WHITE PAPER

Evaluating Remote Data Replication Solutions

Brocade Extension or Array IP Replication Port

A high-level guide focusing on Fibre Channel Storage Area Network (SAN) design and best practices, covering planning, topologies, device sharing in routed topologies, workload monitoring, and detecting server and storage latencies—to help with decisions required for successful SAN design.

Table of Contents

Overview

Are array IP replication ports for remote data replication the right choice? In nearly all cases, Brocade

Extension solutions are the right choice. This tech brief enumerates the values of Brocade Extension and applicable architectures, including both Open Systems and Mainframe. The Brocade 7840 Extension Switch is a purpose-built, enterprise-class product that is characterized by an essential feature set: excellent application performance, wide swath of scale, robust security, high reliability, network integration, comprehensive monitoring, application visibility, and diagnostic tools. Array native IP replication falls short of offering an enterprise-class feature set.

An alternative to array native IP ports are Brocade Extension products connected to array native FC ports. Ultimately, customers want to gain the greatest replication performance, flexibility, and reliability with the least amount of operational and capital expense. Brocade Extension is optimized for a wide range of scale from small to large, which are applicable to nearly every replication environment.

Brocade technology integrates perfectly into any IP network and provides a more efficient data transport capable of full bandwidth utilization across great distances as compared to array native IP. The defining features that bring value to Brocade Extension are described in this tech brief, including Extension Trunking, Extension Hot Code Load (HCL), Adaptive Rate Limiting (ARL), Quality of Service (QoS), Brocade Fabric Vision technology, and Brocade Network Advisor. Brocade also provides a full spectrum of security features and connectivity validation tools. Overall, Brocade Extension products leverage 20 years of distance connectivity innovation and thought leadership, as demonstrated by the fact that they are the market's preferred extension solution.

The Situation

Consider a situation in which there are two or more data centers replicating data for DR (Disaster Recovery). This may be open system LUN replication or mainframe volume replication. This tech brief applies equally to either environment. You might ask if RDR (Remote Data Replication) performance over distance is meeting your requirements, and if the Recovery Point Objective (RPO) objective is optimal. You want to determine if more bandwidth will fix the problem, and how much more that will cost each month. Maybe data replication is taking too much time and occasionally falls behind. Is the replication across multiple arrays unbalanced, and does this negatively affect multisession consistency groups? Compression rate and/or ratio might be insufficient to provide adequate throughput. In addition, when data leaves the protection of your data center, for compliance there may be a requirement to encrypt data in-flight. Other questions might be: Does the combination of encryption and aggressive compression bring throughput to low levels? Is there a better way to configure, manage, monitor, and troubleshoot your RDR network?

There is a clear gap between the IP capabilities that are native on an array and those offered by Brocade Extension. Frequently, customers turn to Brocade seeking performance solutions for their RDR needs. Primarily, these customers are using array native IP ports, ports that are just not providing the expected results or ports that were deployed prior to recent data growth.

Why Brocade? Brocade has been developing RDR technologies for nearly 20 years and is the thought leader in this arena. No other company offers such advanced technology. Array native IP ports are typically an implementation of Internet Small Computer System Interface (iSCSI) and have low performance and inferior Transmission Control Protocol (TCP) stacks, as well as a lack of sophisticated features and valued functionality. Ultimately, the situation with native IP replication involves a lack of performance, visibility, and advanced functionality. Brocade provides a cost-effective high-performance solution that is intuitively managed with robust operations (configuration, monitoring, actions, reporting, and diagnostics).

Architectures

Brocade Extension can scale from small to large. The smallest-scale deployment is two Brocade 7800 4/2 Extension Switches connected directly to storage array replication ports. This offers 2 Gigabits per second (Gbps) of Fibre Channel over IP (FCIP) Wide-Area Network (WAN) bandwidth. Assuming 2:1 data compression,

4 Gbps of replication bandwidth will be seen by the array. At this scale, cost is sensitive, and only a single Brocade Extension network is connected to both controllers. The environment may grow and expand to two parallel RDR networks. Further scaling can be achieved by adding a PoD (Port on Demand) license, which upgrades the Brocade 7800 4/2 to 16/6. This more than doubles the original capacity. This deployment is small-scale, cost-effective, upgradable, and extraordinarily powerful.

On the other hand, a very large scale deployment may use four Brocade 7840s deployed in parallel pairs.

Two parallel Brocade 7840s can accommodate 160 Gbps of mainframe XRC, mainframe tape, disk replication, and open systems tape—all coming from multiple sources. The overall capacity doubles if you merely add a second parallel redundant network. 160 Gbps = 2 x 7840, and each Brocade 7840 has 2 x Data Processor (DP) at 40 Gbps each. This is the bandwidth seen on the FC/FICON side. On the FCIP WAN side, this equates to 80 Gbps (2 x 7840 times 2 x DP@20 Gbps each), assuming 2:1 compression using the Fast Deflate algorithm. There is plenty of failover bandwidth for ARL or Extension HCL. The Brocade 7840 can directly connect to applications via either 24 x 16 Gbps FC ports or a production fabric. This deployment offers an extremely large-scale, redundant, and cost-effective architecture that is extraordinarily powerful.

The Brocade 7840 provides the performance and tools to best transport storage data of all kinds to anywhere in the world, offering high reliability, great efficiency, outstanding performance, and considerably easier management. Converging disparate flows to a purpose-built Brocade Extension device offers many benefits. With growth, individual array native IP ports become increasingly difficult to manage. Brocade Network Advisor is a comprehensive tool that simplifies management and helps users proactively diagnose and troubleshoot issues to maximize uptime and increase operational efficiency. Brocade Network Advisor pulls data from Brocade Fabric Vision technology, including the Monitoring and Alerting Policy Suite (MAPS) and Flow Vision, into customizable dashboard views with deep drill-down capabilities that provide comprehensive visibility into network health and performance of storage replication.

Brocade Extension uniquely fits into large-scale storage deployments. Large-scale deployments often require the following: multimodality (disk, tape, open systems, mainframe, and so on), heterogeneous arrays, huge bandwidths, high throughput, nonstop operations, tools for operations, and robust diagnostics. The RDR network referred to here can be integrated into production fabrics or kept completely separate, depending on what makes most sense. Separation can be achieved logically using Virtual Fabrics (VF) or physically using completely different switches. Either way, the RDR network and all fabrics can be managed from a single pane of glass using Brocade Network Advisor.

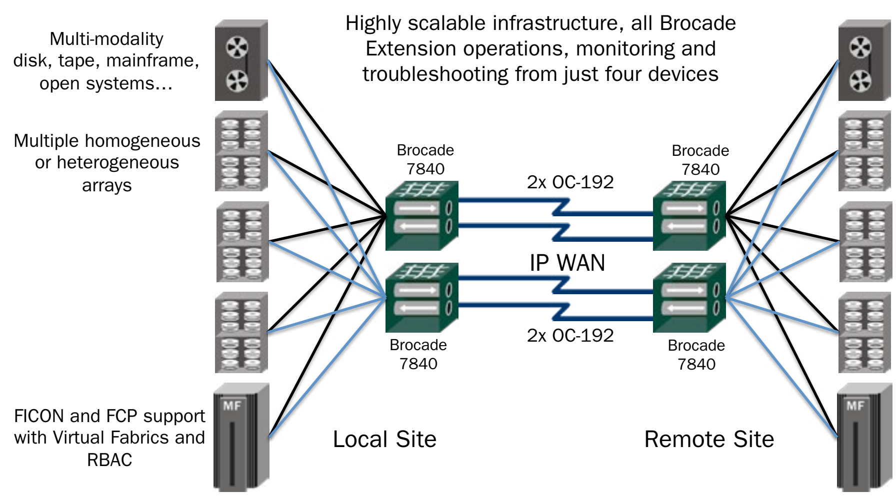

Figure 1: Multi-modality large-scale Brocade Extension deployment.

Enterprise Class

The Brocade 7840 is indisputably the storage industry's highest throughput extension device compared to competing extension products and methods. Native 8 Gbps or 16 Gbps FC ports on most arrays far outperform their native IP 10 GbE counterparts. These much faster FC ports are the ports that connect to Brocade Extension. Brocade Extension provides ample capacity for multiple FC ports, including those coming from multiple sources. The Brocade 7840 contains the world's fastest FC switching ASIC, 64 network processors, and 128 GB of high-speed RAM, an impressive offering. Even systems that distribute processing across multiple ports cannot claim near equivalent processing power. Brocade Extension provides tremendous value.

Performance

Purpose-Built Hardware

There are two ways to build hardware for applications such as extension. You can build a general-purpose processor-based device that is defined by the software it runs. This is the least expensive option to design and build and is the most flexible. By changing the software, you can entirely change the application. Extension networking is by far the most common application for such service processors and, once deployed, it is rarely changed out during the life of the product. The downside of this approach is lower performance and higher latency. Generally, array-based native IP replication takes this approach.

Alternatively, you can develop purpose-built hardware and firmware. This is a more expensive option to build and design, but it offers the benefit of extreme performance and ultralow latency. The Brocade 7840 has such a design and runs Brocade Fabric OS® (Brocade FOS, or Fabric Operating System) to provide flexibility of code updates and new features, and to reprogram the hardware (Field Programmable Gate Array [FPGA]) if needed. The FPGA performs line-rate FCcomp (Fast Deflate), line-rate IP security (IPsec) (AES 256), and other high-speed data transmission functions.

The architecture of the Brocade 7840 is elegantly simple and enables ultralow latency for synchronous applications. The data path within the Brocade 7840 is concise and is implemented in a few fast but highly effective components.

Synchronous Replication

Brocade offers the only extension products with ultralow added latency. The added propagation time gained through Brocade Extension products is well within the tolerance for synchronous applications. In today's array native IP ports, the TCP/IP processing cannot provide ultralow added latency and cannot be used efficiently with synchronous applications. There is no reduced latency advantage to using array-based native IP ports; in fact, in relative terms, it is a disadvantage.

For synchronous applications, there is more value-add than just low added latency. Purpose-built hardware for compression and IPsec enables use with synchronous replication. This is made possible with native FC on Brocade Gen5 switches as well; however, combine this capability with Extension Trunking, and the synchronous solution becomes very compelling. Extension Trunking, which is explained in more detail later in this tech brief, is a technology exclusive to Brocade for establishing multiple circuits between two VE_Ports. Each circuit can take its own IP network path, usually Dense Wavelength-Division Multiplexing (DWDM). Multiple paths provide resiliency, fast error recovery from lost links, no data lost in-flight from lost links (Lossless Link Loss [LLL]), data integrity, in-order delivery, and true bandwidth aggregation of each circuit. These are all beneficial to fast response time and network reliability, which are demanded by synchronous storage replication. An example benefit of Extension Trunking in a DWDM environment is the prevention of IFCC (Interface Control Check) in the event of an optic, cable, or optical multiplexor failure.

Today, customers use Brocade Extension for synchronous applications with positive results. Keep in mind that the IP network itself must perform equally. Adding synchronous applications to robust extension over a poor performing IP network will equal a poor performing synchronous application. IP networks are not poor performing by nature. Nevertheless, IP networks can be poor performing if they are not constructed well, as is true with any network not suited for the task. IP networks can be properly built and configured to meet the requirements of synchronous storage. Brocade has the IP networking products that meet these requirements but that discussion requires a different tech brief.

As mentioned above, LLL and Extension Trunking are both exclusive to Brocade and are described in the High Availability (HA) section of this tech brief.

Encapsulation Method (FC FCIP)

FCIP encapsulation has the following headers: the replication application's header, FCP, FCIP, TCP, IP, and Ethernet. Many arrays use the underpinnings of iSCSI for their replication transport. This means that the array's native IP transport consists of the storage payload plus the following headers: the replication application's header, iSCSI, TCP, IP, and Ethernet. The native IP claim is that the replication data goes directly into TCP/IP without the need for FCP and FCIP headers. This is only partially true. What is not true is the implication that this process is more efficient. FCIP is switched out for iSCSI, and the FCP header is removed, thus only the removal of the FCP header saves a slight amount of overhead. TCP, IP, and Ethernet are the same across both models.

Depending on compression, array-based native IP replication can be much less efficient at producing IP datagrams than the supported Maximum Transmission Unit (MTU). The arrays are doing one iSCSI frame at a time, and IP compression may reduce that frame below the MTU. Just because the MTU is set to a particular size does not mean the array will fill it. The smaller the datagram is relative to the supported MTU, the less efficient it is. The goal is to reduce overhead by filling datagrams to the supported MTU. This creates maximum payload per unit of overhead.

Brocade uses a unique method of forming streams of bytes from storage I/O. There is no concept of individual FC frame discrete encapsulation, which would be far too inefficient (yet is used by many competing products). Brocade forms a stream of bytes, which is transported by WAN-Optimized TCP (WO-TCP). 16 data frames form a stream called a "batch." Each batch has a single FCIP header, which reduces headers by 16:1. The batch is then compressed. By compressing the entire batch, it is possible to gain higher compression ratios. Brocade has developed various Deflate-based compression algorithms, namely Fast Deflate, Deflate, and Aggressive Deflate. Each algorithm has a different trade-off of speed vs. compression ratio. The stream fills TCP segments to their maximum segment size. The maximum segment size is the IP MTU minus the IP and TCP headers (IP + TCP headers is about 40 bytes). The result is full-size IP datagrams and minimal overhead, no matter what the compression is. Relative to other competing replication transports, the Brocade encapsulation method excels in efficiency and has no disadvantage compared to array-based native IP transports.

WAN-Optimized TCP

Transmission Control Protocol (TCP) is centric to the high-speed transport of large data sets that are common in storage extension. Through years of experience, Brocade has developed an aggressive TCP stack called WAN-Optimized TCP (WO-TCP). WO-TCP is a transport that cannot be outperformed by competing WAN optimization products. In other words, you receive negligible benefit from WAN optimization when using the Brocade 7840 Extension Switch. Overall, Brocade technology is comparable from the perspective of the data transport bottom line. The total bytes transferred within the same period of time, over the same bandwidth, will be virtually the same compared to competing WAN optimization products. All of these benefits are provided, plus the added satisfaction that the cost of purchasing Brocade Extension is considerably less compared to WAN optimization products.

Often array-based native IP requires WAN optimization to be comparable. WAN optimization may be indicated for array-based native IP replication in one or more circumstances:

- WAN latency exceeds 100 milliseconds (ms).
- WAN quality is poor.
- WAN is prone to errors.
- WAN has excessive jitter.

WAN optimization equipment is tremendously expensive compared to Brocade Extension. Brocade Extension makes WAN optimization totally unnecessary. Adding WAN optimization introduces complexity, another point of failure, and another asset to configure, manage, monitor, and troubleshoot. If WAN optimization already exists, Brocade Extension will unnecessarily consume that resource, which other applications can use instead.

WO-TCP integrates with ARL, and the synergy of these two technologies creates an industry-dominating transport for storage. No similar transport exists on any storage array-based native IP replication. Clearly, WO-TCP demonstrates the enterprise-class Brocade 7840.

Refer to Figure 2. Enterprise-class Brocade Extension exclusively offers the following:

- The industry's highest performance
- Ultralow latency Brocade Extension devices with IPsec, supporting synchronous applications
- LLL, which prevents IFCC when a circuit is disrupted
- Extension HCL for nonstop operations during firmware updates
- WO-TCP, the industry's highest-performing TCP stack

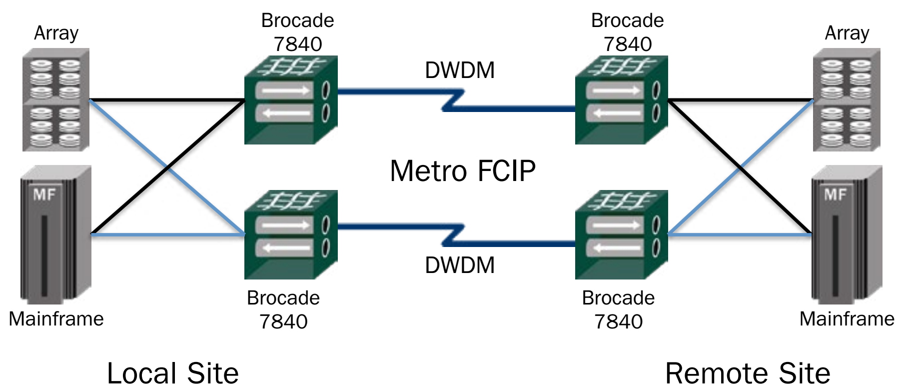

Figure 2. Enterprise-class Brocade Extension.

Bandwidth Delay Product (BDP)

In TCP, the amount of outstanding unacknowledged data that is needed to fully utilize a WAN connection depends on what the Bandwidth Delay Product (BDP) is. BDP is easily calculated by multiplying the bandwidth by the IP network's Round Trip Time (RTT). For example, a 1 Gbps WAN link with 160 ms RTT has a BDP of 20 megabytes (MB). This means that you must have at least 20 MB of data in-flight to fully utilize the bandwidth of this link. Any less than 20 MB results in "droop." Droop is the inability to fully saturate the WAN connection.

How does BDP apply to the Brocade 7840 vs. array-based native IP replication? Brocade Extension has superior BDP capacity built into WO-TCP, and it can maintain line-rate for a 10 Gbps WAN connection across 160 ms RTT without droop—outstanding! Consider that this connection could be an OC-192 between the USA and Hong Kong, which has a RTT of about 160 ms. This works out to 1250 MBps times 0.16 seconds = 200 MB. The Brocade 7840 actually has more than 200 MB of BDP capacity, if you take into account added memory for retransmits and other TCP window elasticity. You certainly cannot find this level of BDP in any array native IP ports. In fact, some arrays have only 2 MB available for their TCP stack. Insufficient BDP results in data transmission droop. The Brocade 7840 may experience droop, but not until after 160 ms RTT. Even if you do not need an OC-192 that extends halfway across the world, this is a testament to the potent architecture of Brocade Extension.

Some array native IP ports can scale BDP by adding more TCP sessions. If four native IP ports were used with four TCP sessions on each port, it would yield a BDP of 16 x 2 MB = 32 MB. If the WAN connection

was an OC-192 (10 Gbps), the maximum RTT without droop is 26 ms under perfect circumstances. 40 Gbps (4 x 10 Gbps) worth of array Ethernet ports are consumed, producing only 10 Gbps of data rate. This indicates a lack of serious technology supporting array-based native IP replication. Two 16 Gbps FC ports on an array should be able to feed 10 Gbps of data to a Brocade Extension device. This saves valuable ports on the array to be prudently used for hosts.

What about “dirty link” performance? It is not reasonable to think that all links run error-free all the time. What happens to the BDP when there is a transmission error (bit error, out-of-order segment, dropped segment, and so on), and data has to be recovered and put back in-order? These situations effectively lengthen the RTT by the number of round trips it takes to recover from the transmission problem. It could be 1 RTT, or it could be multiple RTTs. If all you have is minimal BDP capacity in your native IP port TCP session, data transfer is going to become abysmal until the situation resolves itself. Because it has such a large BDP capacity, the Brocade 7840 has the ability to maintain high throughput in such situations. WO-TCP is very aggressive, and its aggressiveness is applied directly to the Storage Administrator 's mission to expedite the protection of data.

Maximum Transmission Unit (MTU)

MTU is the largest-size IP datagram that an IP network can support end-to-end. If you are unsure what your

Path MTU (PMTU) is, the Brocade 7840 can automatically determine the path MTU by using the PMTU feature.

Perhaps your IP network has a ring or multiple links with an active/passive architecture. For example, your storage applications might be using one of two links. The other link either remains passive, and your service provider cuts you a low-cost deal, or nonstorage applications use this path. Storage is permitted to take the alternate link only when the primary link goes down. Brocade Extension circuits have metrics and failover groups to automate failover for such architectures. When an extension circuit goes down, another circuit within the same failover group comes online. No data in-flight is lost, and all data remains in-order. All circuits, including backup circuits, are independent and can be uniquely configured for each environment—in this example, primary and alternate.

Protocol Optimization

Array-based native IP replication has no protocol advantage. Claims are made that native IP ports have the benefit of protocol optimization because there is no need for unnecessary round trips when transporting data. This is no longer an issue, because many storage replication applications, whether native FC or native IP, perform unsolicited writes. Unsolicited writes are writes sent without first requiring a Transfer Ready from the remote side. Unsolicited writes indicate the immediate sending of data across the WAN, usually on the heels of the SCSI write command. This means many FC-based RDR applications have built-in protocol optimization, and the behavior is identical to array-based native IP replication. In the case of FC-based replication that does not have single round-trip functionality, Brocade has innovated FastWrite. FastWrite is a protocol optimization technique that eliminates the command/transfer-ready round-trip needed to start the data-out of an I/O. The effect is the same as unsolicited writes.

When using extension external to storage arrays, combining other storage applications such as tape and mainframe over the same tunnel (one VE_Port) or different tunnels (different VE_Ports) is cost-effective. Brocade Extension has protocol optimization for Open Systems Tape Pipelining (OSTP) read/write and FICON Acceleration (XRC, tape read/write, and Teradata); furthermore, optimization of all these protocols simultaneously is supported. Brocade Extension can discern these different applications and apply protocol optimization accordingly. These applications can be extended great distances, mitigating the effects of latency while maintaining full bandwidth utilization.

Scale and Operations

There are many aspects to scale that you might need to address. How small is your environment? How large? Can it be managed at scale? Will it be cost-effective at scale? As outlined in this tech brief, all aspects of scale are addressed by Brocade Extension, which offers a wide range of solutions that are cost-effective. In addition, Brocade has customers and experience that span the entire spectrum of these scales.

Throughput

Brocade Extension products scale on the FCIP WAN side from 100 Megabits per second (Mbps) up to 40 Gbps. On the FC/FICON side, they can scale from 2 Gbps to 80 Gbps, depending on compressibility of the data.

Compression

Brocade has developed specialized compression algorithms for the Brocade 7840 switch. These algorithms vary in processing rate and compression ratio and are the most aggressive compression algorithms available in the industry. They cannot be found on any array-based native IP ports or competing products.

On the Brocade 7840, there are three compression algorithms:

- Fast Deflate
  - Rate: FC/FICON maximum ingress rate is 40 Gbps precompressed per DP
  - Ratio: Typical is approximately 2:1
- Deflate
  - Rate: FC/FICON maximum ingress rate is 16 Gbps precompressed per DP
  - Ratio: Typical is approximately 3:1
- Aggressive Deflate
  - Rate: FC/FICON maximum ingress rate is 10 Gbps precompressed per DP
  - Ratio: Typical is approximately 4:1

Note: Brocade makes no promises, guarantees, or claims to compression ratios achievable on customer specific data.

Compression ratios will vary and can be higher or lower.

Multimodality

An investment in high-performance extension technology usually means it must be leveraged across the enterprise to include the various modalities, for example, mainframe volume replication, various open systems disk replication, mainframe tape, open systems tape, and so on. All of these can easily be accommodated by Brocade Extension and managed by different administrator groups within an enterprise, by using Virtual Fabrics and Role Based Access Control (RBAC).

Special features can be applied to these modalities to ensure proper operation. In the case of FICON, you can apply FICON Accelerator, FICON CUP, and FICON Management Server. In the case of open systems disk and tape, you can apply FCIP-based FastWrite and OSTP. You can use Virtual Fabrics to separate ports into their own Logical Switch (LS). You can configure Logical Switches for FICON traffic and Logical Switches for FC traffic with different required settings. Member circuits of VE_Ports located in various LSs can all share the same Ethernet interfaces. This is vital, considering that an Ethernet interface may be 10 GbE or 40 GbE and is meant to supply connectivity for many trunks across the different Virtual Fabric Logical Switches. In addition, to parse out the various circuits coming through that Ethernet link at the next hop DC Local Area Network (LAN) switch, Virtual LAN (VLAN) tagging is used. All the functionality needed to support multimodality environments is available on Brocade Extension.

Configuration Simplicity

Configuration of Brocade Extension is considerably simple compared to alternative solutions. Most customers choose to configure Brocade Extension themselves, whereas with some arrays it is necessary to schedule a customer engineer to come out and configure the platforms or make any subsequent changes.

You can configure Brocade Extension in two ways: One method uses the Command-Line Interface (CLI), and the other method uses Brocade Network Advisor. Brocade Network Advisor is a graphical user interface method of configuration for users that prefer GUI methods.

The example configuration shown below includes the following: First, the Ethernet interfaces are set to 10 GbE (10 Gbps SFP+ are required). An IPsec policy is created. There is one trunk, defined by VE_Port 24. The trunk has two circuits. Two logical IP interfaces (ipif) are created, one for each circuit (192.168.0.2 and 192.168.0.3), and the 10 GbE interfaces 0 and 1 are used (GE0 and GE1). Two IP routes are created to point to the local router gateway (192.168.0.1), one for each circuit. The remote side is 192.168.1.0/24. The MTU of this IP network supports jumbo frames at 9216 bytes. There are two data processors (DPs) per Brocade 7840 (DP0 and DP1). This trunk uses DP0. A tunnel is created with a minimum bandwidth (-b) of 5 Gbps and a maximum (-B) of 10 Gbps, specified in kilobits per second (Kbps). The trunk uses Fast Deflate compression, and IPsec is enabled. The first circuit (numbered "0" is not entered in the CLI command) is added automatically when the tunnel is created. A second circuit (numbered "1" is specified in the CLI command) is added next, which turns the tunnel into a trunk. Of course, many features and functions can be deployed, which adds to any configuration. In most cases, adding functionality (VLANs, QoS) is as simple as adding additional arguments to the commands shown below.

Below is a Brocade 7840 configuration example of a two-circuit trunk:

```
portcfgge ge0 --set -speed 10G
portcfgge ge1 --set -speed 10G

portcfg ipsec-policy pol1 create –k "think up some pre shared key for both sides"

portcfg ipif ge0.dp0 create 192.168.0.2/24 netmask 255.255.255.0 mtu 9216
portcfg iproute ge0.dp0 create 192.168.1.2 netmask 255.255.255.255 192.168.0.1

portcfg ipif ge1.dp0 create 192.168.0.3/24 netmask 255.255.255.0 mtu 9216
portcfg iproute ge1.dp0 create 192.168.1.3 netmask 255.255.255.255 192.168.0.1

portcfg fciptunnel 24 create --local-ip 192.168.0.2 --remote-ip 192.168.1.2 -b 5000000
-B 10000000 -c –fast-def -i enable pol1
portcfg fcipcircuit 24 create 1 --local-ip 192.168.0.3 --remote-ip 192.168.1.3 -b
5000000 -B 10000000
```

Doing the mirror of this on the remote side creates a trunk. Overall, the configuration of Brocade Extension is fairly simple—even with some of the advanced features discussed in this document (ARL, compression, IPsec, Extension Trunking, jumbo frames).

As for the IP network itself, the SLA (Service Level Agreement) with your IP networking department is not more stringent for Brocade Extension relative to array native IP requirements. In fact, the IP network SLA for Brocade Extension is less, if not the same, due to the robust ability to drive across less capable IP networks. Considerable planning might be involved to obtain the right IP network deployment for RDR; however, this is a similar requirement when deploying both array native IP and Brocade Extension.

Brocade offers both engineering services and IP networking products to help with this type of infrastructure planning, network design, equipment acquisition, implementation, and acceptance testing. Brocade Network Advisor is a comprehensive management tool enabling Storage Administrators to manage their infrastructure end-to-end, including both the Brocade storage network and Brocade IP network.

Security

IPsec

Any data leaving the safe confines of the data center should be protected using encryption. Encryption does not only apply to the public Internet. Private WAN connections are not secure outside of your data center. Unsecured data leaving your data center potentially could cause data breaches and even unwanted publicity for an enterprise. Brocade has developed hardware-based IPsec for secure data in-flight across Brocade Extension Inter-Switch Links (ISLs). Brocade IPsec operates at line-rate and introduces only a couple of microseconds (µs) of added latency, making it useful for synchronous applications. Brocade IPsec uses AES-GCM-256, Diffie-Hellman 2048-bit Modular Exponential (MODP), Internet Key Exchange version 2 (IKEv2), Hashed Message Authentication Mode Secure Hash Algorithm 512 (HMAC-SHA2-512), and Transport Mode, and it is rekeyed every few hours without disruption. A Pre-Shared Key (PSK) is configured per tunnel and trunk on each side.

Best practice is to use Brocade IPsec for Extension. Brocade IPsec is part of circuit formation and protects data from virtually every type of attack, including sniffers, data modification, identity spoofing, man-in-the-middle, and denial of service. Brocade IPsec requires no additional licenses or costs and is very simple to configure. IPsec plus Extension Trunking gives you the ability to granularly load balance encrypted storage flows across all the trunk's member circuits. Up to 20 Gbps is supported for a single trunk, and two such trunks are supported per Brocade 7840. This is a large amount of encrypted load balanced data bandwidth (40 Gbps) for a single box. IPsec provides prudent security for most organizations and costs nothing extra with Brocade Extension.

Brocade FOS Security Features

Brocade FOS offers a large number of security features, such as RBAC. These features are beyond the scope of this document, but you should know that they exist.

High Availability

There are many aspects to building a highly available RDR and tape network. Availability can be enhanced by network redundancy, resiliency of components, failover/failback functionality, continuous operations during firmware updates, and preservation of bandwidth.

Extension HCL

Extension HCL (Extension Hot Code Load) was introduced to the storage industry with the Brocade 7840. Firmware upgrades can be done without tunnel disruption. A firmware update can take considerable time, too much time to have a large extension connection down. Years ago, WAN links had much less bandwidth, and it was not paramount to maintain connectivity during firmware updates. The interim backlog of data was acceptably small. However, by today's standards the amount of backlog data during a firmware upgrade can be significant, on the order of half a terabyte or more when using one 10 Gbps connection. At many enterprises, to comply with RPO policy and to maintain a comfort level for Storage Administrator s, nonstop operations are required. The Brocade 7840 is the only product on the market that maintains extension connectivity during a firmware upgrade.

Extension HCL from Brocade is lossless and always keeps data in-order. During the firmware update process no data is lost, and all data sent to ULP is consistent and in-order. This means that Extension HCL can be used in mainframe environments without causing IFCC, which is a testament to the underlying advancements to this technology.

Extension Trunking

With Extension Trunking from Brocade, each storage I/O accesses all the WAN bandwidth that is seen by all the circuits belonging to a tunnel. An extension tunnel is defined by its VE_Port endpoint. The tunnel has a maximum bandwidth of 20 Gbps on the Brocade 7840 Extension Switch, 10 Gbps on the Brocade FX8-24 Extension Blade, and 6 Gbps on the Brocade 7800 Extension Switch.

Having multiple circuits per tunnel enables high availability. Extension Trunking spreads data across all circuits, and those circuits can be dispersed across various paths and service providers; there is no requirement for equal bandwidth or latency among the circuits. Load balancing uses Deficit Weighted Round Robin (DWRR) on a per-batch basis. This is a granular load balancing method with the ability to failover/failback without data loss or out-of-order data. This capability is essential for mainframe environments and makes for more durable open system RDR environments as well.

If an IP path goes down at any level (service provider, local or remote, switches or routers, optics, cables, and so on)—and circuits are dispersed across different service providers, routers, switches and paths—then no outage will occur, provided at least one path remains up. ARL will optimally readjust the bandwidth usage based on the remaining path or paths. Extension Trunking is lossless: No data will be lost, and all data will be received by the Upper Layer Protocols (ULP) in-order. The storage applications will not time out and will not perform error recovery.

ARL (Adaptive Rate Limiting)

Where rate limiting occurs in the network is important, and that point is after storage flows have been aggregated and before the IP network. Brocade Extension should be connected as close to the WAN as possible. This way, the aggregate of all data flows is managed holistically with security and QoS effectively applied.

Array “auto-adjust” rate limiting pertains to just the array itself. More than one array renders auto-adjust rate limiting ineffective. Moreover, auto-adjust rate limiting cannot take into account changes occurring in the WAN. As an example, consider a degraded situation in which a primary OC-192 (10 Gbps) goes offline and is backed up by two secondary OC-48s (5 Gbps) that are shared with nonstorage applications. There is no way for array auto-adjust rate limiting to compensate for this outage. The overall bandwidth has been reduced in half, forcing the native IP ports to use TCP flow control to manage inevitable congestion. TCP does not efficiently manage flow control while providing performance, and the result is poor storage throughput, which is worse than the bandwidth outage itself.

This is not the case with Brocade ARL. ARL automatically adjusts the rate limiting on all associated circuits replicating across the IP network, regardless of the ingress FC device and the WAN path or paths. ARL automatically adjusts rate limiting when other Brocade Extension circuits go online/offline or the available IP bandwidth that is being experienced changes. ARL works across all Brocade Extension products using the same WAN infrastructure.

Shared WAN connections with nonstorage applications are very common. ARL is designed to work on WAN connections that are shared with other IP storage and/or nonstorage applications. Array auto-adjust rate limiting was not designed for such instances. In fact, the Brocade 7840 can be configured so that during an outage, high-priority applications maintain their bandwidth while lower-priority devices sacrifice theirs. ARL dynamically adjusts rate limits independent to each circuit, permitting efficient use vof WO-TCP across a variety of ever-changing WAN environments. In this example, during the WAN service outage the overall bandwidth is halved, and the Brocade ARL, integrated with WO-TCP, best utilizes the available bandwidth while maintaining nonstop operations.

ARL is a function of primary importance for optimal operation. If auto-adjust rate limiting on array-based native IP ports cannot efficiently adapt to changes in WAN bandwidth, either too much rate or too little rate will cause nonoptimal performance. Consider how constantly varying workloads make it impractical to tune individual arrays. Additionally, LUN/volumes cannot practically be relocated to remedy the array auto-adjust rate limiting problem. Dedicating WAN bandwidth to specific array native IP ports causes imbalanced issues. The only answer is to have Brocade ARL located downstream from the arrays.

Metrics and DF Bit

Not all WAN connections are provisioned equally, and this may be due to the capabilities of the service provider or intermediate devices, and/or due to the cost associated with various connections. This means that backup circuits may need to have different configurations compared to primary circuits. Consider an example of two WAN connections. The primary connection supports jumbo frames of 9216 bytes (MTU = 9216). The other is a less expensive secondary connection that does not support jumbo frames (MTU = 1500).

There are two ways to deal with this. The first example described here is not the Brocade way. Do not set the DF bit (Don't Fragment bit) in FCIP datagrams. The double negative (don't set the Don't Fragment bit) means that it is permitted to fragment these packets. If IP datagrams exceed the network's supported MTU, routers will fragment datagrams in order to conform to the supported MTU. Fragmentation is a resource-expensive operation and is not done in router hardware. It fact, it is done in software, upon arriving at the destination device. The destination is forced to reassemble these fragments, which takes time and processor resources. Generally speaking, IP fragmentation is a highly inefficient process that is not intended for high-speed high-rate data transfers. Fortunately, there is a better way to handle this situation.

The Brocade way is as follows. Set the DF bit in FCIP datagrams. Setting the DF bit is not a configuration option. Brocade FCIP datagrams always have the DF bit set for optimal operation. Simply put, if datagrams do not pass across the network, then the circuit is misconfigured. Oversized datagrams will not be fragmented and are dropped. In this example, two circuits are configured in the same failover group and with different metrics. The primary circuit is configured with metric 0 and an MTU of 9216 bytes. The backup circuit is configured with metric 1 and an MTU of 1500 bytes. A WAN path change occurs, resulting in jumbo frame to standard frame support (9216   1500 bytes). This prevents the passage of FCIP datagrams on the primary circuit. When these datagrams stop, the circuit (metric 0) goes down as soon as the Keepalive Time Out Value (KATOV) expires (which is set for 1 second). At this point, the IP network has already converged to the secondary path; now the backup circuit (metric 1) must be brought back online. The backup circuit is brought online, and data resumes without any data loss and before the RDR application times out.

The Brocade 7840 automatically, immediately, and repeatedly tries to bring circuits back online after going down. When the primary path returns to an online state, the primary circuit with metric 0 will retry, succeed, and come back online. When a metric 0 circuit comes online and is in the same failover group as a metric 1 circuit, the metric 1 circuit will go offline. The transitions from metric 0   1 and 1   0 is a lossless process due to the Brocade LLL.

The Brocade 7840 can reroute between different MTU paths without any disruption, frame loss, or out-of-order frames. In this example, during interims of degraded IP network operation, storage is forced to use a less optimal MTU path. Nevertheless, operations stay online, and no data is lost in transit. During degraded IP network operations there is no need for IP fragmentation, which is inefficient. During normal IP network operations there is no need for smaller-than-supported MTU packet sizes, which is inefficient. It is most efficient to use circuits configured specifically for the primary and backup environments.

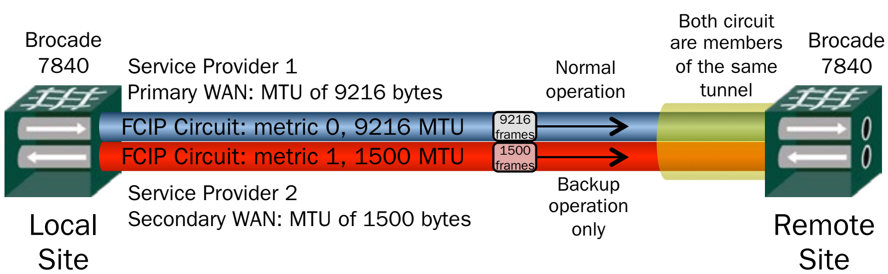

Figure 3. Brocade reroute between different MTU paths.

Reliability, Availability, and Serviceability (RAS)

The Brocade 7840 is one component of an overall system that works together to guard against disruption. The Brocade 7840 has certain features that can facilitate the quick resolution of support issues and the root cause determination of faults or degradation.

Brocade Fabric Vision technology is supported on Brocade Extension products to help maximize uptime, simplify management, and provide unprecedented insight and visibility across the storage network. With powerful built-in monitoring, management, and diagnostic tools, organizations can proactively monitor, increase availability, and dramatically reduce costs.

Monitoring and Alerting Policy Suite (MAPS)

Brocade customers ask, "How can we resolve support issues more quickly and effectively?" OEM support organizations struggle to resolve cases before they become critical issues and before the Remote Data Replication (RDR) application is already down. This situation is further aggravated by the inability to quickly pinpoint whether the problem is a network or storage issue. Both customers and OEMs are greatly interested in the ability provided by Brocade to proactively monitor and effectively troubleshoot the local FC connections and network device health—as well as the ability of the IP network to meet its SLA. Storage arrays are not in a position to provide proactive warnings or identify network problems.

It is important to build intelligence into these networking systems. When a data connection starts to experience errors of any kind, the proper action may not be readily apparent until the situation becomes a major outage. Years of practical experience must be applied, because there is a large permutation of errors and effects. Brocade provides operational excellence by leveraging 20 years of Extension experience, both in open systems and mainframe. Brocade introduced MAPS for Brocade FOS and Brocade Network Advisor to provide a comprehensive suite of monitors, alerts, actions, and reporting. MAPS assists operations in achieving higher availability, quicker troubleshooting, and infrastructure planning. It provides a prebuilt, policy-based threshold monitoring and alerting tool that proactively monitors the storage extension network health, based on a comprehensive set of metrics at tunnel, circuit, and QoS layers. Administrators can configure multiple fabrics at one time using predefined or customized rules and policies for specific ports or switch elements.

MAPS monitors utilization, packet loss, RTT, jitter, and state changes for tunnels/trunks, circuits, and PerPriority-TCP-QoS (PTQ). Each PTQ priority (class-F, low, medium, high) is monitored independently and includes throughput, duplicate Acknowledgments (ACKs), packet count, packet loss, and slow-starts.

MAPS can be used in many situations. One example is the fencing of circuits that exhibit errors. MAPS is simple and easy to deploy with preset threshold levels and responses (Conservative, Moderate, and Aggressive) based on Brocade best practices. As needed—though not required—virtually every element is customizable in MAPS. This type of configuration, monitoring, reporting, and diagnosis system is not available on array-based native IP replication.

Flow Vision

There are advantages to using Brocade Extension. Scale is certainly one of those advantages, but there are others. Visualization of flows through tunnels is an advantage. Not all flows are created equal, and a tunnel managed by Brocade allows administrators to visualize each application. To ensure SLAs are being met, Storage Administrators monitor network and flow behavior. This would be very difficult to accomplish if managed from each originating device and port.

Troubleshooting network flows is often a difficult and daunting endeavor. Making matters worse, Storage Administrators are not familiar with IP networks, and the IP Network Administrators are not familiar with storage. These two groups have very different cultures and operating guidelines. It is difficult for Storage Administrators to depend solely on Network Administrators to maintain their replication environment, which makes flow, TCP, circuit, and tunnel monitoring and visualization considerably more important.

When troubleshooting storage flows, imagine that the flows fall into one of two categories: victims or perpetrators. If something goes wrong in the network, every flow becomes a victim. However, sometimes there is nothing wrong with the network, and flows fall victim to perpetrators. Perpetrator flows are flows that utilize excessive resources to the point that other flows fall victim. This frequently happens downstream from the storage handoff to IP networking. Brocade Extension provides features, functionality, and tools to deal with storage SLAs. Flows within the protection of Brocade Extension tunnels meet their SLAs when they come up against perpetrator flows. Flows from array native IP ports may encounter insurmountable challenges, because they do not have the best technology for the task.

Flow Vision enables administrators to identify, monitor, and analyze specific application flows in order to simplify troubleshooting, maximize performance, avoid congestion, and optimize resources. The Brocade 7840 has the capability to monitor specific LUN flows between F_Ports that are communicating end-to-end across the extension network. It is also possible to monitor flows coming in from an E_Port. This feature set is called Flow Vision, a component of Brocade FOS and the Gen5 ASIC. At LUN level granularity, I/O Operations Per Second (IOPS) and data rate can be monitored. Flow Vision includes the following features:

- **Flow Monitor:** This provides comprehensive visibility into flows across storage extension networks, including the ability to automatically learn flows and nondisruptively monitor flow performance. Administrators can monitor all flows from a specific storage device that are writing to or reading from a destination storage device or LUNs, or across a storage extension network. Additionally, administrators can perform LUN-level monitoring of specific frame types to identify resource contention or congestion that is affecting application performance.
- **Flow Generator:** This is a built-in traffic generator for pretesting and validating storage extension infrastructure—including route verification, QoS zone setup, Extension Trunking configuration, WAN access, IPsec policy setting, and integrity of optics, cables, and ports—for robustness before deploying applications.

For more information on Brocade Fabric Vision technology, please visit:

http://www.brocade.com/solutions-technology/technology/san-fabric-technology/fabric-vision.page

Refer to Figure 4. Brocade Extension meets all network integration needs:

- What are the interface speeds? 40 GbE, 10 GbE, or GbE
- What is the path's MTU? Use Brocade PMTU.
- What about primary and backup paths? They are lossless when IP paths switch.
- Do you need QoS? PTQ assigns each priority to an autonomous TCP session.
- Do you need Ethernet-based QoS? Use 802.1P Layer 2 (L2) Class of Service (CoS).
- Do you need IP-based QoS? Use DSCP.
- Do you need to determine Full Duplex and Pause Frames? Enable/disable GbE Autonegotiation.
- Does one physical connection support multiple circuits over different VLANs? Use 802.1Q VLAN tagging.
- Is Network Address Translation (NAT) required? Assign inside vs. outside devices.

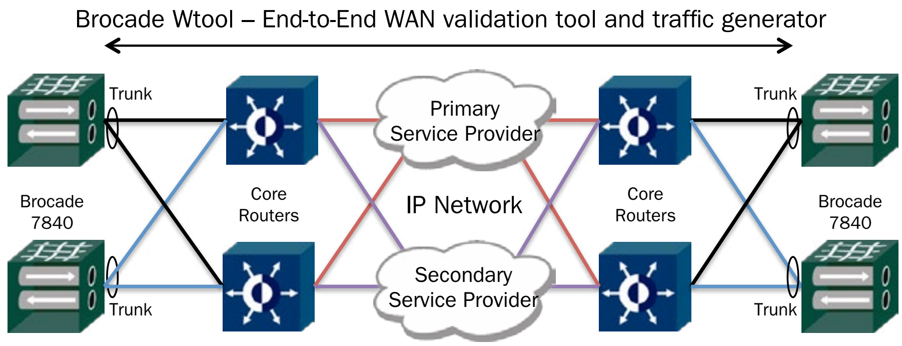

Figure 4. Brocade Extension integrates into the IP network.

Qualification and Validation Tools

Brocade Extension offers a variety of tools to validate and troubleshoot IP networks.

Wtool

Wtool was introduced with the Brocade 7840 and accurately tests multiple IP network paths. Wtool generates traffic at specified rates between a pair of IP addresses. Wtool reports achieved throughputs, jitter, experienced latencies, congestion, packet losses, and network reordering. Wtool supports pertinent circuit characteristics, including PMTU, VLAN tagging, IPv4/IPv6, IPsec, and jumbo frames. The main purpose of Wtool is to validate the IP network before deploying a circuit. It is also useful as a diagnostics tool when you have a reliability issue with a circuit.

Wtool simulates Extension traffic exactly how the IP network would see it, such that the test results are truly relevant. Wtool runs in the background and allows multiple simultaneous test sessions to coexist, up to eight sessions (four sessions per DP). Each test session equates to a single circuit. The total concurrent test capacity is eight circuits or two fully loaded tunnels/trunks. These connections are a UDP-like simulation to facilitate detection of congestion, out-of-order delivery, and packet loss; however, Wtool runs the same TCP as the circuits do, so that IP network security mechanisms do not prevent testing, and IP network security devices are tested too.

Ping and Trace Route

Brocade Extension supports both ping and trace route, which are well-known IP networking tools. Ping is an Internet Control Message Protocol (ICMP) echo that is used to determine if an IP datagram can successfully reach the destination and subsequently return. This is typically the first tool you use to validate end-to-end connectivity.

Trace route is similar to ping, except that the time to live (TTL) on the IP datagram is incremented by one from a starting value of 1 with each iteration. When a router receives a datagram with a TTL of 1, it drops the datagram and returns an ICMP message to the source, indicating the drop. That message has the IP address of the router responding to the drop, thereby informing trace route of the path along which the drops occurred. This is how a trace route is obtained, and it is a very useful tool for troubleshooting.

Brocade FOS RAS Features

Brocade FOS offers a large number of RAS features, such as RASlog. These features are beyond the scope of this document, but you should know they exist.

IP Network Integration

Array-based native IP replication provides retransmits, error recovery, in-order delivery, and aggregated bandwidth (by adding more controller ports). Array ports are significantly more expensive—close to an order of magnitude more expensive—compared to Brocade Extension ports. Committing many array ports to replication makes meeting high data-rate demands and infrastructure expansion considerably more expensive.

IP networks have no requirement to deliver datagrams in-order. How well can array-based native IP replication deal with out-of-order packets? Effectively, most cannot. Array-based FC replication with Brocade Extension can easily handle this. Brocade clearly fills this technology gap.

Interfaces

What interface speeds do you need? The Brocade 7840 has 16 x 1 GbE/10 GbE, 2 x 40 GbE, 24 x 16 Gbps FC ports. You can run multiple 10 Gbps circuits across the two 40 GbE interfaces. Port and optic redundancy is now a reality with 20 Gbps VE_Ports that can span multiple Ethernet interfaces. Depending on the available interfaces that already exist in your data center, the most appropriate speed and number of interfaces are available on the Brocade 7840.

PerPriority-TCP-QoS (PTQ)

Where QoS is enforced does matter. An optimal place to enforce QoS is at the point at which different storage applications converge just prior to being directed into the WAN. Frequently, the IP network does not have QoS configured, at least for the storage applications. Therefore, at a minimum it is important to deliver data to the IP network sequenced according to the Storage Administrator's priorities. Brocade Extension is located at the endpoints of the data transport, the TCP points of origin and termination. These endpoints are the most effective place to QoS-mark data and apply it to various applications. The association of proper QoS markings to specific circuits, either primary or backup, is easily done here. Brocade has PTQ, in which each priority receives its own autonomous WO-TCP session. Cooperating with IP Network Administrators, QoS values for 802.1P/DSCP can be vetted and deployed if QoS is being enforced in the IP network. Using Brocade Extension makes this an easy and automated process.

Brocade PTQ assigns each QoS priority its own autonomous TCP session within each circuit. Additionally, each priority within each circuit can be independently marked with 802.1P (L2 CoS), DSCP (Layer 3 [L3] DiffServ), or both, as needed. This robust QoS schema permits storage connections to traverse different paths, because the connections have different QoS characteristics and requirements. It is not effective to enforce multiple QoS priorities within a single TCP session. Using multiple TCP sessions not under the same supervisor tunnel is also not effective. PTQ is jointly supervised across all WO-TCP sessions for all circuit members of a tunnel. PTQ is monitored by MAPS; refer to the MAPS section of this tech brief for more detail.

Array-based native IP replication cannot offer this level of QoS functionality. For array native IP ports, enforcing flows from disparate sources is not possible. Changing QoS markings during periods of failover is difficult or impossible.

Pause Frames (IEEE 802.3X)

Brocade Extension supports autonegotiation on Gigabit Ethernet (GbE) interfaces. Gigabit Ethernet autonegotiation is not used to negotiate link speed; the speed is always Gigabit Ethernet. Gigabit Ethernet autonegotiation is used to determine link duplex (Full or Half) and Pause Frames (802.3x) support. By default, Brocade enables autonegotiation and supports Pause Frames either on or off, and only full duplex. Autonegotiation is enabled by default on most data center Ethernet switches, and Pause Frames are disabled by default on those same switches. Some Ethernet switches do not support IEEE 802.3x Pause Frames. In practice, it is unlikely that Network Administrators will enable Pause Frames on their Ethernet switches. Pause Frames can lead to Head of Line Blocking (HoLB) on DC LAN switches, causing all flows to sporadically stop on an Ethernet ISL, resulting in poor performance.

On some arrays, the use of native IP port Pause Frames is essential to perform flow control, and it is required for proper operation. Unfortunately, in nearly all cases Pause Frames are disabled by default on the connecting DC LAN switches, and for good reason. There is a much better way to deal with storage flow control across an IP infrastructure using Brocade ARL. Refer to the ARL section in this tech brief for more detail. This is another example of Brocade network integration.

Ethernet Sharing and VLAN Tagging (IEEE 802.1Q)

Brocade Extension supports VLAN tagging (802.1Q). VLAN tagging is frequently used when a single physical connection carries data from different VE_Ports (VE_Ports define trunk endpoints, therefore, these are different trunks); most likely those different VE_Ports live in different Virtual Fabric Logical Switches. These tunnels will share a common Ethernet interface, because the interface bandwidth is large—10 Gbps or 40 Gbps—and can easily accommodate multiple trunks. This is a smarter solution that better utilizes a small number of interfaces, rather than barely utilizing a large number of interfaces. In this case, the Ethernet connection cannot be placed into one particular VLAN on the data center LAN switch port. By using VLAN tagging, each destination VLAN can be sorted upon receipt within the LAN switch. Each circuit from the Brocade 7840 will specify its VLAN for a distinct path through the IP network. Multiple circuits from various tunnels can share the same large Ethernet interface if desired. This is an important integration feature, used by many customers.

Network Address Translation (NAT)

Brocade Extension supports NAT within the IP network. There are specific facilitating functions for proper integration in these environments.

Link Aggregation

Link Aggregation (IEEE 802.1ax LACP, LAG) is not supported on Brocade Extension products. LAG is not needed if you are using Extension Trunking, because the purpose of LAG is accomplished in a more effective way using Extension Trunking. Extension Trunking performs not only the link aggregation, but a number of other important storage specific functions as well (for example, single logical link, LLL, and In-Order Delivery [IOD]). Extension Trunking is integrated into both the FC side and the LAN side, making it superior to LAG for storage applications. LAG solves only part of the problem that Extension Trunking solves, and LAG does it less effectively. LAG is flow-based. Extension Trunking is batch-based, which is more granular. If a link disconnect occurs, LAG is not lossless for data in-flight; Extension Trunking is lossless (LLL). All links in a LAG have to be in the same configuration; circuits in a trunk are not restricted and can be unique.

Summary

This tech brief has covered intrinsic advantages of Brocade Extension relative to array-based IP replication at similar or lower TCO (Total Cost of Ownership). This applies equally to open system or mainframe environments that use array-to-array replication. The choice is supported by citing numerous innovative technological advantages found in Brocade purpose-built hardware and firmware. Brocade Extension leverages nearly 20 years of experience and innovation. Brocade Extension solutions with this technology prove to be of enterprise-class quality and demonstrate excellence in performance, reliability and availability, security, scale, and operational management. The ability to integrate into any IP network is enabled through a plethora of features and validation tools. There is a comprehensive management platform with Brocade Fabric Vision technology that provides unprecedented insight and visibility across the storage network. The platform uses powerful built-in monitoring, management, and diagnostic tools that enable organizations to simplify monitoring, increase availability, and dramatically reduce costs. These unique management capabilities are not offered by any other company. In conclusion, to best satisfy the array-based RDR mission and achieve superior solution value, Brocade Extension is the right choice.

Corporate Headquarters
San Jose, CA USA
T: +1-408-333-8000
info@brocade.com

European Headquarters
Geneva, Switzerland
T: +41-22-799-56-40
emea-info@brocade.com

Asia Pacific Headquarters
Singapore
T: +65-6538-4700
apac-info@brocade.com

BROCADE