STORAGE NETWORKING BUSINESS CONTINUITY SOLUTIONS

BY STEVE GUENDERT, PH.D.

The most common reason for extending storage networks over geographical distances is to safeguard critical business data and provide near-continuous access to applications and services in the event of a localized disaster. Designing a distance connectivity solution involves a number of considerations. The enterprise must classify stored data and determine how important it is for business operation, how often it must be backed up and how quickly it needs to be recovered in the event of failure.

This is the first of a two-part series on long-distance storage networking extension for business continuity and Disaster Recovery (DR). This article focuses on some of the basics of fiber optic technology and presents the alternative technologies available today for long-distance extension for business continuity. The second article will discuss network design for high availability, best practices and take a look at Fiber Channel over IP (FCIP) trunking.

Classifying Data and Applications

From the business perspective, applications and their data need to be classified by how critical they are for business operation, how often data must be backed up and how quickly it needs to be recovered in the event of failure.

Two important concepts in the design process are:

- **Recovery Point Objective (RPO)** is the period of time between backup points that describes the acceptable age of the data that must be restored after a failure has occurred. For example, if a remote backup occurs every day at midnight and a site failure occurs at 11 p.m., changes to data made within the last 23 hours won't be recoverable.
- **Recovery Time Objective (RTO)** is the time it takes to recover from the disaster; this determines the acceptable length of time a break in continuity can occur with minimal or no impact to business services.

Options for replication generally fall into one of several categories. A business continuity solution with strict RTO and RPO may require high-speed synchronous or near-synchronous replication between sites as well as application clustering for immediate service recovery. A medium-level DR solution may require high-speed replication that could be synchronous or asynchronous with an RTO from several minutes to a few hours. Backup of non-critical application data that doesn't require immediate access after a failure can be accomplished via tape vaulting. Recovery from tape has the greatest RTO. Other technologies, such as Continuous Data Protection (CDP), can be used to find the appropriate RPO and RTO.

From a technology perspective, there are several choices for the optical transport network and configuration options for the Fibre Channel (FC) Storage Area Network (SAN) when it's extended over distance. Applications with strict RTO and RPO require high-speed synchronous or near-synchronous replication between sites with application clustering over distance for immediate service recovery. Less critical applications may only require high-speed replication that could be asynchronous to meet the RPO/RTO metrics. Lower priority applications that don't need immediate recovery after a failure can be restored from backup tapes from remote vaults.

Business Requirements

As more applications drive business value, and the associated data becomes key to competitive advantage, cost-effective protection of the applications and data from site disasters and extended outages has become the norm. Modern storage arrays provide synchronous as well as asynchronous array-to-array replication over extended distances.

When the array provides block-level storage for applications, FC is the primary network technology used to connect the storage arrays to servers, both physical and virtual. For this reason, cost-effective DR designs leverage FC to transport replicated data between arrays in different data centers over distances spanning a few to more than 100 kilometers. Therefore, SAN distance extension using FC is an important part of a comprehensive, cost-effective and effective DR design.

It's important to understand the FC protocol and the optical transport technology and how they interact. In a discussion of long-distance configuration, it's also useful to note the formal structure of the FC protocol and the specific standards that define the operation of the protocol:

- FC-0 Fibre Channel Physical Media
- FC-1 Fibre Channel Encode and Decode
- FC-2 Fibre Channel Framing and Flow Control
- FC-3 Fibre Channel Common Services
- FC-4 Fibre Channel Upper Level Protocol Mapping.

Fiber Optics

There are two basic types of optical fiber: Multimode Fiber (MMF) and Single-Mode Fiber (SMF). MMF has a larger core diameter of 50 µm or 62.5 µm (the latter was common for Fiber Distributed Data Interface [FDDI] and carries numerous modes of light through the waveguide). It's generally less expensive than SMF, but its characteristics make it unsuitable for distances greater than several hundred meters. Because of this, MMF is generally used for short distance spans and is common for interconnecting SAN equipment within the data center.

SMF has a smaller core diameter of 9 µm and carries only a single mode of light through the waveguide. It's better at retaining the fidelity of each light pulse over long distances and results in lower attenuation. SMF is always used for long-distance extension over optical networks and is often used even within the data center for FICON installations. Figure 1 describes various types of optical fiber and operating distances at different speeds.

There are several types of SMF, each with different characteristics that should be taken into consideration when a SAN extension solution is deployed. Non-Dispersion Shifted Fiber (NDSF) is the oldest type of fiber and was optimized for wavelengths operating at 1310 nm, but performed poorly in the 1550 nm range, limiting maximum transmission rate and distance. To address this problem, Dispersion Shifted Fiber (DSF) was introduced.

Fiber Type	50 µm OM3 MMF				50 µm OM2 MMF			
Data Rate (MB/Sec)	100	200	400	800	100	200	400	800
Operating Distance (m)	0.5 – 860	0.5 – 500	0.5 – 380	0.5 - 150	0.5 – 500	0.5 – 300	0.5 – 150	0.5 - 50
Fiber Type	62.5 µm OM1 MMF				9 µm SMF			
Data Rate (MB/Sec)	100	200	400	800	100	200	400	800
Operating Distance (m)	0.5 – 300	0.5 – 150	0.5 – 70	0.5 - 21	2.0+	2.0+	2.0+	2.0+

Figure 1: Optical Fiber Supported Distances as Described in FC-PI

DSF was optimized for 1550 nm, but introduced additional problems when deployed in Dense Wavelength Division Multiplexing (DWDM) environments. The most recent type of SMF, Non-Zero Dispersion Shifted Fiber (NZ-DSF) addresses the problems associated with the previous types and is the fiber of choice in new deployments.

As light travels through fiber, the intensity of the signal degrades; this is called attenuation. The three main transmission windows in which loss is minimal are in the 850, 1310 and 1550 nm ranges. Figure 2 lists common fiber types and the average optical loss incurred by distance.

Fiber Loss and Link Budgets

A key part of designing SANs over long-distance optical networks involves analyzing fiber loss and power budgets. The decibel (dB) unit of measurement describes loss mechanisms in the optical path of a fiber link. Decibel loss is usually determined by comparing the launch power of a device to the receive power. Launch and receive power are expressed as decibel milliwatt (dBm) units, which is the measure of signal power in relation to 1 mW.

The link power budget identifies how much attenuation can occur across a fiber span while sufficient output power for the receiver is maintained. It's determined by finding the difference between "worst-case" launch power and receiver sensitivity. Transceiver and other optical equipment vendors typically provide these specifications for their equipment. A loss value of 0.5 dB can be used to approximate attenuation caused by a connector/patch panel. An additional 2 dB is subtracted as a safety margin:

Power Budget = (Worst Case Launch Power) – (Worst Case Receiver Sensitivity) + (Connector Attenuation)

Signal loss is the total sum of all losses due to attenuation across the fiber span. This value should be within the power budget to maintain a valid connection between devices. To calculate the maximum signal loss across an existing fiber segment, use the following equation:

*Signal Loss = (Fiber Attenuation/km * Distance in km) + (Connector Attenuation) + (Safety Margin)*

Figure 2 provides average optical loss characteristics of various fiber types that can be used in this equation, although loss may vary depending on fiber type and quality. It's always better to measure the actual optical loss of the fiber with an optical power meter.

Some receivers may have a maximum receiver sensitivity. If the optical signal is greater than the maximum receiver sensitivity, the receiver may become oversaturated and not be able to decode the signal, causing link errors or even total failure of the connection. Fiber attenuators can be used to resolve the problem. This is often necessary when connecting FC switches to DWDM equipment using single mode FC transceivers.

FC Transceivers for Extended Distances

Optical Small Form-factor Pluggable (SFP) transceivers are available in short- and long-wavelength types. Short-wavelength transceivers transmit at 850 nm and are used with 50 or 62.5 µm multimode fiber cabling. For fiber spans greater than several hundred meters without regeneration, use long-wavelength transceivers with 9 µm single-mode fiber. Long-wavelength SFP transceivers typically operate in the 1310 or 1550 nm range.

Optical transceivers often provide monitoring capabilities that can be viewed through FC switch management tools, allowing some level of diagnostics of the actual optical transceiver itself.

Distance Connectivity Options

FC SANs can be extended over long-distance optical networks in different ways. Any of the following technologies can provide a viable long-distance connectivity solution, but choosing the appropriate one depends on a number of variables—including technological, cost or scalability needs.

Note that terms are often misused or used in a generic way. In addition, products can be configured and used in any of the different ways discussed in the following sections. Ensure there's no confusion or uncertainty about the type of equipment being used. If connectivity is being provided by a service provider, in addition to equipment deployed at your site, it's important to understand all devices in the network.

Native FC Over Dark Fiber

The term "dark fiber" typically refers to fiber optic cabling that has been laid but remains unlit or unused. The simplest, but not necessarily most cost-effective or scalable method for extending SANs over distance, is to connect FC switches directly to the dark fiber using long-wavelength SFP transceivers. An optional Brocade Extended Fabrics license can be used to provide additional buffer credits to long-distance E_Ports in order to maintain FC performance across the network.

Wave Division Multiplexing

DWDM is optimized for high-speed, high-capacity networks and long distances. DWDM is suitable for large enterprises and service providers that lease wavelengths to customers. Most equipment vendors can support 32, 64 or more channels over a fiber pair with each running at speeds up to 10 Gbit/sec or more. Fiber distances between nodes can generally extend up to 100 km or farther. DWDM equipment can be configured to provide a path protection scheme in case of link failure or in ring topologies that also provide protection. Switching from the active path to the protected path typically occurs in less than 50 ms.

Fiber		Optical Loss (dB/km)		
Size	Type	850 nm	1310 nm	1550 nm
9/125 µm	SM	-	0.35	0.2
50/125 µm	MM	3.0	-	-
62.5/125 µm	MM	3.0	-	-

Figure 2: Average Attenuation Caused by Distance

Coarse Wavelength Division Multiplexing (CWDM) provides the same optical transport and features of DWDM, but at a lower capacity, which allows for lower cost. CWDM is generally designed for shorter distances (typically 50 to 80 km) and thus doesn't require specialized amplifiers and high-precision lasers (lower cost). Most CWDM devices support up to eight or 16 channels. CWDM generally operates at a lower bit rate than higher-end DWDM systems—typically up to 4 Gbit/sec.

There are two basic types of Wavelength Division Multiplexing (WDM) solutions—both are available for CWDM and DWDM implementations, depending on customer requirements:

- **Transponder-based solutions** allow connectivity to switches with standard 850 or 1310 nm optical SFP transceivers. A transponder is used to convert these signals using Optical-to-Electrical-to-Optical (O-E-O) conversion to WDM frequencies for transport across a single fiber. By converting each input to a different frequency, multiple signals can be carried over the same fiber.
- **SFP-based solutions** eliminate the need for transponders by requiring switch equipment to utilize special WDM transceivers (also known as "colored optics"), reducing the overall cost. Coarse or Dense WDM SFPs are similar to any standard transceiver used in FC switches, except they transmit on a particular frequency within a WDM band. Each wavelength is then placed onto a single fiber through the use of a passive multiplexer.

Traditionally, SFP-based solutions were used as low-cost solutions and were mostly CWDM-based. Due to compliance requirements, some customers are using these solutions to minimize the number of active, or powered, components in the infrastructure. Along with the need for increasing bandwidth and the use of such solutions to support Ethernet as well as FC connectivity, some customers are now using DWDM SFP-based implementations and require DWDM colored optics rather than CWDM colored optics to allow sufficient connections through a single fiber.

Time Division Multiplexing

Time Division Multiplexing (TDM) takes multiple client-side data channels, such as FC, and maps them onto a single higher-bit-rate channel for transmission on a single wavelength. TDM can be used in conjunction with a WDM solution to provide additional scalability and bandwidth utilization. Because TDM sometimes relies on certain FC primitives to maintain synchronization, it may require special configuration on SAN switches. Most TDM devices require IDLE primitives as fill words. Specific configuration modes are used on SAN switches to support the use of IDLE as fill words.

Additionally, it should be noted that TDM-based systems can result in a level of jitter or variable latency. As such, it isn't possible to make broad statements about the ability to use frame-based trunking and, in general, best practice is to avoid frame-based trunking on a TDM-based configuration. The need to use IDLE primitives may impact the availability of other vendor-specific features. The specific details often depend on switch firmware levels and which configuration mode is used for compatibility.

FC-SONET/SDH

Synchronous Optical NETwork (SONET) is a standard for optical telecommunications transport, developed by the Exchange Carriers Standards Association for ANSI. SONET defines a technology for carrying different capacity signals through a synchronous optical network. The standard defines a byte-interleaved multiplexed transport occupying the physical layer of the OSI model.

SONET and Synchronous Digital Hierarchy (SDH) are standards for transmission of digital information over optical networks and are often the underlying transport protocols that carry enterprise voice, video, data and storage traffic across Metropolitan and Wide Area Networks (MANs and WANs). SONET/SDH is particularly well-suited to carrying enterprise, mission-critical storage traffic, because it's connection-oriented and latency is deterministic and consistent. FC-SONET/SDH is the protocol that provides the means for transporting FC frames over SONET/SDH networks. FC frames are commonly mapped onto a SONET or SDH payload using an International Telecommunications Union (ITU) standard called Generic Framing Procedure (GFP).

SONET is useful in a SAN for consolidating multiple low-frequency channels (such as ESCON and 1, 2 Gb FC) into a single higher-speed connection. This can reduce DWDM wavelength requirements in an existing SAN infrastructure. It can also allow a distance solution to be provided from any SONET service carrier, saving the expense of running private optical cable over long distances.

The basic SONET building block is an STS-1 (Synchronous Transport Signal), composed of the transport overhead plus a Synchronous Payload Envelope (SPE), totaling 810 bytes. The 27-byte transport overhead is used for operations, administration, maintenance and provisioning. The remaining bytes make up the SPE, of which an additional 9 bytes are path overhead.

An STS-1 operates at 51.84 Mb/sec, so multiple STS-1s are required to provide the necessary bandwidth for ESCON, FC, and Ethernet, as shown in Figure 3. Multiply the rate by 95 percent to obtain the usable bandwidth in an STS-1 (reduction due to overhead bytes). One OC-48 can carry approximately 2.5 channels of 1 Gb/sec traffic, as shown in Figure 3. To achieve higher data rates for client connections, multiple STS-1s are byte-interleaved to create an STS-N. SONET defines this as byte-interleaving three STS-1s into an STS-3, and subsequently interleaving STS-3s. By definition, each STS is still visible and available for ADD/DROP multiplexing in SONET, although most SAN requirements can be met with less complex point-to-point connections. The addition of DWDM can even further consolidate multiple SONET connections (OC-48), while also providing distance extension.

STS	Optical carrier	Optical carrier rate (Mb/s)
STS-1	OC-1	51.840
STS-3	OC-3	155.520
STS-12	OC-12	622.080
STS-48	OC-48	2488.320
STS-192	OC-192	9953.280

Figure 3: SONET/Synchronous Digital Hierarchy (SDH) Carrier and Rate

Like TDM, FC-SONET devices typically require special switch configuration to ensure the use of IDLE rather than ARB primitives for compatibility.

Note again that using FC-SONET/SDH-based systems can result in a level of jitter, or variable latency. As a result, it's impossible to make broad statements about the ability to use frame-based trunking and, in general, it's best practice to avoid frame-based trunking on an FC-SONET/SDH-based configuration.

TCP/IP and Gigabit Ethernet (GbE) and 10GbE

Gigabit Ethernet (GbE) is a terminology describing a family of technologies involved in the transmission of Ethernet packets at the rate of 1024 megabits (Mb/sec or 1 gigabit per second) or multiples thereof such as 10GbE and 100GbE. GbE is defined by the IEEE publication 802.3z, which was standardized in June 1998. This is a physical layer standard following elements of the ANSI Fibre Channel's physical layer. This standard is one of many additions to the original Ethernet standard (802.3 – Ethernet Frame) published in 1985 by the IEEE organization. GbE/10GbE is mainly used in distance extension products as the transport layer for protocol such as TCP/IP. However, in some cases, the product is based on a vendor-unique protocol. Distance products using GbE/10GbE may offer features such as compression, write acceleration and buffer credit spoofing.

The Transmission Control Protocol (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 achieves this by assigning each byte of data a unique sequence number, maintaining timers, acknowledging received data through the use of acknowledgements (ACKs) and retransmission of data, if necessary. Once a connection is established between the endpoints, data can be transferred. The datastream that passes across the connection is considered a single sequence of 8-bit bytes, each of which is given a sequence number.

Conclusion

This article has discussed fiber optic characteristics and the different technologies used in long-distance storage network extension. A follow-up article will delve into network design for high availability, best practices and FCIP trunking technology. **ETJ**

Dr. Steve Guendert is a principal engineer and global solutions architect for Brocade Communications, where he leads the mainframe-related business efforts. A senior member of both the Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM), he serves on the board of directors for the Computer Measurement Group (CMG). He is also a former member of the SHARE board of directors.
Email: stephen.guendert@brocade.com
Twitter: @BRCD_DrSteve