BROCADE
A Broadcom Company

Extension Design and Best Practices

Design Guide

Table of Contents

Chapter 1: Overview and Purpose

To address disaster recovery requirements, Broadcom offers an Extension solution portfolio designed to provide organizations with flexible deployment options for data replication. Both Brocade Extension Switches and Blades are robust platforms for large-scale, multi-site data center environments implementing block, file, and tape data protection solutions. Brocade Extension is a purpose-built solution that securely moves more data over distances faster, while minimizing the impact of disruptions. These platforms deliver unprecedented performance, security, and availability to handle the unrelenting growth of data traffic between data centers in Fibre Channel, FICON, and IP storage environments.

Brocade, a Broadcom company, has always been the leader in Extension innovation and was the first to develop technologies such as Extension Trunking, FC Routing (FCR), Advanced Accelerator for FICON, Adaptive Rate Limiting (ARL), FastWrite, Open Systems Tape Pipelining (OSTP), FC compression, FCIP encryption, Per-Priority TCP QoS, IP Extension, and more. These capabilities deliver unmatched value and drive down capital and operational expenses.

This document covers deployment best practices for storage and tape extension.

Chapter 2: The Three Sides of Extension

Brocade Extension has three sides: the Fibre Channel side, the WAN side, and the LAN side. The Fibre Channel side uses a full-fledged Brocade switch for connectivity. Fibre Channel and FICON traverse the WAN side using FCIP. The WAN side connects to the WAN network and includes tunnels, circuits, encryption, compression, QoS, etc. The LAN side connects to the data center LAN and is used for IP Extension. Tunnels and circuits carry Fibre Channel, FICON, and IP traffic.

The Fibre Channel and FICON side of extension is not within the scope of this document. Please refer to the Brocade Fabric OS Administrators Guide specific to your version.

Fibre Channel devices need to be zoned across extension the same as any other fabric. Best practice is to set the `defzone` to `noaccess` on all SAN switches or logical switches that replication ports connect to. A Fibre Channel replication port should only be able to access other Fibre Channel replication ports if explicitly zoned.

The legacy default `defzone` mode has previously been `allaccess`. Due to `allaccess` issues related to RSCN storms and end-device connectivity, for FOS v10.0.0 and later `defzone` has been changed to `noaccess`.

Additionally, `defzone` is set to `noaccess` in the following scenarios:

- Logical switch creation: `defzone` is set to `noaccess` upon creating a logical switch.
- `firmwarecleaninstall`
- `factoryreset`
- `configremoveall` when the remove zone option is set to '`y`'.

```
Remove zone/AD database (no recovery) for fid 128: (yes, y, no, n): [no] y
```

Chapter 3: The WAN Side: Tunnel and Circuits

The majority of extension features and technology are on the WAN side. An extension tunnel forms an inter-switch link (ISL) between two domains. ISLs that live within a data center experience ultra-low latencies and very high bandwidths; ISLs across a WAN do not. Compared to a fiber cable between devices in a data center, extension technology has to be robust enough to mitigate orders of magnitude higher latencies and much lower bandwidths, not to mention the need for added security.

The following sections are technology components paramount to reliably and securely transmitting storage data between distant data centers.

3.1 The WAN Network

The IP WAN network must be able to transport extension datagrams with adequate performance and service levels. By today's standards, 0.1% is considered high packet loss for a network, but Brocade Extension can actually support up to 1% packet loss. This means that you can achieve high replication performance even when your WAN circuits are in a degraded state, having more packet loss than normal – or – you may be able to use a less expensive and less reliable WAN connection and still achieve decent replication performance. Brocade Extension can also run circuits across disparate service provider networks with different I/O characteristics. WAN circuits in a tunnel can have different bandwidth characteristics. However, it is highly recommended that a tunnel's WAN circuits have a bandwidth difference no greater than a 4:1 ratio. For example, a 250Mb/s and 1Gb/s circuit in the same tunnel work well. A 100Mb/s and 1Gb/s circuit in the same tunnel do not work well. The bandwidth differences are not enforced, but highly recommended; otherwise, performance issues may result.

From end to end, the IP network must allow specific protocols to pass. Brocade Extension selects a random ephemeral port (source port) between 49152 and 65535. Brocade Extension uses TCP destination ports 3225 and 3226. When IPsec is enabled, it encrypts up through the TCP header, only the IP header is unencrypted, and an ESP header is added. Brocade Extension always uses these TCP ports, and with IPsec enabled the network cannot see which TCP ports are being used. The same is true for the TCP URG flag; it is always used. With IPsec enabled, it is encrypted and the network cannot see it.

The TCP URG flag is required on Brocade Extension and must not be dropped or modified by any intermediate network device, such as a firewall. If firewall security requires modification of URG flags, IPsec must be used to hide TCP headers from network devices. See Section 3.12, IPsec: Encryption for additional information concerning the network path when using encryption.

3.2 GE Interfaces: RJ-45, SFP, SFP+, and QSFP

On Brocade Extension platforms, Ethernet interfaces are referred to as GE (Gigabit Ethernet) interfaces. Every effort should be made to connect the GE interfaces as close to the WAN as possible to prevent excessive hops, added latency, potential congestion, and performance degradation. The supported propagation delay across an IP network is up to 250 ms RTT, which in most cases will reach just about anywhere on Earth. Brocade products do not support IP network devices that have oversubscribed or blocking switch ports. Some Ethernet switches have host access ports that are oversubscribed or blocking, which degrades performance.

Each Extension platform has a variety of GE speeds. The Brocade 7850 has 1GbE, 10GbE, 25GbE, and 100GbE interfaces; the Brocade SX6 has 1GbE, 10GbE, and 40GbE interfaces; and the Brocade 7810 has 1GbE and 10GbE interfaces. A GE interface must be capable of carrying the bandwidth of its circuits; for example, you cannot put a 25GbE circuit on a 10GbE interface.

On the other hand, you can put two 25GbE circuits on a 100GbE interface, which is commonly done. For example, a Brocade 7850 tunnel can be built with four circuits and two failover groups, where each failover group has one production and one backup circuit. Place the production circuits across both 100GbE interfaces and the backup circuits across the opposite 100GbE interfaces. By doing this, if one of the 100GbE pathways is interrupted, the backup circuit becomes active, and the remaining pathway functions without losing bandwidth.

By default, GE interfaces are set for the WAN side. A limited number can be configured for LAN-side (IP Extension) operation. A GE interface cannot operate in both LAN and WAN modes. In the case of the Brocade SX6 and Brocade 7850, the 40GbE and 100GbE interfaces do not support LAN-side (IP Extension) operation.

Brocade SX6 1GbE/10GbE interfaces and Brocade 7850 1GbE/10GbE/25GbE interfaces are in groups, and the speed should be consistent within each group. Table 1 lists ports that block each other when their speed is different. Best practice is to use the first interface in the group before proceeding to double up interfaces within a group. The Brocade 7810 does not have a GE interface group concern. Refer to the Brocade Fabric OS Extension User Guide for additional information.

Table 1: Brocade SX6 and Brocade 7850 GE Groups

GE Interface Group	Brocade SX6 GE Interfaces in the Same Group	Brocade 7850 GE Interfaces in the Same Group
1	0, 1, 13, 17	0, 4
2	2, 6	1, 5
3	3, 7	2, 6
4	4, 8	3, 7
5	5, 9	8, 12
6	10, 14	9, 13
7	11, 15	10, 14
8	12, 16	11, 15, 16, 17

GE interfaces do not perform speed negotiation. The user must configure the desired speed on a GE interface supporting multiple speeds if it is not the default speed. The optics of the Brocade 7850 are dual speed, 1GbE and 10GbE capable or 10GbE and 25GbE capable. The interface speed is not automatically negotiated and must be configured in Brocade Fabric OS by the user.

Specific to the Brocade 7810 and Brocade SX6, a 1GbE SFP and a 10GbE SFP+ are different optics, and neither can change its speed to the other. If you need 1GbE connections to the data center LAN, you must order 1GbE optics; likewise, if you need 10GbE connections to the data center LAN, you must order 10GbE optics.

The Brocade 7810 has 1GbE and 10GbE capable interfaces. GE0 and GE1 are RJ-45 copper ports (1Gb/s only) and are enabled by default. These copper interfaces are mutually exclusive to the SFP interfaces GE2 and GE3. Either copper is enabled or the SFP bays are enabled, but not both. These interfaces (GE0/GE1 and GE2/GE3) are enabled or disabled as a set, not individually. There is no disadvantage or advantage to using RJ-45 interfaces over SFPs. Use the most appropriate interface type according to the available data center switch interfaces.

MAPS is enhanced to monitor the Ethernet SFPs on the extension platforms. This feature was introduced in FOS v10.0.0. MAPS monitors the following smart data of the FC SFPs:

- Voltage
- Current
- Temperature
- RXP
- TXP

In line with FC SFP monitoring, MAPS has been enhanced to support reading threshold values for GE SFPs at runtime. Instead of using static threshold values, the new rule definitions define that at runtime, the MAPS module will read optimal threshold values for each SFP as prescribed by its manufacturer. The ALL_EXT_GE_PORTS logical group will monitor all SFP stats. GE SFP monitoring applies only to Extension platforms.

In-band management is not supported on Brocade Extension GE interfaces.

3.3 LLDP

Best practice is to use Link Layer Discovery Protocol (LLDP) as an Ethernet connectivity indicator. For troubleshooting, adding, changing, and removing Ethernet connections, LLDP is a handy tool. LLDP information is logged on both ends, therefore, LLDP is used by both SAN administrators and network administrators. Network administrators see what is connected to their switch ports, and storage administrators know which switch ports they have connected to. LLDP is enabled by default and operates on all extension GE interfaces. It does not operate on the Brocade management ports. Most data center LAN switches support LLDP.

There are specific LLDP TLV (type, length, value) that should be enabled and disabled when connecting an extension platform to a data center LAN switch:

- Enable
  - Chassis ID
  - System Name
  - System Capabilities
  - System Description
  - Port ID
  - Port Description
  - Management Address
- Disable
  - DCBX
  - FCoE App
  - FCoE lls
  - Dot1
  - Dot3

3.4 VE_Ports

VE_Ports are tied to a specific data processor (DP). All extension platforms have a DP0; the Brocade 7850 and Brocade SX6 also have a DP1.

The Brocade 7810 has only one VE mode; the number of available VEs depends on if the device is a base or upgraded unit. The Brocade 7850 has two VE modes with 6VE and 18VE, and the Brocade SX6 has two VE modes with 10VE and 20VE. The VE mode with fewer ports supports more bandwidth per tunnel than the mode supporting more ports. If the environment does not require 18 or 20 tunnels, it is best practice to use the mode with fewer VE_Ports. In the mode with fewer VE_Ports, the unavailable VE_Ports are shown as persistently disabled.

Best practice is to start at the first VE_Port on DP0. If another tunnel is required, use the first VE_Port on DP1. In the case of the Brocade 7810, move on to the next VE_Port on DP0.

The following is a list of the Brocade Extension platforms, their DPs, and the DP-associated VE_Ports:

- Brocade 7850
  - 6VE Mode
    - DP0: VE24 to 26
    - DP1: VE33 to 35
- Brocade 7850
  - 18VE Mode
    - DP0: VE24 to 32
    - DP1: VE33 to 41
- Brocade SX6
  - 10VE Mode
    - DP0: VE16 to 20
    - DP1: VE26 to 30
- Brocade SX6
  - 20VE Mode
    - DP0: VE16 to 25
    - DP1: VE26 to 35
- Brocade 7810
  - 2VE (Base)
    - DP0: VE12 to 13
  - 4VE (Full)
    - DP0: VE12 to 15

3.5 Virtual Fabrics

Virtual Fabrics (VF) comprise one or more logical switches (LS) within a physical chassis. There are different types of logical switches (Default, Base, FICON, and Fibre Channel), each with a unique Fabric ID (FID). An FID must be unique within each platform; two logical switches in the same platform cannot have the same FID. Multiple connected logical switches with the same FID form a logical fabric across the physical fabric. A logical fabric plays a vital role in achieving a deterministic path required for protocol optimization, such as Advanced Accelerator for FICON, FastWrite, and OSTP.

NOTE: Advanced Accelerator for FICON has been removed from Fabric OS starting with version 10.x.

Additionally, in mixed FICON and FCP environments, the implementation for each protocol is typically different, and each has unique requirements, configuration settings, and management. On the same platform, these differences are only achievable using logical switches.

Protocol optimization requires that an exchange, sequences, and frames pass through the same outbound and return VE_Ports. A deterministic, optimized extension path requires that the logical switch contain only a single VE_Port. By putting a single VE_Port in a logical switch, one path to the remote switch is defined. Logical switches sufficiently scale Fibre Channel port connectivity, which is operationally simplistic and creates a stable environment.

The Brocade 7850 and Brocade SX6 support Virtual Fabrics; the Brocade 7810 does not.

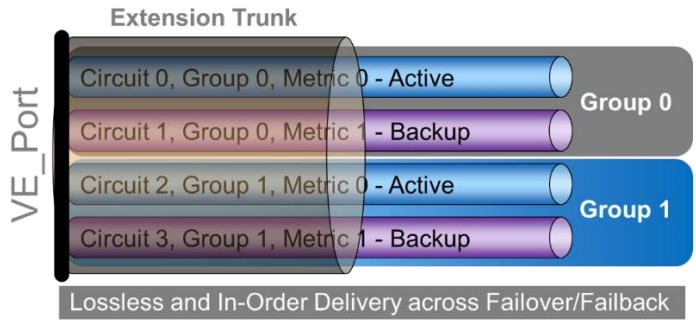
A tunnel cannot be shared across different logical switches; the VE_Port must live within a single logical switch. Ethernet interfaces should stay in the default switch, which can be used by circuits originating from any logical switch. See Section 3.11, GE Interface Sharing.

Best practice is to disable unused VE_Ports and GE interfaces persistently.

3.6 Failover and Failback: Metrics and Groups

All circuits have a metric of 0 or 1, as shown in Figure 1. Metric 0 is the default and is preferred over metric 1. Metric 0 circuits are used until all metric 0 circuits within the failover group have gone offline, after which the metric 1 circuits become active. Metric 0 and metric 1 circuits belong to the same VE_Port (the same extension trunk), which means that during failover and failback, no data in flight is lost or sent out of order to the upper-layer protocol.

Figure 1: Circuit Failover Metrics and Groups



There are two primary cases for using metrics. First, when a backup circuit needs to be passive until a production circuit goes offline. Passive means there is a low quantity of traffic, such as keepalives; nevertheless, there is some traffic. Second, use metric 1 backup circuits when licensing doesn't permit the aggregate maximum needed to retain bandwidth after a production circuit has gone offline. For example, the Brocade 7810 calculates aggregate bandwidth based on currently active circuits, and only metric 0 or metric 1 can be active at any one time. It is common to use a failover group for each circuit, therefore, each circuit has a backup circuit.

3.7 Keepalive Timeout Value

Brocade Extension circuits use keepalives to determine circuit health and to ensure that higher-level FC or TCP/IP extended flows do not experience protocol timeouts caused by WAN outages. Five keepalives are sent during the keepalive timeout value unless it would cause a keepalive to be greater than 1 second, in which keepalives are sent at a maximum of 1-second intervals. Each keepalive arrival resets the timer. If the timer expires, the circuit is dropped.

Keepalives are injected into the same transport as data. Data is never lost in transit, therefore keepalives are never lost in transit. Massive IP network congestion and dropped packets can cause keepalives to be delayed, which could cause a circuit to be dropped. Do not set the keepalive timeout value too short because WO-TCP can quickly recover through brief outages or bouts of congestion. When the timer is too short, circuits can be inadvertently dropped when they should not have been. Conversely, a longer keepalive timeout value will take more time to detect a failed circuit.

FICON and FCP circuits have different default keepalive values. FICON has stricter timing than FCP and must have no more than a 1-second keepalive timeout value. Specify `--ficon` when creating a tunnel with circuits that carry FICON.

If there is only one circuit in a tunnel, the proper keepalive timeout value is probably the application timeout value plus 1 second. A circuit may drop if the keepalive timeout value is set too short when the IP network has oversubscription, congestion, long convergence times, or deep buffers.

If there are multiple circuits in a tunnel, set the keepalive timeout value to a value less than the overall application protocol timeout, so each circuit could fail with a keepalive timeout value before the protocol timer expires. Most supported RDR applications have a 6-second application protocol timeout. If there are two circuits, set a 2.5-second or 3-second keepalive timeout value for each circuit. If there are three circuits, set a 2-second keepalive timeout value for each circuit, and so on.

3.8 Brocade Extension Trunking

Brocade Extension Trunking is an exclusive feature that offers powerful benefits:

- In-order delivery
- Remediation of data lost in flight
- Bandwidth aggregation
- Granular load balancing
- Lossless failover and failback

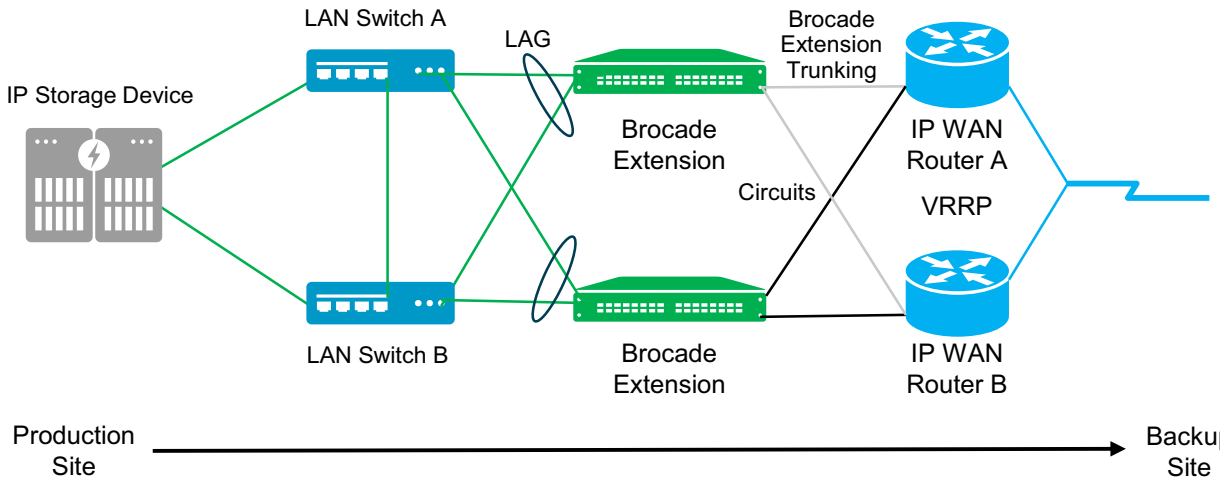
Tunnels and circuits are on the WAN side. A tunnel that contains more than one circuit is a Brocade extension trunk. A VE_Port defines a tunnel endpoint. Brocade Extension Trunking forms a single ISL between two VE_Ports. Brocade Extension Trunking circuits terminate at the VE_Port on each side, therefore, only one VE_Port load balances across its member circuits. A circuit is a connection defined by a source and destination IP address plus other configuration parameters such as minimum and maximum rate limits, eHCL, QoS marking, and keepalive timeout value.

NOTE: Compression and IPsec are at the tunnel level, not the circuit level.

A circuit is assigned to a GE interface by an IPIF assigned to that GE interface. Often, circuits are assigned to a dedicated interface; this depends on factors like interface speed, port redundancy, and, if there are multiple circuits, cumulative max-comm-rates.

GE interfaces physically connect to data center LAN switches for IP Extension or WAN routers for circuits. For IP Extension, connect one interface to LAN switch A and the other to LAN switch B. The LAN switches must be logically one switch, and the connections must form a Link Aggregation (LAG). WAN-side circuits connect to WAN routers A and B. BET circuits load balance with a high degree of granularity. If a single subnet is being used, typically there is a single WAN-side gateway because the routers implement Virtual Router Redundancy Protocol (VRRP) or Hot Standby Router Protocol (HSRP). Brocade Extension Trunking prevents data transmission loss when a network device fails or goes offline, such as from optic instability, cable damage, human error, router failure, or service-provider disruption.

Figure 2: Brocade Extension Trunking on the WAN Side



Circuits may have varying characteristics. For instance, circuits may experience different latency, take separate paths such as data center LAN switches and WAN routers, and belong to different service providers. Circuits that belong to a VE_Port can have bandwidth differences up to 4x the lowest bandwidth circuit. For example, if the VE24-cir0 min-comm-rate is 1Gb/s, the min-comm-rate on VE24-cir1 can be no more than 4Gb/s.

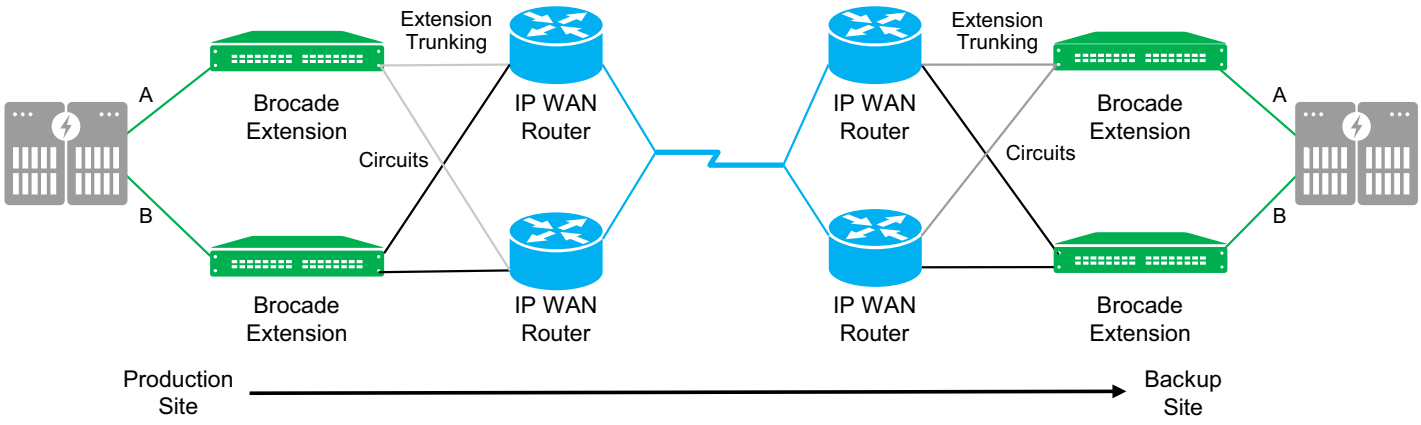
Using Brocade Extension Trunking is considered best practice, rather than using multiple parallel VE_Ports. Multiple VE_Ports between the same two domains use one of the following methods to route traffic across the VE_Ports:

- Exchange-based routing (EBR): default
- Device-based routing (DBR)
- Port-based routing (PBR)

The architecture shown in Figure 3 is a dedicated high-availability replication SAN with no connections to the production SAN. Array replication ports are dedicated to replication; do not connect these ports to a production fabric.

With Brocade Extension Trunking, there is a trunk for the A path (grey top-to-top extension boxes) and a trunk for the B path (black bottom-to-bottom extension boxes). The IP network merely connects these point-to-point circuits to each associated endpoint.

Figure 3: Four-Box High-Availability Replication SAN



All data centers have redundant routers and switches. Best practice is to connect each VE_Port to each router or switch for redundancy and resiliency. Without Brocade Extension Trunking, every parallel connection would require a VE_Port. With Brocade Extension Trunking, one VE_Port with multiple circuits provides performance, resiliency, and redundancy. It is best practice to use one VE_Port with multiple circuits to form a trunk between the local and remote domains.

Typically, tape takes a single SAN path; there are exceptions. Logically, trunking is a single path. With multiple circuits, Brocade Extension Trunking takes advantage of redundant network paths. Tape must transparently failover and failback paths without data loss while maintaining in-order delivery; otherwise, tape jobs fail. Brocade Extension Trunking allows failover and failback without data loss and out-of-order data.

For disk and tape protocol optimization (FastWrite for disk and Open Systems Tape Pipelining for tape), a single tunnel between domains is required. Use multiple circuits for redundancy, resiliency, failover protection, and performance. If multiple tunnels are necessary, Virtual Fabrics logical switches (VF LS) must be configured to ensure that all sequences from every exchange traverse the same VE_Port in both directions. A deterministic path means one VE_Port per LS.

3.9 Compression

Fast-Deflate, Deflate, and Aggressive-Deflate are the compression algorithms implemented on the Brocade 7850 and Brocade SX6. The Brocade 7810 has Deflate and Aggressive-Deflate; it does not have Fast-Deflate. The Brocade 7810 does not support the throughput warranting Fast-Deflate hardware and can achieve higher compression ratios using Deflate and Aggr-Deflate while meeting throughput capacity. Fast-Deflate is unavailable for IP Extension on any extension platform; IP Extension compression is limited to Deflate and Aggressive-Deflate.

Compression operates in the tunnel scope, not per circuit. Compression must be configured identically on both ends of the tunnel; asymmetrical compression is not supported.

Compression can be configured specifically for each protocol (FCIP and IP Extension) running over a specific tunnel. For example, on a Brocade 7850, configure Fast-Deflate for FCIP and Deflate for IP Extension. Using these algorithms for each protocol is considered best practice because the Fast-Deflate and Deflate compression engines are different. Fast-Deflate uses a FPGA engine, and Deflate uses a hardware engine in a DP processor. 20Gb/s of Fibre Channel ingress to the Fast-Deflate engine does not consume any IP Extension Deflate or Aggressive-Deflate compression engine resources.

Compression is recommended with RDR applications, including RDR/S. Tape data is commonly compressed, and compressing data again is not helpful.

3.9.1 Fast-Deflate

Fast-Deflate typically gets about a 2:1 compression ratio. Fast-Deflate is a hardware-implemented field-programmable gate array (FPGA) compression algorithm suitable for synchronous applications. It accommodates maximum circuit line rates and adds a mere 10 µs of propagation delay. Fast-Deflate can only be used for FCIP; it cannot be used for IP Extension. The Brocade 7810 does not support Fast-Deflate.

3.9.2 Deflate

Ratio and rate are tradeoffs in compression, and the Deflate algorithm was implemented to provide medium speed and ratio. Deflate typically gets about a 3:1 compression ratio. Deflate is a hardware-assist processor-based algorithm and may not be suitable for synchronous applications.

3.9.3 Aggressive-Deflate

Aggressive-Deflate furthers the tradeoff between the compression ratio and the compression rate. Aggressive-Deflate typically gets about a 4:1 compression ratio. Aggressive-Deflate is a hardware-assist processor-based algorithm and may not be suitable for synchronous applications.

Use the guidelines in Table 2, Table 3, and Table 4 to assign a compression algorithm for a tunnel.

Table 2: Assign Compression on the Brocade 7810

Total Tunnel (WAN-Side) Bandwidth	Compression Algorithm	Protocol
1.5Gb/s and higher	Deflate	FCIP and IP Extension
1.5Gb/s or less	Aggressive deflate	FCIP and IP Extension

Table 3: Assign Compression on the Brocade 7850

Total Tunnel (WAN-Side) Bandwidth	Compression Algorithm	Protocol
10Gb/s and higher	Fast Deflate	FCIP only
5Gb/s to 10Gb/s	Deflate	FCIP and IP Extension
5Gb/s or less	Aggressive deflate	FCIP and IP Extension

Table 4: Assign Compression on the Brocade SX6

Total Tunnel (WAN-Side) Bandwidth	Compression Algorithm	Protocol
4Gb/s and higher	Fast Deflate	FCIP only
2Gb/s to 4Gb/s	Deflate	FCIP and IP Extension
2Gb/s or less	Aggressive deflate	FCIP and IP Extension

NOTE: Compression ratios are approximate. Broadcom makes no warranties, guarantees, or claims regarding the actual compression ratio achieved with customer-specific data.

3.10 Protocol Optimization

Protocol optimization is used to accelerate data transmission across wide-area connections. Protocols are optimized by leveraging local acknowledgments and generating spoofed responses. Local acknowledgments expedite data into the local extension platform, after which it is aggressively forwarded to the remote side. On the remote side, data is fed to the destination device as if it were coming from the initiator. There are three protocol optimizations for FCIP: FastWrite, OSTP, and Advanced Accelerator for FICON. FastWrite and OSTP are covered in the sections that follow. Advanced Accelerator for FICON applies to FICON Tape, FICON z/OS Global Mirror (Extended Remote Copy or XRC), and FICON Teradata flows.

NOTE: Advanced Accelerator for FICON has been removed from Fabric OS starting with version 10.x.

3.10.1 FastWrite

FastWrite (FCIP-FW) is a SCSI write protocol optimization technique. FCIP-FW is managed at the VE_Port level. Therefore, it is paramount that all traffic traversing in both directions always pass through the same set of VE_Ports. The VE_Ports maintain state for the optimization algorithm and a non-deterministic path will break the I/O. Best practice is to either have a single tunnel between sites or place each tunnel into its own VF LS if multiple tunnels are necessary. A single tunnel can have multiple circuits; see Section 3.8, Brocade Extension Trunking.

FCIP-FW does not apply to replication over IP Extension.

The following array replication applications do not benefit from FCIP-FW:

- Dell EMC SRDF/S (uses SiRT)
- SRDF/A and SRDF/Adaptive Copy do benefit from FCIP-FW
- Hitachi Vantara HUR
- IBM products

NOTE: FCIP-FW must be enabled before OSTP is enabled.

3.10.2 Open Systems Tape Pipelining

Open Systems Tape Pipelining (FCIP-OSTP) is a SCSI/FCP optimization protocol for reading and writing to tape. FCIP-OSTP takes advantage of the sequential nature of tape blocks. FCIP-OSTP requires FastWrite to be enabled before it can be enabled. FCIP-OSTP requires the same rules as FCIP-FW concerning traversing a deterministic single set of VE_Ports and maintaining I/O state.

FCIP-OSTP does not apply to tape over IP Extension or FICON Tape.

3.11 GE Interface Sharing

GE interface sharing technology allows a physical GE interface to be used by circuits coming from tunnels in various logical switches. Allowing circuits from different logical switches to share physical Ethernet interfaces optimizes the usefulness of the interface, particularly low port-count, high-bandwidth interfaces like 100GbE.

Brocade Extension Trunking uses a single VE_Port with multiple circuits. Frequently, circuits are assigned to dedicated GE interfaces for port, optics, and cable redundancy. The VE_Port resides in the LS. The GE interfaces remain in the default switch.

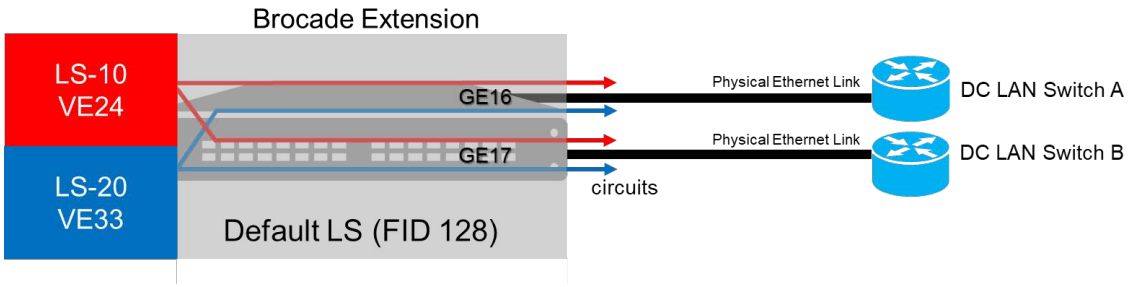
An IP interface (IPIF) designates the DP, IP address, subnet mask, GE interface, VLAN ID (default is no tagging), and MTU (default is 1500 bytes). IPIFs are created in the same LS as the VE_Port. Creating an IPIF assigns an IP address to a specific GE interface and DP. To use the IP address for a circuit, the circuit must belong to any one of the VE_Ports on the DP specified when creating the IPIF.

Each IPIF can only be used by one circuit. It is possible to configure multiple IPIFs on the same GE interface. The same DP, VLAN ID, or MTU need not be used for each IPIF; each IPIF can be different. The IPIFs on an interface do not need to be assigned circuits from any particular VE_Port. A circuit from a VE_Port in any logical switch can use an IPIF on an interface as long as the IPIF was created specifying that the VE_Port's DP and the interfaces are in the default switch. By assigning multiple IPIFs to the same interface, it can accommodate multiple circuits from the same or different VE_Ports.

An IPIF, IP address, and circuit can be associated with only one GE interface. When more than one interface is needed, multiple circuits must be created. The same IP address cannot be configured more than once in the same box. Sharing an interface with multiple circuits from different logical switches is possible. The interface must be resident in the default switch (context FID128). The IPIF and IP routes are configured on the default switch. The VE_Port is in the logical switch to be extended; the tunnel and circuits are configured in that context. GE interface sharing facilitates the efficient use of the 40GbE and 100GbE interfaces.

Figure 4 shows an example of two logical switches (10 and 20) in addition to the default switch. GE16 and GE17 are 100GbE interfaces, and in the default switch (FID 128). There are two 25Gb/s circuits per Brocade Extension Trunking. There is a red (LS-10) and a blue (LS-20) extension trunk. VE24 is on DP0 and has one circuit that goes to GE16 and one to GE17. VE33 is on DP1 and has one circuit that goes to GE16 and one to GE17. There are two 25Gb/s circuits in each 100GbE interface. The 100GbE interfaces are physically connected to respective A and B data center switches or routers.

Figure 4: Virtual Fabrics Logical Switches Sharing Ethernet Interfaces



3.12 IPsec: Encryption

Would your company operate Wi-Fi with no encryption? Of course not! Best practice is to use Brocade IPsec to protect data end-to-end. When you implement Brocade Extension, it is always prudent to enable IPsec. Data leaving its secure confines into a service provider's infrastructure is vulnerable to opportunistic attack, and for data in flight, there is no service provider security guarantee. Links must be authenticated and data must be encrypted to prevent eavesdropping and attacks.

In FOS v10.0.0, quantum-resistant algorithms per the CNSA 2.0 classification are supported for all encryption purposes. In parallel, for backward compatibility, encryption algorithms that are not considered quantum-resistant are still supported as customers transition to utilize only post-quantum cryptography (PQC) algorithms. In FOS v10.0.0, IKEv2 used for IPsec is enhanced to support ML-KEM-768 and ML-DSA-65 when the PQC profile is selected during configuration.

A new IPsec profile type, a PQC profile, supports ML-KEM-768 and ML-DSA-65. The existing legacy shared-key profile (PSK) and public-key infrastructure profile (PKI) continue to be supported:

- PSK security profile with no changes
- PKI security profile with no changes
- New PQC profile supporting ML-KEM-768 and ML-DSA-65

The command `SecCertMgmt` functionality is enhanced to perform the following actions:

- Creation of CSR of the type mldsa65
- Creation of a self-signed certificate of the type mldsa65

A new IPSec profile named PQC with corresponding changes to CLI and REST allows configuration of an IPSec policy to use this new profile. Enhancements include passing the entire certificate across the network to verify PQC-based certificates:

- IKEv2 is enhanced to exchange the full certificate with the peer when using the PQC profile.
- PKI-based remote-switch certificates must still be imported through secCertMgmt (no changes).
- PQC self-signed remote certificates will still need to be imported with secCertMgmt (no changes).
- PQC CA-signed remote certificates do not need to be imported with secCertMgmt; however, the signing CA certificate for that remote certificate must still be imported. Standard practice would be to use the same signing CA certificate for the local and remote switch certificates.

Brocade IPsec is a hardware implementation that can operate at line rate. IPsec is included in all base extension platforms; no additional licenses or costs. Encryption adds a propagation delay of approximately 5 µs. Preshared key configuration is easy.

Using Brocade IPsec removes the need for a firewall. Best practice is to connect extension directly to the WAN routers or switches and to avoid intermediate devices such as firewalls. IPsec IKEv2 uses UDP port 500 as its destination.

Brocade IPsec complies with Suite B, CNSAv1, and CNSAv2 and implements the latest encryption technologies and ciphers. CNSAv1 implements AES 256, SHA-512 HMAC, IKEv2, and Diffie-Hellman. Rekeying occurs in the background approximately every 2 billion frames or every 4 hours, and the process is non-disruptive.

Firewalls are not considered best practice for the following reasons:

- Per DP, Brocade IPsec operates at line rate on the WAN side, which on the Brocade 7850 is 50Gb/s. Most firewalls cannot meet this throughput requirement.
- Storage traffic requires minimal propagation delay.
- Brocade IPsec encrypts data closer to the data source and destination, which is considered best practice.
- Firewalls and WAN optimization devices may proxy TCP sessions, resulting in remote IP packets not being identical to the originals, which is not supported. These devices are not supported.

3.13 Bandwidth

For critical remote data replication (RDR), the best practice is to use a separate and dedicated IP connection between the production data center and the backup site. Even so, often a dedicated IP connection between data centers is not practical. In this case, bandwidth must be, at a minimum, logically dedicated to extension. There are a few ways this can be done:

1. Use QoS and give extension a higher priority, which logically dedicates enough bandwidth to extension over other competing traffic.
2. Use committed access rate (CAR) to identify and rate-limit certain traffic types. Use CAR on non-extension traffic to apportion and limit that traffic to a maximum bandwidth, leaving the remainder of the bandwidth to extension. Set the aggregate circuit min-comm-rates to use the remaining dedicated bandwidth, logically dedicating bandwidth to extension.
3. For all traffic to coexist without congestion, massively overprovisioning bandwidth is the easiest, most costly, and most common practice in extension deployments.

Brocade Extension uses an aggressive TCP stack called WAN Optimized TCP (WO-TCP), which dominates other TCP flows on the path causing them to dramatically back off. UDP-based flows may result in considerable congestion and excessive packet drops for all traffic.

The best practice is to rate-limit and flow-control traffic on the extension platforms at the source and destination and not rate-limit in the IP network. Rate limiting extension in the IP network leads to performance problems and complex troubleshooting issues. Brocade Extension rate limiting is advanced, accurate, and consistent; there is no need to double rate limit.

To determine the network bandwidth needed, record the number of bytes written over a month (or more). A granular recording capable of calculating rolling averages of varying lengths is helpful. It is essential to understand the number of bytes written to volumes during various interims of the day, night, weekends, end of quarter, end of fiscal year, holidays, etc. These data rates must be available for replication to maintain an adequate recovery point objective (RPO).

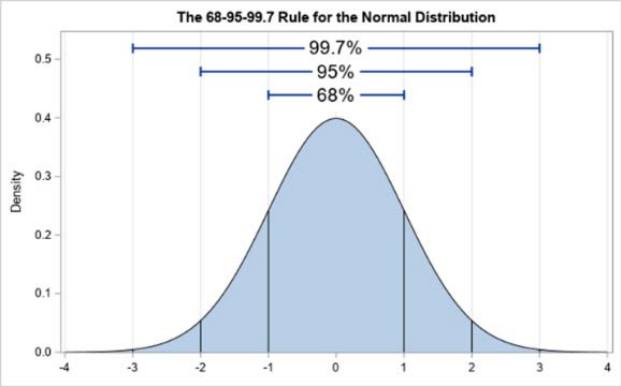
If replicating asynchronous RDR (RDR/A) across a tunnel, calculate the average value over a finite number of minutes throughout the day and night to determine the maximum average. Remember that business transactions may increase, and replication may require significantly more bandwidth during the quarter end, the fiscal end, and certain holidays. RDR/A needs enough bandwidth to accommodate the high averages discovered over a finite period. RDR/A performs traffic shaping, which moves peaks into troughs. Averaging over too long a period may cause a backlog of writes when the troughs do not occur frequently enough to relieve the peaks. Excessive journaling is challenging to recover from, depending on available bandwidth.

The best practice for synchronous replication (RDR/S) is to dedicate the IP network without sharing bandwidth. A 10Gb/s DWDM λ is an example of dedicated bandwidth. Dedicated bandwidth virtually eliminates packet drops and reduces the need for TCP retransmits. Record the peak traffic rates if you plan to replicate RDR/S across an extension tunnel. RDR/S must have enough capacity to send writes immediately, accommodating the entire demand at any time.

Plot the recorded values into a histogram. Suppose you have calculated 5-minute rolling averages from your recorded data. Bytes written over each period will be an integer value. The x-axis is the number of bytes written in each of the 5-minute averages. The left x-axis starts at zero bytes. The right x-axis is the largest number of bytes written during an interim. Averages with the same number of bytes will occur multiple times. The y-axis is the number of times that a particular average occurred. The smallest size occurred relatively rarely. The largest size occurred relatively rarely. As shown in Figure 5, 68% (68% = 1 standard deviation = 1σ) of the averages fall into the middle section, which is the largest group, and the group you are most concerned with. The resulting curve is a bell curve.

Based on cost, plan your WAN bandwidth to accommodate at least the first standard deviation (68%), and if possible, include the second standard deviation (95%). In most cases beyond the second deviation, occurrences are so rare that they can be disregarded in bandwidth planning.

Figure 5: Gaussian Standard Normal Distribution



You can plan for a certain amount of compression, such as 2:1 if your data is compressible, however, data compressibility is a moving target. The best practice is to use compression as a safety margin to address unusual and unforeseen situations. Also, achieving some compression can provide headroom for potential future growth. If there is no margin, periodic demand increases may not be advantageous.

For remote tape backups, extension is used to extend a fabric to a remote VTL. Measuring tape volume in MB/h and the number of simultaneous drives in use is crucial. Bandwidth will limit the backup window even if an adequate number of drives are available.

3.14 Adaptive Rate Limiting

Adaptive Rate Limiting (ARL) is integral to most extension designs. When there is more than one circuit feeding the same WAN or when the WAN is shared with other traffic, ARL is an essential component. ARL manages data sent into the IP network based on minimum and maximum rate settings and the available WAN bandwidth in that range. There may be single or multiple WAN connections; the cumulative bandwidth of the circuits assigned to a particular link matters. Multiple WAN connections are evaluated independently.

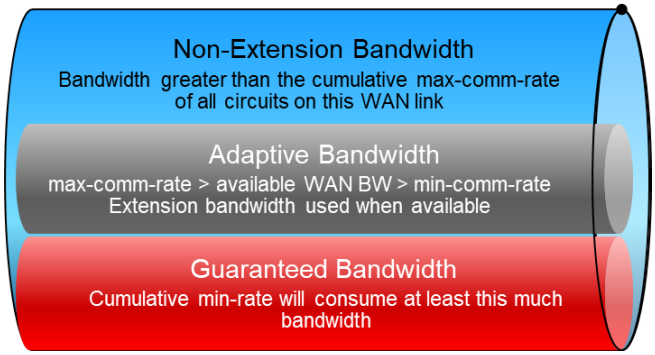
See Figure 3 and assume a 1Gb/s WAN. The ARL max-comm-rate is set to either the GE interface line rate or the maximum available WAN bandwidth, whichever is lowest. In this case, the max-comm-rate is set to 1Gb/s.

Each circuit is configured with a floor (min-comm-rate) and ceiling (max-comm-rate) bandwidth value in bits per second (b/s). The minimum circuit bandwidth will never be less than the min-comm-rate and will never be more than the max-comm-rate. The available bandwidth to the circuit will adjust automatically between the minimum and maximum based on IP network conditions. A congestion event causes the rate limit to readjust toward the minimum. An absence of congestion events causes it to rise toward the maximum. If the current rate is not at the maximum, ARL will periodically attempt to adjust upward; if another congestion event is detected, the rate will remain stable.

The ARL min-comm-rate is set to the max-comm-rate divided by the number of circuits feeding the WAN connection. In this example, 1Gb/s ÷ 4 = 250Mb/s. The min-comm-rate is set to 250Mb/s. When all the circuits are up, each will run at 250Mb/s. In an extreme case where three circuits have gone offline, the remaining circuit will run at 1Gb/s. At 1Gb/s, all the WAN bandwidth continues to be consumed and the replication application remains satisfied.

When more than one circuit feeds a WAN link, the two circuits equalize and utilize the available bandwidth, as shown in Figure 6. When an interface or the entire platform goes offline, ARL will readjust to utilize bandwidth that is no longer used by the offline circuits. Readjusting bandwidth maintains WAN utilization during periods of maintenance or failures.

Figure 6: ARL Min-Comm-Rates and Max-Comm-Rates



In a shared WAN, consider bandwidth as separated into three distinct areas:

- Red guaranteed bandwidth segment of Figure 6 – Reserved extension bandwidth (0 to min-comm-rate)
- Gray adaptive bandwidth segment of Figure 6 – Extension's adaptive bandwidth (min-comm-rate to max-comm-rate)
- Blue non-extension segment of Figure 6– Non-extension bandwidth (max-comm-rate to WAN bandwidth)

ARL manages the bandwidth in the red and gray sections. Red is the minimum bandwidth used by extension and is reserved exclusively for extension. Note that the minimum bandwidth will be the aggregate of all circuit minimums assigned to the WAN link. Blue is reserved exclusively for other traffic that is sharing the WAN link. The bandwidth where the blue section starts is the aggregate of all maximum circuit values on the WAN link. Gray is the area between the top of the red section and the bottom of the blue section. Extension circuits may use this bandwidth when available only when other applications that share the WAN are not currently using it. There are many ways in which ARL can be leveraged.

3.15 Quality of Service

Brocade Extension has a comprehensive Quality of Service (QoS) suite:

- Protocol Distribution (enforcement)
- Priority - High/Medium/Low (enforcement)
- DSCP (marking)
- 802.1P (marking)
- PTQ (segregation)

NOTE: Best practice is not to alter F-class (control-traffic) QoS markings unless it is required to differentiate and expedite the F-class traffic across the IP network. Failure of F-class traffic to arrive promptly will cause replication fabric instability.

3.16 Per-Priority-TCP-QoS

Per-Priority-TCP-QoS (PTQ) is an exclusive, patented Brocade technology that is crucial when prioritizing data across extension. PTQ is not configurable. Any FC zone with a QoS prefix will use the corresponding priority virtual circuits within the fabric and automatically map traffic to the corresponding priority TCP session in extension circuits. Extension traffic that does not come from within a fabric can also be prioritized.

One of the hallmarks of TCP is ensuring in-order delivery. QoS, by its nature, marks flows with varying expediency, thus designating to the IP network its handling. TCP and QoS would work against each other if QoS expedited a flow and TCP returned it to the original sending order. Therefore, PTQ applies autonomous TCP sessions for each QoS priority. Expediting a particular priority using PTQ is not an issue because all traffic for that priority uses an independent TCP session.

3.17 FCIP and IP Extension Distribution Percentages

Protocol distribution is the highest tier of QoS enforcement in Brocade Extension. There are two protocols, FCIP and IP Extension. FCIP and IP Extension protocol distribution percentages are rate limiting to bandwidth. If bandwidth is utilized in a distribution path, the total bandwidth percentage is reserved and unavailable to the other distribution.

The default distribution is 50/50, meaning that each protocol gets 50% of the tunnel's cumulative circuit bandwidth. If only one protocol is being used, the protocol being used receives 100% of the distribution. For example, FCIP would get 100% of the bandwidth if there was no IP Extension traffic at that instance. Because distribution is not limited when only one of the two protocols is implemented, distribution does not need to be adjusted.

3.18 QoS Priority Bandwidth Percentages

Brocade Extension supports three levels of priority: high, medium, and low. The default amount of bandwidth that the scheduler apportions is 50%/30%/20%. FCIP and IP Extension circuit QoS percentages are rate limiting to bandwidth. If there is bandwidth utilization in a QoS path, the total bandwidth percentage for that path is reserved and not available to another QoS path. It is possible to change the default portions to any values you wish, as long as they are 10% or greater and add up to 100%.

There are seven QoS sessions per circuit. The default percentages are shown:

- Class-F (strict)
- FCIP (50%)
  - High (50%)
  - Medium (30%)
  - Low (20%)
- IP Extension (50%)
  - High (50%)
  - Medium (30%)
  - Low (20%)

Class-F uses strict queuing, meaning all class-F traffic is sent before sending any other traffic. Relative to storage traffic, there is a negligible amount of class-F traffic.

Each QoS session is autonomous and has no reliance on other QoS sessions. Each QoS session can be configured with its own DSCP, VLAN tagging, and 802.1P values, permitting QoS sessions to be treated independently along the IP path from site to site based on the SLA for that QoS priority.

QoS is implemented in Brocade Fibre Channel/FICON fabrics and across Fibre Channel ISLs via virtual channels (VCs). There are different VCs for H/M/L and class-F. Each VC has buffer-to-buffer credits and flow control. There are five VCs for high, four VCs for medium, and two VCs for low. Devices are assigned to QoS VCs by enabling QoS on the fabric and putting QOSH_ or QOSL_ as the prefix to a zone name. The default is QOSM_; there is no need to designate medium zones, although it is supported explicitly. Devices use VCs throughout the fabric, including extension. If data ingresses via a QOSH_ prefix zone, that data will use high VCs within the fabric and automatically be assigned to high TCP sessions in extension circuits. There is no requirement to go through the fabric. Devices connected to an extension platform are assigned to a priority TCP session based on the zone name prefix. The default TCP session is medium.

3.19 QoS Marking: DSCP

Differentiated Services Code Point (DSCP) is an IP-based (L3) QoS marking protocol, therefore, it is an end-to-end QoS marking protocol. DSCP has 64 values; however, the values 0 through 63 do not denote lowest to highest priority; the marking schema is different. First, all odd values are for private use (similar to RFC 1918 IP addresses) and can be used in any way an enterprise sees fit. For example, 192.168.0.1 and DSCP 3 are private values and can be used in any way.

For non-private DSCP values, DSCP value 46 is referred to as expedited forwarding and is the highest priority. Zero is the default, and it is the lowest priority. There are four groups of high, medium, and low values referred to as assured forwarding. Another group of numbers has backward compatibility with the legacy Type of Service schema.

DSCP selection is the IP Network Administrators' responsibility. Without their support and configuration of per-hop behavior per QoS marking, no QoS can happen. Ethernet switches' default behavior is to replace ingress QoS values with the default value (0) unless the data coming in on that interface is explicitly deemed QoS trusted, which prevents end users from setting QoS values unannounced to their network administrators.

DSCP marking is configured per circuit + per protocol + per priority (H/M/L).

If IPsec is enabled, it is recommended that all QoS priorities are marked with the same DSCP value.

3.20 QoS Marking - L2 Class of Service: 802.1P

802.1P is a data-link-based (L2) QoS marking protocol; the scope extends from the interface of the extension platform to the interface of the directly attached switch. Network devices enforcing 802.1P provide QoS across their ISLs. The 802.1P header resides within the 802.1Q VLAN header, therefore, VLAN tagging is required to get 802.1P QoS marking. Brocade Fabric OS refers to 802.1P as L2CoS (Layer 2 Class of Service). There are only eight 802.1P values, from 0 to 7. Zero is the default and lowest priority. Seven is the highest priority. As with DSCP, the IP network must be configured to implement QoS before the markings are honored.

L2CoS is configured per circuit + per protocol + per priority (H/M/L).

If IPsec is enabled, it is recommended that all QoS priorities are marked with the same L2CoS value.

Chapter 4: FCIP Architectures

Data preservation permits an organization to recover, and extension is commonly used for business continuity via disaster recovery (DR). Preserve data by leveraging RDR and remote tape applications to transport critical data beyond a potentially catastrophic event.

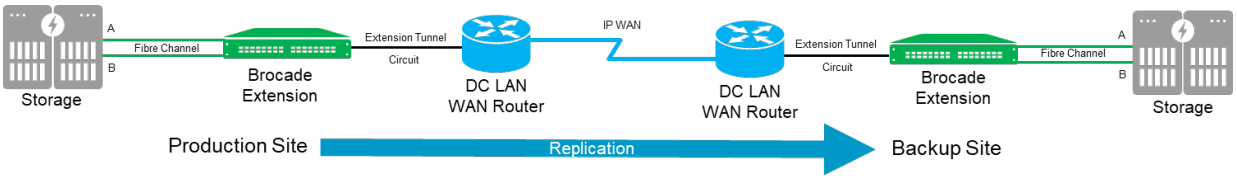
RDR is typically array-to-array communications. The local storage array at the production site sends data to the array at the backup site. RDR can be done via native Fibre Channel if the backup site is within the same metropolitan area and there is a fiber service between the sites. However, cost-sensitive, ubiquitous IP infrastructure is often available, not native Fibre Channel.

Brocade Extension is high speed and adds only about 75 µs of propagation delay per pass-through of an extension platform (four passes round trip = 0.3 ms). It is appropriate for both asynchronous RDR and synchronous RDR applications. A best-practice deployment connects array N_Ports directly to extension F_Ports and does not connect through a production fabric. Storage replication ports are dedicated to RDR and should have no host traffic. Nevertheless, there remain valid reasons to connect via the production fabric, such as tape applications and when there are more replication ports than the extension platform can accommodate.

4.1 Two-Box Solution

A single extension platform in each data center is referred to as a two-box solution, as shown in Figure 7. A single extension platform can be directly connected to both A and B storage controllers.

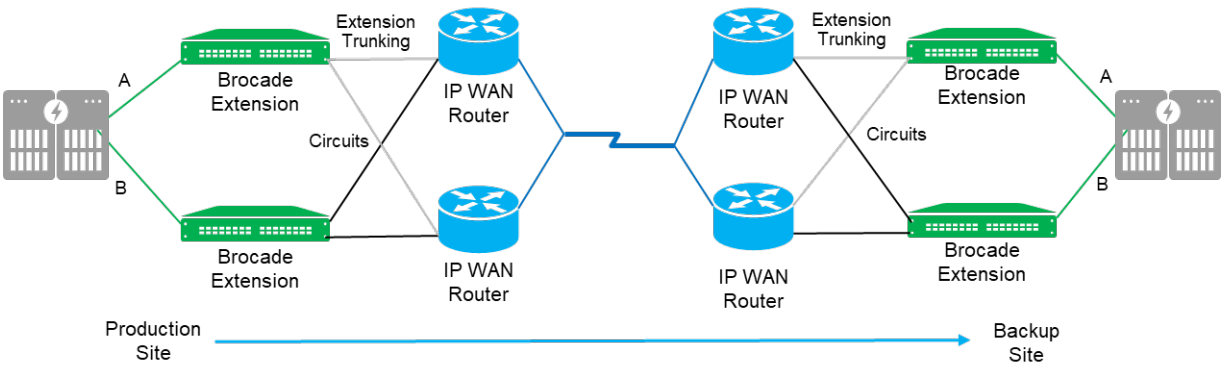
Figure 7: Basic Non-redundant Extension Architecture



4.2 Four-Box Solution

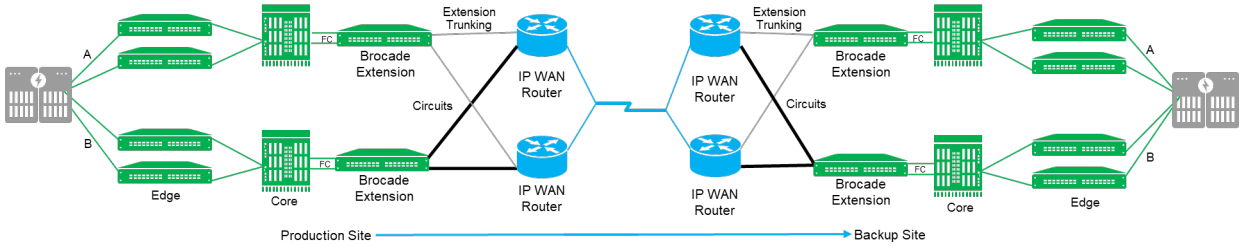
When the extension platform is dedicated to the A fabric or controller and a physically different extension platform is dedicated to the B fabric or controller, this is referred to as a four-box solution as shown in Figure 8. A single WAN link for both paths may be used, or different service providers may be used depending on the tolerance to an outage and cost.

Figure 8: FCIP Architecture with Dedicated Extension for Each Controller



A production fabric can be extended, as shown in Figure 9, although this is only recommended if there is a compelling reason. A common reason is distributed tape, which connects many devices and pipelines the traffic to a DR site. In a tape scenario, OSTP may be used in a logical switch that isolates the tape connections and VE_Port.

Figure 9: Extension of a Routed SAN

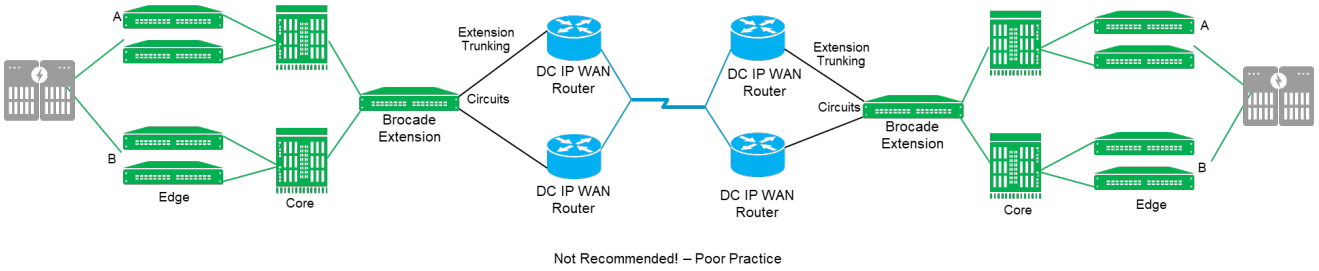


In environments that require extending a production SAN, do not interconnect the same extension platform to both the A and B fabrics. Best practice is to have two separate and redundant fabrics in a production environment, especially if the organization could suffer financial losses during a SAN outage. Even momentary SAN outages can cause servers to stop responding, forcing a reboot and consistency check, which in most situations takes significant time.

For maximum availability, it is best practice to build a redundant SAN with A and B fabrics, which implies that there is an air gap between the two autonomous fabrics from the server to the storage. There are no physical links between the two fabrics. Servers, storage, and VMs are equipped with drivers that monitor pathways and send data accordingly. When a path is detected as down, the driver fails over the traffic to a remaining path.

When using extension and FCR to connect production edge fabrics, do not connect both the A and B fabrics, as shown in Figure 10. Without FCR, the fabrics would merge into one big fabric, which destroys any notion of redundant autonomous fabrics. If FCR is used, the fabrics do not merge, however, a common Linux device is still attached to both fabrics. If maximum availability is the goal, this architecture is unacceptable and considered poor practice due to its high risk. A SAN with a common device to A and B fabrics is susceptible to human error, which can bring down the entire SAN.

Figure 10: Two-Box Solution Connected to Both Production Fabrics—Poor Practice



When connecting extension to production fabrics, design each using best-practice, traditional, core-edge concepts. Since storage connects directly to the core in a core-edge design, extension switches connect to the core, or an extension blade is placed in a core director. For redundancy, connect standalone extension platforms to a fabric with at least two inter-switch links (ISL).

In a four-box solution, it is inappropriate to make ISL cross-connections between the two extension platforms and the A and B fabrics because of the same reasons discussed above, a common Linux instance, and human error.

Cross-connecting circuits from a tunnel to various Ethernet data center LAN switches or IP network devices is encouraged. Circuits that traverse the IP network are point-to-point and can take alternate resilient and redundant paths without merging the A and B fabrics.

4.3 Extension with FCR

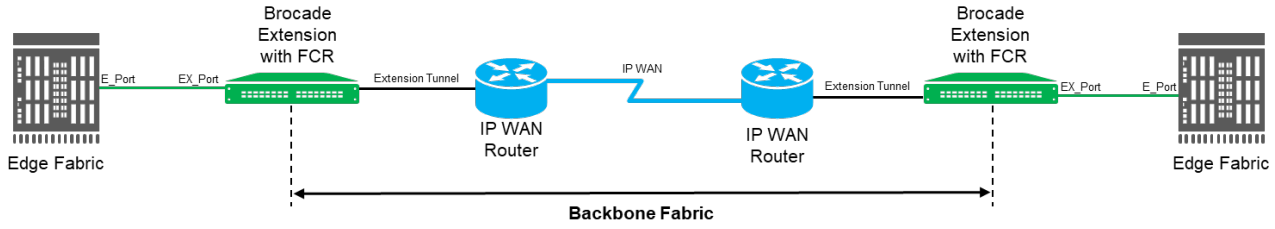
An extension tunnel traverses a WAN, and a WAN has very different characteristics compared to a Fibre Channel ISL. A FCIP link is considered a Fibre Channel ISL. An extension tunnel that crosses a WAN is essentially a Fibre Channel ISL over a WAN. A WAN link that experiences flapping can disrupt a fabric. Use FCR if a production SAN needs to be isolated from potential WAN instability. Disruption comes from fabric services that attempt to converge with each repeated flap. Convergence requires CPU processing, and excessive convergence may lead to high utilization. If the CPU can no longer process tasks promptly, instability ensues. Limiting fabric services to a local edge fabric and not permitting services to span the WAN prevents large-scale and taxing convergence. Best practice is to construct completely separate production and replication SANs. FCR is not needed when replication ports are connected to a dedicated replication SAN.

Isolated fabrics connected to FCR EX_Ports are called edge fabrics. EX_Ports are the demarcation points used to contain fabric services. Fabric services do not pass beyond an EX_Port, which forms an edge. FCR constrains fabric services to within an edge fabric or a backbone.

The following are basic architectures with and without FCR:

- The simplest architecture is no FCR. An independent replication SAN is the most common implementation for distributed systems and mainframes, and is the easiest to manage. Directly connecting storage replication ports to extension is highly reliable and easy to implement, manage, and troubleshoot. See Figure 8.
- An edge-backbone-edge architecture using FCR in which edge fabrics bookend a transit backbone fabric; see Figure 11. The backbone fabric has EX_Ports that are outward-facing to the edge fabrics. Extension is located within the backbone segment using VE_Ports.

Figure 11: Extension with FCR (Edge-Backbone-Edge)



NOTE:

- Mainframe FICON environments do not support FCR.
- When a mainframe host writes FICON to a volume on a direct-access storage device (DASD) and the DASD performs RDR to a remote DASD, the array-to-array replication does not use FICON; the array-to-array replication is FCP based.

Chapter 5: The LAN Side: IP Extension

IP Extension technology accelerates, secures, and manages bandwidth for supported IP storage applications. FCIP and IP Extension share the same tunnel; the transport technology is the same. Latency and packet loss are mortal enemies to TCP/IP replication and cause poor performance. IP Extension overcomes performance degradation caused by characteristics inherent to most WAN. Plus, IP Extension provides encryption.

Brocade Extension can be represented by having three sides, and one of those sides is the LAN side. The LAN side is specific to IP Extension and used to connect IP storage via the data center LAN. IP Extension supports the connectivity of multiple data center LANs, which can be done with VLAN tagging (802.1Q trunk) or without (using different independent links). The Ethernet trunks between the extension platform and the LAN switches identify each VLAN using tagging. An IP Extension gateway (`ipif lan.dp#`) must be configured for each VLAN.

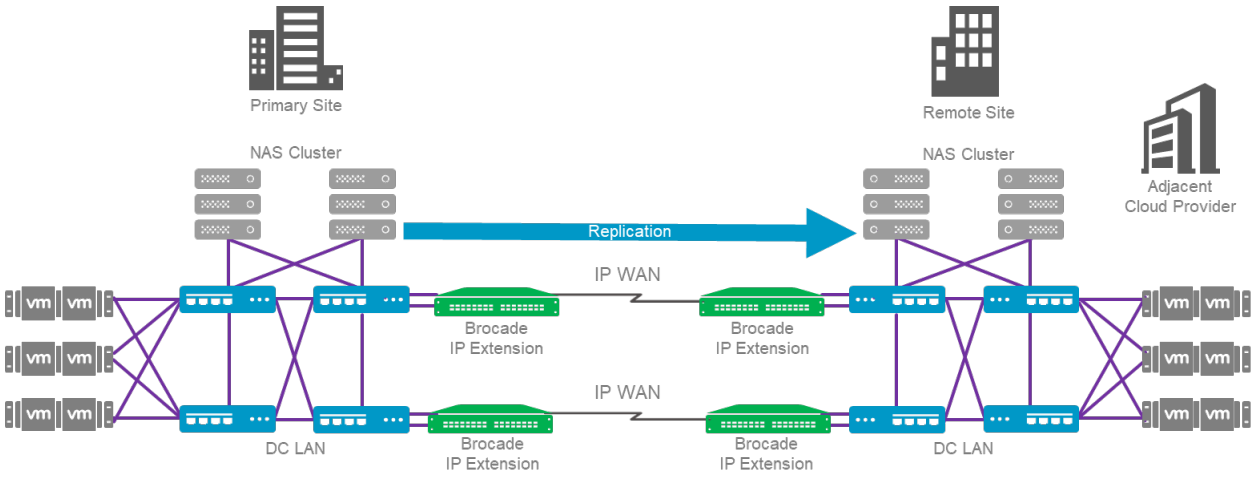
This section covers points unique to IP Extension design, best practices, and architectures.

5.1 Use Cases

IP Extension use cases include replication acceleration, data encryption, bandwidth management, data migration, network visibility, troubleshooting tools, high availability, congestion avoidance, and tape grids.

Figure 12 shows a high-availability architecture used in a VM-NAS environment that replicates traffic between a primary site and a cloud site. The cloud site provides compute elasticity and DR. The data must be replicated quickly to maintain a coherent RPO.

Figure 12: IP Storage Replication over IP Extension



5.2 VE_Ports (The LAN Side)

VE_Ports for IP Extension are merely representative of the tunnel ID. Unlike FCIP, no IP Extension data flows through a VE_Port, although disabling a VE_Port while doing only IP Extension will indeed disable the tunnel.

Best practice when running IP extension and FCIP in the same switch is to use the same VE_Port for FCIP and IP Extension. The desire is to have the VE_Port manage the bandwidth of both protocols across the tunnel. Managing bandwidth cannot be accomplished when different VE_Ports are used for each protocol.

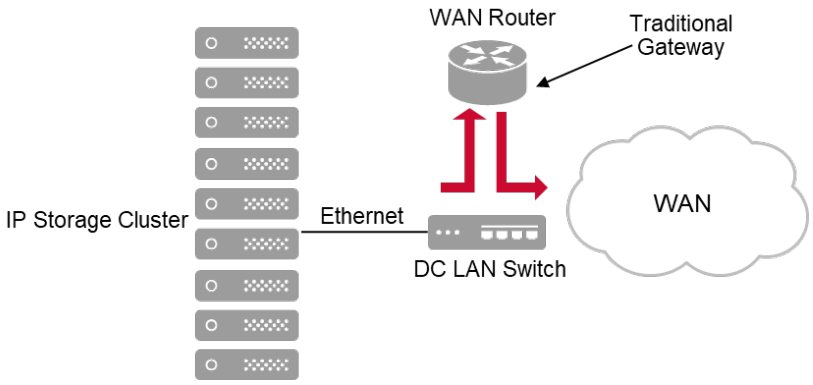
5.3 IP Extension Gateway

IP Extension acts as the gateway for traffic intending to cross the extension tunnel. If IP storage traffic is not forwarded to the IP Extension gateway, it will not utilize IP Extension.

In Figure 13, traffic from the IP storage cluster comes into the data center LAN switch. The data center LAN switch forwards traffic to the traditional router gateway, which could be an inherent part of the data center LAN switch. The router sends the traffic toward the destination.

NOTE: The LAN side subnet used for the IP Extension Gateway must be different on the local and remote ends.

Figure 13: Traditional Data Center Gateway

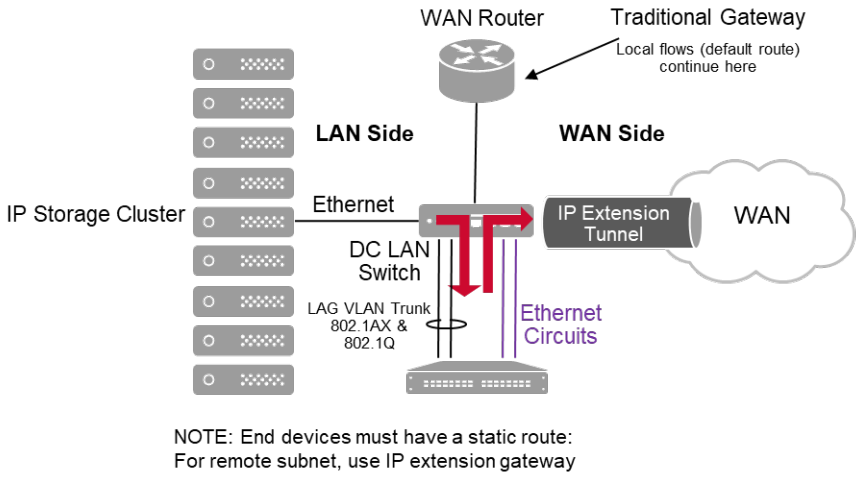


With IP Extension, it is required that the end device has a static route or some route that is more specific than the default route. The more specific static route forwards traffic destined for the remote subnet to the IP Extension gateway. The remote end device is on the remote subnet. The default route stays pointed to the traditional router gateway and is used only when a more specific route does not exist.

Remember that putting IP Extension in path and removing it from path involves activating or deactivating the end devices' static routes. Traffic goes to the IP Extension gateway with static routes; IP Extension is in path. The traffic goes to the traditional router without static routes; IP Extension is out of path.

Figure 14 shows that replication traffic to the remote IP storage cluster is directed to the IP Extension gateway. The TCL evaluates the incoming TCP 3-way handshake. If the traffic matches a rule that allows the traffic to enter a specified tunnel, it is sent to the tunnel (target). The IP Extension traffic is now on the WAN side.

Figure 14: IP Extension Gateway



5.4 GE Interfaces: The LAN Side

GE interfaces are either WAN (tunnel) facing or LAN (IP Extension) facing. An interface cannot do both and must be configured for one or the other. The default is WAN facing. LAN-side connectivity can be made from 1GbE, 10GbE, and 25GbE interfaces.

Specific to the Brocade SX6, it must be in the hybrid (FCIP and IP Extension) application mode before a GE interface can be configured as LAN. The Brocade 7810 and Brocade 7850 have no app-mode setting; it only has hybrid mode. The maximum number of LAN side interfaces on the Brocade 7850 and Brocade SX6 is 8 out of the 16 GE interfaces. On the Brocade 7810, the maximum number of LAN side interfaces is 4 out of the 6 GE interfaces.

The Brocade 7810 has two copper (RJ-45) ports that operate only at 1Gb/s. There is no advantage or disadvantage to using the copper ports (GE0 and GE1) over the SFP ports (GE2 and GE3). Speed is the only copper port limitation.

5.5 Logical Switches: The LAN Side

VF Logical Switches are specific to Fibre Channel and FICON. VF-LS are a function of Brocade Fibre Channel Switching ASICs. IP Extension traffic does not pass through these Brocade ASICs, and these ASICs have no involvement in IP Extension. LAN side GE interfaces must be in the default switch. Therefore, IP Extension LAN side interfaces do not participate in logical switches. Nonetheless, there is no reason why a VE_Port cannot be resident in any VF-LS for Fibre Channel /FICON use while also accommodating IP Extension transport.

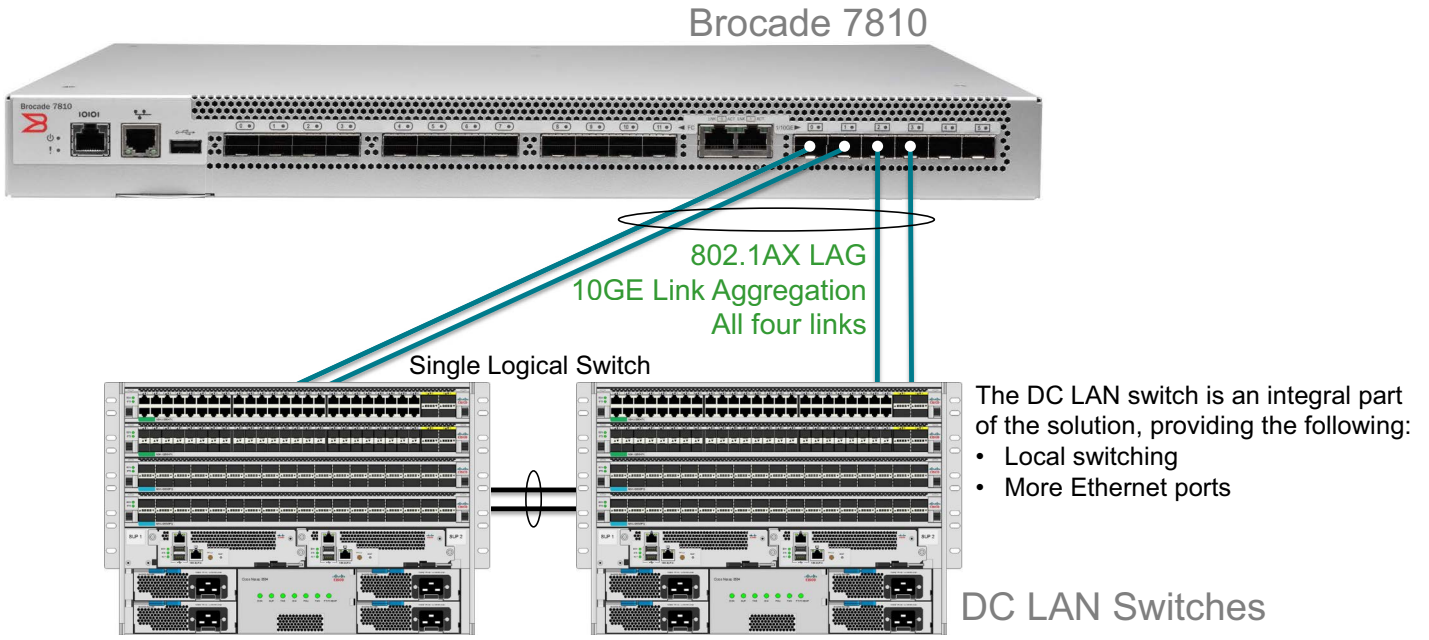
5.6 The Data Center LAN and Portchannels

Connecting to the data center LAN is an integral part of an IP Extension solution. The data center LAN provides more Ethernet interfaces for connecting end devices and offers local switching to other devices within the data center. Brocade IP Extension LAN ports are neither Ethernet switch ports nor IP router ports. One of two things will happen when traffic arrives at an IP Extension LAN port; either it matches a TCL rule and is forwarded into a tunnel, or the traffic is dropped. Incoming traffic will never be switched back out of a LAN interface on the same box. L2 loops cannot occur, and there is no need for Spanning Tree Protocol.

Data center LAN connectivity may vary from one link to multiple links; logically, only one link is supported. However, the link may be a link aggregation (LAG) or portchannel, a widely supported standard (802.1AX). A LAG makes multiple links appear as a single link.

Additionally, it is possible to create a LAG across multiple data center LAN switches, provided the switches appear logically as a single switch, which is common; see Figure 15. A portchannel with multiple links to multiple data center LAN switches creates the greatest level of resiliency, even if not all of the link bandwidth is required.

Figure 15: Data Center LAN Connectivity and Port Channels



IP Extension LAN-side Ethernet links support VLAN tagging (802.1Q); connect multiple LANs within the data center through the same links. An IP Extension gateway (`ipif lan.dp#`) must be configured for each VLAN in which IP Extension traffic flows. Up to eight gateways per DP are supported. Best practice is to have the network administrator prune all unused VLANs off their Ethernet ISL trunks, preventing spurious traffic from consuming the Ethernet links.

Best practice is to have at least two links, one to data center LAN switch A and one to switch B. Assuming the data center LAN switches are logically a single switch, configure the two links as a portchannel. If either data center LAN switch goes offline, IP Extension traffic continues to use the remaining LAN side link. Four links can be used, as shown in Figure 15.

NOTE:

- Using more than one link, in which the links do not form a LAG, will result in instability and a potential outage. Connectivity may operate for a period, however, it will eventually stop when the forwarding tables update.
- Portchannels do not apply to the WAN side. Brocade Extension Trunking is exclusively used on the WAN side because it provides everything portchannels do, plus, when data is lost in flight, extension trunking recovers the lost data and puts it back in order before sending it to the upper-layer protocol.

5.7 Traffic Control Lists

IP Extension uses a traffic control list (TCL) to direct incoming LAN-side IP Extension traffic to a specific tunnel. Interfaces starting with `lan.dp#` are IP Extension gateways and traffic flows if the incoming traffic matches an active (admin-enabled) TCL priority, which directs the traffic to the specified target (tunnel/VE_Port). A TCL is specific to IP Extension and has no bearing on IP routing or Fibre Channel/FICON functions. The TCL default is disabled; ensure that the priority number has an asterisk (*) next to it, which indicates that it is active.

For IP Extension outgoing traffic to the LAN side, an `iproute lan.dp# <router gateway>` statement can be configured, which is typically used in a Layer 3 deployment where the end device is on a subnet in which a router is between the IP Extension LAN and the end device. With `iproute` configured, IP Extension sends the data to the next-hop router, which forwards it to the destination subnet.

Give TCL rules a meaningful, human-readable name. TCL rules are assigned priority numbers. Rules are processed by priority number, from low to high. Upon the first match, the processing of rules terminates. Best practice is to pick priority numbers, leaving space between each, in case more rules need to be added. For example, count by tens or hundreds if using many rules.

Do not create a wide-open TCL in which everything can communicate with everything, including broadcast and multicasts. Sending broadcasts across the WAN is not considered best practice. On the other hand, it is not required to close communications down to exact device IP addresses. Specifying the end-device source and destination subnets, for example, `-S 192.168.10.0/24 -D 10.10.10.0/24`, works reliably.

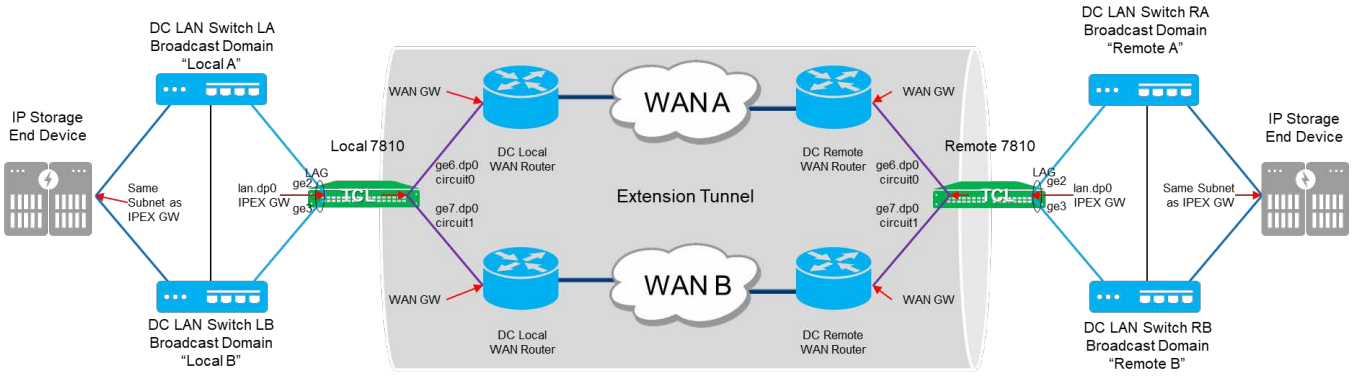
TCL rules are evaluated one time, when a TCP session forms its initial three-way handshake. After removing, adding, or making changes to TCL rules, the VE_Port for the tunnel must be disabled (`portdisable`) and enabled (`portenable`) to cycle the tunnel, which forces all the proxied TCP sessions to reform their connections, and forces the TCL to be reevaluated. If bringing down the entire tunnel is disruptive, for example if FCIP replication coexists, it is possible to disable all the LAN-side GE interfaces. Wait until all the TCP sessions have timed out, and then re-enable the GE interfaces. Another alternative is disabling the end devices' IP storage replication ports. The goal is to get the TCP sessions to initiate new connections across the TCL.

The hit counter for each TCL rule identifies matching incoming LAN traffic. The hit counter will increase by one for every new TCP session that forms. The hit counter will count by one for every matching UDP, ICMP, or BUM (broadcast, unknown, multicast) datagram. Eventually, traffic that does not match any configured rule falls to the bottom. The last rule is a deny-all rule and cannot be removed or modified. Any traffic that falls to the last rule will be dropped. Other than BUM traffic, no traffic should arrive at the IP Extension gateway unintended.

Chapter 6: IP Extension Architectures

The data flow through an IP Extension architecture starts at the originating end device. A static route on that device indicates if the destination is subnet x, use the IP Extension gateway. The IP Extension gateway is on the same L2 network, and the end device forwards the traffic to the gateway. IP Extension puts the traffic in the tunnel and forwards it to the remote side via the IP WAN network. On the remote side, it is removed from the tunnel and forwarded to the destination end device. See Figure 16.

Figure 16: Data Flow through an IP Extension Architecture

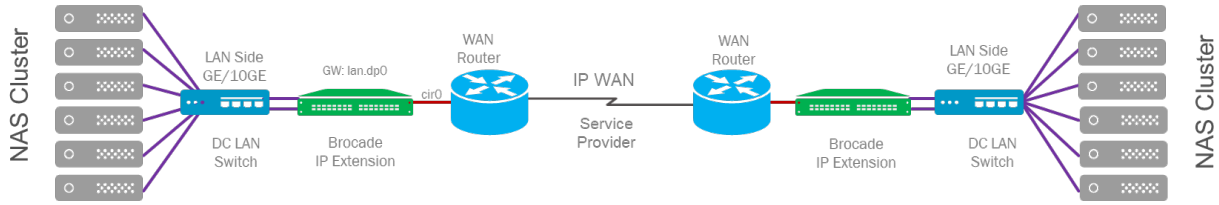


Various IP Extension architectures can be built, ranging from straightforward and cost-effective to complex with high availability and capacity.

6.1 Two-Box Solutions

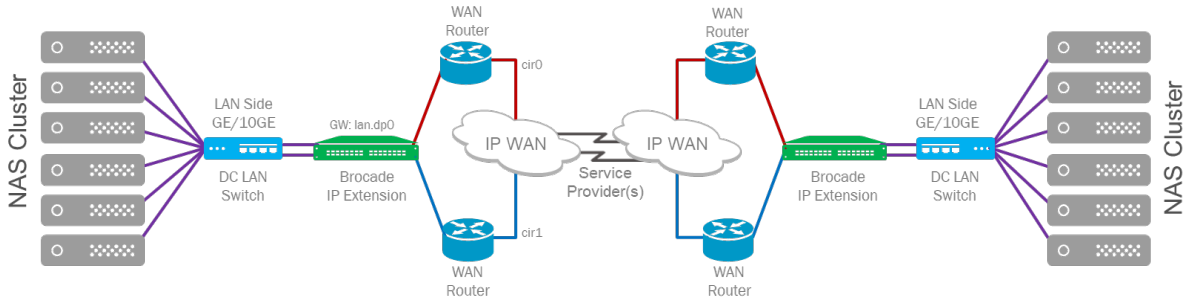
The base Brocade 7810 forms a typical two-box solution that does not use Brocade Extension Trunking. The base Brocade 7810 is a cost-sensitive platform that does not have Brocade Extension Trunking enabled; however, it can be enabled with the upgrade license. Since enabling Brocade Extension Trunking is required to create more than one circuit per tunnel, the architecture depicted in Figure 17 shows only a single circuit.

Figure 17: Brocade 7810 Base Unit Architecture



A two-box architecture using Brocade Extension Trunking is more common than not. Connecting the WAN side to the A and B switches increases availability; see Figure 18. The tunnel remains undisturbed when a switch, router, optic, cable, or WAN connection goes offline. All traffic will be delivered and delivered in order, although a portion of the bandwidth may no longer be available. In this architecture, the data center LAN switches and Brocade Extension platforms remain a single point of failure.

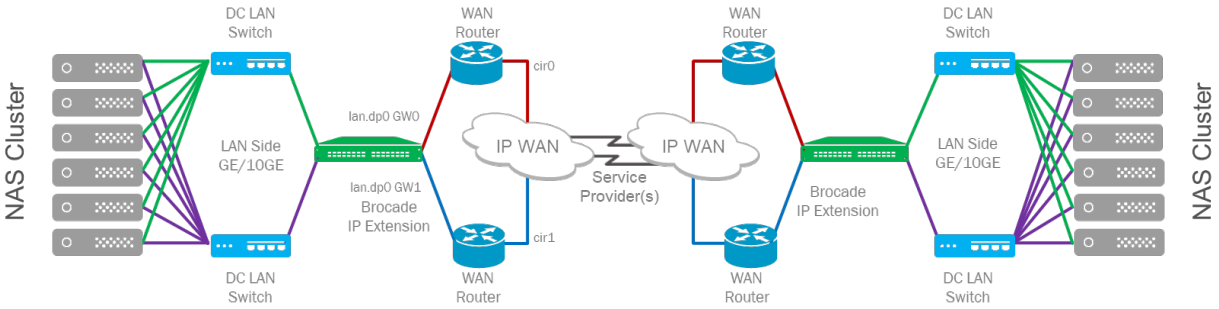
Figure 18: Two-Box Solution Using Brocade Extension Trunking, Single DC LAN Switch



The two-box solution with Brocade Extension Trunking and dual data center LAN switches eliminates them as a single point of failure. This architecture requires that the IP storage devices be capable of multiple gateways, preferably a gateway specific to each replication port or set of ports.

As shown in Figure 19, each end device has dual NICs connected to a different data center LAN switch. The data center LAN switch, in turn, connects to the extension platform. The same IP Extension gateway cannot accommodate the two separate data center LAN switches. The green path requires its own VLAN and IP Extension gateway, as does the purple path. The Brocade Extension platform is a single point of failure in this architecture.

Figure 19: Two-Box Solution Using Brocade Extension Trunking, Dual DC LAN Switches



The two-box solution with Brocade Extension Trunking and a single logical data center LAN switch eliminates the switches as a single point of failure; see Figure 20. The data center LAN switches have been joined together and logically appear to the network as a single switch, however, they have the redundancy of two devices, allowing for a single IP Extension gateway. End devices connect to the VLAN spanning the switches. The attached end-device replication ports are configured using the same subnet and IP Extension gateway. The Brocade Extension platform is a single point of failure in this architecture.

Figure 20: Two-Box Solution with Port Channels

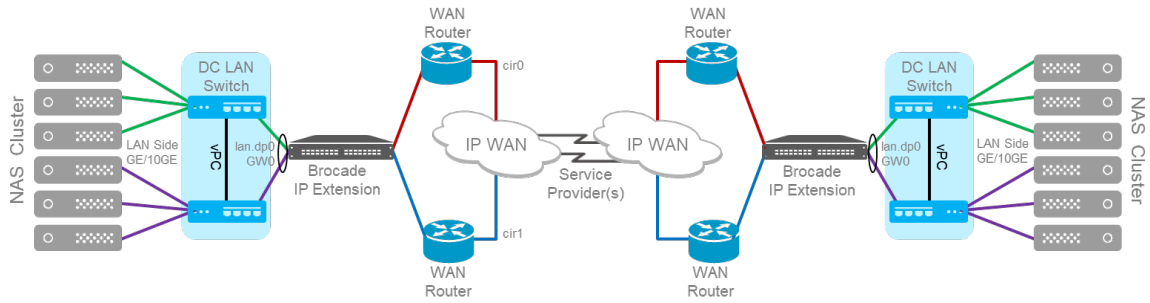
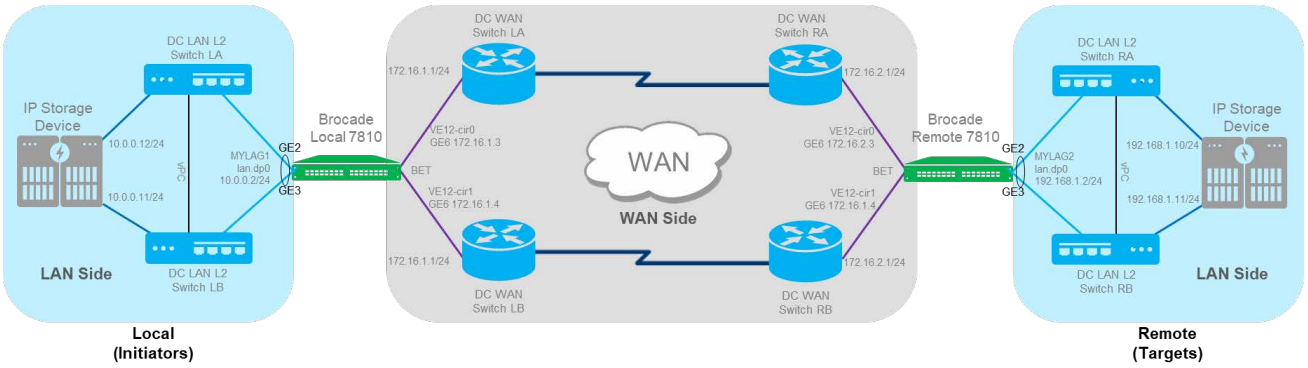


Figure 21 complements Figure 20 with additional detail. The example shows IP subnets and addresses for the end devices, IP Extension gateways, and WAN circuits.

Figure 21: Popular IP Extension Architecture with One Box per Data Center



6.2 Four-Box Solutions

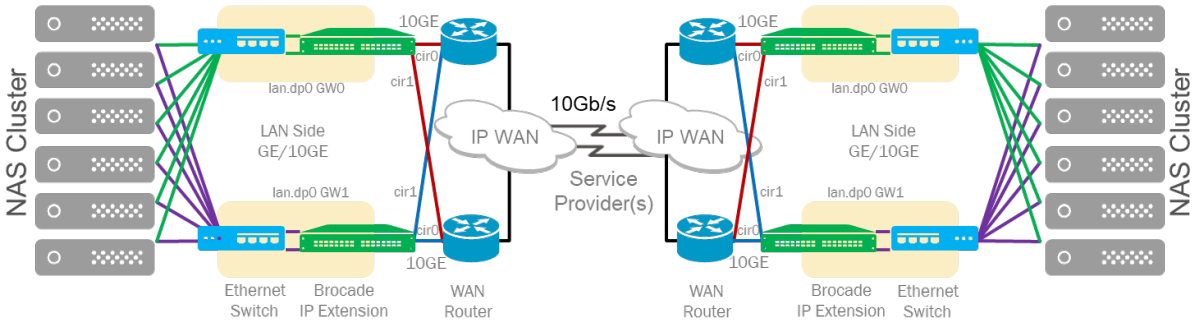
Avoiding a single point of failure may be paramount in some environments. A four-box Brocade Extension solution is possible when the end device supports the ability to accommodate more than one gateway. Deploying more than one local IP Extension platform requires more than one gateway. Brocade IP extension on the LAN side does not offer VRRP or HSRP, which provide a single virtual gateway. Therefore, the end device must be capable of the following:

- The end device performing replication must be capable of different subnets, static routes, or gateways per Ethernet interface or set of Ethernet interfaces.
- The end device performing replication using one subnet must be capable of unique static routes or gateways per Ethernet interface or set of Ethernet interfaces.

The data center LAN switches in this architecture are not configured to form a single logical Ethernet switch; see Figure 22. The two data center LAN switches are autonomous. The NAS cluster can configure multiple unique gateways, not including the default gateway. The IP Extension gateway is not the default gateway. When sending replication traffic to the remote NAS cluster, the green connections forward traffic to lan.dp0 GW0 (top), and the purple connections forward traffic to lan.dp0 GW1 (bottom).

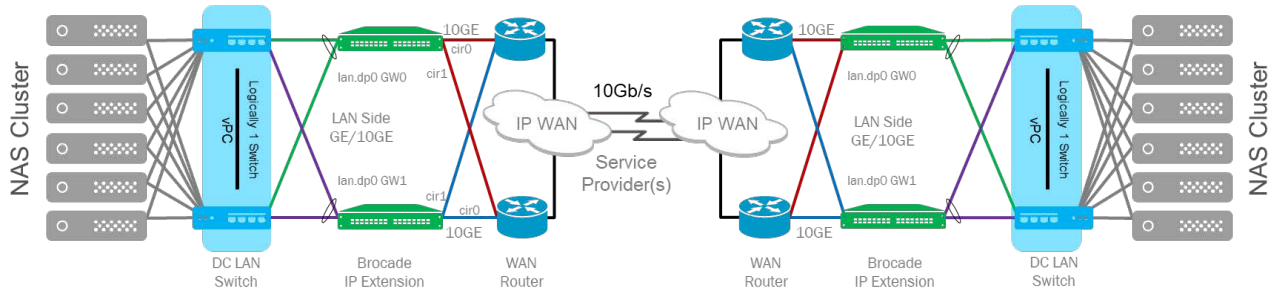
On the WAN side, there are two extension trunks (red and blue) with two circuits each. Each circuit connects to a different WAN switch or router. While most data centers implement VRRP, connecting to different WAN routers is still necessary if a network device goes down, an optic fails, or a cable is damaged. The gateway used on the WAN side is often the virtual gateway created by VRRP or HSRP and is accessible from both data center switches. The Brocade Extension platforms are not a single point of failure in this architecture.

Figure 22: Dual Connected High-Availability Architecture



The data center LAN switches shown in Figure 23 are logically a single Ethernet switch. A virtual portchannel connection allows them to join and appear as one. On the LAN side, Brocade Extension can form portchannels with a data center LAN switch or across a pair of data center LAN switches. To do this, the data center LAN switches must be configured as one logical switch. The purpose of this architecture is to gain redundant data center LAN switches. When one switch goes offline, the other forwards traffic to the IP Extension gateway. The Brocade Extension platforms are not a single point of failure in this architecture.

Figure 23: Four-Box Solution with a Single Logical DC LAN Switch



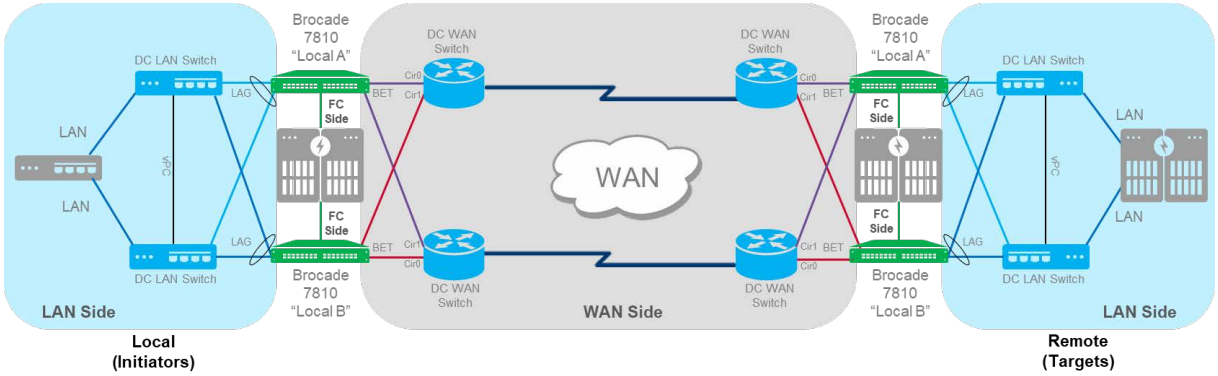
6.3 Mixed FCIP and IP Extension Solution

FCIP architectures are nearly always four-box solutions, and redundancy is of great concern. Adding IP Extension to a four-box architecture depends on the end device and the previously mentioned limitations. The WAN side remains unchanged because both protocols run through the same tunnel. When adding a protocol to an existing tunnel, ensure that the available bandwidth will adequately accommodate the new traffic. Circuit bandwidth may need to be increased.

The best practice when running IP Extension and FCIP in the same switch is to use the same VE_Port for FCIP and IP Extension. The desire is to have the VE_Port manage the bandwidth of both protocols across the tunnel. Managing bandwidth cannot be accomplished when different VE_Ports are used for each protocol.

Figure 24 is an expanded version of Figure 23, with the addition of directly attached Fibre Channel replication ports. Directly attached Fibre Channel replication ports are best practice, as is an independent replication SAN.

Figure 24: Mixed FCIP and Brocade IP Extension Architecture



Chapter 7: Connectivity and WAN Validation Tools

Connectivity test tools are available to validate and measure end-to-end IP path performance characteristics between a pair of Brocade Extension endpoints.

Table 5: Test Tools

Test Tool	Functionality
`portcmd --wtool`	A test tool for circuits that generates traffic using the same circuit configuration used in production and the same FCIP TCP ports and IP addresses. It sends test data to determine the characteristics and reliability of the IP network used by the circuit.
`portcmd --ping`	Tests connections between a local IP interface (ipif) and a destination IP address.
`portcmd --traceroute`	Traces hop by hop from a local IP interface (ipif) to a destination IP address.
`portshow fciptunnel -c --perf`	Displays tunnel and circuit performance statistics.
`portcfg sla`	The Brocade Extension SLA feature has been discontinued and removed from FOS as of version 10.x.

7.1 WAN Test Tool

It is best practice to evaluate the IP network characteristics before placing an extension tunnel into production. Often the IP network is different from what you expect. WAN Test Tool (Wtool) is a utility that can generate line-rate traffic between a local and remote extension platform via one or more of the tunnel's circuits. The Wtool session bit rate is user-specified; ARL settings are not used.

Existing circuits are not required to use Wtool. Wtool sessions can be configured before tunnel/circuits are configured, which may be desirable on new installations because it eliminates the need to disable existing circuits at both ends before Wtool is set up.

If there are existing circuits, they must be taken offline before Wtool sessions can be enabled. Other circuits that remain online will stay in production. Wtool is configured on both ends, as with any tunnel. After Wtool testing is complete, the sessions can be disabled. There is no need to delete sessions, which can be used again later. Enable the circuit again to put it back into production.

Wtool reports the amount of local data sent and the amount of remote data received, to determine the IP network's characteristics and reliability. This data includes packet drops, latency (RTT), jitter, out-of-order segments, and more. Typically, one end is the data source and the other is the data sink, however, bidirectional traffic is supported. The process continues for the user-specified duration or until terminated (Ctrl+C).

7.2 Ping

The `portcmd --ping` command tests simple connectivity between a local IP address (IPIF) on a particular Ethernet interface and a destination IP address. The protocol ICMP is used.

If no IP routes exist, ping a destination IP address on the same subnet, such as the IP address of the subnet's local gateway. Pinging on the same subnet demonstrates Layer 2 connectivity. If an IP route exists, you can ping a foreign subnet's destination; however, the remote device must have a compliment IP route to return the ping reply. Pinging across subnets demonstrates Layer 3 connectivity.

7.3 Traceroute

The `portcmd --traceroute` command traces the datagram path from a local IP address to a remote IP address through a particular Ethernet interface. The protocol ICMP is used.

Traceroute makes sense only on Layer 3 routed networks with one or more hops. Intermediate and remote devices must have complement IP routes to return the traceroute response. It is not unusual for service provider devices to not return traceroute responses; nonetheless, traceroute continues without the service provider responses until the time-to-live increments beyond the service provider's infrastructure.

7.4 Portshow fciptunnel

The `portshow fciptunnel` command is commonly used to show tunnel and circuit state, traffic, and configuration details. Some arguments modify the output to highlight specific aspects of the tunnels and circuits.

These are the different available output arguments for the `portshow fciptunnel` command:

```
--circuits
--detail
--summary
--reset
--config
--perf
--ha
--hcl
--tcp
--qos
```

The `--circuits` argument adds the circuits to the tunnel output; see the --perf example that follows. The `--perf` argument shows retransmits (ReTx), round trip time (RTT) in milliseconds, compression ratio (CmpRtio), transmit and receive rates in MB/s, transmit WAN percentage (TxWAN%), and transmit queue percentage (TxQ% for the tunnel) or BW (ARL settings for the circuits).

The following view is the summary view. If the output does not default to the summary view and you wish to see the summary, use the `--summary` argument.

```
switch:FID128:admin> portshow fciptunnel --perf
 Tunnel Circuit St Flg TxMBps RxMBps CmpRtio RTTms  ReTx  TxWAN%  TxQ%/BW  Met/G
-------------------------------------------------------------------------------
 24    -        Up -I-    0.0    0.0   1.0:1   - 1572382     0        0       -
-------------------------------------------------------------------------------
   Flg (tunnel): I=IP-Ext, s=Spillover
              St: High level state, Up or Dn
 TxWAN (tunnel): Tx WAN utilization high of primary circuits (--qos for range)
```

```
   TxQ (tunnel): Tx data buffering utilization high (--qos for range)
switch:FID128:admin> portshow fciptunnel --perf --circuits
 Tunnel Circuit St Flg TxMBps RxMBps CmpRtio RTTms  ReTx  TxWAN%  TxQ%/BW  Met/G
-------------------------------------------------------------------------------
 24    -        Up -I-    0.0    0.0   1.0:1   - 1572381     0        0       -
 24    0 ge2    Up ---    0.0    0.0       -  20  790667     0     100/150   0/-
 24    1 ge4    Up ---    0.0    0.0       -  20  781714     0     100/150   0/-
-------------------------------------------------------------------------------
   Flg (tunnel): I=IP-Ext, s=Spillover
              St: High level state, Up or Dn
 TxWAN (tunnel): Tx WAN utilization high of primary circuits (--qos for range)
      (circuit): Tx WAN utilization high (--qos for range)
   TxQ (tunnel): Tx data buffering utilization high (--qos for range)
```

The `--hcl` argument shows the Extension Hot Code Load (eHCL) state and whether traffic will be disrupted during a firmware update.

```
switch:FID128:admin> portshow fciptunnel --hcl
Checking FCIP Tunnel HA Status.
Current Status         : Ready
CP Version             : v8.2.2b
DP0 Status:
   State               : Online - Inactive
   Version             : v8.2.2b
   Current FC HA Stage : IDLE
   Current IP HA Stage : IDLE
   IP SVI Swapped      : NO
   DP COMM Status      : UP
DP1 Status:
   State               : Online - Inactive
   Version             : v8.2.2b
   Current FC HA Stage : IDLE
   Current IP HA Stage : IDLE
   IP SVI Swapped      : NO
   DP COMM Status      : UP
 Tunnel 24 (FID:128) FC:HA Offline IP:HA Offline - FC and IP traffic will be disrupted.
```

The `--config` and `--circuits` arguments show the local and remote IP addresses used to build each circuit for a tunnel.

For eHCL, the `--ha` argument shows the IP addresses used to build the automatically cloned remote and local backup tunnels when configuring circuits with the `--local-ha-ip` and `--remote-ha-ip` values. The circuits are designated as main tunnel (`M`), remote backup tunnel (`R`), and local backup tunnel (`L`) under `Flags`; refer to the output shown below.

eHCL is supported only on the Brocade 7850 and the Brocade SX6. The Brocade 7810 does not perform eHCL, however, it supports eHCL-enabled tunnels from the Brocade 7850 and Brocade SX6.

```
switch:admin> portshow fciptunnel --config --circuits --ha
 Tunnel Circuit   AdminSt Flags   Local IP             Remote Ip
-------------------------------------------------------------------------------
 24    -         Enabled -Mi----PI
 24    0 ge2     Enabled ----ah-i4 10.85.0.210         10.175.0.202
 24    1 ge3     Enabled ----ah-i4 10.85.0.211         10.175.0.203
 24    -         Enabled -Ri----PI
 24    0 ge2     Enabled ----ah-i4 10.85.0.210         10.175.0.226
 24    1 ge3     Enabled ----ah-i4 10.85.0.211         10.175.0.227
 24    -         Enabled -Li----PI
 24    0 ge2     Enabled ----ah-i4 10.85.0.230         10.175.0.202
```

```
 24    1 ge3     Enabled ----ah-i4 10.85.0.231         10.175.0.203
-------------------------------------------------------------------------------
 Flags (tunnel): l=Legacy QOS Mode
                 M=MainTunnel  L=LocalBackup  R=RemoteBackup
                 i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

The `--tcp` argument shows the TCP sessions for the circuits. In this example, every three rows are tunnel, circuit, and circuit. Although it is the same tunnel, VE24, the tunnel is logically separated into QoS classifications. Those classifications include control traffic, FCIP (High, Med, Low), and IP Extension (High, Med, Low). Traffic rates can be monitored specifically for each QoS classification.

```
switch:FID128:admin> portshow fciptunnel --tcp --qos --summary --circuits
 Tunnel Circuit  OpStatus  Flags    Uptime  TxMBps  RxMBps ConnCnt CommRt Met/G
-------------------------------------------------------------------------------
 26    -         Up       c--fT--a-  2h1m23s    0.00    0.00   2      -       -
 26    0 ge5     Up       ----ah--4  2h1m23s    0.00    0.00   2     0/4000  0/0
 26    1 ge6     Up       ----ah--4  2h1m23s    0.00    0.00   2     0/4000  0/0
 26    -         Up       h--fT--a-  2h1m23s    0.00    0.00   2      -       -
 26    0 ge5     Up       ----ah--4  2h1m23s    0.00    0.00   2  1000/4000  0/0
 26    1 ge6     Up       ----ah--4  2h1m23s    0.00    0.00   2  1000/4000  0/0
 26    -         Up       m--fT--a-  2h1m23s  668.21    5.43   2      -       -
 26    0 ge5     Up       ----ah--4  2h1m23s  334.10    2.72   2   600/4000  0/0
 26    1 ge6     Up       ----ah--4  2h1m23s  334.10    2.72   2   600/4000  0/0
 26    -         Up       l--fT--a-  2h1m23s    0.00    0.00   2      -       -
 26    0 ge5     Up       ----ah--4  2h1m23s    0.00    0.00   2   400/4000  0/0
 26    1 ge6     Up       ----ah--4  2h1m23s    0.00    0.00   2   400/4000  0/0
-------------------------------------------------------------------------------
 Flags (tunnel): c=Control h=HighPri m=MedPri l=LowPri
                 i=IPSec f=Fastwrite T=TapePipelining F=FICON r=ReservedBW
                 a=FastDeflate d=Deflate D=AggrDeflate P=Protocol
                 I=IP-Ext
      (circuit): h=HA-Configured v=VLAN-Tagged p=PMTU i=IPSec 4=IPv4 6=IPv6
                 ARL a=Auto r=Reset s=StepDown t=TimedStepDown  S=SLA
```

The `--qos` argument without `--summary` shows detailed information about the various QoS flows. In the following example, the first section of the output is the VE_Port level, tunnel scope. The second section is the scope of the first circuit (circuit0 on DP0). The control traffic section is the scope of the first QoS classification. This output has many sections, and using a pipe to more (`| more`) becomes handy for pagination. The following sections will be all the other QoS classifications for circuit0 and then all the classifications for circuit1. The output will continue for all circuits in the tunnel; the output is limited to the specified tunnel (24).

```
switch:FID128:admin> portshow fciptunnel 24 --tcp --qos --circuits | more
 Tunnel: VE-Port:24 (idx:0, DP0) Broadcom Extension
 ==================================================
  Oper State           : Online
  TID                  : 24
  Flags                : 0x00000000
  Priority             : Control
  Compression          : Aggressive Deflate
  QoS BW Ratio         : 0%
  Fastwrite            : Disabled
  Tape Pipelining      : Disabled
```

```
  IPSec                 : Enabled
  IPSec-Policy          : MyTunnel
  Legacy QOS Mode       : Disabled
  Load-Level (Cfg/Peer): Failover (Failover / Failover)
  Local WWN             : 10:00:c4:f5:7c:3b:24:86
  Peer WWN              : 10:00:c4:f5:7c:3b:25:06
  RemWWN (config)       : 00:00:00:00:00:00:00:00
  cfgmask               : 0x0000001f 0x4001034e
  Flow Status           : 0
  ConCount/Duration     : 1 / 57m1s
  Uptime                : 16d19h
  Stats Duration        : 57m1s
  Receiver Stats        : 251460 bytes / 987 pkts /   58.00 Bps Avg
  Sender Stats          : 266544 bytes / 987 pkts /   59.00 Bps Avg
  TCP Bytes In/Out      : 5011204 / 6206168
  ReTx/OOO/SloSt/DupAck: 42 / 23 / 38 / 0
  RTT (min/avg/max)     : 20 / 20 / 30 ms
  Wan Util (low/high)   : 0.0% / 0.0%
  TxQ Util (low/high)   : 0.0% / 0.0%
 Circuit 24.0 (DP0)
 ==================================================
  Admin/Oper State      : Enabled / Online
  Priority              : Control
  Flags                 : 0x00000000
  IP Addr (L/R)         : 10.1.0.10 ge2 <-> 10.2.0.10
  HA IP Addr (L/R)      : 0.0.0.0  <-> 0.0.0.0
  Configured Comm Rates: 0 / 150000 kbps
  Peer Comm Rates       : 0 / 150000 kbps
  Actual Comm Rates     : 0 / 150000 kbps
  Keepalive (Cfg/Peer) : 6000 (6000 / 6000) ms
  Metric                : 0
  Connection Type       : Default
  ARL-Type              : Auto
  PMTU                  : Disabled
  HA PMTU               : Disabled
  SLA                   : (none)
  IPSEC                 : Enabled (0)
  Failover Group        : 0
  VLAN-ID               : NONE
  L2Cos (Ctrl)          : 0
   DSCP (Ctrl)          : 0
  cfgmask               : 0x40000000 0x01e10def
  Flow Status           : 0
  ConCount/Duration     : 1 / 57m1s
  Uptime                : 16d19h
  Stats Duration        : 57m1s
  Receiver Stats        : 117016 bytes / 490 pkts /   44.00 Bps Avg
  Sender Stats          : 130544 bytes / 490 pkts /   48.00 Bps Avg
  TCP Bytes In/Out      : 2479184 / 3092620
  ReTx/OOO/SloSt/DupAck: 22 / 14 / 20 / 0
  RTT (min/avg/max)     : 20 / 20 / 30 ms
  Wan Util (low/high)   : 0.0% / 0.0%
  TCP Connection 24.0 HA-Type:Main Pri:Control Conn:0x0a8a9434
  ==================================================
   Local / Remote Port      : 3225 / 55334
   Duration                 : 16d19h
   MSS                      : 1362 bytes
   ARL Min / Cur / Max      : 0 / 0 / 75000
```

```
   ARL Reset Algo           : StepDown
   Send Window
    Size / Scale            : 1874432 / 9
    Slow Start Threshold    : 16777216
    Congestion Window       : 16778576
    Pkts InFlight           : 0
   Recv Window
    Size / Scale            : 1874944 (Max:1874944) / 9
   SendQ Nxt / Min / Max    : 0xc2123cef / 0xc2123cef / 0xc2123cef
   RecvQ Nxt / Min / Max    : 0x2d40784e / 0x2d40784e / 0x2d5d13ae
   RecvQ Pkts               : 0
   Sender Stats
    Sent Bytes / Pkts       : 659788860 / 4475663
    Unacked Data            : 0
    Retransmits Slow / Fast : 4207 / 0 (High:0)
    SlowStart               : 3740
   Receiver Stats
    Recv Bytes / Pkts       : 534668512 / 4468668
    Out-of-Order            : 0 (High:24)
    Duplicate ACKs          : 0
   RTT / Variance (High)    : 19 ms (30 ms) / 0 ms (20 ms)
```

The `--reset` argument resets the output counters. Note that when the command is executed, the counters do not display as reset until the subsequent execution; see the following example.

```
switch:FID128:admin> portshow fciptunnel -c --perf --summary --reset
 Tunnel Circuit St Flg TxMBps RxMBps CmpRtio RTTms  ReTx  TxWAN%  TxQ%/BW  Met/G
-------------------------------------------------------------------------------
 24    -        Up -I-    0.0    0.0   1.0:1   -     352     0        0       -
 24    0 ge2    Up ---    0.0    0.0       -  20     172     0     100/150   0/-
 24    1 ge4    Up ---    0.0    0.0       -  20     180     0     100/150   0/-
-------------------------------------------------------------------------------
   Flg (tunnel): I=IP-Ext, s=Spillover
              St: High level state, Up or Dn
 TxWAN (tunnel): Tx WAN utilization high of primary circuits (--qos for range)
       (circuit): Tx WAN utilization high (--qos for range)
   TxQ (tunnel): Tx data buffering utilization high (--qos for range)
switch:FID128:admin> portshow fciptunnel -c --perf --summary
 Tunnel Circuit St Flg TxMBps RxMBps CmpRtio RTTms  ReTx  TxWAN%  TxQ%/BW  Met/G
-------------------------------------------------------------------------------
 24    -        Up -I-    0.0    0.0   1.0:1   -       4     0        0       -
 24    0 ge2    Up ---    0.0    0.0       -  20       4     0     100/150   0/-
 24    1 ge4    Up ---    0.0    0.0       -  20       0     0     100/150   0/-
-------------------------------------------------------------------------------
   Flg (tunnel): I=IP-Ext, s=Spillover
              St: High level state, Up or Dn
 TxWAN (tunnel): Tx WAN utilization high of primary circuits (--qos for range)
       (circuit): Tx WAN utilization high (--qos for range)
   TxQ (tunnel): Tx data buffering utilization high (--qos for range)
```

Chapter 8: Unsupported Features

Brocade products do not support the following features:

- Brocade products do not support Cisco E_Port (ISL) interoperability or OEM efforts to support Cisco ISL interoperability.
- VE_Port lay-in-wait is not supported for FICON, OSTP, and FastWrite optimizations. Conceptually, by manually adjusting Fabric Shortest Path First (FSPF) costs to be unequal, a higher-cost VE_Port becomes active when another lower-cost VE_Port goes offline. For instance, imagine two Brocade SX6 Extension Blades, in which one with the lowest cost VE_Ports goes offline, allowing the backup blade to become active. The problem is that considerable state information involved with FC protocol optimization (FCP/SCSI and FICON flows) are neither maintained nor recovered when rerouting traffic to the newly active lay-in-wait VE_Port. The jobs will fail and must be restarted.
- Brocade products do not support per-packet load balancing (PPLB). PPLB is often the cause of chronic and sometimes severe out-of-order (OOO) packets. To some degree, TCP can easily handle OOO packets; however, PPLB will cause TCP to go into overdrive, putting a dramatic strain on resources. Also, PPLB increases I/O response times from incomplete sequences because TCP is waiting for out-of-order data to arrive. PPLB may also induce a significant number of retransmits that unnecessarily consume valuable bandwidth. There are load-balancing techniques that are flow-based, do not cause problems, and load-balance WAN links equally well.
- There are no supported WAN optimization platforms.
- Starting with Fabric OS 10.x, FICON Acceleration is no longer an Extension feature in Fabric OS.
- Starting with Fabric OS 10.x, SLA is no longer an Extension feature in Fabric OS.

Revision History

Extension-Design-DG102; November 18, 2025

- Update with Gen 8 and FOS 10.

Extension-Design-DG101; August 15, 2023

- Update; remove Brocade 7840, add Brocade 7850.

Extension-Design-DG100; November 20, 2020

Initial document version.