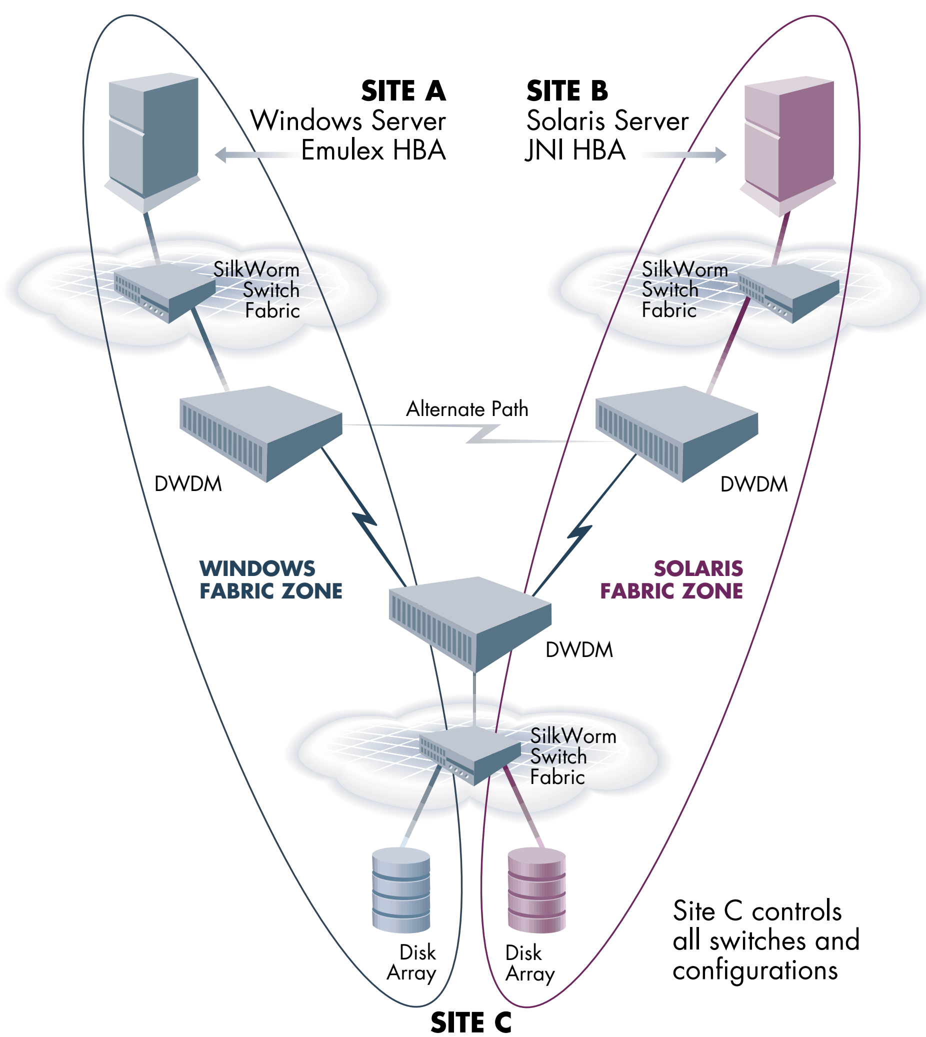
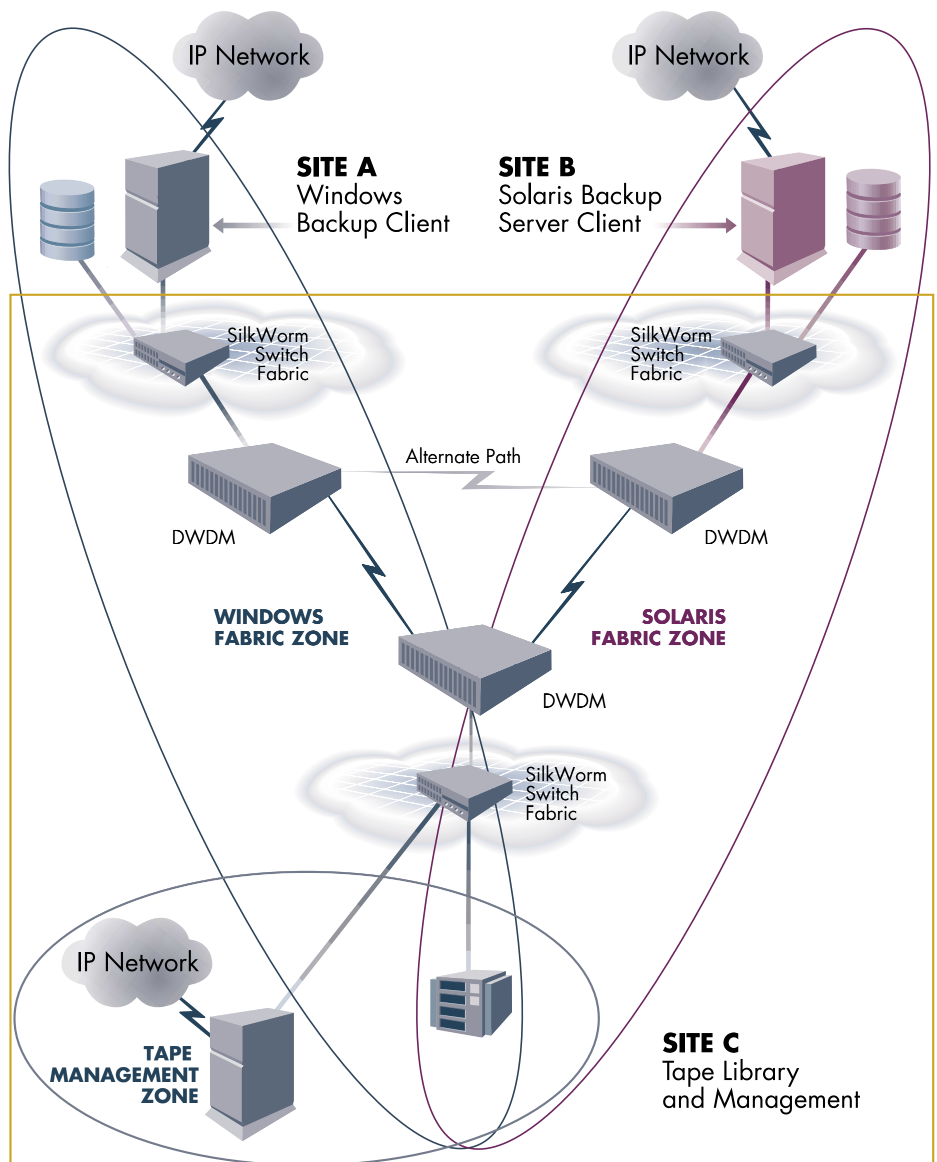


Figure 13.
Remote storage centralization over a MAN (SSP model)

Centralized Backup Over a SAN/DWDM Infrastructure

Centralized remote backup enables multiple sites to back up data to a single shared tape library by using fabric zoning (see Figure 14). Sites A and B share the tape library provided by Site C, which allows the tape library into both sites' respective zones. As a result, each site can perform data backup with any tape device in the library.

Figure 14.
Centralized remote backup over a MAN



Site C controls all devices and configurations within this square region

Business Continuity Over a SAN/DWDM Infrastructure

A business continuity solution provides synchronous data mirroring to a remote location. In the event of a disaster, a redundant system can take over for the main system and access the mirrored data. This solution also facilitates the recovery from the redundant remote system back to the main system after it is operational again. Figure 15 shows how two sites can utilize this type of solution concurrently.

Sites A and B are the primary sites (running different operating systems), and Site C is the remote business continuance site for both Sites A and B. If either Site A or B goes down, it can fail over to Site C.

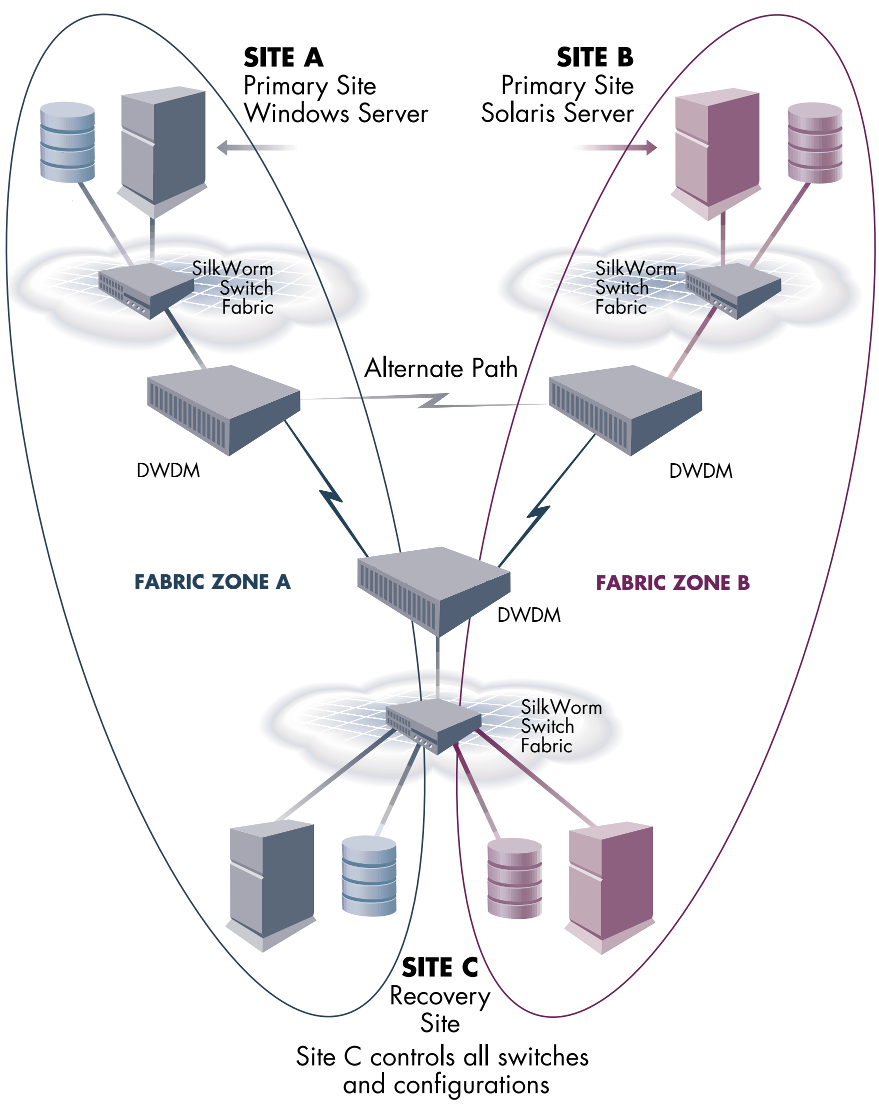


Figure 15.
Business continuity over a MAN

Certified Brocade SOLUTIONware MAN Configurations

Brocade and ONI Systems have recently certified new Brocade SOLUTIONware™ (pretested SAN configurations) offerings designed to enable SAN internetworking over a high-speed optical infrastructure. These long-distance SAN configurations include:

- **Storage consolidation over optical networks:** Uses a Brocade-based SAN and the ONI ONLINE transport platform to create a remote storage consolidation infrastructure that leverages a high-performance optical network. This solution enables any-to-any server and storage connectivity over distances up to 120 km. Storage resources can be consolidated and shared by servers residing at another point on the MAN.
- **Centralized backup over optical networks:** Provides centralized, heterogeneous data backup using the ONI ONLINE transport platform and VERITAS NetBackup DataCenter over a Brocade-based SAN. This solution enables a campus or multisite environment to centralize backups at a single, remote site to optimize tape library utilization and reduce backup administration.
- **Business continuance over optical networks:** Uses Brocade SilkWorm 2800 Fibre Channel fabric switches, the ONI ONLINE transport platform, UNIX and Windows 2000 hosts, and VERITAS remote mirroring software to protect information from disaster by mirroring it to a remote location. It enables a redundant remote system, up to 120 km away, to take over for the main system by accessing the mirrored data in an emergency.

For more information about available SOLUTIONware configurations, visit the Brocade SAN Solution Center at **www.brocade.com/san.**

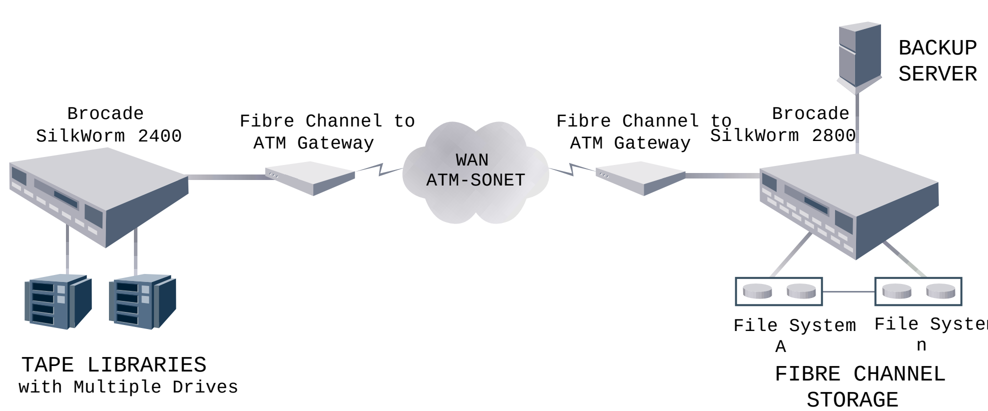
Connecting SANs Over Wide Area Networks (WANs)

To connect areas that surpass the reach of ELWL GBICs, link extenders, and DWDM, protocols. The following case studies demonstrate how organizations can deploy Remote Device Mapping and Remote Switch configurations over existing WAN infrastructures.

Remote Device Mapping Configuration: Performing Global Tape Backup Over DiskLink Gateways and a WAN

In the following example, a Fibre Channel-to-ATM gateway provided access to remote tape libraries. These remote devices were imported to the local SAN over the gateway, and the tape drives appeared as local LUNs on the local SAN (see Figure 16).

Figure 16. Global remote backup over a WAN



To maximize link bandwidth, this configuration employed two tape libraries, each with two tape drives. At the local end, the SAN was configured with a Windows 2000 server and JBOD connected to one switch. The JBOD contained 12 disks divided into four file systems, each striped across three disks. The remote end included another Brocade switch with two tape libraries: an ATL P1000 with two DLT 7000 tape drives, and an ADIC Scalar 218 with two DLT 8000 tape drives (attached to the switch over an ADIC FCR100 Fibre Channel-to-SCSI bridge).

Data throughput of the remote backup process was tested over long distances by using DiskLink Fibre Channel-to-ATM gateways. A special device introduced latency to simulate distance over ATM. Several backup tests were performed over distances of 0, 3000, and 5000 km by using four DLT tape drives. Both the OC-3 and OC-12 versions of the DiskLink gateway were tested.

The throughput of the data being backed up was determined by the information contained in the VERITAS NetBackup Datacenter Activity Monitor utility. Throughput was also measured by the output of the Brocade switch PortPerfShow utility. 10 GB of data files were stored on each of the four file systems, with individual files ranging in size from 500 KB to 40 MB. NetBackup was configured to back up each file system concurrently to improve performance. The same configuration was tested at distances of 0, 3000, and 5000 km. For comparison purposes, an identical local backup was performed with all devices connected to one switch. Although the configuration used built-in DLT hardware compression, no software compression was used.

Test results revealed that throughput declined only slightly at the OC-3 rate when distances increased between the DiskLink gateways—from 16 MB/sec at 3000 km to 13 MB/sec at 5000 km (see Table 1). The maximum throughput achieved with the OC-3 version of the DiskLink gateway was 16 MB/sec, while the OC-12 version reached 38 MB/sec at zero distance. The OC-3 link was saturated using four DLT tape drives working concurrently, to determine the maximum throughput at various distances for the same file backups. The OC-12 link was saturated with six DLT drives.

Table 1.
Combined throughput for various protocols and distances

PROTOCOL	**Number of Tape Drives**	**Number of Media Servers**	**Distance** (km)	**NetBackup Combined Throughput** (MB/sec)
Fibre Channel (native)	4	1	0	27
Fibre Channel to OC-3	4	1	0	16
Fibre Channel to OC-3	4	1	3,000	16
Fibre Channel to OC-3	4	1	5,000	13
Fibre Channel (native)	6	2	0	46
Fibre Channel to OC-12	6	2	0	38
Fibre Channel to OC-12	6	2	3,000	32
Fibre Channel to OC-12	6	2	5,000	32
Fibre Channel to OC-12	6	2	10,000	28

The low overhead of Fibre Channel-to-ATM conversion—38 MB/sec out of 46 MB/sec—resulted in 82.6 percent bandwidth utilization. The better-than-expected throughput for the six tape drives at 46 MB/sec compared to four drives at 27 MB/sec was the result of an additional media server on the host to generate more traffic. Using an additional media server enabled the configuration to exceed the maximum DiskLink OC-12 line speed of 40 MB/sec (DiskLink OC-12 performance is approximately half of the full OC-12 capacity of 622 Mbits/sec in the current release).

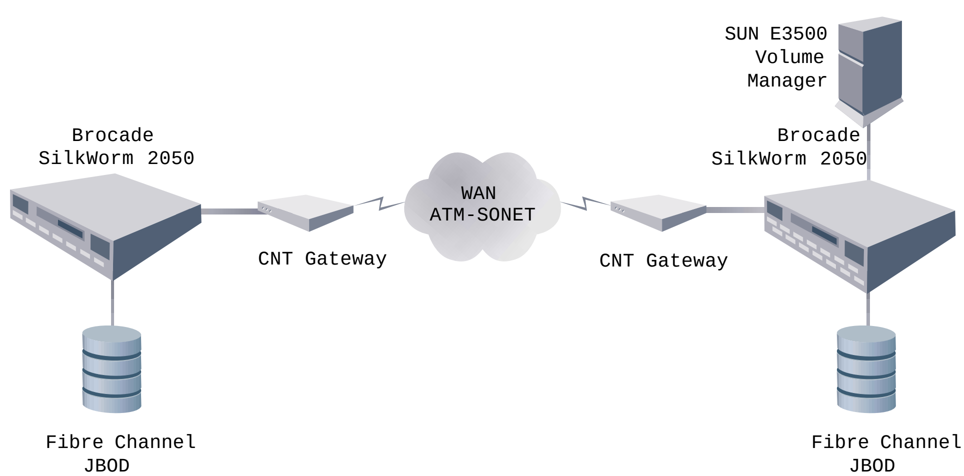
Remote Switch: Connecting SAN Islands Over Open Systems Gateways and a WAN

Connecting SAN islands over CNT's Open Systems Gateway (OSG) Fibre Channel-to-ATM device enables organizations to extend their solutions over a WAN. The OSGs interconnect remote SANs by acting like Fibre Channel E_Ports to create a single large fabric. The optional Remote Switch feature within the Brocade SilkWorm switch provides this capability. This type of configuration can be used for solutions such as remote disk mirroring and remote tape backup, as described in the following sections.

Remote Disk Mirroring Over a WAN

Figure 17 shows a configuration for mirroring file systems over a WAN through the CNT gateway. A Sun E3500 running Solaris 8 provided the host system, and VERITAS Volume Manager 3.1 created the striped volume mirrored at a remote site. Each side of the mirror (also known as a plex) in the volume consisted of six disks, and VERITAS VxFS was used as the file system. Mirroring performance was monitored by using the UNIX dd command, which tested file system I/O at various block sizes.

Figure 17.
Remote disk mirroring with CNT gateways



Testing revealed that using the dd command to write to the file system makes it possible to achieve I/O throughput (at 0 km) of approximately 15 MB/sec—maximizing use of the CNT OC-3 data link. Block sizes between 2 KB and 256 KB produced the best results.

Remote Backup Over a WAN

Another test used the CNT gateway devices to concurrently back up four local file systems—each containing approximately 10 GB of data—to a remote tape library. After the file systems were backed up, a restore was performed to ensure data integrity (see Figure 18).

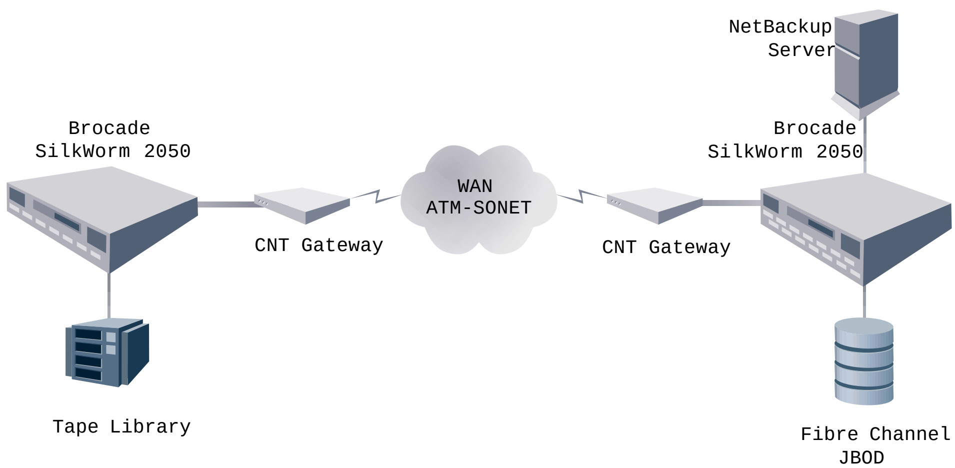


Figure 18. Remote tape backup with CNT gateways

The remote tape library was an ATL P1000 with four DLT7000s, and VERITAS NetBackup DataCenter 3.0 was used to perform the backups on a Windows 2000 Advanced Server system. NetBackup was configured to stream each file system to a separate DLT drive within the library.

The DLT library's built-in hardware compression was used, but no software compression was used. Throughput was determined by the information contained in the NetBackup Activity Monitor utility.

During the test, NetBackup reported a throughput (at 0 km) of approximately 13 MB/sec, a figure consistent with the output of the PortPerfShow command provided by the switch. Data restoration also completed successfully, with equivalent performance.

Reliable SAN Solutions for Real-World Business Requirements

Today, there are cost-effective, reliable methods for extending SAN fabrics beyond standard Fibre Channel distances. Organizations can choose to utilize dark fiber connections for MAN-based configurations up to 120 km. They can also employ existing WAN infrastructures (and translation protocols) to extend Fibre Channel over even longer distances.

Regardless of what approach organizations take, a reliable SAN infrastructure based on flexible Brocade SilkWorm Fibre Channel fabric switches can provide significant advantages for a variety of long-distance SAN applications. Brocade and its partners have thoroughly tested these types of SAN applications, and leading-edge companies such as SSPs are already implementing them to solve real-world business challenges.

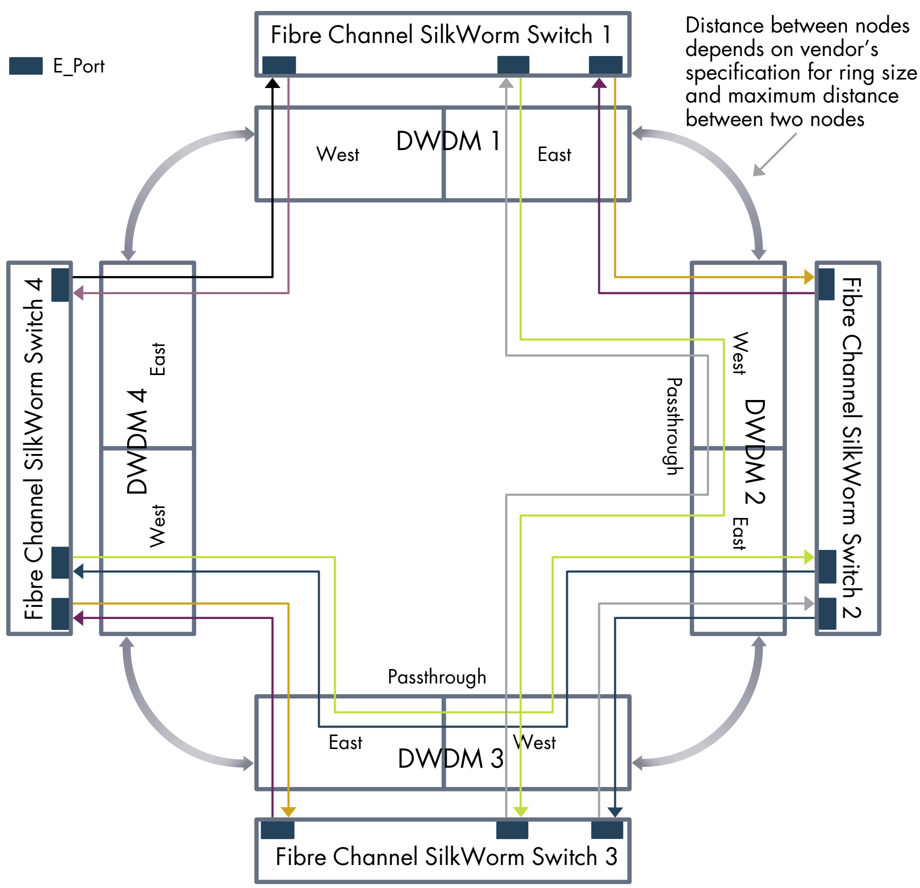
For information, about certified Brocade SOLUTIONware SAN configurations, **visit www.brocade.com/san.**

For information about Brocade education, visit **www.brocade.com/education_services.**

Appendix A. A Multinode DWDM Configuration

This appendix describes a multinode DWDM configuration that spans four sites and provisions optical services (see Figure 19). There are four switches, with each switch's E_Ports connected over a DWDM channel that includes dual paths for transmitting and receiving. Each path has its own wavelength. The DWDM passthrough feature enables non-contiguous sites to connect over an intermediate site as if they were directly connected. The only additional overhead of the passthrough is the minimal latency (5 usec/km) of the second link. The passthrough has no overhead since it is a passive device. The logical view of this configuration is shown in Figure 20.

Figure 19.
Service layout over DWDM



Each of the links can operate in protected mode, which provides a redundant path in the event of a link failure. In most cases, link failures are automatically detected within 50 msec. In this case, the two wavelengths of the failed link reverse directions and reach the target port at the opposite side of the ring. If the link between DWDM 1 and 4 fails, the transmitted wavelength from 4 to 1 would reverse direction and reach 1 through 3 and 2. The transmitted wavelength from 1 to 4 would also reverse direction and reach 4 through 2 and 3.

Calculating the distance between nodes in a ring depends on the implementation of the protected path scheme. For instance, if the link between DWDM 2 and 3 fails, the path from 1 to 3 would be 1 to 2, back from 2 to 1 (due to the failed link), 1 to 4, and finally 4 to 3. This illustrates the need to utilize the entire ring circumference (and more, in a configuration with over four nodes) for failover.

Another way to calculate distance between nodes is to set up the protected path in advance (in the reverse direction) so the distance is limited to the number of hops between the two nodes. In either case, the maximum distance between nodes determines the maximum optical reach. An example of this specification is 80 to 100 km for a maximum distance between nodes and 160 to 400 km for maximum ring size. These distances should continue to increase as fiber optic technology advances.

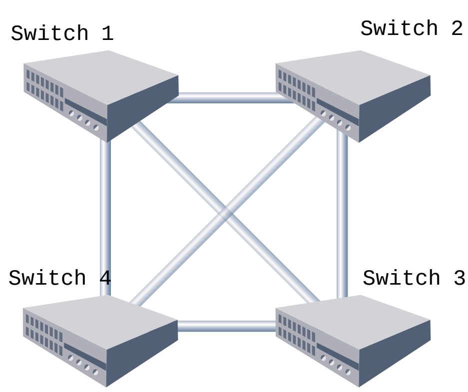


Figure 20.
Logical view of the fabric

BROCADE
The intelligent platform for networking storage

Corporate Headquarters
1745 Technology Drive
San Jose, CA 95110
T: (408) 487-8000
F: (408) 487-8101
info@brocade.com

European Headquarters
29, route de l'Aéroport
Case Postale 105
1211 Geneva 15, Switzerland
T: +41 22 799 56 40
F: +41 22 799 56 41
europe-info@brocade.com

Asia Pacific Headquarters
The Imperial Tower 15th Fl.
1-1-1 Uchisaiwaicho,
Chiyoda-ku, Tokyo 100-0011
Japan
T: +81 3 5219 1510
F: +81 3 3507 5900
apac-info@brocade.com