



FROM THE SOLUTIONS CENTER

Solution Guide: SAN Distance Extension Reference

A technical reference for technologies and components associated with long-distance SAN solutions, including recommended B-Series switch configurations

BROCADE

CONTENTS

Introduction.....	3
Disaster Recovery and Business Continuance Solutions	3
Fiber Optics	3
Fiber Cabling	3
Fiber Loss and Link Budgets	4
FC Transceivers for Extended Distances	5
Distance Connectivity Options	5
Native FC over Dark Fiber	5
Wave Division Multiplexing.....	5
Time Division Multiplexing.....	6
FC-SONET/SDH	6
Extended Distance Solutions	6
Redundancy in Optical Networks	7
Brocade B-Series Switch Configuration for Distance Extension.....	9
FC over Dark Fiber	9
Coarse and Dense WDM Devices	9
TDM and FC-SONET Devices	9
Extended Distance Solutions	9
Brocade FOS Series Command Reference.....	10
CLI Commands.....	10
Extended Fabrics.....	10
Fabric-Wide Long-Distance Parameter.....	13
Trunking.....	13
Diagnostics and Troubleshooting	14
Brocade M-Series Switches.....	16
Distance and B-Series/M-Series Interoperability.....	18
FC Protocol and Application Operation over Distance.....	18
Flow Control	18
Buffer Allocation	19
SCSI Protocol.....	19
FICON Protocol	20
Appendix A: FOS Switch Platform Buffer Allocation.....	21
Brocade 2Gbit/sec Products.....	21
Brocade 4Gbit/sec Products.....	21
Condor.....	21
Goldeneye	22
Appendix B: SAN Design using Router Technology.....	23

INTRODUCTION

This document is a technical reference for the technologies and components associated with long-distance Storage Area Network (SAN) solutions, as well as specific Brocade switch configuration requirements in long-distance scenarios. The document focuses on Fabric OS® (FOS)-based switches, called “B-Series” switches; references will be made to switches running Brocade M-Enterprise OS (M-EOS), called “M-Series switches,” and the differences in operation to B-Series switches. To get the most out of this document, you should have a basic understanding of Fibre Channel (FC) protocol and experience with Brocade switch configuration.

In a discussion of long-distance configuration, it is useful to note the formal structure of the FC protocol and the specific standards that define the operation of the protocol, listed below:

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

NOTE: Brocade ASIC names, such as “Bloom,” “Condor,” and “Goldeneye” are used in this guide to distinguish between different levels of ASIC functionality.

DISASTER RECOVERY AND BUSINESS CONTINUANCE SOLUTIONS

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.

Two important objectives in the designing process are the Recovery Point Objective (RPO) and the Recovery Time Objective (RTO). The RPO is the time period between backup points and 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 pm, changes to data made within the last 23 hours will not be recoverable. The RTO describes the time needed to recover from the disaster. The RTO 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 Disaster Recovery (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 does not require immediate access after a failure can be accomplished via tape vaulting. Recovery from tape has the greatest RTO. In addition other technologies such as Continuous Data Protection (CDP) can be used to find the appropriate RPO and RTO.

FIBER OPTICS

Fiber Cabling

There are two basic types of optical fiber: Multimode Fiber (MMF) and Single-Mode Fiber (SMF). *Multimode fiber* 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 is less expensive than single-mode fiber, but its characteristics make it unsuitable for distances greater than several hundred

meters. Because of this, multimode fiber is generally used for short distance spans and is common for interconnecting SAN equipment within a data center.

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

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

Table 1. Optical fiber distances as described in FC-PI.

There are several types of single-mode fiber, each with different characteristics that you must take into consideration when deploying a SAN extension solution. 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. DSF was optimized for 1550 nm, but introduced additional problems when deployed in Dense Wavelength Division Multiplexing (DWDM) environments. The most recent type of single-mode fiber, 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, called *attenuation*. The three main transmission windows in which loss is minimal are in the 850, 1310, and 1550 nm ranges. The table below lists common fiber types and the average optical loss incurred by distance.

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	-	-

Table 2. Average attenuation caused 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 is used to describe loss mechanisms in the 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 still maintaining sufficient output power for the receiver. It is 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. It is useful to subtract an additional 2 dB for 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)

Table 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. A better solution is 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, which often occurs when single mode FC transceivers are connected to DWDM equipment.

FC Transceivers for Extended Distances

Optical Small Form-factor Pluggable (SFP) transceivers are available in short- and long-wavelength versions. 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

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

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

Dense Wavelength Division Multiplexing is optimized for high-speed, high-capacity networks and long distances. DWDM is suitable for large enterprises and service providers who 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. 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.

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 does not require specialized amplifiers and high-precision lasers (lower cost). Most CWDM devices will support up to 8 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:

- **Transponder-Based Solutions.** Allows 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 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.** These eliminate the need for transponders by requiring switch equipment to utilize special WDM transceivers (also known as colored optics), reducing the overall cost. CWDM SFPs are like any standard transceiver used in Fibre Channel switches, except that they transmit on a particular CWDM frequency. Each wavelength is then placed onto a single fiber through the use of a passive multiplexer.

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 is 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 B-Series switches when Extended Fabrics is enabled. By default, Extended Fabrics E_Ports use ARB primitives (specific to Virtual Channels) as fill words between frames. Most TDM devices require idles as fill words. Configuring a B-Series switch to use R_RDY flow control will remedy this problem and enable interoperability. Brocade M-Series switches do not use VC_RDY flow control so no special configuration is required.

FC-SONET/SDH

Synchronous Optical Network (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 area networks. SONET/SDH is particularly well suited to carry enterprise mission-critical storage traffic because it is 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).

Like TDM, FC-SONET devices typically require enabling R_RDY flow control on FOS based switches.

Extended Distance Solutions

Some FC distance extension equipment (WDM, FC-SONET, FC-IP, and so on) can participate in FC buffer-to-buffer flow control to increase the distance to be greater than what is possible with Extended Fabrics. Such devices typically participate in E_Port link initialization with an FC switch or snoop the receive-buffer field in the Exchange Link Parameters (ELP) payload. Since these devices are not aware of Brocade VC_RDY flow control, R_RDY flow control must be enabled on a B-Series switch. These devices return R_RDY credit to the switch in order to maintain performance over hundreds or even thousands of kilometers. Flow control

and error correction between distance extension nodes are performed independent of the switch and are usually dependent on the long haul network protocol.

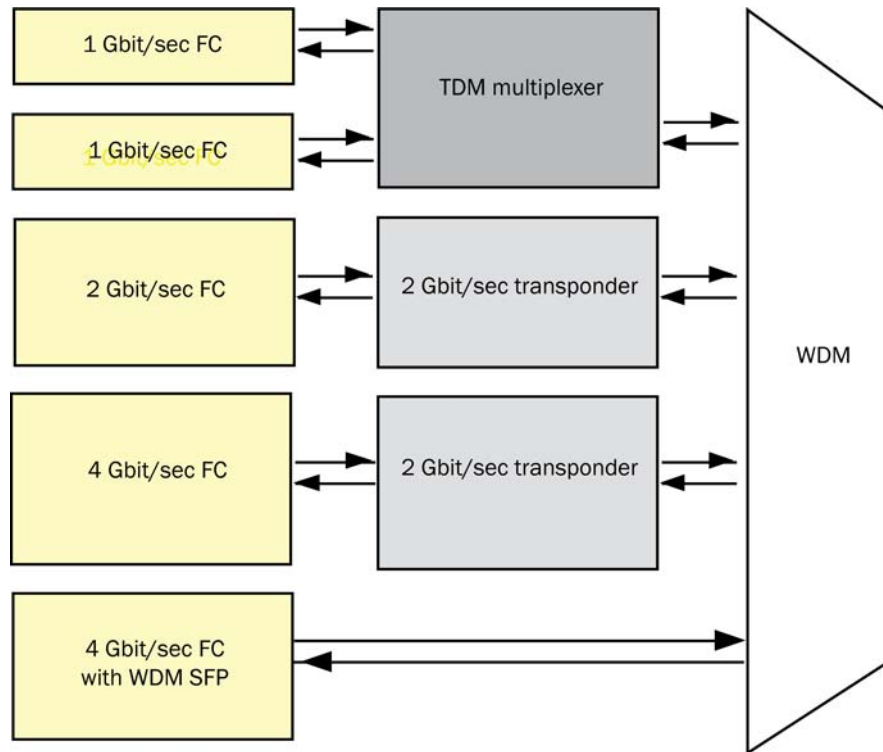


Figure 1. TDM and WDM Technologies

Redundancy in Optical Networks

Many distance extension devices are capable of providing several options for maintaining fault tolerance and availability of Inter-Switch Links (ISLs) spanning the optical network. Different methods provide different levels of availability and can be used alone or together for increased availability requirements.

The simplest form of protection is from the FC switch itself. Multiple ISLs are interconnected over the long-distance network to provide both additional bandwidth and fault tolerance in the case of a port failure on the switch or distance extension product.



Figure 2. Client-protected topology with multiple ISLs

The second form of protection is through an optical Y-Cable. This provides protection from a port or line card failure in the distance extension product, but does not provide any protection from port failure on the FC switch.



Figure 3. Y-Cable/line card protection

Optical extension products often provide protection switching in the case of optical fiber failure on the long-haul side of the network. Most WDM and SONET/SDH equipment has the ability to perform a protection switch from the active path to the standby path in less than 50 ms in order to maintain connectivity between sites.

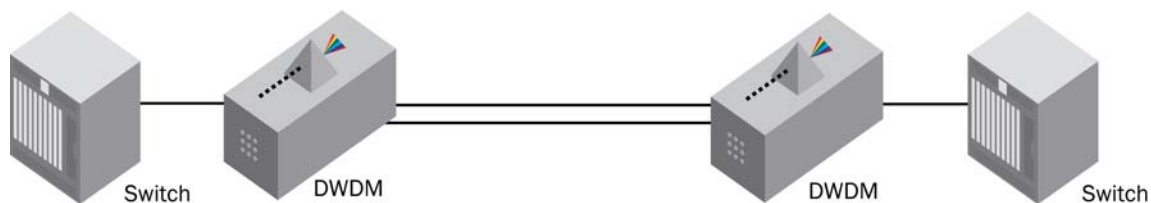


Figure 4. WDM line protection

Previous generation Bloom-based platforms were designed with short loss-of-sync timer that would cause an E_Port to reinitialize if a protection switch occurred. This would in turn cause a non-disruptive reconfiguration of the fabric, although traffic communicating with local devices would not be impacted.

Brocade has increased the loss-of-sync timer to 100 ms beginning with Condor-based platforms. For most distance extension devices, an ISL remains online during a protection switch. It is important to note that frames in flight during a protection switch will be lost and must be retried by the initiating end device.

It is also important to note that when frame-based trunking is used, the deskew value is calculated when the trunk is formed. (For more about the deskew value, see the *Fabric OS Administrator's Guide*.) If a protection switch results in a change of latency, then these deskew values are no longer valid and unpredictable behavior could result. Therefore a key design principle is to ensure that if any rerouting is possible, then entire trunks must be rerouted as homogenous groups.

Finally, from a High Availability (HA) perspective, at the FC level most SANs are designed using dual redundant fabrics. It is necessary and appropriate to ensure that the availability is not impacted through either a single point of hardware failure within the optical network nor from the cable runs following a common route, which would lead to an event that damages one optical cable most likely damages many or all of them.

BROCADE B-SERIES SWITCH CONFIGURATION FOR DISTANCE EXTENSION

Configuration requirements for B-Series switches may vary depending on the type of distance extension device used. The following information provides generic guidelines for determining how a switch should be configured. Many extension products, including WDM, can support a number of technologies within the same chassis through the use of removable blades or line cards. It is important to take note of the model and type of line card connected to the FC switch in order to determine the proper configuration, as incorrect configuration may result in errors or link failure.

FC over Dark Fiber

Connecting FC switches directly to dark fiber requires the use of long-wavelength SFP transceivers. Use Extended Fabric to allocate sufficient buffers to the long distance E_Ports.

- Configure Brocade Extended Fabrics for the appropriate distance.

Coarse and Dense WDM Devices

Because CWDM and DWDM products are protocol and bit-rate transparent, FOS configuration is identical to connecting FC switches directly to dark fiber, although you should “hard” set the FC port speed to the desired rate to ensure that WDM transponders can lock onto the bit-rate of the ISL.

1. Configure Brocade Extended Fabrics for the appropriate distance.
2. Set FC port speed on long distance E_Ports.

TDM and FC-SONET Devices

TDM and FC-SONET/SDH devices that do not actively participate in buffer credit management also require Brocade Extended Fabrics for optimal configuration. In addition, it is usually required that you configure the E_Port to operate with R_RDY flow control to maintain synchronization. Some devices may also require ports to be configured to G_Port mode so that loop initialization is not attempted.

1. Configure Brocade Extended Fabrics for the appropriate distance.
2. Set FC port speed on long-distance E_Ports.
3. Configure E_Ports to use R_RDY flow control.
4. Configure E_Ports to G_Port mode.

Extended Distance Solutions

In this document, extended distance solutions refer to products that actively participate in FC buffer-to-buffer management or extend the distance between switches further than that which is supportable by internal ASIC buffering. Many products can support distances of up to 1000 km or greater while maintaining FC speeds. FC frames are often encapsulated within another protocol, such as SONET or IP, before transmission over the long-haul network.

1. Configure E_Ports to use R_RDY flow control.
2. Set FC port speed.

The following table describes the recommended guidelines for configuring a B-Series switch for various distance extension devices. Some caveats may apply, so contact an appropriate support resource for verification.

Technology	Extended Fabrics	R_RDY Mode	Port Speed Hard Set	G_Port Mode
CWDM	√		√	
DWDM	√		√	
FC-SONET	√	√ *	√	√
TDM	√	√ *	√	√
Any of above with buffer extension		√	√	

Table 3. SAN Extension interoperability matrix

* Alternate configuration (prior to Fabric OS 5.1): fabric.ops.mode.longDistance = 1

BROCADE FOS SERIES COMMAND REFERENCE

Although this document highlights the Fabric OS Command-Line Interface (CLI) commands, all information can also be viewed through GUI management tools, including Brocade Web Tools and Brocade Fabric Manager. (See the product documentation for details.)

CLI Commands

This section details Fabric OS Command-Line Interface (CLI) commands related to long-distance SAN configuration and link monitoring tools.

Extended Fabrics

Brocade Extended Fabrics is a licensable feature that allocates additional buffering to E_Ports for increased performance over long distances. Extended Fabrics is configured using the PortCfgLongDistance command. See Appendix A for switch platform buffer allocation.

```

mw101:root> portcfglongdistance
Usage: portCfgLongDistance PortNumber Distance_Level <VC Translation Link Init> <Desired
Distance for LD or LS mode>

distance_level:      L0 : normal
distance_level:      LE : <= 10km
                    LD : auto
                    LS : static
vc_trans_link_init:  0  normal
                    1  VC translation
desired_distance:    in Kilo-meters
                    valid for LD or LS level only

```

The *VC Translation Link Init* option enables the long-distance link initialization sequence. This option should be set for long-distance links configured for VC_RDY flow control (default). For an ISL configured for R_RDY mode, it should be set to 0.

The *Desired Distance* is a required parameter to configure a port as an LS or LD mode link. For an LD mode link, the desired distance is used as the upper limit of the link distance to calculate buffer availability for other ports in the same port group. When the distance measured by the switch round-trip timer is greater than desired distance, the desired distance is used to allocate the buffers. In this case, the port operates in degraded mode instead of being disabled due to insufficient buffers. For an LS mode link, the actual distance is not measured; instead the desired distance is used to calculate the buffers required for the port.

Basic guidelines for choosing an Extended Fabrics distance level:

- L0 mode (the default E_Port mode) is used for interconnecting switches in the same data center. If the distance between two switches is greater than several kilometers, performance may be degraded because of insufficient buffer credits.
- LE mode is used to support distances up to 10 km and does not require an Extended Fabrics license. Enough buffers are allocated to the port to support full bandwidth up to 10 km regardless of port speed.
- L0.5, L1, and L2 modes operate similar to LE mode and support distances up to 25 km, 50 km, and 100 km respectively. These modes have been deprecated and are supported only for backward compatibility with older generation B-Series software. Use LD or LS Extended Fabrics in place of these modes.
 - LD mode provides dynamic distance discovery. A round trip timer determines the latency between two connected switches and automatically allocates the desired number of buffer credits needed to sustain full bandwidth on the ISL using full-size 2112-byte FC frames. The switch will never allocate more buffers than the maximum desired distance specified by the administrator. Up to a total of 250 buffers can be reserved per port.
 - LS mode is used for static allocation of buffer credits. The administrator must specify the distance of the ISL in kilometers and the switch allocates the correct number of full-size frame buffers based on the currently configured port speed. Up to a total of 250 buffers can be reserved per port.

Extended Fabrics Mode	Supported Distance	License Requirement
L0	2 km	No
LE	10 km	No
L0.5	25 km	Yes
L1	50 km	Yes
L2	100 km	Yes
LD	Dynamic Discovery	Yes
LS	Static Configuration	Yes

Table 4. Extended Fabrics distance levels

portCfgISLMode

The portCfgISLMode command is used to enable or disable R_RDY flow control on an E_Port. Both ends of an ISL should be configured for the same flow control mode. This mode cannot detect some fabric-wide inconsistencies in the fabric, including fabric-wide, long-distance mode and Port ID (PID) format, so it is important to verify identical configuration on all switches before connecting. View the portCfgISLMode help page for additional details.

```
mw101:root> portcfgislmode
Usage: portCfgISLMode PortNumber Mode
Mode:   1 - Enable ISL R_RDY Mode on port
        0 - Disable ISL R_RDY Mode on port
```

portCfgShow

The portCfgShow command is used to verify Extended Fabrics port configurations. In the example below, port 1 is configured for LS mode and VC Translation Link Initialization is set to 1.

```
mw102:root> portcfgshow
Ports of Slot 0   0  1  2  3   4  5  6  7   8  9 10 11  12 13 14 15
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
Speed           AN AN AN AN  AN AN AN AN  AN AN AN AN  2G AN AN AN
Trunk Port      ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
Long Distance   LD . . . .  . . . . . . . . . .  LS . . . .
VC Link Init    ON . . . .  . . . . . . . . . .  ON . . . .
Locked L_Port   . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Locked G_Port   . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Disabled E_Port . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ISL R_RDY Mode  ON . . . .  . . . . . . . . . .  . . . . . . . . . .
RSCN Suppressed . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Persistent Disable . . . . . . . . . . . . . . . . . . . . . . . . . . . .
NPIV capability ON ON ON ON  ON ON ON ON  ON ON ON ON  ON ON ON ON
FC Fastwrite    . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

```
where AN:AutoNegotiate, .:OFF, ?:INVALID,
SN:Software controlled AutoNegotiation.
LM:L0.5
```

portBufferShow

Use the `portBufferShow` command to determine how many buffers a port is using or determine how many buffers are free to be configured for other ports on the switch.

```

mw102:root> portbuffershow

```

User	Port	Lx	Max/Resv	Buffer	Needed	Link
Remaining	Port	Type	Mode	Usage	Buffers	Distance
Buffers	Buffers					
0			8	0	-	-
1			8	0	-	-
2			8	0	-	-
3			8	0	-	-
4			8	0	-	-
5			8	0	-	-
6			8	0	-	-
7			8	0	-	-
8			8	0	-	-
9			8	0	-	-
10			8	0	-	-
11			8	0	-	-
12	E	LS	31	31	31	25km
13			8	0	-	-
14			8	0	-	-
15			8	0	-	-
121						

Fabric-Wide Long-Distance Parameter

The fabric-wide long-distance parameter `fabric.ops.mode.longDistance` has been deprecated and should not be used with current Fabric OS release. For FOS releases prior to version 5.1 that do not support Extended Fabrics and R_RDY flow control on the same port, it may be necessary to configure this option when connecting to TDM or FC-SONET devices that do not actively participate in buffer-to-buffer flow control. Because this parameter is fabric wide, it must be set on all switches in the fabric. Enabling this option requires the switch to be disabled.

Trunking

Long-distance links using VC_RDY flow control can be part of a trunk group when they are configured for the same speed, distance level, and their link distances are nearly equal. When R_RDY flow control is used, frame-based trunking becomes disabled. Exchanged-based routing policy, used to interleave FC exchanges across multiple ISLs, can be used in either mode.

Diagnostics and Troubleshooting

Problems that arise in long-distance SAN configurations are often the result of bad or dirty fiber optic components or attenuation problems. The resulting symptom is links going offline at unexpected times or not coming up at all. B-Series switches have various tools to troubleshoot marginal links. The CLI command `portStatsShow` is used to monitor port statistics related to bad links, such as encoding errors, CRC errors, and truncated frames.

```

mw102:root> portstatsshow 12
stat_wtx          6633460      4-byte words transmitted
stat_wrx          4378728      4-byte words received
stat_ftx          321079       Frames transmitted
stat_frx          321079       Frames received
stat_c2_frx       0            Class 2 frames received
stat_c3_frx       0            Class 3 frames received
stat_lc_rx        181777       Link control frames received
stat_mc_rx        0            Multicast frames received
stat_mc_to        0            Multicast timeouts
stat_mc_tx        0            Multicast frames transmitted
tim_rdy_pri       23           Time R_RDY high priority
tim_txcrd_z       251371       Time BB credit zero
er_enc_in         0            Encoding errors inside of frames
er_crc            0            Frames with CRC errors
er_trunc          0            Frames shorter than minimum
er_toolong        0            Frames longer than maximum
er_bad_eof        0            Frames with bad end-of-frame
er_enc_out        0            Encoding error outside of frames
er_bad_os         103821       Invalid ordered set
er_c3_timeout     0            Class 3 frames discarded due to timeout
er_c3_dest_unreach 0            Class 3 frames discarded due to
destination unreachable
er_other_discard  0            Other discards
er_zone_discard   0            Class 3 frames discarded due to zone
mismatch
er_crc_good_eof   0            Crc error with good eof
er_inv_arb        0            Invalid ARB
open              0            loop_open
transfer          0            loop_transfer
opened           0            FL_Port opened
starve_stop       0            tenancies stopped due to starvation
fl_tenancy        0            number of times FL has the tenancy
nl_tenancy        0            number of times NL has the tenancy
zero_tenancy      0            zero tenancy

```

The portErrShow command is similar and displays error statistics for all ports on the switch. It is useful to run these commands several times in a row or over various time periods to see if error counters are incrementing at unexpected rates

```

mw102:root> porterrshow
      frames  enc  crc  too  too  bad  enc  disc  link  loss  loss  frjt  fbsy
      tx   rx   in  err shrt long  eof  out   c3 fail sync  sig
=====
0:    0    0    0    0    0    0    0    0    0    0    0    13    0    0
1:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
2:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
3:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
4:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
5:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
6:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
7:  664   758    0    0    0    0    0    0    94    0    1    2    0    0
8:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
9:    0    0    0    0    0    0    0    0    0    0    0    1    0    0
10:   0    0    0    0    0    0    0    0    0    0    0    1    0    0
11:   0    0    0    0    0    0    0    0    0    0    0    1    0    0
12: 321k 321k    0    0    0    0    0    0    0    0    13   22    0    0
13:   0    0    0    0    0    0    0    0    0    0    0    1    0    0
14:   0    0    0    0    0    0    0    0    0    0    0    1    0    0
15:   0    0    0    0    0    0    0    0    0    0    0    1    0    0

```

BROCADE M-SERIES SWITCHES

Brocade M-Series switches do not use virtual channels, frame-based trunking, or exchanged-based trunking (also called DPS). Open Trunking balances by manipulating Fabric Shortest Path First (FSPF) routing based on throughput and BB credit utilization. This means that most of the discussion of this document in terms of compatibility issues among certain behaviors of some optical equipment and Fibre Channel are not relevant for M-EOS fabrics.

Extended Fabrics is not a licensed feature on M-Series products, therefore where additional Buffer-to-Buffer (BB) credits are required, they are simply configured on specific ports. Some M-Series products have credits pooled (Brocade M4300, M4400, M4500, M4700, and 2 – 10 Gbit/sec blades in the Brocade Mi10K); others have credits dedicated to specific ports (Brocade M3016, M3216, M3032, M3232, M6064, M6140, and 4Gbit/sec blades on the Mi10K). In either case, the GUI or CLI must be used to configure availability of additional credits for long distance.

M-Series Open Trunking, a licensed feature, monitors and rebalances based upon thresholds on both utilization and throughput. Before rebalancing, the system calculates the net affect of a rebalance to ensure that improved performance will result. Rebalancing can occur either because the throughput threshold is exceeded or because of a shortage of BB credits. Preferred and Prohibit path are sometimes used instead or as well as Open Trunking in bandwidth allocation. Refer to specific product manuals and white papers for details.

As with B Series products more recent firmware revisions on all these platforms have Advanced Port Diagnostic capabilities which can be used to better monitor/manage ports particularly where used on metro optical connections. Most importantly port fencing can isolate a port that is being impacted by multiple fault conditions, to ensure that the fabric as a whole is not destabilized. Similarly Advanced Performance Monitoring capabilities are available using Brocade Enterprise Fabric Connectivity Manager (EFCM) along with the Performance and Event optional modules—analogueous to many of the capabilities in B-Series APM and FW FOS-based features.

At metro optical distances it should not be necessary to make changes to the E_D_TOV (default 2 seconds, range 0.2 to 60 seconds) or R_A_TOV (default 10 seconds, range 1 to 120 seconds) configuration values in the fabric.

As with B-Series switches, ports can be manually set to a specific FC speed or left on auto-negotiation. For metro optical connections, use a fixed speed matched to the underlying optical equipment or appropriate for the dark fiber cable length.

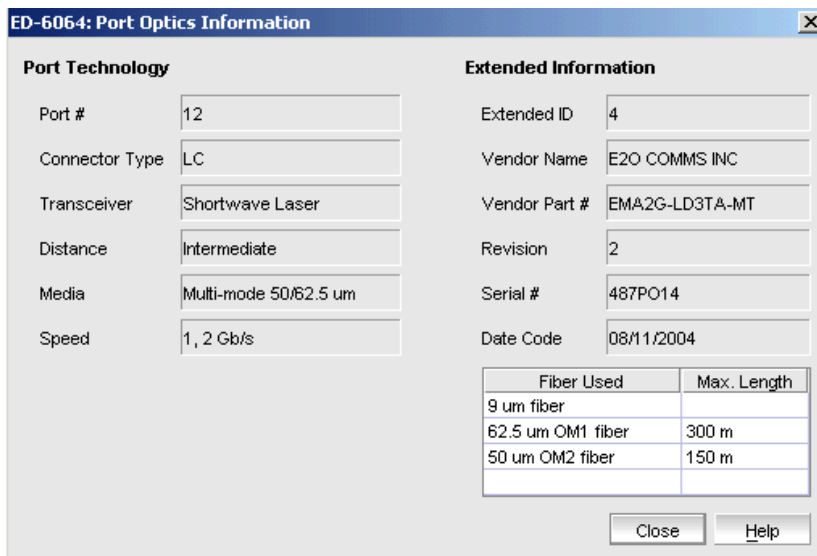
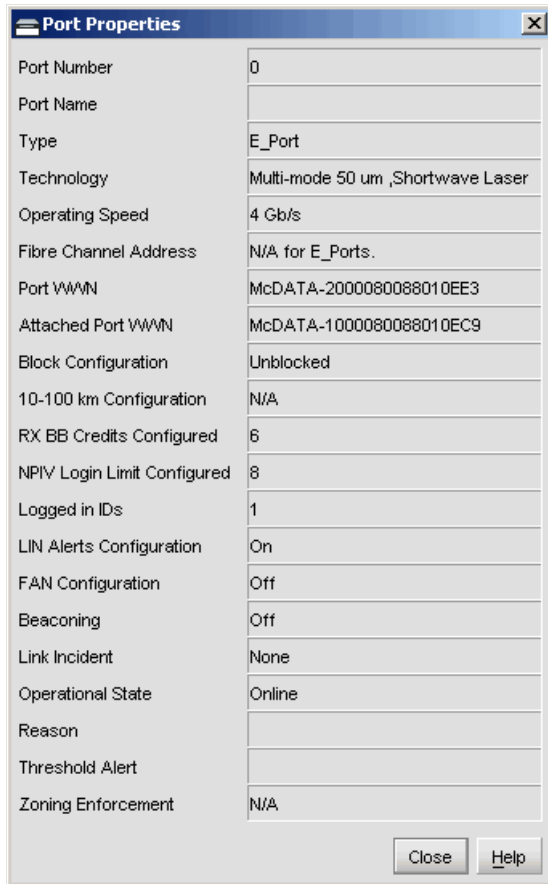
BB credits can be changed per port between ranges that are dependent on the specific M-Series product using the following procedure from the CLI (shown) or GUI:

- Examine available and current receive BB credit: `> config port showCredits`
- Block / offline port: `> config port blocked 4 true`
- Change port receive BB credit: `> config port rxCredits 4 40`
- Unblock / online port: `> config port blocked 4 false`

NOTE: M-series switches have a setting “reroute delay,” which is equivalent to IOD on B-Series switches in that it delays frame movement by E_D_TOV time period during fabric reconfiguration to ensure frames are in order. This setting is OEM dependent.

M-Series products include internal and external loopback test capability. The external loopback capability is designed to test the optics of the port as well as the port logic by connecting the RX and TX of the GBIC/SFP

together. In principle, it is used to test a long cable run. Note that the loopback distance would be double the one-way distance, which means that the failure of such a test may not indicate a problem.



DISTANCE AND B-SERIES/M-SERIES INTEROPERABILITY

Currently interoperability is supported only using McDATA mode and Fabric OS 5.2.2_NI on the B-Series Brocade 5000 switch and M-EOS 9.2 or 9.6 on any 1-, 2-, or 4-Gbit/sec port across the M-Series range of Sphereon and Intrepid products. This will be extended across most Brocade 4-Gbit/sec products with the release of FOS 6.0. Also not supported at this time is 10-Gbit/sec interoperability. Support is planned for interoperability using McDATA open fabric mode in a future firmware release. Unlike the current Brocade open mode (mode 1), the new mode will be a more feature-rich, SW-2 open fabric, although it will be supported only for interoperability between B-Series and M-Series products. When the new mode is available, mode 1 will be removed. Refer to the appropriate interoperability material for details.

An ISL between a B-Series and an M-Series product does not use virtual channels, but instead normal R_RDY flow control. Neither will multiple ISLs form frame-based trunk groups, as this is an ASIC-dependent capability and can be used only between B-Series switches. These restrictions eliminate many of the potential challenges in compatibility between FC and the underlying optical network. DPS and Open Trunking are not precluded from interoperability, as they work independently on outgoing transmission (also known as exchange-based trunking) from B-Series to M-Series switches and Open Trunking from M-Series to B-Series switches. Similarly BB credits are direction specific; so on each side, the normal mechanism can be used to configure greater-than-standard BB credits for each direction. (The mechanisms for configuring and monitoring the distance-specific components for each side were discussed in the previous section.)

FC PROTOCOL AND APPLICATION OPERATION OVER DISTANCE

Flow Control

Brocade B-Series switches can support two methods of flow control over an ISL: Virtual Channel (VC_RDY) and Receiver Ready (R_RDY) flow control. VC_RDY is the default method and uses multiple lanes or channels, each with different buffer credit allocations, to prioritize traffic types and prevent head-of-line blocking.

VC_RDY flow control differentiates traffic across an ISL. It serves two main purposes: to differentiate fabric internal traffic from end-to-end device traffic, and to differentiate different data flows of end-to-end device traffic to avoid head-of-line blocking. Fabric internal traffic is generated by switches that communicate with each other to exchange state information (such as link state information for routing and device information for Name Service). This type of traffic is given a higher priority so that switches can distribute the most up-to-date information across the fabric even under heavy device traffic. Additionally, multiple I/Os are multiplexed over a single ISL by assigning different VCs to different I/Os and giving them the same priority (unless QoS is enabled). Each I/O can have a fair share of the bandwidth, so that a large-size I/O will not consume the whole bandwidth and starve a small-size I/O, thus balancing the performance of different devices communicating across the ISL.

The second method, R_RDY flow control, is defined in FC standards and has only a single lane or channel for all frame types. It should be noted however that even when B-Series switches are configured to use R_RDY flow control (as used by M-Series switches), there are other mechanisms to enable QoS and avoid head-of-line blocking problems.

When connecting switches across dark fiber or WDM communication links, VC_RDY is the preferred method, but there are some distance extension devices that require the E_Port to be configured for R_RDY. In order to configure R_RDY flow control on B-Series switches, use the portCfgISLMode command.

Buffer Allocation

Before discussing FC-level buffer allocation, it is important to note that the availability of sufficient FC-level buffering is not itself sufficient to guarantee bandwidth utilization. Specifically other limitations, particularly at the SCSI level of the storage initiator and/or target, are often the limiting factor. The I/O size, IO per second limit, and concurrent/outstanding IO capability at the SCSI level of the initiators/targets can be and often are gating factors.

While exact calculations are possible, a simple rule of thumb is used in the calculation of the BB credit requirement of a given link. Based on the speed of light in an optical cable, a full-size FC frame spans approximately 4 km at 1 Gbit/sec, 2 km at 2 Gbit/sec, 1 km at 4 Gbit/sec, or 400m at 10 Gbit/sec. To calculate the BB credit requirement, simply take the round-trip cable distance and divide by the frame length. For example, at 1 Gbit/sec at 10km, 5 BB credits are required to allow wire speed data transfer. As the FC speed or distance increases the number of BB credits increases in a linear fashion.

In addition, as a rule of thumb, a link with a round-trip latency of 1 ms spans a distance of approximately 100 km. With a round-trip time of 200 km, 200 buffer credits would be required at 4 Gbit/sec.

Having insufficient BB credits will not cause link failure, but will reduce the maximum throughput. So the aforementioned 1-ms link running at 4 Gbit/sec with only 100 BB credits can achieve a maximum throughput of only approximately 2Gbit/sec.

Finally, these calculations use approximate metrics. The most important thing to remember is that if a given implementation had in a high percentage of FC frames significantly less than maximum size, then more BB credits may be required.

SCSI Protocol

The performance of SCSI write operations over a long-distance ISL can depend on several factors, including the size of the I/O operation, the number of outstanding I/Os, and target response latency. A typical write operation requires at least four trips across the network: 1) the SCSI write request, 2) a Transfer Ready (XFER_RDY) response from the target, 3) data transmission, and 4) SCSI completion status (FCP_RSP). Because the initiator must wait until XFER_RDY is received from the target before sending actual data, the latency of a long-distance ISL can reduce response times and overall throughput of data transmission over the ISL.

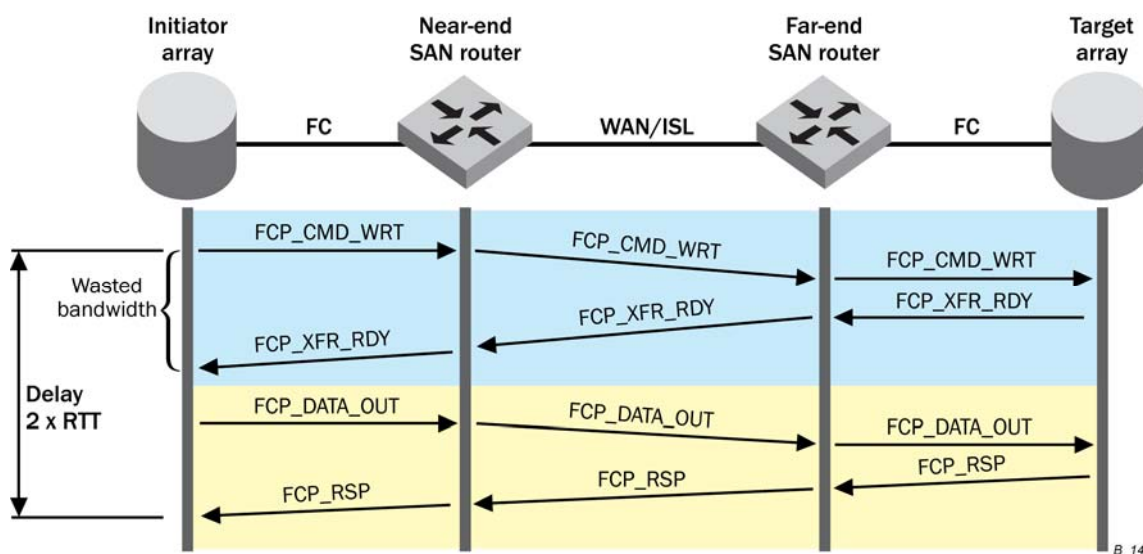


Figure 5. SCSI write operation without Brocade FastWrite

FOS 5.3.x and later support FC FastWrite to improve SCSI write operations over long-distance ISLs. This is accomplished by the creation of a Proxy Target (PT) local to the initiator host, and a Proxy Initiator (PI) local to the target storage device. When a SCSI write is received, the PT responds with XFER_RDY, enabling the host to begin transmission of data over the network without having to wait for response from the target device, thus eliminating latency caused by the long-distance link. Once the PI intercepts the XFER_RDY from the target device, data is sent to the target.

FastWrite is supported on the Brocade 7500 SAN Router and the Brocade 48000 Director with the FR4-18i Routing Blade.

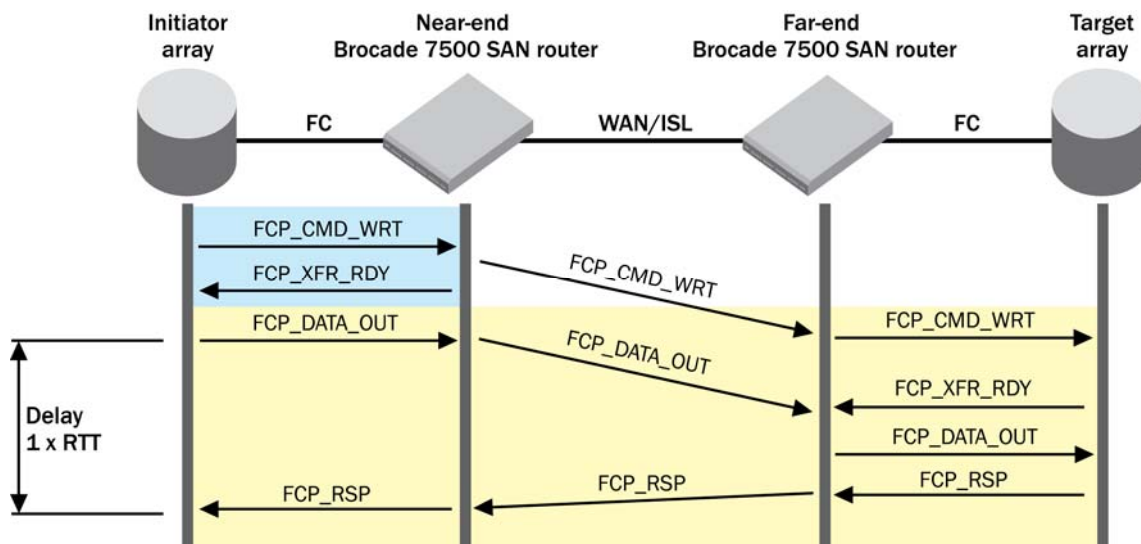


Figure 6. SCSI write operation with Brocade FastWrite

The following steps are required to configure and enable FC Fastwrite on a Brocade switch. Refer to the FOS Administrators Guide for additional information on configuration and additional considerations.

1. Create a zone set to filter FC FastWrite flows. FC FastWrite flows are configured by creating a zone with "fcacc_" as the prefix. SAN zones are required to use "lsan_fcacc" as the prefix.
2. Enable FastWrite using the fastwriteCfg command. Enabling or disabling FC FastWrite disrupts data traffic, and a Brocade 7500 will be rebooted. In the case of the FC4-18i in a Brocade 48000, the blade is powered off and then back on.
3. Once enabled on the switch, FC FastWrite can be configured on a per-port basis using the portCfg command.

FICON Protocol

In a number of cases, the FC traffic will be FICON rather than SCSI/FCP. Most array-based replication solutions produce SCSI/FCP traffic for replication, although some arrays (particularly those supported for mainframe systems only) produce FICON traffic. In addition, in FICON installations, the use of XRC for replication and tape both result in FICON traffic across the distance connections.

APPENDIX A: FOS SWITCH PLATFORM BUFFER ALLOCATION

Buffer allocation functions differently on Brocade 2 Gbit/sec products and 4 Gbit/sec products.

Brocade 2Gbit/sec Products

With Brocade's 2 Gbit/sec switches, a total of 108 buffers are available per quad. By default, 16 buffer credits are assigned to any F_Port or FL_Port, and 26 are assigned to local (not long-distance) E_Ports. The extra buffers are available to all ports in the quad as a pool. Any node that reaches 100 percent utilization automatically goes to the pool for additional buffer credits before releasing its dedicated credits. The pool is also used to support long-distance ISLs. A single port can receive all 108 credits, but this will result in the remaining 3 ports in the quad being disabled.

Brocade 4Gbit/sec Products

Brocade's 4 Gbit/sec products use two different ASICs and each one has a different amount of buffer credit as described below. Local E_ports still receive the same amount of credits as the 2 Gbit/sec products, that is, 26 credits.

Condor

Brocade's fourth generation ASIC has a total of 32 ports and 1000 usable buffer credits that are shared among the entire chip. The Condor ASIC is used in the Brocade 4100, 4900, and 7500 Switches and the Brocade 48000 Director. In each of these products, 8 buffer credits are reserved per port to avoid buffer starvation on a given port. If the port is configured as an E_port, then the 8 reserved credits are released to the pool. Some credits are used for internal links in the Brocade 4900, 7500, and 48000.

The following lists the available credits for each product in addition to what is already reserved on each port to prevent buffer starvation (8 credits per port), as well as internal links.

- The Brocade 4100 has a single group of 32 ports; it has 744 additional available credits.
- The Brocade 4900 has four 16-port groups, and each has 584 additional available credits. The four port groups are: ports 0 - 15, 16 - 31, 32 - 47, and 48 - 63.
- The Brocade 7500 has two 8-port groups, and each has 472 additional available credits. The two port groups are: ports 0 - 7 and 8 - 15.
- The Brocade 48000 currently has three different port blades, and each has the following additional available credits:
 - The FC4-16 port blade has 584 additional available credits.
 - The FC4-32 port blade has two groups of 16 ports, each of the 16-port groups has 624 additional available credits. The two groups are: ports 0 - 15 and 16 - 31.
 - The FR4-18i Routing Blade has two groups of 8 ports, and each of the 8-port groups has 472 additional available credits. The two groups are: ports 0 - 7 and 8 - 15.

Goldeneye

Brocade's fifth generation ASIC is used primarily in embedded products (The Brocade 4012, 4016, 4018, 4020, and 4024) and also in the Brocade 200E Switch. Goldeneye has 24 ports—all capable of running at 1, 2, or 4 Gbit/sec; it has 96 Gbit/sec (192 Gbit/sec full duplex) of cross-sectional bandwidth. It also has 272 available credits across the chip. In each of the previously mentioned products, 8 buffer credits are reserved per port to avoid buffer starvation on a given port; if the port is configured to E_port, then the 8 reserved credits are released to the pool.

Even though each of the embedded products is slightly different when it comes to port count, it is easy to figure out how many credits are leftover for long distance. The last two digits represent the total number of ports available on the embedded switch. For example, the Brocade 4012 has 12 total ports. This means that 96 buffer credits ($12 * 8 = 96$) are reserved out of the total 272, and 192 are leftover for long-distance ports.

Brocade added Extended Fabrics support to the Brocade 200E Switch in Fabric OS 5.1.0; prior to that, the Brocade 200E had only LO and LE modes. Since the Brocade 200E has a total of 16 ports, 128 credits are reserved and 144 are leftover for long-distance ports.

APPENDIX B: SAN DESIGN USING ROUTER TECHNOLOGY

Currently there are three use cases for using routers over metro optical infrastructures as follows:

- **IP over metro optical.** Prior to FOS 5.3 and with the Eclipse Routers many customers still used routing technology for both fault isolation and the write acceleration from FastWrite technology which allowed synchronous replication at extended distance
- **FC over metro optical.**
 - Prior to FOS 5.3, some customers used routers for scalability and fault isolation when extending FC over metro optical networks.
 - **Running FOS 5.3 and later on the Brocade 7500 SAN Router,** customers now use FC over distance and gain the benefits of write acceleration. By using the Brocade 48000 Director chassis with a mix of routing blades and 10 Gbit/sec FC blades, customers can get fault isolation, 10-Gbit/sec connectivity site to site, and write acceleration.

The use of routers for fault isolation is valid only for open systems FCP, because the address translation capability of the router that occurs as part of the fault isolation is not compatible with FICON.

See “Multiprotocol Routing for SANs” by Josh Judd (available for download from the www.brocade.com > Brocade Bookshelf) for router-specific rules, restrictions, and best practices.

Given the commonality between 10 Gbit/sec Ethernet and 10 Gbit/sec FC, many customers have 10-Gbit/sec WDM connectivity available. Indeed this may be more readily available than 4 Gbit/sec FC, particularly if the connections are provided by a telephone company rather than from customer-owned equipment. Therefore in some cases there may be a cost advantage to the use of 10-Gbit/sec connectivity. Similarly, with dark fiber, a higher speed naturally reduces the effective cost of a link.

While leveraging 10 Gbit/sec is relatively simple in non-router configurations, since all B-Series and M-Series directors have a 10 Gbit/sec FC option, it is possible to use 10 Gbit/sec in a routed infrastructure as well. Leveraging a Brocade 48000 chassis, a number of routing blades can be installed along with a number of 10 Gbit/sec FC blades. The routing blades can be connected to the local fabrics and the 10 Gbit/sec ports can be used to form the backbone fabric. Subject to the configuration rules on the use of FastWrite, it is even possible to then leverage FastWrite over 10-Gbit/sec metro optical networks.

The routers can be used for FICON acceleration if distances require, however, for best practices in this area, consult FICON-specific documentation and white papers.

© 2007 Brocade Communications Systems, Inc. All Rights Reserved. 11/07 GA-SG-061-00

Brocade, the Brocade B-weave logo, McDATA, Fabric OS, File Lifecycle Manager, MyView, SilkWorm, and StorageX are registered trademarks and the Brocade B-wing symbol, SAN Health, and Tapestry are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. FICON is a registered trademark of IBM Corporation in the U.S. and other countries. All other brands, products, or service names are or may be trademarks or service marks of, and are used to identify, products or services of their respective owners.

Notice: This document is for informational purposes only and does not set forth any warranty, expressed or implied, concerning any equipment, equipment feature, or service offered or to be offered by Brocade. Brocade reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact a Brocade sales office for information on feature and product availability. Export of technical data contained in this document may require an export license from the United States government.