SAN Fabric Resiliency and Administration Best Practices User Guide

Table of Contents

Chapter 1: Introduction

This document is intended for SAN administrators, Brocade® certified systems engineers, vendor and Brocade technical support personnel, IT architects, and system integrators who provide value-added management solutions based on the latest product releases from Broadcom.

As the leading provider of data center storage networking solutions, Broadcom helps organizations around the world connect, share, and manage their information in the most efficient manner. For over 20 years, Broadcom has been developing, installing, and supporting customers on Fibre Channel (FC) storage area networks (SANs) and, over time, has developed deep technical knowledge in administering resilient SANs. This document provides a high-level description of the most commonly experienced detrimental network and device behaviors, it explains how to use Brocade products and features to protect your data center, and it recommends best-practice configuration and administration guidelines.

Many factors in a SAN environment, such as inadequate design, faulty or improperly configured devices, misbehaving hosts, and faulty or substandard FC media, can significantly affect the performance of FC fabrics and the applications that they support. Broadcom offers many tools and features to assist with managing a robust SAN.

The scope of this document is to address common issues that are faced by administrators in managing their SANs. The goal is to reduce the time and cost needed for troubleshooting and dealing with application anomalies by using available Brocade technology to minimize fabric-wide disruptions. The details outlined in this document are for Gen 6 (32Gb/s) and Gen 7 (64Gb/s) devices only.

Although certain aspects of today's data centers are common in most environments, no two data centers are exactly alike, and no single set of configuration parameters apply universally to all environments. Broadcom works directly with customers in every type of SAN deployment to develop recommendations and guidelines that apply to most environments. However, you should always validate all recommendations for your particular needs and consult with your vendor or Broadcom representative to ensure that you have the most robust design and configuration available to fit your situation.

Broadcom also offers extensive professional services to assist with tuning and optimizing the features discussed in this document for a customized deployment in your data center. For details, visit:

https://www.broadcom.com/support/fibre-channel-networking

About Brocade

Networking solutions from Brocade, a Broadcom® company, help the world’s leading organizations transition smoothly to a world where applications and information reside anywhere. This vision is designed to deliver key business benefits such as unmatched simplicity, nonstop networking, application optimization, and investment protection.

Innovative storage networking solutions for data center storage networks help reduce complexity and cost while enabling virtualization and cloud computing to increase business agility.

NOTE: The features and functions covered in this document apply only to products based on Brocade Fabric OS® software. This document is not a replacement for product-specific manuals or detailed training on Brocade Fabric OS or Brocade SANnav™ Management Portal.

Chapter 2: Trends in Data Center Storage Networking

Enterprise data centers continue to push virtualization at higher density and support a much broader set of workloads. More virtual machines (VMs) are being deployed on the server end, and more storage LUNs are provisioned on storage arrays. At the same time, adoption of all flash storage has accelerated. Customers are starting to adopt next generation NVMe based storage that offer much lower latency access to data. The use of virtualization, flash storage, and automation tools has allowed applications and services to be deployed faster while shattering performance barriers. The unprecedented number of application and service interactions has also increased the complexity, risk, and instability of mission-critical operations. For example, increasing workload virtualization leads to a higher number of application flows but reduces the visibility of the flows. At the same time, storage virtualization increases the number of short FC frames and the potential for congestion. As a result, IT organizations require storage networks that deliver ultra-low latency, higher-capacity bandwidth, and greater reliability. To meet service-level agreements (SLAs) or achieve service-level objectives (SLOs), IT administrators also need new tools that can help ensure nonstop operations, quickly identify potential points of congestion, and maximize application performance, while simplifying administration.

In the past, the process of problem detection and mitigation relied on the diligence, skill, and experience of the fabric administrator. Usually, a manual task of searching and tracing error conditions was the regimen. The operational cost and potential disruptive impact to critical applications are simply not acceptable in today's data center. To assist in administrative automation and improve network uptime, Broadcom has developed Autonomous SAN technology that offers self-learning, self-optimizing, and self-healing capabilities. Autonomous SAN is realized through Fabric Vision® technology that combines built-in capabilities in Brocade Gen 6 and Gen 7 platforms, Brocade Fabric OS, and Brocade SANnav Management Portal features. Fabric Vision provides detailed monitoring and alerting, as well as response and mitigation, that vastly improve the fabric administrator's insight and response time.

This document discusses these capabilities and their application to the factors that affect fabric resiliency by detailing the processes of detection and mitigation for each factor. The fabric-specific tools and design recommendations that follow provide you with a robust approach to fabric resiliency.

Chapter 3: Brocade Fabric OS and Fabric Vision

Brocade Fabric Vision technology combines integrated technology in Brocade Gen 6 and Gen 7 platforms, Fabric OS, and Brocade SANnav Management Portal with deep know-how from 20 years of storage networking best practices to help customers manage a large-scale storage network. Organizations that use Brocade products and services with Brocade Fabric Vision technology greatly simplify monitoring, increase operational stability, and dramatically reduce costs associated with network administration.

In real-world scenarios, the factors that affect the resiliency of a network can be inside or outside the fabric itself. In both cases, Brocade Fabric Vision technology has features to help identify and troubleshoot these factors, mitigate their impact, and increase overall uptime.

This document is divided into the following main sections that cover the topics of resiliency:

- Factors affecting fabric resiliency
- Faulty media (description, detection, and mitigation)
- Different types of congestion:
  - Oversubscription (description, detection, and mitigation)
  - Credit stalled devices (description, detection, and mitigation)
  - Loss of buffer credits (description, detection, and mitigation)
- Available tools to aid the fabric administrator
- Designing resiliency into the fabric
- Appendices detailing selected topics

This document also covers Fabric OS configuration best practices to improve the reliability, availability, and supportability of a fabric:

- Fabric configuration
- Routing policy configuration

The topics covered in this document are complex. Additional details on individual feature settings and configuration are included in the appendices to allow the main body of this document to flow more easily. You are encouraged to review the main body of the document and then refer to the appendices after the general concepts and information are understood.

Every effort has been taken to ensure the accuracy of the information in this document. However, it is essential to consult the appropriate version of the *Brocade Fabric OS Command Reference Manual* (covering the CLI) to confirm the correct syntax for all commands used.

The following is a partial list of material that supplements the information provided here (all documents are available to registered users at www.broadcom.com/support/download-search):

- *Brocade SAN Design and Best Practices*
- *Brocade Fabric OS Administration Guide* (produced for each major Brocade Fabric OS release)
- *Brocade Fabric OS Command Reference Manual* (produced for each major Brocade Fabric OS release)
- *Brocade Fabric OS Monitoring and Alerting Policy Suite User Guide* (produced for each major Brocade Fabric OS release)
- *Brocade Fabric OS Flow Vision User Guide* (produced for each major Brocade Fabric OS release)
- *Brocade SANnav Management Portal User Guide* (produced for each major Brocade SANnav release)

This document covers Brocade Fabric OS releases from version 8.0 to version 9.0. New revisions are produced covering subsequent Brocade Fabric OS releases.

Chapter 4: Feature Availability

This document concentrates specifically on Brocade Fabric Vision features (and related capabilities) that help provide optimum fabric resiliency. While many Fabric Vision features were introduced in Fabric OS 7.x, Fabric OS 8.1 is the minimum supported version for gaining the most benefits from these features on Gen 5 and Gen 6 platforms. For the latest Gen 7 platforms, Fabric OS 9.0.0a or later is required. (Download the *Brocade Software Release Support and Posting Matrices* or consult your vendor for the latest supported Brocade Fabric OS releases.)

Throughout this document, special requirements including required licenses, minimum release levels, and platform limitations are noted. Review the additional documentation noted above to understand all of the tools that are available for maintaining an FC SAN environment. Also, read the Brocade Fabric OS release notes for important information related to your specific version of Brocade Fabric OS.

Chapter 5: Fabric Resiliency

Fabric resiliency is the ability of an FC fabric to tolerate unusual conditions that might place the fabric's operation at risk. Examples of such risks include server or storage ports that cannot process or respond fast enough, marginal links that degrade performance, or long-distance links that are at a greater risk of experiencing degradation in signal quality. Unless mitigated, these risks can lead to instability and unexpected behavior in the fabric. In some cases, such behavior can affect the I/O response for many devices that are connected to the fabric.

There are several common types of abnormal behavior that originate from fabric components or attached devices:

- Faulty media (fiber-optic cables and small form-factor pluggable [SFP]/quad small form-factor pluggable [QSFP] optics)
  - Faulty media increase bit errors, leading to excessive cyclic redundancy check (CRC) errors, invalid transmission words, and other conditions that cause frame loss. This may result in I/O failure and application performance degradation.
- Congestion – Congestion is defined as reduced network performance that occurs when the traffic being carried on a network link or node exceeds its capacity. Congestion in a fabric can come from the following sources:
  - Loss of buffer credits –Buffer credits are permanently lost when the acknowledgment signal from the receiving end of a link fails to reach sending end. The loss of buffer credits results in degraded or, in extreme cases, zero throughput on a link.
  - Credit-stalled devices – End devices (servers or storage arrays) that do not respond as quickly as expected by returning buffer credits can cause the fabric to hold frames for excessive periods of time. Credit-stalled devices cause increased latency, application performance degradation, or, in extreme cases, I/O failure. If not immediately addressed, credit-stalled devices may result in severe stress on the fabric.
  - Oversubscription – Oversubscription is caused by insufficient link bandwidth or oversubscribed end devices. Link or device oversubscription can result in application performance degradation or, in extreme cases, I/O failure.

This document examines each of these types of behavior in depth.

Chapter 6: Faulty Media

6.1 Description

Faulty media is one of the most common sources of fabric problems and eventual data center disruption. Faulty media can include damaged or substandard cables, optical transceivers including SFPs and QSFPs, and patch panels or receptacles; improper connections; malfunctioning extension equipment; and other types of external issues. Media can fault and fail on any port type, E_Port or F_Port, often intermittently and unpredictably, making it even harder to diagnose.

Faulty media that involve F_Ports impact the end device attached to the F_Port and the devices communicating with that device. This can lead to broader issues, because the effects of the media fault are propagated through the fabric.

Failures on E_Ports have the potential for even greater impact. Multiple flows (host/target pairs), as well as inter-switch control traffic, simultaneously traverse a single E_Port. In large fabrics, the number of flows passing through an E_Port can be very high. In the event of a media failure involving one of these links, it is possible to disrupt some or all of the flows that use that link.

A severe media fault or complete media failure can disrupt the port or even take the port offline. When an F_Port fails completely, the condition is usually detected by the connected device (storage or host). The device is usually configured to continue working through an alternative connection to the fabric. When this occurs on an F_Port, the impact is specific to flows that involve that F_Port.

An E_Port that goes offline causes the fabric to drop all routes that use the failed E_Port. This is usually easy to detect and identify. E_Ports are typically redundant, such that a severe failure results in a minor drop in bandwidth since the fabric automatically utilizes available alternate paths. The error reporting built into Brocade Fabric OS or the MAPS dashboard readily identifies the failed link and port, allowing for simple corrective action and repair.

Moderate levels of media fault cause failures to occur intermittently. However, a port may remain online or, in some cases, may repeatedly transition between online and offline states. This can cause repeated errors that, if left unresolved, can recur indefinitely or until the media fails completely.

Intermittent problems can be more difficult to assess. The fabric does its best to determine if a problem is critical enough to disable a port. Disabling a port causes the host failover software to stop using the port that is attached to the faulty media and to re-establish connectivity through its alternative path. However, problems can occur when the severity of the fault remains undetermined. This leaves the host and the applications that are running on it to cope with the results of the intermittently failing media.

Multipath design is typically deployed in data center SANs to improve performance and resiliency. The multipath driver in server software theoretically should mitigate effects of faulty media. However, intermittent problems caused by moderate media faults result in long retry times without ever triggering a full path failover in multipath driver. This leads to severe application degradation even in a multipath environment.

When this type of failure occurs on E_Ports, the result can be devastating. Repeated errors can affect many flows. This can result in a significant impact to applications that can last for a prolonged period of time.

Brocade Gen 6 and Gen 7 platforms support the following features that simplify the detection of, troubleshooting of, and in some case, recovery from faulty media problems:

- Monitoring and Alerting Policy Suite (MAPS)
- Fabric Notification

- Forward Error Correction (FEC)
- Automatic Credit Loss Detection and Recovery
- ClearLink® Diagnostics

NOTE: FC switches cannot correct for all problems caused by faulty media; ultimately, the problems must be addressed in the host or target devices, cable plant, or media where the fault occurs.

6.2 Detection

The presence of faulty media can manifest with the following symptoms:

- A spike in reported power draw by the media
- CRC errors on frames
- Invalid transmission words (including encoder out errors)
- Loss of sync errors on ports
- Loss of signal errors on ports
- State changes (ports going offline or online repeatedly)
- Credit loss:
  - Complete loss of credit on a virtual channel (VC) of an E_Port prevents traffic from flowing on that VC, which results in frame loss and I/O failures for devices that use the VC.
  - Partial credit loss, while a concern, usually does not significantly affect traffic flow, due to the high link speeds of ISLs today.
- Switch-issued link resets (resets when a device or link fails to respond within two seconds)

You can enable Brocade MAPS to automatically detect these types of faulty media conditions. Brocade MAPS has three predefined policies with rules that generate alerts based on predefined thresholds.

6.3 Mitigation

You must identify and correct faulty media issues as soon as possible. Otherwise, they can lead to severe fabric problems, such as dropped frames, performance impact, and permanent credit loss. At the very least, you must isolate the ports with faulty media.

Brocade Gen 6 (32Gb/s) and Gen 7 (64Gb/s) platforms support Forward Error Correction (FEC), which automatically corrects bit errors. FEC enhances the link reliability and improves resiliency with the presence of marginal media. Enable FEC between all supported devices. FEC is standard and mandatory on Gen 6 (32Gb/s) and Gen 7 (64Gb/s) links.

The Brocade MAPS feature also protects against faulty components with the actions of Port Fencing and Port Decommissioning. Port Fencing and Port Decommissioning are available for each of the previously noted conditions. They are included in all predefined MAPS policies with the thresholds that automatically protect the fabric from these error conditions. The thresholds have been tested and tuned to quarantine misbehaving components that are likely to cause a fabric-wide impact. These thresholds are very unlikely to falsely trigger on normally behaving components.

When you use the Port Decommissioning feature with MAPS, enable the “port decommissioning without disabling” feature for all ISLs. Instead of fencing the port by decommissioning and disabling the link, the port is decommissioned without disabling and is marked as impaired. An impaired port is removed from all the routes from both sides of the link, and it is prevented from being used in further route decisions. An impaired port remains online and is used only if it is the last link available to a neighboring domain. For more information on using Port Decommissioning features, refer to the *Brocade Fabric OS Administration Guide*.

The Fabric Notification feature introduced in FOS v9.0 to Brocade Gen 6 and Gen 7 platforms automatically send Fabric Performance Impact Notifications (FPIN) to end devices when CRC or ITW errors are defected. The FPINs are sent to the devices on the faulty link and their peer devices in the same zone. The Fabric Notification feature enables the affected devices to autonomously mitigate faulty media, for example, to automatically failover immediately to an alternate path. Fabric Notification is discussed in detail in Chapter 8 and in Appendix G.

The Brocade ClearLink Diagnostics feature dramatically simplifies troubleshooting of faulty media. ClearLink Diagnostics run a battery of tests to validate the link integrity between two ports, including connectivity through SFPs and optical cables. If one of the components is faulty, ClearLink Diagnostics can pinpoint that component. ClearLink Diagnostics tests require only a few FOS commands or Brocade SANnav Management Portal menu selections. The ClearLink Diagnostics are much easier and less time-consuming to initiate, with more deterministic results and link-distance estimates that can identify excessive cable length. Whenever symptoms related to faulty media are detected, administrators can use ClearLink Diagnostics to determine which faulty media components need to be replaced and which do not. Refer to your product's hardware reference manual for instructions on the repair of media faults. Furthermore, to proactively prevent marginal media from impacting applications, run ClearLink Diagnostics before any ports are enabled to pass I/O. Administrators can automate these tests with configuration options. ClearLink Diagnostics is discussed in detail in Chapter 8 and in Appendix A.

Chapter 7: Congestion

Congestion is another common class of abnormal fabric behavior that can impact application performance or, in severe cases, cause I/O failure. Congestion occurs when traffic that is carried on a link or network node exceeds its capacity to handle the demand. Congestion can be caused by oversubscription, credit-stalled devices, or loss of buffer credits. Each source of congestion is described in detail below with the appropriate tools to detect and mitigate the different types of congestion.

7.1 Oversubscription

7.1.1 Description

Oversubscription, in terms of source ports to target ports and devices to ISLs, can cause instances of insufficient link capacity, which leads to congestion. As Fibre Channel link bandwidth has increased to 32 or 64 gigabits per second, instances of insufficient link bandwidth capacities have radically decreased within a properly designed SAN. On the other hand, oversubscription can still cause congestion if a SAN fabric is improperly designed. A common example is when the storage port speed is upgraded without upgrading the server port speed. This form of oversubscription can cause serious congestion when servers have many READ data requests but cannot handle the data sent by storage, effectively limiting throughput of the storage port to the lower speed of the server port.

7.1.2 Detection

Brocade MAPS is an important tool to detect congestion by monitoring port bandwidth utilization. When transmit, receive, or overall throughput exceeds the bandwidth utilization thresholds, violations are recorded and notifications are sent to administrators. MAPS also monitors the number of NPIV logins to detect when the number of NPIV hosts approaches the configured limit on an N_Port. More important, Fabric Performance Impact (FPI) monitors congestion on ports and sends notifications to administrators if the congestion is caused by oversubscription. FPI also monitors the zoned device ratio and provides an alert if a zoning configuration allows a disproportionate number of hosts to communicate with a target or a disproportionate number of targets to communicate with a host.

MAPS and FPI are discussed in detail in Chapter 8 and in Appendix B and Appendix C. On Brocade Gen 6 and Gen 7 platforms, administrators can use the IO Insight capability to obtain pending IO metrics to gain insight on the number of IOs in a device queue.

7.1.3 Mitigation

Oversubscription is primarily a problem of improper SAN design. Hence, once the source of congestion is identified, the problem must be addressed by correcting the design. For example, reduce the number of hosts per target or LUN, increase the number of ISLs, or upgrade the server HBA speed. Common SAN design mistakes include link-speed mismatch, high storage target fan-in ratio, and an oversubscribed N_Port in an NPIV environment. Administrators should pay attention to these important ratios during the network design phase or when scaling out the network, in particular when old and new generations of servers, storage, and SAN switches are deployed in the same network as existing infrastructures being upgraded. Refer to the *Brocade SAN Design and Best Practices* document for a detailed discussion on how to properly design a SAN.

If oversubscription is detected, the Fabric Notification feature can automatically send notification to congested devices and their peer devices in a same zone. When devices receive the congestion notification, they can perform a number of actions to automatically mitigate oversubscription, for example, by slowing down requests for data for the fabric. Fabric Notification is discussed in detail in Chapter 8 and in Chapter G.

The Brocade Traffic Optimizer feature supported on Gen 7 platforms automatically isolates traffic flows on an E_port into different Performance Groups that are based on the bandwidth of the destination device. Congestion in one Performance Group will not affect traffic in different Performance Groups on the same ISL. Traffic Optimizer isolates oversubscription congestion impact to the congested devices without slowing down other device traffics that are running in the same fabric. The Traffic Optimizer feature is by default automatically enabled on all Gen 7 platforms and FC32-X7-48 blade. Refer to *Brocade Fabric OS Administration Guide* for a detailed description on the Traffic Optimizer feature.

Best-Practice Recommendations:

- Follow the *Brocade SAN Design and Best Practices* to determine the best source-to-target and device-to-ISL ratios.
- Enable MAPS to monitor port bandwidth.
- Enable MAPS FPI rules to monitor port oversubscription.
- Enable C3 timeout monitoring in Brocade MAPS.
- Enable FPIN notification for congestion due to oversubscription.
- Enable Traffic Optimizer to isolate congested devices.

7.2 Credit-Stalled Devices

7.2.1 Description

Another common class of congestion originates from high-latency end devices (host or storage). A high-latency end device is one that does not respond as quickly as expected and therefore causes the fabric to hold frames for excessive periods of time. Device latency is a major source of congestion in today's fabrics due to a device's inability to promptly return buffer credits to the switch. This can result in application performance degradation or, in extreme cases, I/O failure. These high-latency end devices are referred to as credit-stalled devices.

Common examples of credit-stalled devices include disk arrays that are overloaded and hosts that cannot process data as quickly as requested, a performance issue that is more common as hardware ages. This type of performance issue is usually caused by defective host bus adapter (HBA) hardware, incorrect configuration, or defects in the HBA firmware, and problems with HBA drivers. Storage ports can produce the same symptoms due to defective interface hardware or firmware issues. Some arrays, by design, reset their fabric ports if they are not receiving host responses within their specified timeout periods.

Severe latency is caused by devices that stop receiving, accepting, or acknowledging frames for excessive periods of time.

7.2.1.1 Latency Caused by a Credit-Stalled Device

A credit-stalled device that is experiencing latency responds more slowly than expected. Typically, the device stops returning buffer credits (through R_RDY or VC_RDY primitives) to the transmitting switch for tens or hundreds of milliseconds. The device does not respond fast enough to support the offered load, even though the offered load is less than the maximum physical capacity of the link connected to the device. These devices are commonly referred to as slow-drain devices.

Figure 1: Example of Latency Caused by a Credit-Stalled Device

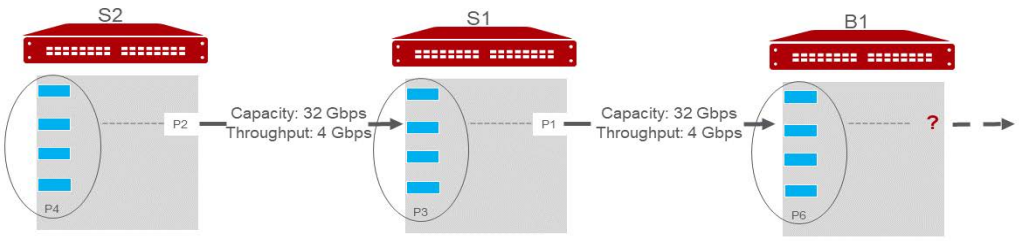
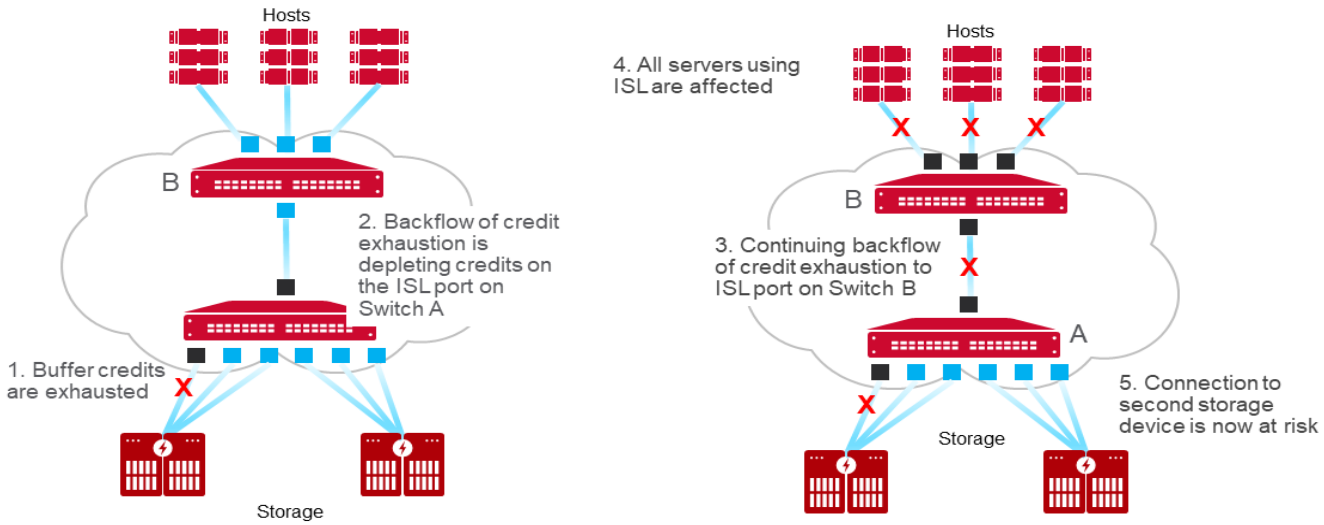


Figure 1 Illustrates the condition where a credit-stalled device on B1 results in buffer backup on ingress port 6 on B1 and causes congestion upstream on S1, port 3. Once all available credits are exhausted, the switch port connected to the device must hold additional outbound frames until a buffer credit is returned by the device. The network underperforms as a result.

When a device does not respond in a timely fashion, the transmitting switch is forced to hold frames for longer periods of time, resulting in high buffer occupancy. This in turn results in the switch lowering the rate at which it returns buffer credits to other transmitting switches. This effect propagates through switches (and potentially multiple switches when devices attempt to send frames to devices that are attached to the switch with the credit-stalled device), and it ultimately affects the fabric.

Figure 2 depicts the process in which latency from a credit-stalled device can propagate through the fabric.

Figure 2: Latency on a Switch Can Propagate through the Fabric



NOTE: The impact to the fabric (and other traffic flows) varies based on the severity of the latency exhibited by the device. The longer the delay caused by the device in returning credits to the switch, the more severe the problem.

7.2.1.2 Moderate Device Latency

Moderate device latency from the fabric perspective is defined as latency that is not severe enough to cause frame loss. If the time between successive credit returns by the device is between a few hundred microseconds to tens of milliseconds, then the device exhibits mild to moderate latency, since this delay is typically not enough to cause frame loss. Moderate device latency does cause a drop in application performance but typically does not cause frame drops or I/O failures.

The effect of moderate device latency on host applications may still be profound, based on the average disk service times expected by the application. Mission-critical applications that expect average disk service times of, for instance, 10 ms, are severely affected by storage latency in excess of the expected service times.

7.2.1.3 Severe Device Latency

Severe device latency results in frame loss, which triggers the host SCSI stack to detect failures and retry I/Os. This process can take tens of seconds (possibly as long as 30 to 60 seconds), which can cause a very noticeable application delay and potentially result in application errors. If the time between successive credit returns by the device is in excess of 80 ms, the device is exhibiting severe latency. When a device exhibits severe latency, the switch is forced to hold frames for excessively long periods of time (on the order of hundreds of milliseconds). When this time becomes greater than the established timeout threshold, the switch drops the frame according to Fibre Channel standards. Frame loss in switches is also known as Class3 (C3) discards or timeouts.

Since the effect of device latency often spreads through the fabric, frames can be dropped due to timeouts. In this case, frames are dropped not just on the F_Port to which the misbehaving device is connected but also on E_Ports that are carrying traffic to the F_Port. Dropped frames typically cause I/O errors that result in a host retry, which can result in significant decreases in application performance. The implications of this behavior are compounded and exacerbated by the fact that frame drops on the affected F_Port (device) result not only in I/O failures to the credit-stalled device (which are expected) but also on E_Ports, which may cause I/O failures for unrelated traffic flows involving other hosts (which typically are not expected).

7.2.1.4 Latency on ISLs

Latency on inter-switch links (ISLs) is usually the result of back pressure from latency elsewhere in the fabric. The cumulative effect of latency on many individual devices can result in slowing down the link. The link itself might be producing latency, if it is a long-distance link with distance delays or if too many flows are using the same ISL. Although each device may not appear to be a problem, the presence of too many flows with some level of latency across a single ISL or trunked ISL may become a problem. Latency on an ISL can ripple through other switches in the fabric and affect unrelated flows.

7.2.2 Detection

Latency on ISLs can affect all flows and multiple switches in the fabric that are connected through the congested ISLs. Hence, it is critical to detect and mitigate the cause of the latency as quickly as possible.

Brocade Fabric OS provides alerts and information indicating possible ISL latency in the fabric through one or more of the following ways:

- Fabric Performance Impact (FPI) alerts (generated by switches in the fabric) the affected E_Ports.
- C3 transmit discards (er_tx_c3_timeout) on the device E_Port or EX_Port carrying the flows to and from the affected F_Port or device.
- Brocade MAPS alerts, if they are configured for C3 timeouts.
- Elevated tim_txcrd_z or transmit queue latency counts on the affected E_Port, which may also indicate congestion.
- C3 receive discards (er_rx_c3_timeout) on E_Ports in the fabric that contains flows of a high-latency F_Port.

Often, ISL latency detection requires investigating the places where the symptoms are detected and then working backward to identify the actual source of the latency. Brocade Fabric Vision technology has made device latency much easier to detect through one of the following ways:

- Fabric Performance Impact (FPI) DEV_LATENCY_IMPACT alerts on the F_port where the credit-stalled device is connected.
- Elevated tim_txcrd_z or transmit queue latency counts on the F_Port where the affected device is connected.
- Potentially elevated tim_txcrd_z or transmit queue latency counts on all E_Ports that are carrying the flows to and from the affected F_Port device.
- Elevated IO latency statistics for the flows reported by IO Insight at the affected device ports.

NOTE: The tim_txcrd_z parameter is defined as the number of times that the port was unable to transmit frames because the transmit buffer-to-buffer credit (BBC) was zero. The purpose of this statistic is to detect congestion or a device that is affected by latency. This parameter is sampled at intervals of 2.5 microseconds, and the counter is incremented if the condition is true. Each sample represents 2.5 microseconds of time with zero Tx BBC. The tim_txcrd_z counts are not an absolute indication of significant congestion or latency and are just one of the factors in determining if real latency or fabric congestion is present. Some level of congestion is to be expected in a large production fabric and is reflected in tx_crd_z counts. The Brocade FPI feature was introduced to remove the uncertainty around identifying congestion in a fabric.

NOTE: Brocade Gen 5 and later platforms have ASIC counter that measures the latency time that a frame incurs in the transmit queue of its corresponding virtual channel (VC). The purpose of transmit queue latency is to directly measure the frame transmit latency of a switch port. The Brocade FPI feature uses this counter to enhance the detection of credit-stalled device and oversubscription.

The Brocade Fabric Performance Impact (FPI) monitoring feature is specifically designed to detect credit-stalled device that can impact performance at the device level or at the fabric level. FPI has replaces the legacy Bottleneck Detection feature, which is deprecated in Fabric OS 8.0. FPI is discussed in detail in Chapter 8 and in Appendix C.

The Brocade Flow Vision feature provides advanced troubleshooting capability that helps administrators isolate device-based latency from a storage fabric and to identify the sources of the problems. As with the issue of credit-stalled devices, the IO Insight capability on Gen 6 and Gen 7 platforms provides visibility of IO latency, which makes it straightforward to detect and isolate the source of latency. Flow Vision and IO Insight are discussed in detail in Chapter 8 and in Appendix D.

Brocade Fabric OS Frame Viewer can be used to detect severely congested flows from C3 discard data. The source ID and destination ID information about the flow can point in the right direction to locate the source of congestion.

7.2.3 Mitigation

Once you detect the congestion due to credit-stalled device, you can identify mitigation steps. Brocade Fabric Vision includes features within a fabric to temporarily mitigate the effect of credit-stalled devices. These features include Port Toggle, Slow Drain Device Quarantine, Port Decommissioning, and Port Fencing. These features are provided as part of FPI and MAPS actions. In particular, Slow Drain Device Quarantine automatically isolates flows destined to a credit-stalled device, which prevents congestion in the healthy path of the fabric.

To completely resolve the effects of a credit-stalled device, actions need to be taken outside of a fabric. Brocade Fabric Notification can send congestion notifications either through hardware or software to credit-stalled devices and their peer devices. These notifications enable the receiving devices to trigger recovery actions programmatically. Therefore, it is important for end devices to register and receive the Fabric Notification for congestion.

Dynamic Path Selection (default switch routing) tries to distribute frames across as many equal-cost paths as possible from the host to the storage. Assuming that all paths are not affected by latency, then providing more paths through the fabric reduces the likelihood that a congested device or link will affect overall traffic flow.

7.2.3.1 Initiators Compared to Targets

Field experience has shown that hosts and storage are about equally responsible for causing severe credit-stall issues. In architectures where hosts and targets are deployed on separate switches, configure the edge hold time (EHT) to a value that is less than the hold time.

Some storage arrays, for example, reset a port if they have not received a credit in 400 ms. Every flow that is connected to the reset port must log back in to that port, causing an interruption on each attached host. Setting an EHT to less than 400 ms on the switches where the hosts are located avoids the resets.

Best-Practice Recommendations:

- Enable Fabric Performance Impact monitoring on all switches to detect credit-stalled device. FPI can also help you determine inter-switch latency.
- Enable the Slow Drain Device Quarantine (SDDQ) or Port Toggle action for FPI monitoring for automatic mitigation of slow drain devices.
- Enable C3 timeout monitoring in Brocade MAPS.
- Enable Fabric Notification for congestion due to credit-stalled device.
- Configure the EHT only on switches with hosts.
- Enable MAPS monitoring of latency metrics on critical workload flows with Gen 6 and Gen 7 IO Insight support.

7.3 Loss of Buffer Credits

7.3.1 Description

Buffer-to-buffer credits (BBCs) form the basis of flow control in a Fibre Channel SAN and determine the number of frames that a port has the buffers to store. Permanent credit loss is very infrequent and is caused when an R_RDY or VC_RDY primitive is not returned (or when a malformed primitive is returned) after a frame is received and processed from one end of a link to the other. If loss of credit happens too often, all the available buffer credits over a link may eventually become depleted, and traffic between the two endpoints ceases. This condition is cleared by issuing a link reset (LR) on the port in question.

Permanent credit loss can be caused by some external condition, such as electrical noise, faulty media, poorly seated blades that can corrupt the primitive, and misbehaving devices not returning R_RDYs. Corrupt primitives are dropped by the receiver as malformed frames. Permanent credit loss can occur on a port (such as an F_Port) or on a VC on ISLs, BE ports, or other links where VCs are supported. Note that traffic still flows as long as there are available credits on a VC or port.

Congestion caused by a credit-stalled device is often mistaken for congestion caused by permanent credit loss, because the initial symptoms are exactly the same for both conditions. The difference is that congestion usually abates, and the buffer credits are eventually freed.

All credit starvation results in effects that appear elsewhere in the fabric due to back pressure from the source of the starvation. The following are some examples of the effects of back pressure:

- I/O timeouts on servers
- SCSI transport errors in server logs that appear to be random and unrelated
- Transmit C3 timeouts on the affected port (and, possibly, receive C3 timeouts on ISLs in other switches)

7.3.1.1 Gen 5 and Later ASIC Enhancements

Starting with the fifth generation, Brocade FC switching ASICs (Condor3 and later) have the capability to automatically detect and recover a single permanently lost credit at the VC level, without the need for an LR, provided that both ends of the link are Condor3 or later ports. Multiple lost credits are recovered automatically by an LR with appropriate user notifications and alerts.

Within the boundary of a switch, there is a potential (although it is an extremely low probability) for credit loss between the BE links, the internal links between two ASICs, on the same switch. This could be due to improperly seated blades or other conditions. The Condor3 and later ASICs have built-in features that further minimize the credit loss on the BE ports. As soon as a port blade is plugged into a chassis and a BE link is brought online, Condor3 and later ASIC ports automatically tune the links for optimal performance. This standards-based automatic tuning mitigates the effect of electrical noise due to incorrectly seated blades or other environmental factors. This capability is automatically enabled on the Condor3 and later based BE links, and it recovers from any loss of buffer credits.

In addition to the autotuning of BE links, Condor3 and later ports are capable of Forward Error Correction (FEC). FEC is optional on Gen 5 platforms but enhances transmission reliability and performance. FEC on Gen 5 platforms provides the ability to correct up to 11 bit errors in every 2112-bit transmission in a 10Gb/s or 16Gb/s data stream in both frames and primitives. FEC is enabled by default on the BE links of Condor3 and later based switches and blades, and it minimizes the loss of credits on BE links. FEC is also enabled by default on FE links that are connected to another FEC-capable device. FEC on Gen 6 platforms uses a more robust coding algorithm that corrects up to seven 10-bit streams and detects up to fourteen 10-bit streams, without requiring that the errors are in a burst. FEC is mandatory on Gen 6 and Gen 7 platforms for 32Gb/s or 64Gb/s speed to ensure that the bit error rate stays within the standard requirement. The Condor4 and later ASIC automatically turns on FEC when a port operates at a speed of 32Gb/s or higher, and FEC cannot be disabled.

7.3.2 Detection

Brocade Fabric OS has evolved through seven generations of Fibre Channel speed transitions and has provided progressive capabilities to monitor traffic flow through the switches. On FE device ports, link credit loss is monitored every time that traffic stops flowing for more than two seconds. Credit loss detection is the means that you use to determine if the credit starvation is due to congestion or actual permanent credit loss. The number of available credits returns to the maximum number of credits that are assigned to the port or VC when credit starvation is due to congestion. Permanent loss of one or more credits results in a value lower than the assigned maximum. Credit recovery is supported on FE E_ports, F_ports, and EX_ports. On the BE links, Brocade Fabric OS provides mechanisms to detect and mitigate the loss of credits.

Lost credit situations result in a variety of symptoms, depending on the Brocade Fabric OS release version and FC generation of the switch, and whether BE credit recovery is activated. The set of RASLog messages that may be generated in the event of lost credits is as follows:

- CX-1012
- CX-1014
- CX-1015
- CX-1016
- CX-1017
- CX-1018

Total credit loss results in flows being blocked across the port or VC in question. Hosts and/or targets are affected, and C3 receive timeouts are observed on other switches in the fabric.

The external conditions that cause permanent loss of credit typically result in link errors such as CRC and ITW. Similar to the detection of faulty media issues, you can use Brocade MAPS to monitor and detect these errors on FE and BE ports.

7.3.3 Mitigation

Brocade Fabric OS cannot control the conditions leading to permanent loss of credit on FE ports. Permanent loss of a single credit is automatically recovered by ASIC with the deployment of Brocade Gen 5 and later platforms on both ends of a link. Permanent loss of all credits on a port (which is very rare) can be handled through either a manual or automatic LR on the port. Automatic credit recovery on the FE ports is enabled by default. It can also be enabled by the `portCfgCreditRecovery` command.

7.3.3.1 Credit Recovery on Back-End Ports

Use the `creditRecovMode` commands to enable or disable credit recovery of back-end ports, and use the `--show` parameter to display the configuration. When this feature is enabled, credit is recovered on back-end ports (ports connected to the core blade or core blade back-end ports) when credit loss is detected on these ports. If complete loss of credit on a back-end port causes frame timeouts, an LR is performed on that port regardless of the configured setting, even if that setting is `-recover off`.

When used with the `-recover onLrOnly` option, the recovery mechanism takes the following escalating actions:

- When the mechanism detects credit loss, it performs an LR and logs a RASLog message (CX-1014).
- If the LR fails to recover the port, the port re-initializes. A RASLog message is generated (CX-1015). Note that the port re-initialization does not fault the blade.
- If the port fails to re-initialize, the port is faulted. A RASLog message (CX-1016) is generated.
- If a port is faulted and there are no more online back-end ports in the trunk, either the port blade or the core blade is faulted, depending on the `--fault` configuration option with the `creditRecovMode` command. A RASLog message (RAS CX-1017) is generated.

When used with the `-recover onLrThresh` option, recovery is attempted through repeated LRs, and a count of the LRs is kept. If the threshold of more than two LRs per hour is reached, the blade is faulted (RAS CX-1018). Note that regardless of whether the LR occurs on the port blade or on the core blade, the port blade is always faulted.

Permanent loss of credit is caused by external conditions such as faulty media or other network components. Hence, regular cable and SFP maintenance help to prevent loss of credit. Furthermore, Brocade ClearLink Diagnostics can be used to troubleshoot and identify faulty media.

Best-Practice Recommendations:

- Enable Automatic Credit Loss Detection and Recovery for lost credits on FE and BE ports.
- Always have at least two member trunks (using Brocade Trunking) on FE links, where possible. This eliminates the potential for stopped traffic until all credits on all trunk members for the VC or port are lost (which is a very rare).
- Enable MAPS to monitor link errors (CRC and ITW errors).

Chapter 8: Tools

8.1 ClearLink Diagnostics

Brocade ClearLink Diagnostics is a Fabric Vision feature on Brocade Gen 5 and later platforms that simplifies the troubleshooting of faulty or marginal media problems. Enabling ClearLink Diagnostics between two Brocade devices does not require a license. Enabling ClearLink Diagnostics between a Brocade device and a third-party device requires a Fabric Vision license.

Refer to Appendix A for more details and the evolution of capabilities. Also, refer to the *Brocade Fabric OS Administration Guide* for complete details on the use of Brocade ClearLink Diagnostics.

8.2 Brocade MAPS

Brocade MAPS is an optional (licensed) feature that monitors various Brocade Fabric OS metrics, statistics, and switch component states. MAPS has predefined thresholds on all counter values and component states, and predefined alerts and actions can be generated when thresholds are exceeded.

Refer to Appendix B for more details and the evolution of MAPS capabilities. Also, refer to the *Brocade Fabric OS Monitoring and Alerting Policy Suite Configuration Guide* for complete details on the use of Brocade MAPS.

8.3 Fabric Performance Impact Monitoring

Fabric Performance Impact (FPI) monitoring is a MAPS feature that monitors and sends alerts about congestions that impact the performance of a SAN. When FPI was introduced in FOS 7.3, it only monitored congestion that was caused by credit-stalled devices. In FOS v9.0, FPI was extended to also monitor congestion that is caused by oversubscription. FPI also added automatic mitigation capabilities through Slow Drain Device Quarantine (SDDQ) and Port Toggle actions.

FPI monitoring is supported on Gen 5 and later platforms with Fabric OS 8.0 without requiring a license. The SDDQ and Port Toggle actions require a Fabric Vision license.

Refer to Appendix C for more details on the evolution of FPI capabilities and recommended best-practice settings.

8.4 Flow Vision and IO Insight

Brocade Flow Vision is an optional (licensed) feature that provides comprehensive and in-depth diagnostics tools for performance-related issues. IO Insight is a Brocade Gen 6 and Gen 7 built-in capability that provides deeper visibility on application and device IO performance and latency.

Refer to Appendix D for an overview of Flow Vision and IO Insight capabilities as well as recommended practices.

8.5 Edge Hold Time

EHT allows an overriding value for hold time (HT) that will be applied to individual F_Ports on Gen 5 and later Fibre Channel platforms. The default HT setting is 500 ms.

In some environments, the highest latency is produced by initiators and not targets. When a host is not responding, back pressure can build up in the fabric and affect many flows between different hosts and their storage ports. Edge hold time can relieve this pressure by dropping frames sooner than they are dropped when using the default hold time setting, therefore allowing frames to flow again.

For complete details and recommendations on setting EHT, refer to Appendix F.

8.6 Frame Viewer

Frame Viewer was introduced in Brocade Fabric OS to allow the fabric administrator more visibility into Class 3 frames that are dropped due to timeouts. Recent Fabric OS versions enhanced Frame Viewer by capturing frames that are dropped due to other conditions.

When frame drops are observed on a switch, the user can use this feature to determine which flows the dropped frames belong to and potentially to determine the affected applications by identifying the endpoints of the dropped frame.

Frames discarded due to timeouts are sent to the CPU for processing. Brocade Fabric OS captures and logs information about the frame, such as source ID (SID), destination ID (DID), and transmit port number. This information is maintained for a limited number of frames.

The end user can use the CLI to retrieve and display this information.

NOTE: Frame Viewer captures only FC frames that are dropped in a receive (Rx) buffer due to a timeout, because the frames are unroutable, or because an unreachable destination was received on an edge ASIC (an ASIC with FE ports). If the frame is dropped for any other reason, it is not captured by Frame Viewer. If the frame is dropped because of a timeout on an Rx buffer on a core ASIC, the frame is not captured by Frame Viewer. A timeout is defined as a frame that lives in an Rx buffer for longer than the hold time default of 500 ms or longer than the edge hold time value custom setting. See Appendix E for an example of a Frame Viewer session.

8.7 Fabric Notification

Fabric Notification is introduced in Brocade Fabric OS v9.0 for a fabric to automatically send hardware or software notifications to SAN devices to signal various conditions that cause congestion. Upon receiving the notification, devices can run appropriate recovery and mitigation actions automatically. Fabric Notification provides a SAN the self-healing capability to increase resiliency. For complete details and recommendations on Fabric Notification, refer to Appendix G.

Chapter 9: Designing Resiliency into the Fabric

This topic is handled in detail in a companion document entitled *Brocade SAN Design and Best Practices*. What follows here is a brief introduction to some of the issues that affect congestion in SAN fabrics.

Modern fabrics have very different requirements than fabrics that existed when Fibre Channel first appeared:

- Special node behaviors, such as server, workload, and storage virtualization, are widely adopted. These applications can blur the difference between initiator and target behavior and can produce large volumes of very short frames. Heartbeat and other control devices may use in-band Fibre Channel and potentially disrupt normal traffic flow.
- Modern fabrics are typically much larger, providing more potential for congestion, credit-stalled devices, media faults, and lost buffer credits.
- Higher-capacity storage arrays allow for much higher initiator fan-in to storage ports than previously seen.
- Many more fabrics are routed (using Layer 3 routing).

9.1 Factors Affecting Congestion

Some node behaviors merit particular attention when introduced into a fabric:

- Storage virtualization – Virtualized storage presents virtualized logical unit numbers (LUNs) to initiators. Usually these devices are composed of some sort of FE controller and BE storage. Virtualized storage tends to increase traffic across ISLs because frames must first be transferred to and from the controller and its BE storage, as well as to and from the initiator. The proper configuration of the controllers and the placement of BE storage are critical to ensure that fabrics are not congested. The controllers are usually in some form of cluster and may communicate through in-band Fibre Channel to maintain storage integrity. This can result in high volumes of short frames that can consume precious buffer credits.
- Workload virtualization – A virtualized workload can present its own challenges to fabric traffic throughput. Migrating virtual machines can increase the overhead for the fabric and storage array because it causes the issuing of high volumes of SCSI reserves, which can create latency that is difficult to trace to the root cause. Many of these products run on virtual machines that are controlled by a hypervisor, which can obscure initiator flow and latency information.
- Clustering – Clustering solutions often impose higher I/O and synchronization requirements on a fabric beyond the volumes typically seen in standalone platforms. Clustering can result in more short frames traversing the fabric when the storage status is being continuously queried.

Congestion from applications is best addressed at the source itself. The fabric cannot completely compensate for node behavior.

9.2 Resiliency

Fabric resiliency is usually threatened by credit-stalled devices or other factors external to the fabric. Many features have been added to Brocade Fabric OS to improve resiliency, but not all failing node conditions can be detected or handled transparently in the fabric.

Redundancy is a very effective means of increasing resiliency in any SAN. The addition of fabric components must, of course, be weighed against the cost of doing so. Those costs should also be weighed against the opportunity cost of losing access to mission-critical applications.

9.3 Redundancy

Fabrics – Redundancy provides a complete failover to a redundant fabric. This requires that all multipathing software on all hosts is in perfect working order and detects the failure.

Cores – Redundancy allows for higher resiliency because only the hosts attached to failing edge switches are required to fail over to the redundant fabric.

Edges – Redundancy can be important in routed environments.

Backbones – Redundancy is particularly important when distance is involved. ISLs – More paths allow for more alternatives for frames to traverse the fabric.

Best-Practice Recommendations:

- Duplicate the core switches in core-edge and edge-core-edge designs. This vastly improves resilience in the fabric and avoids host failovers to alternate fabrics in case a core platform ever fails. Experience has shown that host failover software sometimes does not function as expected, causing outages for applications that were expected to participate in a host failover to a second fabric.
- Duplicate the Fibre Channel Router (FCR) backbone switches to protect against host failover failures. Often the costs associated with a failover failure greatly exceed the cost of the second FCR platform.
- Provide as many different paths through the fabric as possible. This is particularly true for routed fabrics since these are prime points of congestion.

Chapter 10: Fabric Configuration

10.1 Fabric-Wide Parameters

The `configShow` CLI command can be used to list fabric-wide configuration parameters on each switch. Do not change any of these parameters unless you are directed by a Brocade support representative, because misconfiguration can lead to fabric instability and major fabric disruptions.

10.2 Event and Change Log Level Settings

Fabric OS system message types include RASLog messages, audit log messages, and First Failure Data Capture (FFDC) messages. RASLog messages record significant system events and high-level user-initiated actions to nonvolatile memory. Gen 6 platforms can save 8192 RASLog messages. Audit log messages support post-event auditing. FFDC messages capture failure-specific data for debugging and analyzing problems. Debugging level settings can be changed on more than 150 modules for in-depth troubleshooting and diagnostics. Customers should leave the setting at the factory default unless advised by Brocade support personnel. The data collected by the SupportSave process is sufficient for initial diagnosis.

For environments where an audit trail is necessary, configuration changes can be sent to an external syslog server. These configuration changes can include: login failures, zone configuration changes, firmware downloads, changes to security settings and any other critical changes that have a serious effect on the operation and security of the switch.

Auditable events are generated by the switch and streamed to an external host through a configured system message log daemon (syslog), and only the last 512 events are persistently stored on the switch. Generated audit events are specific to the particular switch and have no negative impact on its performance. In case the audit log is too verbose and too many events are generated by the switch, the remote host's system message log may become a bottleneck, and audit events can be dropped by the switch.

Audit logging can be configured and displayed using the `auditcfg` command. This command allows you to set filters by configuring certain classes, to add or remove any of the classes in the filter list, to set severity levels for audit messages, to enable or disable audit filters, and enable or disable a specific message ID. Based on the configuration, certain classes are logged to syslog for auditing. A syslog configuration is required for logging audit messages. Use the `syslogAdmin` command to add the syslog server host name or IP address.

10.3 Zoning

Zoning is a fabric-based service that enables you to partition your SAN into logical groups of devices that can access each other. Zones provide controlled access to fabric segments and establish barriers between operating environments. A device in a zone can communicate only with other devices that are connected to the fabric within the same zone. A device that is not included in the zone is not available to members of that zone. When zoning is enabled, devices that are not included in any zone configuration are inaccessible to all other devices in the fabric.

Two types of device specification in zoning are supported: WWN zoning and port zoning. Registered state change notification (RSCN) messages are limited to the zone in which they occurred.

- WWN zoning – WWN zoning permits connectivity between attached nodes based on WWN. The attached node can be moved anywhere in the fabric and remains in the same zone. WWN zoning is used in open systems environments but does not make sense for FICON channels and FICON control units.

- Port zoning – Port zoning limits port connectivity based on port number—in other words, all devices connected to all ports that are members of the zone can talk to each other. Port zoning is used in FICON configuration. Ports can easily be added to port zones, even if there is nothing attached to the port. Port zoning is specified in the (Domain, Index) format.
- Mixed zoning – A mix of WWN zoning and port zoning is also supported. A zone can be configured to contain members specified by a combination of ports or port aliases and WWNs or WWN aliases.

Zone configuration is managed on a fabric basis. When a new switch is added to the fabric, it automatically takes on the zone configuration information from the fabric. Adding a new fabric that has no zone configuration information to an existing fabric is very similar to adding a new switch. All switches in the new fabric inherit the zone configuration data. If the existing fabric has an effective zone configuration, then the same configuration becomes the effective configuration for the new switches. If a new switch that is already configured for zoning is being added to the fabric, the zone configuration should be cleared on that switch before connecting it to the zoned fabric. When a change in the configuration is saved, enabled, or disabled according to the transaction model, it is automatically distributed to all switches in the fabric (by closing the transaction), preventing a single point of failure for zone information.

Zone changes in a production fabric can result in a disruption of I/O under conditions when an RSCN is issued because of the zone change and, as a result, the HBA is unable to process the RSCN fast enough. Although RSCNs are a normal part of a functioning SAN, the pause in I/O might not be acceptable. For these reasons, you should perform zone changes only when the resulting behavior is predictable and acceptable. Ensuring that the HBA drivers are current can shorten the response time in relation to the RSCN.

Peer zoning is a new type of zoning supported since FOS v8.1. Peer zoning simplifies single-initiator/single-target zoning functionality without the need to define the zones individually. Peer zoning allows one or more *principal* devices to communicate with the rest of the devices (*nonprincipal* devices) in the zone as if they were a single-initiator zone. Nonprincipal devices in the zone can communicate with the principal devices only, but they cannot communicate with each other—nor can principal devices communicate with each other. This approach establishes zoning connections that are effectuated as single-initiator zoning with the operational simplicity of one-to-many zoning and reduced zone database usage.

A peer zone can have one or multiple principals. In general, storage ports are assigned as principals. Multiple principal members in a peer zone are used when all the nonprincipals (initiators) in the zone are to share the same target (storage) ports.

The peer zone members can be defined as WWNs and aliases specifying WWNs or “domain, port” and aliases specifying “domain, port” (“domain, index”), but you cannot mix WWNs and “domain, port” or corresponding aliases when defining peer zoning.

Peer zones can also be created automatically by storage arrays through FC in-band commands. The peer zones created by target devices are called target-driven peer zones.

10.4 Advanced Zoning Considerations

Brocade Fabric OS supports the following types of zones for advanced functionality:

- LSAN zones – LSAN zones provide device connectivity between fabrics without merging the fabrics. A logical SAN (LSAN) consists of zones in two or more edge or backbone fabrics that contain the same devices. LSANs essentially provide selective device connectivity between fabrics without forcing you to merge those fabrics. FC routers provide multiple mechanisms to manage inter-fabric device connectivity through extensions to existing switch management interfaces. If you want to share devices between any two fabrics, you must create an LSAN zone in both fabrics that contain the port WWNs of the devices to be shared.

- QoS zones – A Quality of Service (QoS) zone is a special zone that indicates the priority of the traffic flow between a given host/target pair. The members of a QoS zone are the host/target pairs. The switch automatically sets the priority for the host/target pairs specified in the zones based on the priority level (H or L) in the zone name. QoS zones are regular zones with additional QoS attributes specified by adding a QoS prefix to the zone name. WWN and port QoS zones are supported in Fabric OS.

10.5 Zoning Recommendations

- Use peer zones with a single initiator as a principle member and multiple targets as non-principle members, or with multiple initiators as non-principle members and a single target as a principle member. Peer zones that follow this convention achieve the same effect as single initiator and single target using standard zone sets. A peer zone's ability to restrict communication to between only a principle member and a non-principle member reduces the complexity of managing hundreds of zones in a large fabric.
- Define zones using device WWPNs (World Wide Port Names).
- Monitor the zone database size using the `cfgSize CLI` command. The maximum supported zone database size is 4MB in Brocade Fabric OS v9.0. If the fabric includes a switch with an earlier Fabric OS version, the maximum supported size of a zone database is 1 MB or 2 MB depending on the switch platform type.
- Periodically back up the zone database and remove database entries that are no longer in the fabric.
- The default zone setting (what happens when zoning is disabled) should be set to No Access, which means that devices will be isolated when zoning is disabled.
- Always zone using the highest Fabric OS level switch. Switches with earlier Fabric OS versions do not have the capability to view all the functionality that a newer version of Fabric OS provides, since functionality is backward-compatible but not forward-compatible. Fabric OS v9.0 and later supports Zoning Fabric Lock that automatically protects against concurrent transactions fabric wide.
- Zone using an enterprise-class platform rather than a switch. An enterprise-class platform has more resources to handle zoning changes and implementations.
- Follow vendor guidelines for preventing the generation of duplicate WWNs in a virtual environment.
- Zoning changes affect the entire fabric. Thus, when you are performing fabric-level configuration tasks, allow time for the changes to propagate across the fabric before issuing subsequent commands. For a large fabric, you should wait several minutes between commands.

10.6 Firmware Management

Broadcom offers customers a choice in selecting Brocade Fabric OS releases with new features or versions with extensive field deployment. Before upgrading, check the release notes of the selected Brocade Fabric OS release to see if all the switches in your fabric are supported. Review the following questions before determining which Brocade Fabric OS release to use:

- Why am I upgrading?
- Is the Fabric OS release part of my server/SAN/storage firmware upgrade to keep current support?
- Do I need to upgrade to address a technical support bulletin or bug fix that I encountered?
- Do I need the new features to monitor some fabric anomalies?
- Do any new switches in the fabric require the latest firmware?

Switches can be upgraded sequentially or in parallel using Brocade Network Advisor or custom scripts. If you have a single resilient or dual redundant fabric, you can upgrade in parallel if application uptime is critical and can be assured.

10.7 Firmware Recommendations

- Upgrade firmware during nonbusiness or peak hours if possible.
- If new features are not required, upgrade using the same release or minor release train.
- Do not perform a concurrent firmware upgrade on two switches that are physically connected by E_Ports.
- When hosts and storage are connected by two redundant fabrics, upgrade the firmware one fabric at a time. Never upgrade firmware on both fabrics at the same time.
- For a FICON environment, install firmware sequentially on switches in the FICON fabric. The control unit port (CUP) should be varied offline before a FICON switch firmware upgrade.
- Determine the support status of Fabric OS versions according to the software Software Release Support and Posting Matrices. Only use the supported Fabric OS versions specified in the matrices.

NOTE: Brocade Fabric OS Target Path releases are recommended code levels for Brocade Fibre Channel switch platforms in both Open Systems and FICON environments. Use these Target Path recommendations to determine the ideal version of Brocade Fabric OS software, and consider them in conjunction with other requirements that may be unique to a particular environment. Refer to the Target Path section in the *Brocade Software Release Support and Posting Matrices* for the recommended Brocade Fabric OS version appropriate for the environment.

Chapter 11: Routing Policies

Data moves through a fabric from switch to switch and from storage device to server along one or more paths that make up a route. Routing policies determine the path for each frame of data. Before the fabric can begin routing traffic, it must discover the route that the frame should take to reach the intended destination. The routing policy configured on the switch determines the route selection, based on one of three user-selected routing policies:

- Port-based routing
- Device-based routing
- Exchange-based routing

Each switch can have its own routing policy, and different policies can exist in the same fabric.

NOTE: Setting a routing policy is a disruptive process. Use the Advanced Performance Tuning (APT) command `aptPolicy` to configure the desired routing policy. For most configurations, the default routing policy is optimal and provides the best performance. The routing policy should be changed only if there is a performance issue that is of concern or if a particular fabric configuration or application requires it.

11.1 Port-Based Routing

In port-based routing, the choice of routing path is based only on the incoming port and the destination domain. Thus, all frames destined to a particular domain that ingress on a particular port follow the same route, as long as Dynamic Load Sharing (DLS) is not enabled. This routing policy minimizes disruption caused by changes in the fabric (events not directly impacting the ports in the route); but it represents a less efficient use of available bandwidth. To optimize port-based routing, DLS can be enabled to balance the load across the available output ports within a domain.

Port-based routing supports specific use cases:

- With devices that do not tolerate out-of-order exchanges
- In FICON environments when Lossless DLS is not enabled

11.2 Device-Based Routing

In device-based routing, the choice of routing path is based on the incoming port, the source ID (SID), and the destination ID (DID). Path utilization is optimized by allowing I/O traffic between different source-device and destination-device pairs to use different paths. Device- based routing is another form of Dynamic Path Selection.

Device-based routing supports specific use cases:

- In FICON environments when lossless Dynamic Load Sharing is enabled

11.3 Exchange-Based Routing

In exchange-based routing, the choice of routing path is based on the source ID (SID), destination ID (DID), and Fibre Channel originator exchange ID (OXID), optimizing path utilization for the best performance. Thus, every exchange can take a different path through the fabric. Exchange-based routing requires the use of the DLS feature.

Exchange-based routing is also known as Dynamic Path Selection (DPS). DPS is where exchanges or communication between end devices in a fabric are assigned to egress ports in ratios proportional to the potential bandwidth of the ISL or trunk group. When there are multiple paths to a destination, the input traffic is distributed across the different paths in proportion to the bandwidth available on each of the paths. This distribution improves utilization of the available paths, thus reducing possible congestion on the paths. Every time there is a change in the network (which changes the available paths), the input traffic can be redistributed across the available paths. This is a nondisruptive process when the exchange-based routing policy is engaged.

The trunking feature allows a group of physical links to merge into a single logical link, called a trunk group. Traffic is distributed dynamically and in order over this trunk group, achieving greater performance with fewer links, thus optimizing the use of bandwidth. Within the trunk group, multiple physical ports appear as a single port, thus simplifying management. Refer to the *Brocade Fabric OS Administration Guide* for details on trunking.

11.4 Dynamic Load Sharing

With Dynamic Load Sharing (DLS) enabled, Brocade Fabric OS balances ingress ports as evenly as possible across available ISL or trunk links. The exchange-based routing policy depends on the Brocade Fabric OS DLS feature for dynamic routing path selection. When you are using the exchange-based routing policy, DLS is enabled by default and cannot be disabled. When the port-based policy is in force, DLS can be enabled to optimize routing. When DLS is enabled, it shares traffic among multiple equivalent paths between switches.

DLS recomputes load sharing when any of the following events occur:

- A switch boots up.
- An E_Port goes offline and online.
- An EX_Port goes offline.
- A device goes offline.
- There is a zone change in the fabric.

DLS can be configured using the `dlsSet`, `dlsReset`, and `dlsShow` commands from the Fabric OS CLI and Web Tools.

11.5 Lossless Dynamic Load Sharing

Lossless DLS enables DLS for optimal utilization of the ISLs without causing any frame loss. In other words, lossless DLS enables rebalancing port paths without causing input/output (I/O) failures. Lossless mode ensures no frame loss during a rebalance and takes effect only if DLS is enabled. Note that zero frame loss can be guaranteed only when a new additional path is used to do load rebalancing; it cannot be guaranteed on an existing data path that encounters the failure.

Lossless DLS can be enabled on a fabric topology in order to have zero frame drops during rebalance operations. If the end device also requires the order of frames to be maintained during the rebalance operation, then In-Order Delivery (IOD) must be enabled. However, this combination of lossless DLS and IOD is supported only in specific topologies, such as in a FICON environment.

Lossless DLS can be configured using the `dlsSet`, `dlsReset`, and `dlsShow` commands from the Fabric OS CLI and Web Tools.

11.6 In-Order Delivery (IOD)

In a stable fabric, frames are always delivered in order, even when the traffic between switches is shared among multiple paths. However, when topology changes occur in the fabric (for example, if an ISL goes down), traffic is rerouted around the failure, and some frames can be delivered out of order. Most destination devices tolerate frames that are delivered out of order, but some do not.

By default, out-of-order frame-based delivery is allowed to minimize the number of frames dropped. Enabling IOD guarantees that frames are either delivered in order or dropped. You should enforce in-order frame delivery across topology changes if the fabric contains destination devices that cannot tolerate occasional out-of-order frame delivery.

The order of delivery of frames is maintained within a switch and is determined by the routing policy in effect. The frame delivery behavior for each routing policy is as follows:

- Port-based routing – All frames that are received on an incoming port and destined for a destination domain are guaranteed to exit the switch in the same order in which they were received.
- Exchange-based routing – All frames that are received on an incoming port for a given exchange are guaranteed to exit the switch in the same order in which they were received. Because different paths are chosen for different exchanges, this policy does not maintain the order of frames across exchanges.
- Device-based routing – All frames that are received on an incoming port for a given pair of source and destination devices are guaranteed to exit the switch in the same order in which they were received.

In-Order Delivery can be configured using the `iodSet`, `iodReset`, and `iodShow` commands.

NOTE: Some devices do not tolerate out-of-order exchanges; in such cases, use the port-based routing policy. Refer to the *Brocade Fabric OS Administration Guide* for more details on routing policies.

Chapter 12: Summary and Recommendations

There are several common classes of abnormal behavior that originate from fabric components or attached devices:

- Faulty media (fiber-optic cables and SFPs/optics) – Faulty media can cause frame loss due to excessive CRC errors, invalid transmission words, and other conditions. This may result in I/O failure and application performance degradation.
- Credit-stalled devices – Credit-stalled devices (host or storage) do not return buffer credits as quickly as expected. This device latency causes the fabric to hold frames for excessive periods of time. Device latency is the major source of congestion in fabrics and results in application performance degradation or, in extreme cases, I/O failure.
- Oversubscription – Improperly designed oversubscription of source ports to target ports and devices to ISLs can cause instances of insufficient link capacity, which leads to congestion.
- Permanent or temporary loss of buffer credits – Buffer credit loss is caused by the other end of a link failing to acknowledge a request to transfer a frame because no buffers are available to receive the frame.

You can use the following features and capabilities to improve the overall resiliency of FC fabric environments based on Brocade Fabric OS:

- Enable Brocade MAPS to detect frame timeouts.
- Enable Port Fencing for transmit timeouts on F_Ports.
- Enable the Edge Hold Time feature.
- Enable Brocade MAPS to monitor (alert) for CRC errors, invalid words, and state changes and to decommission or to fence on extreme behavior.
- Enable Fabric Performance Impact monitoring for congestion conditions.
- Enable Forward Error Correction on 10Gb/s and 16Gb/s connections if both ends of the link support FEC.
- Run ClearLink Diagnostics before deploying a new port into production.

Best-Practice Recommendations:

- Enable Fabric Performance Impact monitoring on all switches to detect congestion.
- Enable Automatic Credit Loss Detection and Recovery on platforms earlier than Gen 5 and Gen 6.
- Enable Forward Error Correction on 10Gb/s and 16Gb/s connections if both ends of the link support FEC.
- Monitor C3 timeouts with Brocade MAPS.
- Deploy EHT on switches that connect to servers only.
- Always have at least two member trunks on FE links where possible. This eliminates the potential for stopped traffic until all credits on all trunk members for the VC or port are lost (which is a very rare situation).
- Duplicate the core switches in core-edge and edge-core-edge designs. This vastly improves resilience in the fabric and avoids host failovers to alternate fabrics if a core platform ever fails. Experience has shown that host failover software sometimes does not function as expected, causing outages to applications that were expected to participate in a host failover to a second fabric.
- Duplicate the FCR backbone switches to protect against host failover failures. Often the cost associated with a failover failure greatly exceed the cost of the second FCR platform.
- Provide as many different paths through the fabric as possible. This is particularly true for routed fabrics, since these are prime points of congestion.

Appendix A: ClearLink Diagnostics

A.1 Introduction to ClearLink Diagnostics

ClearLink Diagnostics is a suite of tests introduced in Fabric OS 7.0 on Brocade Gen 5 platforms to validate deployment and isolate faults that are associated with cables and optics that connect a network. Brocade Gen 5 and later platforms and optics are required for ClearLink Diagnostics. ClearLink Diagnostics is supported between Brocade switches, between a Brocade switch and an Access Gateway, and between HBAs and a Brocade switch. ClearLink Diagnostics between Brocade devices is supported in the base Fabric OS without an additional software license. ClearLink Diagnostics between an HBA and a Brocade device requires a Fabric Vision license. The following section describes the fault isolation process used with Brocade ClearLink technology.

A.2 Background

When a Fibre Channel link fails to come up, isolating the broken component is fairly straightforward. A light meter can be used to determine where the connection is broken, and the appropriate corrective action can be quickly applied. However, if the connection is not fully restored by this corrective action, that can cause issues. Marginal connections can result in errors that severely affect the operation of the link and effectively render it nonoperational.

Brocade ClearLink Diagnostics is easy to use and effective in identifying issues with marginal links. Using ClearLink Diagnostics significantly reduces the time needed to determine which component is faulty when debugging a Fibre Channel link. The process described below, when coupled with ClearLink Diagnostics, minimizes the time and effort needed to isolate a failed component.

ClearLink is intended to diagnose marginal errors on links that can successfully complete link initialization. Phases are arranged to exercise the electrical connection, then the optical connection, and lastly the link capacity. The components exercised during the Diagnostics Port (D_Port) tests are: SFP seating (connection between the SFP and the port); fiber seating/contamination or damaged fiber cables (connection between the SFP and the fiber); and faulty components (cable, SFP, or port).

A.3 Brocade Solutions

Brocade ClearLink Diagnostics functionality helps you to isolate errors that result from marginal operation or marginal components. The various diagnostics included in the ClearLink suite reduce the isolation process to the most likely cause, as described in the following section (see Figure 3 for reference points):

Electrical Loopback – The electrical loopback phase of the diagnostic suite isolates issues associated with the seating of the SFP (1) into the port receptacle.

Electrical Loopback – The electrical loopback phase of the diagnostic suite isolates issues that are associated with the seating of the SFP (1) into the port receptacle.

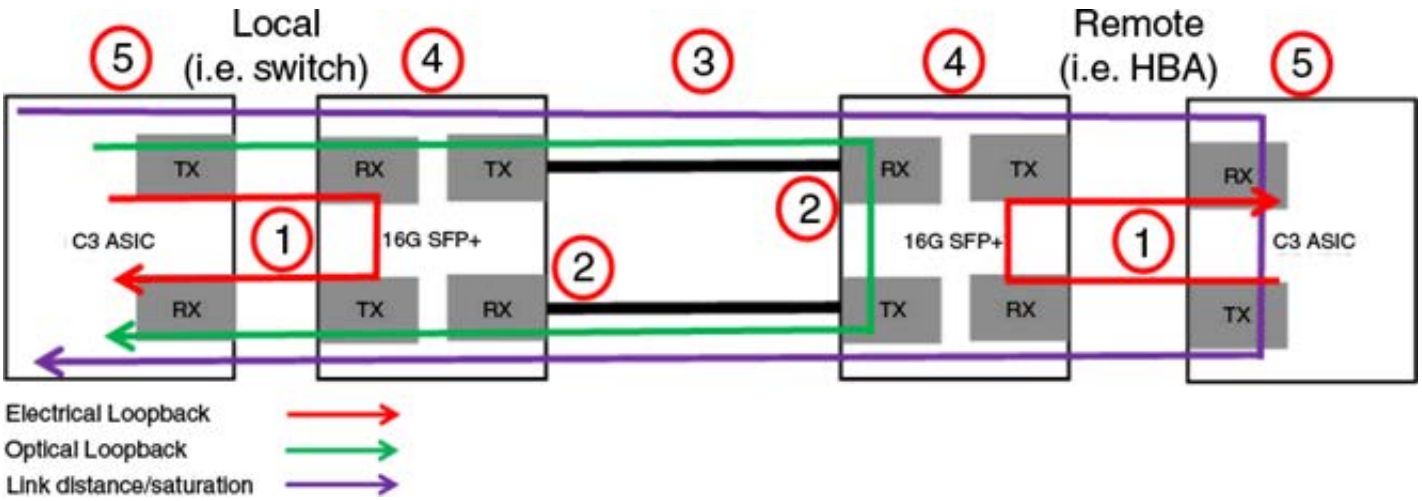
Optical Loopback – The optical loopback phase of the diagnostic suite isolates issues that are associated with contamination or seating of the fiber optic cable (2) into the SFP.

Link Saturation – The link saturation phase of the diagnostic suite isolates issues that are associated with marginal component performance (3-cable, 4-SFP, and 5-port).

Link Latency, Distance, and Power Loss Measurement – At the end of the link saturation phase of the diagnostic suite, ClearLink Diagnostics reports the link latency, link distance, and power loss (dB) measurements. These measurements provide input for administrators to properly design cable layout and configure buffer-to-buffer credits.

The following figure shows the isolation steps associated with the results of each phase of the ClearLink Diagnostics suite. The section that follows this image discusses how to use these steps to identify and isolate the marginal elements of the link.

Figure 3: ClearLink Diagnostics Test Procedures



A.3.1 Electrical Loopback

The Electrical Loopback diagnostic of the ClearLink suite tests the electrical characteristics of the connection between the SFP (number 1 in Figure 3) and the port. Generally, this test identifies seating issues that can be resolved by simply removing the SFP and reseating it in the optical port to establish a proper connection between the SFP and the port. In rare instances, this test may identify a faulty component (SFP or port), which can be determined by following the procedure listed below.

A.3.1.1 Observation

During the execution of the Electrical Loopback diagnostic, both sides of the link place the local SFP in electrical loopback mode. The test then initiates a stream of diagnostic frames that exercises the electrical connection between the SFP and the port. If either Electrical Loopback test fails, the most likely cause is a poor connection between the SFP and the port.

A.3.1.2 Action

The corrective action is to reseat the SFP at the end of the link that reports the Electrical Loopback diagnostic failure and rerun the ClearLink diagnostic suite. If the error persists on the same side of the link as the previous failure, replace the SFP and rerun the ClearLink diagnostic suite. Continued failure of the Electrical Loopback diagnostic indicates that the port is probably marginal and needs to be replaced. Schedule the appropriate service action to address the failed port.

A.3.2 Optical Loopback

The Optical Loopback diagnostic of the ClearLink suite tests the optical characteristics of the SFPs involved in forming the link. Typically, this test identifies fiber-seating issues or occurrences of fiber contamination (number 2 in Figure 3). In unusual cases, this test may identify component failures (SFP, cable, or port), which can be identified by following the procedure listed below.

A.3.2.1 Observation

During the execution of the Optical Loopback diagnostic, the remote SFP is placed in optical loopback mode and a stream of diagnostic frames is transmitted to exercise the optical behavior of the local port and the SFPs that make up the connections of the link. If the Optical Loopback test fails, the most likely cause is a poor connection between the fiber and the SFP.

A.3.2.2 Action

The corrective action is to remove the fiber from the local SFP and clean the fiber and SFP according to standard optical cleaning procedures. Repeat this step for the remote side of the link as well and rerun the ClearLink Diagnostics suite.

If the error persists, replace the local SFP (using the standard cleaning procedure for the cable and the SFP) and rerun the ClearLink diagnostic suite. If the error still persists, replace the remote SFP (using the standard cleaning procedure for the cable and the SFP). Continued failure of the Optical Loopback diagnostic suggests that the cable is faulty and should be replaced. Using the standard cleaning procedures, replace and reseat the fiber cable and rerun the ClearLink diagnostic suite.

At this point, if the Optical Loopback diagnostic still fails, the local port is the most likely cause of the marginal behavior and needs to be replaced. Schedule the appropriate service action to address the failed port.

A.3.3 Link Saturation

The Link Saturation diagnostic of the ClearLink suite tests the operational characteristics of the ports (number 5 in Figure 3), SFPs (number 4 in Figure 3), and cable (number 3 in Figure 3) involved in forming the link. This test typically identifies marginal components in the link, which requires a systematic approach to isolate the offending element. The objective is to attempt to mitigate the errors starting with the most likely component to the least likely component.

A.3.3.1 Observation

During execution of the Link Saturation diagnostic, the remote port is placed in electrical loopback mode and a stream of diagnostic frames is transmitted to exercise the electrical and optical characteristics of the components of the link. If the Link Saturation test fails, the error report should indicate a higher level of detected errors at one end of the link or the other. The suspect side of the link is identified as the side opposite the side with the highest reported counters, and the corrective actions should start at the opposite end of the link from the reported error (for example, if the switch port reports the highest error counts, then focus on the HBA port).

A.3.3.2 Action

The corrective action is to begin with the least disruptive action possible. Reseat the remote SFP, clean and reseat the fiber cable in the remote SFP according to standard optical cleaning procedures, and then rerun the ClearLink Diagnostics suite. If the error persists, repeat this procedure for the local SFP and fiber cable, and rerun the ClearLink Diagnostics suite.

If the Link Saturation diagnostic continues to fail, begin component fault isolation at the end of the link opposite the side with the highest reported error counters. Replace the SFP (cleaning the connections appropriately), and rerun the ClearLink Diagnostics suite. If the errors persist, replace the near side SFP (cleaning the connections appropriately), and rerun the ClearLink Diagnostics suite. If the errors still persist, replace the cable using standard cleaning procedures, and rerun the ClearLink Diagnostics suite.

At this point in the isolation process, a marginal port is the most likely culprit. Specifically, it is either the transmitter of the side opposite the highest error counters or the receiver of the side with the highest error counters. Schedule the replacement of one side or the other, and rerun the ClearLink Diagnostics suite. If the error remains, schedule the replacement of the remaining port, and rerun the ClearLink Diagnostics suite.

A.4 Summary

The Brocade ClearLink Diagnostics suite provides the tools necessary to ensure that Fibre Channel links are properly installed and robust. The tools and procedures outlined above support the quick evaluation and isolation of marginal issues that can be difficult to find under normal operational conditions. In addition to troubleshooting faulty media issues, you should also run ClearLink Diagnostics before you deploy new ports in a fabric. Administrators should follow these general steps:

- Configure dynamic or on-demand mode for ClearLink Diagnostics using the configure command to allow the ClearLink Diagnostics (D_Port) tests to automatically run as ports come online.
- For ports that exhibit faulty media symptoms in networks with active traffic, plan to take the port offline and use the portdporttest command to run static ClearLink Diagnostics. Alternatively, use the Brocade SANnav Management Portal to initiate ClearLink.

Appendix B: Brocade Monitoring and Alerting Policy Suite (MAPS)

B.1 Introduction to MAPS

Brocade Monitoring and Alerting Policy Suite (MAPS) has comprehensive and advanced monitoring capabilities. Combined with simple yet flexible configuration and intuitive reporting, MAPS is a powerful and versatile tool to ensure the resiliency of SAN infrastructures of all variations of scale and topology.

The key benefits of MAPS easy-to-use monitoring are achieved through out-of-the-box predefined system settings and dashboards.

- Predefined groups define all common objects that are commonly monitored in most environments. Predefined groups simplify defining and applying rules to common objects. Thresholds configured for these predefined groups monitor all objects supported by MAPS.
- Predefined policies provide the settings that you can use to monitor your switches. These predefined policies contain hundreds of rules with predetermined thresholds, conditions, and alerting actions that have taken the guesswork and complexity out of configuring the correct monitoring settings.
- Dashboards summarize the reporting of switch conditions. All MAPS violations and errors are clustered in different categories, and specific rule violations and timestamps are recorded. The combined conditions of different categories determine the overall health status of a switch. From the dashboard, you can know the health status of a switch very quickly and drill down to areas of violations.

B.1.1 MAPS Command Examples

The following Fabric OS command output lists the predefined MAPS policies and the number of rules in each policy.

```
:admin> mapspolicy --show -summary
Policy Name                                 Number of Rules
-----------------------------------------------------------
dflt_aggressive_policy           :              208
dflt_conservative_policy         :              210
dflt_moderate_policy             :              210
```

The following Fabric OS CLI output lists all predefined groups, the type of each group, member counts of each group, and the members in each group.

```
:admin> logicalgroup --show
-------------------------------------------------------------------------------------------------------------------
Group Name            |Predefined    |Type             |Member Count      |Members
-------------------------------------------------------------------------------------------------------------------
ALL_PORTS             |Yes           |Port             |444               |1/0-63,2/0-11,3/0-47,4/0-31,5/0-
63,8/0-63,9/0-31,10/0-31,11/0-47,12/0-47
NON_E_F_PORTS         |Yes           |Port             |196               |
1/1,1/5-7,1/9-11,1/15,1/24,1/52,1/54-56,2/0-11,3/0,3/2,3/25-26,4/7,4/23-31,5/0-63,8/0-63,9/1-2,9/4-8,9/27,
11/0-1,11/5,11/13,11/15,11/23,11/25,11/41,11/46,12/4-5,12/15-23,12/39
ALL_E_PORTS           |Yes           |Port             |31                |1/26-39,9/0,9/3,9/9-15,12/40-47
ALL_F_PORTS           |Yes           |Port             |217               |
1/0,1/2-4,1/8,1/12-14,1/16-23,1/25,1/40-51,1/53,1/57-63,3/1,3/3-24,3/27-47,4/0-6,4/8-22,9/16-26,9/28-31,
10/0-31,11/2-4,11/6-12,11/14,11/16-22,11/24,11/26-40,11/42-45,11/47,12/0-3,12/6-14,12/24-38
ALL_OTHER_F_PORTS     |Yes           |Port             |8                 |1/2,1/17,1/44,1/46,1/49,1/59-60,1/63
ALL_HOST_PORTS        |Yes           |Port             |166               |
1/20,1/22-23,1/25,1/47-48,1/50,3/3-23,3/27-47,4/0-5,4/10-21,9/16-26,9/28-31,10/0-30,11/2-4,11/6-12,11/14,
11/16-22,11/24,11/26-40,11/42-45,11/47,12/6-8,12/10-12,12/14,12/32-38
ALL_TARGET_PORTS      |Yes           |Port             |43                |
```

```
1/0,1/3-4,1/8,1/12-14,1/16,1/18-19,1/21,1/40-43,1/45,1/51,1/53,1/57-58,1/61-62,3/1,3/24,4/6,4/8-9,4/22,10/31,
12/0-3,12/9,12/13,12/24-31
ALL_TS                |Yes           |Temperature Sensor|57               |1/0-7,2/0-2,3/0-4,4/0-4,6/0-3,
7/0-3,5/0-3,8/0-3,9/0-4,10/0-4,11/0-4,12/0-4
ALL_FAN               |Yes           |Fan               |3                |1,2,3
ALL_PS                |Yes           |Power Supply      |4                |1,2,3,4
ALL_WWN               |Yes           |WWN               |2                |1,2
ALL_SFP               |Yes           |Sfp               |390              |
1/0-63,2/0-11,3/0-47,4/0-22,4/24-31,5/0-7,5/20-31,5/40-55,8/0-7,8/12-19,8/24-35,8/40-43,8/52-55,8/60-63,
9/0-26,9/28-31,10/0-31,11/0-47,12/0-47
ALL_10GSWL_SFP        |Yes           |Sfp               |0                |
ALL_10GLWL_SFP        |Yes           |Sfp               |0                |
ALL_16GSWL_SFP        |Yes           |Sfp               |142              |
3/3,3/5-23,3/27-47,4/0-5,4/8,4/11,4/16-22,4/24-31,9/8-18,9/20-23,9/26,9/28-31,10/0-2,10/4-31,11/2,11/6,11/9,
11/11-12,11/14,11/17,12/10,12/12-14,12/16-23,12/32-39
ALL_16GLWL_SFP        |Yes           |Sfp               |0                |
ALL_QSFP              |Yes           |Sfp               |68               |5/0-7,5/20-31,5/40-55,8/0-7,
8/12-15,8/28-35,8/40-43,8/52-55,8/60-63
ALL_OTHER_SFP         |Yes           |Sfp               |180              |
1/0-63,2/0-11,3/0-2,3/4,3/24-26,4/6-7,4/9-10,4/12-15,8/16-19,8/24-27,9/0-7,9/19,9/24-25,10/3,11/0-1,11/3-5,
11/7-8,11/10,11/13,11/15-16,11/18-47,12/0-9,12/11,12/15,12/24-31,12/40-47
ALL_SLOTS             |Yes           |Blade             |12               |1,2,3,4,5,6,7,8,9,10,11,12
ALL_SW_BLADES         |Yes           |Blade             |8                |1,2,3,4,9,10,11,12
ALL_CORE_BLADES       |Yes           |Blade             |2                |5,8
ALL_FLASH             |Yes           |Flash             |1                |0
ALL_CIRCUITS          |Yes           |Circuit           |0                |
SWITCH                |Yes           |                  |1                |0
CHASSIS               |Yes           |                  |1                |0
ALL_D_PORTS           |Yes           |Port              |0                |
```

The following Fabric OS command output shows the MAPS dashboard information.

```
:admin> mapsdb --show all

1 Dashboard Information:
========================

DB start time:                  Thu Jun 26 15:19:30 2014
Active policy:                  dflt_conservative_policy
Configured Notifications:       RASLOG,FENCE
Fenced Ports:                   1/7,1/9,1/10,4/7,11/41
Decommissioned Ports:           None

2 Switch Health Report:
=======================

Current Switch Policy Status:  MARGINAL
Contributing Factors:
---------------------
*FAULTY_BLADE (MARGINAL).

3.1 Summary Report:
===================

Category                        |Today                          |Last 7 days                    |
--------------------------------------------------------------------------------------------------
Port Health                     |Out of operating range         |Out of operating range         |
Fru Health                      |In operating range             |Out of operating range         |
Security Violations             |No Errors                      |No Errors                      |
Fabric State Changes            |Out of operating range         |Out of operating range         |
Switch Resource                 |Out of operating range         |Out of operating range         |
Traffic Performance             |In operating range             |In operating range             |
FCIP Health                     |No Errors                      |No Errors                      |
Fabric Performance Impact       |In operating range             |Out of operating range         |

3.2 Rules Affecting Health:
```

```
==========================

Category(Rule Count) |RepeatCount   |Rule Name                | Execution Time   |Object    |Triggered |Value(Units)|
---------------------------------------------------------------------------------------------------------------------
Port Health(22106)   |2             |defALL_OTHER_F_PORTSLF_5 |07/22/14 14:50:55 |Port 1/2  |326       |            |
                     |              |                         |                  |Port 1/17 |10        |            |
...

4 History Data:
===============

Stats(Units)   Current     07/21/14    07/20/14    07/19/14    07/18/14    07/17/14    07/16/14
               Port(val)   Port(val)   Port(val)   Port(val)   Port(val)   Port(val)   Port(val)
-------------------------------------------------------------------------------------------------
CRC(CRCs)      9/26(26)    9/26(16)    9/26(40)    9/26(34)    11/40(42)   11/40(40)   11/40(38)
               11/40(24)   11/40(13)   11/40(39)   11/40(33)   -           -           -
```

B.2 Default MAPS Policy Recommendations

The following three predefined policies provide the default rules for different SAN environments. The rules in the predefined policies have been validated and represent the best practices from Broadcom for managing SANs over 20 years.

- aggressive policy (dflt_aggressive_policy)
- conservative policy (dflt_conservative_policy)
- moderate policy (dflt_moderate_policy)

The predefined policies monitor the same sets of objects. But the thresholds and actions associated with the rules are different among the predefined policies. The default aggressive policy is designed for strict SAN environments where small errors cannot be tolerated. For example, the default aggressive policy might be used for FICON environments. The conservative policy has more lenient settings. It is designed for environments that can tolerate some errors. The moderate policy settings are between those of the aggressive policy and the conservative policy. It is designed by Broadcom as the default for Open Systems environments. Based on the requirements of your specific environment, you can choose one of the predefined policies to activate MAPS monitoring.

B.3 Activating MAPS Actions

You can use the following alerting actions in MAPS to be notified or to take corrective actions when the thresholds are exceeded:

- Send an SNMP trap
- Log a RASLog message
- Send an email alert
- Fence a port
- Decommission a port
- Send a notification to the FICON control unit (CUP)
- Quarantine a slow draining device
- Toggle a port
- Send FPIN to devices that are registered to receive Fabric Notification

Any of these actions can be enabled or disabled on a switch. Disabling an action globally on a switch overrides the configured actions associated with individual rules. Enabling an action on a switch allows the action to be triggered by individual rules that have the actions configured. All actions are disabled by default when MAPS is enabled for the first time.

You must enable the actions using the `mapsConfig –actions` command. Before doing so, you can check the MAPS dashboard to make sure that the enabled policy provides an adequate level of monitoring. You can also globally disable actions for switches that are in a maintenance window to avoid false alerts and incorrect actions.

B.4 Customizing MAPS Monitoring

If the predefined policies are not sufficient for your environment, you can customize MAPS settings in a number of ways.

For most environments, you can use the `mapsPolicy --clone` command to copy one of the predefined policies as a template for customization. If thresholds in the default rules are not adequate, you can use the `mapsRule --clone` command to copy the rules and adjust the parameters. If you receive too many alerts that are generated from a particular rule, you can adjust the threshold settings so that they are higher. Conversely, you should examine the History Data section of the MAPS dashboard output displayed by the `mapsdb --show all` command. If events and errors are recorded for certain counters that are within the thresholds but you would like to see alerts generated for these counters, you can adjust the threshold settings lower to catch these conditions.

For certain objects that must be monitored differently from other objects, you can use the `logicalGroup --create` command to create user-defined groups for these objects. For example, certain storage devices attached to the switch may support a high data rate or host critical application data. For the F_Ports attached to these devices you may need to have different thresholds or rules from other F_Ports that belong to one of the predefined groups. You can create a static or dynamic group for these F_Ports. Then you can define the thresholds in rules that apply specifically to this user-defined group of objects.

You can pause and resume MAPS monitoring for specific members of elements or for a category of similar elements. This is particularly useful during maintenance operations of storage devices and servers. If a few servers or storage devices connected to the SAN are being upgraded, administrators can pause monitoring in advance on the ports that connect these servers or storage devices. All rules that apply to these ports will pause the monitoring. This avoids generating unnecessary out-of-range violations during the maintenance operations. Once maintenance operations are complete, the paused monitoring can be resumed.

B.5 Using MAPS with Brocade SANnav

If you need to configure MAPS policies on multiple switches or multiple fabrics, you can use Brocade SANnav Management Portal to achieve this easily with an option to apply a MAPS policy to selected switches or all switches managed by a SANnav instance. With very large environments that require multiple Brocade SANnav instances to manage, you can use the policy export and import feature to apply the same MAPS settings to all switches in these environments.

B.6 Summary

Because of the simple, flexible, comprehensive, and advanced monitoring capabilities of MAPS, it is essential for administrators to use MAPS to ensure the resiliency of SAN environments. Follow these steps:

- Choose one of the three default policies for the environment.
- Enable MAPS with one of the default policies by using the `mapsConfig --enablemaps` command.
- Check the out-of-range violations and errors using the `mapsdb -show` command.
- If the chosen policy is satisfactory, enable alerting actions using the `mapsConfig --actions` command.
- If you need to customize the settings, clone an existing MAPS policy using the `mapsPolicy --clone` command.
- Change existing rules or create new rules in the cloned policy.

B.7 Evolution of MAPS Features

The following table lists the enhancements of MAPS features and capabilities since Fabric OS 8.0.

Table 1: Evolution of MAPS Features

FOS 8.0	FOS 8.1	FOS 8.2	FOS 9.0
Added monitoring for: 32G optics, IO Insight flow metrics, security certificate expiration, Gigabit Ethernet ports, FCIP tunnel counters, IP Extension circuit and tunnel counters, system airflow direction, and initiator and target device ratio. Added support of Quiet Time for SNMP traps.	Added Rule-on-Rule monitoring.	Support custom message IDs for RASlog alert. Add port decommissioning with impair option. Add monitoring for UCS Uplink monitoring.	Support global quiet time that can be supported by all rules. Support pause monitoring of category of objects. Support FPIN action to be triggered by MAPS threshold violation.

Appendix C: Fabric Performance Impact (FPI) Monitoring

C.1 Introduction to FPI

FPI monitoring is a dedicated MAPS monitoring category to detect and alert congestion. Detecting and pinpointing a source of congestion to a fabric is critically important in ensuring fabric performance. As with other monitoring in MAPS, FPI requires zero user configuration yet provides advanced monitoring and intuitive reporting abilities.

C.2 FPI Monitoring

FPI uses the tx_c3_timeout and tim_txcrd_z counters in a Brocade ASIC to detect a credit-stalled device. When a frame has been unable to transmit due to the lack of buffer credits and has exceeded the timeout value, the frame is discarded. Each frame discard increases the tx_c3_timeout counter. Each increment of the tim_txcrd_z counter represents 2.5 microseconds of time with zero transmit buffer-to- buffer credit. FPI monitors both counters to detect a credit-stalled device. The detection logic uses thresholds that are internally calibrated so that the user does not have to decide the correct thresholds and configure them.

FPI uses the transmit queue (TXQ) latency counter within a Brocade ASIC. The TXQ latency counter measures the length of time a frame has been waiting in the transmit queue. FPI monitors the transmit queue latency to detect ports that may potentially cause congestion. Elevated transmit queue latency and high transmit bandwidth utilization without a significant increase of the tim_txcred_z counter indicates that the port is oversubscribed.

FPI detects different severity levels of latency and reports two latency states. The IO_FRAME_LOSS state indicates a severe level of latency. In this state, frame timeouts either have already occurred or are very likely to occur. Administrators should take immediate action to prevent application interruption. The IO_PERF_IMPACT state indicates a moderate level of latency. In this state, device-based latency can negatively impact the overall network performance. Administrators should take action to mitigate the effect of latency devices. The separate latency states enable administrators to apply different MAPS actions for different severity levels. For example, administrators can configure the Port Fencing action for the IO_FRAME_LOSS state and the email alert action for the IO_PERF_IMPACT state.

C.3 IO_PERF_IMPACT State

FPI evaluates the tim_txcred_z counter increment for all F_Ports at a fixed time interval. For each evaluation, the cumulative increments of the tim_txcred_z counters over three time windows within the fixed interval are calculated simultaneously. The time windows are short, medium, and long duration, respectively. The cumulative increment for each time window is compared to an internal threshold predefined for that time window. The short window uses the high threshold; the medium window uses the medium threshold; and the long window uses the low threshold. If any of these thresholds is exceeded for the corresponding time windows, the F_Port is put in the IO_PERF_IMPACT state. With internally predefined and calibrated thresholds, FPI removes from administrators the complexity of determining appropriate latency thresholds. Furthermore, with simultaneous evaluation against different thresholds for three different time windows, FPI can capture both short spikes of severe latency and sustained medium latency. FPI can capture a range of latency levels without users having to configure the thresholds.

C.4 IO_FRAME_LOSS State

FPI puts an F_Port in the IO_FRAME_LOSS state if an actual timeout occurs on the port. This timeout is indicated with the C3TX_TO counters on the port. In addition, FPI calculates the average R_RDY delay to predict the likely timeout on an F_Port. This calculation is done by evaluating the number of 2.5-microsecond increments of the tim_txcred_z counter in a 60-second interval over the total number of frames during those 60 seconds. If this average R_RDY delay is greater than 80 milliseconds, the corresponding F_Port is also marked as IO_FRAME_LOSS.

C.5 IO_LATENCY_CLEAR

For ports detected in the IO_FRAME_LOSS or IO_PERF_IMPACT state, when the latency conditions are cleared and the corresponding counter evaluations are within the internal thresholds, FPI puts the ports in the IO_LATENCY_CLEAR state. Since this is a different state, customers can receive a separate notification when the latency conditions are cleared.

C.6 OVERSUBSCRIBED

FPI includes detection of congestion due to oversubscription. Elevated transmit queue latency and high transmit bandwidth utilization without a significant increase of the tim_txcrd_z counter indicates that the port is oversubscribed. When this is detected, FPI sets the port to the OVERSUBSCRIBED state.

C.7 Mitigation of Device-Based Latency

Once the source of device latency is detected, manual actions are usually required to mitigate the problem outside of the fabric, such as moving slow responding hosts or arrays to a different path. In addition, since these actions are disruptive, they can be done only during a maintenance window. This implies that the mitigation can happen only after some time has elapsed since the problem was discovered. To apply automatic mitigation on the source of device latency, you can take advantage of two new actions that are associated with FPI rules. One new action is Slow Drain Device Quarantine (SDDQ). This action is automatically triggered when FPI detects an F_Port in either the IO_PERF_IMPACT state or the IO_FRAME_LOSS state. The SDDQ action sends the PID associated with the slow draining device to all switches in a fabric. All switches move the traffic destined for the slow draining device into a low-priority virtual channel (VC). As a result, buffer credits on the regular, medium-priority VC are freed for traffic destined to other devices. This effectively removes the impact of a slow drain device to the fabric performance without disruption to traffic. The result of this action is that slow drain devices are isolated in a quarantine but remain online. This gives administrators more time to find permanent solutions for the slow draining device problem. To use the SDDQ action, the switches in the fabric are required have a Fabric Vision license and you must enable QoS on all ports in the flow path.

Another action that you can use is Port Toggle. This action disables a port for a short, user-configurable duration and then re-enables the port. The Port Toggle action can recover slow draining devices such as those caused by a faulty host adapter. In addition, the Port Toggle action can induce multi-pathing software (MPIO) to trigger traffic failover to an alternate path to prevent severe performance degradation. By using the SDDQ or Port Toggle actions, administrators not only can monitor for device-based latency but can also automatically mitigate the problem when such conditions are detected by FPI.

C.8 Congestion Dashboard

The Congestion Dashboard displays a summary of the congestion states of ports, which includes the port number and the congestion severity level. If you need to troubleshoot potential congestion issues, use the congestion dashboard as a first step to know the issues that are already detected by FPI.

C.9 Summary

Brocade FPI monitoring offers advanced device latency detection and mitigation capabilities that are easy to deploy and use. The detection offers a clear indication that the fabric may be experiencing performance impact due to slow draining devices or other device behavior anomaly. It includes automatic mitigation actions without disruption to application traffic. For the best-practice recommendation, administrators should enable FPI monitoring and automatic mitigation actions by following these general steps:

- Slow Drain Device Quarantine and Port Toggle actions are the default actions for FPI monitoring in the predefined conservative, moderate, and aggressive rules. But in order to have them triggered, the actions must be enabled for a switch using the `mapsconfig --actions` command.

C.10 Evolution of FPI

The following table lists the enhancements of FPI features and capabilities since Fabric OS 8.0.

Table 2: Evolution of FPI Features

FOS 8.0	FOS 8.1	FOS 8.2	FOS 9.0
FPI is available in the dflt_base_policy of MAPS, which does not require a Fabric Vision license. (Only the RASLog notification is supported with the dflt_base_policy.) Monitors the zoned-device ratio to alert when the ratio of initiators to a target or targets to an initiator in a zone exceeds the recommended settings.	FPI supports automatic un-quarantine action.	FPI supports SDDQ on devices connected through Access Gateway.	FPI supports FPI profile to adjust default thresholds. FPI monitor and alert congestion due to oversubscription.

Appendix D: Flow Vision and IO Insight

D.1 Introduction to Flow Vision and IO Insight

Brocade Flow Vision is a comprehensive Fibre Channel diagnostic tool that, without any disruption, provides vision and insight into application flows for analysis of throughput-related, congestion-related, and latency-related performance problems. Flow Vision also has the capability to generate test traffic in a SAN. Flow Vision requires a Brocade Fabric Vision license.

Brocade IO Insight is a Gen 6 and Gen 7 platform capability that provides built-in instrumentation to directly measure device IO performance, specifically the SCSI IO latency metrics. With IO Insight, administrators can gain deeper insight into the application-level and device-level performance of a SAN so that critical SLAs can be achieved. IO Insight is a capability obtained through the Flow Vision feature. Administrators can obtain IO first response time, completion time, and pending IO metrics from device ports for specific application data flows.

Flow Vision includes the following functionality:

- Flow Monitor provides the ability to monitor Fibre Channel flows and gather frame statistics for the monitored flows.
- Flow Learning provides the ability to automatically and dynamically discover flow parameters, such as source device and destination device, within a fabric.
- Flow Generator generates a simulated traffic load for a specific flow to pretest a SAN infrastructure path for hardware, connectivity, and performance validation before applications are deployed.
- Flow Mirror provides the ability to nondisruptively create copies of application flow frames for deeper analysis.

D.2 Understanding and Defining Flows

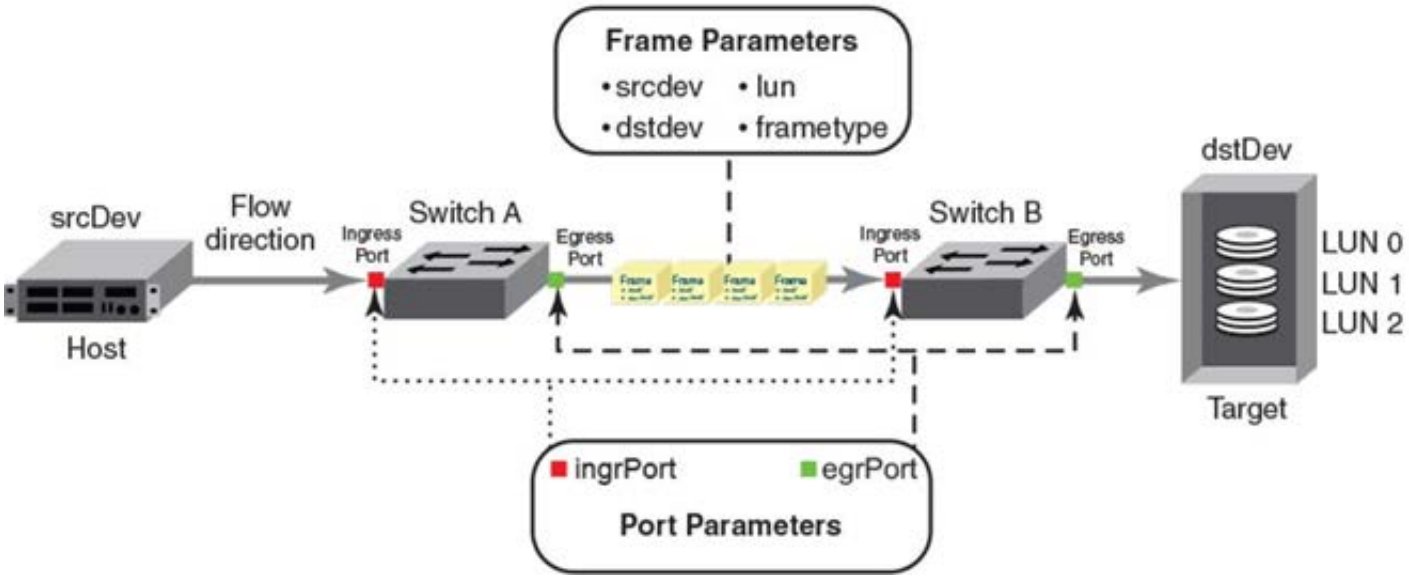
In the broadest sense, a flow is a collection of Fibre Channel frames that share certain traits, such as a source device, a destination device, an ingress port, or an egress port. These common traits help distinguish these frames from all other frames in a network, and they provide useful data for analysis. In the majority of use cases, initiator-target or initiator-target-LUN flows on a specific ingress or egress port are of interest.

To define a flow in Flow Vision, you specify the parameters of interest for filtering frames. The following parameters can be specified in Flow Vision:

- Port parameters – Port parameters specify an FC port, either the ingress direction or the egress direction, as a point of interest for the flow.
- Frame parameters – Frame parameters specify the source device, destination device, LUN ID, or frame type for the flow.
- Direction – This specifies the traffic flow direction. The direction parameter can be implicitly inferred from the frame parameters when it is not specified.

Figure 4 illustrates how the port and frame parameters apply to a flow.

Figure 4: Flow Definition with Frame and Port Parameters



D.3 Using Flow Vision for Troubleshooting

As an advanced diagnostic tool, Flow Vision requires some level of knowledge and understanding of the Fibre Channel network. There are also several administrative steps, such as identifying source and destination devices in order to define a flow. Flow Vision may not be the first tool that administrators need to troubleshoot issues impacting fabric resiliency. But in a large network with a complex mix of servers, storage, and workload, administrators are likely find it necessary to use the various features in Flow Vision to troubleshoot complex issues, particularly those related to misbehaving devices and congestion.

When administrators find it necessary to use Flow Vision to troubleshoot problems, mechanisms exist to simplify the usage of Flow Vision. The following are a few recommendations:

- Predefined flows – With predefined flows, the only administrative step is to activate the flow when necessary. For example, the sys_flow_monitor flow automatically learns all traffic flows across all device ports on a switch. This is a very convenient mechanism for administrators to learn the device traffic patterns and to know where to drill down further.
- Learning flows – Learning flows provide the ability to automatically discover flow parameters such as a source device or a destination device. This knowledge is useful when administrators have only partial information about a problem, information that was maybe obtained from other tools such as MAPS. Administrators can define flows with the partially known information, for example, only a port number, and discover other flow parameters and gain deeper visibility.
- Brocade SANnav – Brocade SANnav 2.1 introduces comprehensive Flow Management features. SANnav Flow Management provides an inventory view of flows that are monitored by each switch. From the inventory view, you can select specific flows to investigate for historical and real-time performance statistics. You can use SANnav to manage flow collections, which is a set of flows that you group together as related flows. SANnav also provides a view of threshold violation events for individual flows and flow collections. Refer to the *Brocade SANnav Flow Management User Guide for SANnav Flow Management* for feature details.

- Fabric Flow Dashboard – The Fabric Flow Dashboard is an enhancement in the Fabric OS Flow Vision command that presents the flow topology and collects relevant data for troubleshooting each flow path. The flow topology presents all flow paths taken between a source device and a destination device. For each flow path, the Fabric Flow Dashboard displays the MAPS violation summaries of each switch along the paths that are relevant. Administrators can use this tool on a switch to logically follow a flow path and to obtain relevant information without going through multiple switches in the fabric.

D.4 Using Flow Vision for Monitoring

Administrators can use Flow Vision, in particular, the Flow Monitor functionality, to monitor performance. A large SAN environment can have tens of thousands of flows. It is impractical and unnecessary to constantly monitor every flow. Therefore, it is a better strategy to monitoring specific flows is recommended to ensure that SLAs for key servers and applications are met or to be vigilant against certain known device behaviors. For this set of use cases, administrators can import a Flow Monitor flow into MAPS as a custom group and define thresholds on the flow metrics. If the thresholds are violated, MAPS will send an alert and apply other actions defined in the monitoring rule. On Brocade Gen 6 platforms, MAPS supports rules with the IO Insight metrics as thresholds. Administrators can use this mechanism to ensure that latency SLAs for critical servers and applications can be maintained.

D.5 Evolution of Flow Vision Features

The following table lists the enhancements of Flow Vision features and capabilities since Fabric OS 8.0.

Table 3: Evolution of Flow Vision Features

FOS 8.0	FOS 8.1	FOS 8.2	FOS 9.0
Flow Monitor adds IO Insight capability support on Brocade Gen 6 platforms. Flow Monitor adds VE_Port support for FCIP. Flow Monitor adds the new frame type Sequence Retransmission Request (SRR)	Flow Mirror adds support to Access Gateway. Flow Monitor adds monitoring of flow with VM tagging.	Flow Monitor adds the predefined system flow sys_mon_all_fports. Flow Mirror supports the predefined system sys_analytics_vtap remote flow mirroring (RFM) to the Brocade Analytics Monitoring Platform. The number of subflows supported per platform is increased.	Flow monitor adds sys_flow_monitor as system predefined learning flow to provide Fibre Channel and IO Insight metrics.

Appendix E: Sample Frame Viewer Session

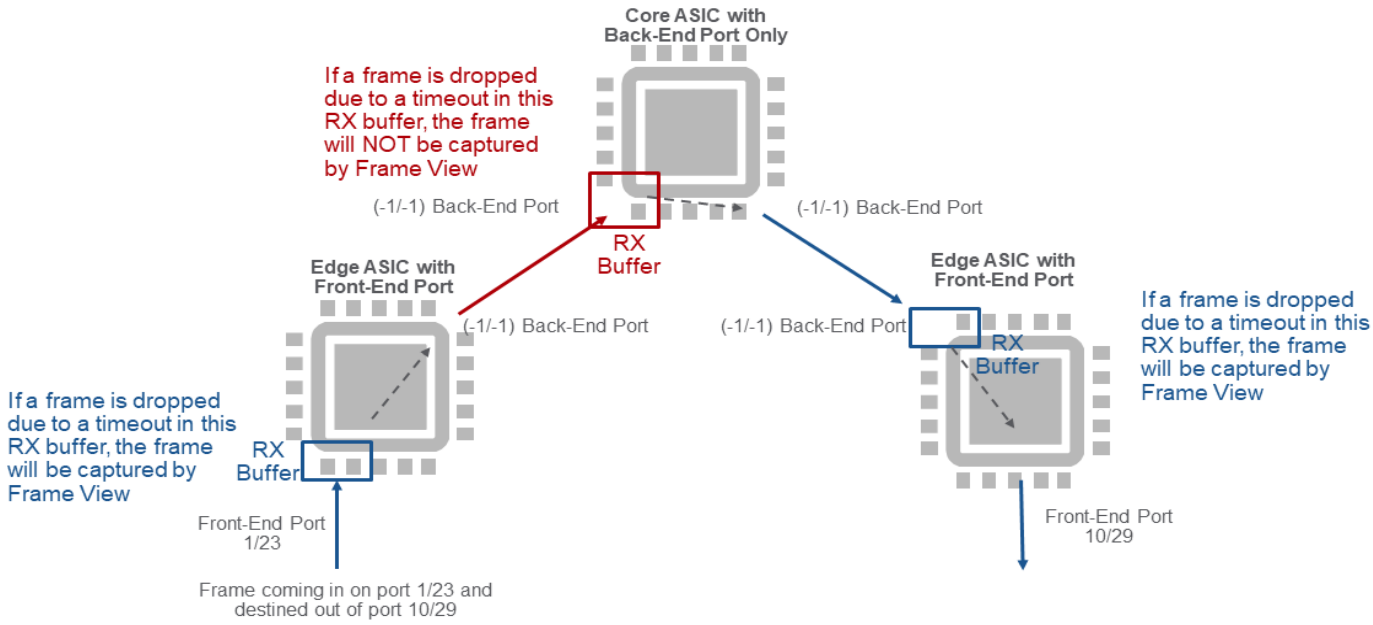
Frames discarded due to a hold time timeout are sent to the CPU for processing. During subsequent CPU processing, information about the frame such as the SID, DID, and transmit port number is retrieved and logged. This information is maintained for a certain fixed number of frames.

Frame Viewer captures only FC frames that are dropped due to a timeout received on an edge ASIC (an ASIC with FE ports). If the frame is dropped for any other reason, it is not captured by Frame Viewer. If the frame is dropped due to a timeout on an Rx buffer on a core ASIC, the frame is not captured by Frame Viewer. Timeout is defined as a frame that lives in an Rx buffer for longer than the hold time default of 500 ms or the edge hold time custom setting.

The user is provided a CLI command (frame log) to retrieve and display this information. Figure 6 provides an overview of the collection process for a bladed director.

Figure 5 provides an overview of the collection process for a bladed director.

Figure 5: Frame Viewer Capture Capability for a Director or Switch with Multiple ASICs



NOTE: If the switch is a single-ASIC switch, such as an embedded switch, Brocade 300 Switch, Brocade 6505 Switch, Brocade 6510 Switch, Brocade G610, Brocade G620, or Brocade G720, there is no core ASIC or back-end ports, and Frame Viewer captures dropped frames due to timeouts. The number of frames captured depends on available switch resources. A core ASIC has only back-end ports and UltraScale inter-chassis link (ICL) ports. If a frame is dropped and captured by Frame Viewer, it displays the frame (FC header and payload) with a timestamp of when the frame was dropped.

```
framelog --show -n 1200:
================================================================================================
Wed Dec 28 08:51:02 EST 2012
================================================================================================
Log              TX        RX
timestamp        port      port      SID          DID          SFID      DFID      Type     Count
================================================================================================
Dec 19 11:37:00  -1/-1     10/29     0x01dd40     0x018758     129       129       timeout  8
Dec 19 11:37:00  -1/-1     1/29      0x018d40     0x01874b     129       129       timeout  8
Dec 19 11:37:00  -1/-1     12/5      0x017500     0x018758     129       129       timeout  8
Dec 19 11:37:00  -1/-1     10/5      0x015500     0x018758     129       129       timeout  8
Dec 19 11:37:00  -1/-1     3/5       0x012500     0x01874b     129       129       timeout  6
Dec 19 11:37:00  -1/-1     3/5       0x012500     0x018541     129       129       timeout  4
Dec 19 11:37:00  -1/-1     1/5       0x010500     0x01874b     129       129       timeout  12
Dec 19 11:37:00  1/23      -1/-1     0x01dd40     0x018758     128       128       timeout  4
Dec 19 11:37:00  1/23      -1/-1     0x015500     0x01874b     128       128       timeout  2
Dec 19 11:37:00  1/23      -1/-1     0x012500     0x018758     128       128       timeout  4
Dec 19 11:37:00  1/23      -1/-1     0x010500     0x018758     128       128       timeout  10
Dec 19 11:36:59  -1/-1     3/5       0x012500     0x01874b     129       129       timeout  8
Dec 19 11:30:51  -1/-1     10/29     0x01dd40     0x01874b     129       129       timeout  8
Dec 19 11:30:51  -1/-1     1/29      0x018d40     0x01874b     129       129       timeout  8
Dec 19 11:30:51  -1/-1     10/5      0x015500     0x018756     129       129       timeout  8
Dec 19 11:30:51  -1/-1     3/5       0x012500     0x01874b     129       129       timeout  8
Dec 19 11:30:51  -1/-1     1/5       0x010500     0x01874b     129       129       timeout  8
Dec 19 11:30:50  1/23      -1/-1     0x01dd40     0x018756     128       128       timeout  6
Dec 19 11:30:50  1/23      -1/-1     0x018d40     0x018756     128       128       timeout  8
Dec 19 11:30:50  1/23      -1/-1     0x012500     0x018756     128       128       timeout  6
```

NOTE:

- TX port: The port that discarded the frame.
- SID: The source port ID (PID).
- DID: The destination PID.
- -1/-1: The port columns refers to a BE port.

Appendix F: Edge Hold Time (EHT)

F.1 Introduction to EHT

Edge Hold time (EHT) is a Fabric OS capability that allows an overriding value for hold time (HT). Hold time is the amount of time a Class 3 frame may remain in a queue, while waiting for credit to be given for transmission, before being dropped.

The default HT is calculated from the RA_TOV, ED_TOV, and maximum hop count values configured on a switch. If you are using the default configuration, which sets the standard 10 seconds for RA_TOV, two seconds for ED_TOV, and a maximum hop count of seven, a hold time value of 500 ms is calculated.

Extensive field experience has shown that when high latency occurs, even on a single initiator or device in a fabric, not only does the F_Port attached to this device see Class 3 frame discards, but the resulting back pressure due to the lack of credit can build up in the fabric and cause other flows not directly related to the high-latency device to have their frames discarded at ISLs.

Edge hold time can be used to reduce the likelihood of this back pressure in the fabric by assigning a lower hold time value only for edge ports (initiators or devices). The lower EHT value will ensure that frames are dropped at the F_Port where the credit is lacking, before the higher default hold time value used at the ISLs expires, allowing these frames to begin moving again. This localizes the impact of a high-latency F_Port to just the single edge where the F_Port resides and prevents it from spreading into the fabric and impacting other unrelated flows.

Like hold time, edge hold time is configured for the entire switch and is not configurable on individual ports or ASICs. Whether the EHT values or the HT values are used on a port depends on the particular platform and ASIC, on the type of port, and also on other ports that reside on the same ASIC. This behavior is described in further detail in the following sections.

F.2 Supported Releases and Licensing Requirements

The EHT behaviors in Fabric OS 8.x and Fabric OS 9.0 are consistent. The behaviors of these earlier releases are noted in later sections for reference purposes.

There is no license required to configure the edge hold time setting. EHT is enabled by default.

F.3 Behavior

F.3.1 Gen 5 and Later Platforms

All Brocade Gen 5 platforms (16G) and later are capable of setting the hold time value on a port-by-port basis for ports that are on Gen 5 or later ASICs.

- All F_Ports will be programmed with the alternate edge hold time value.
- All E_Ports will be programmed with the default hold time value (500 ms).

The same EHT value that is set for the switch will be programmed into all F_Ports on that switch. Different EHT values cannot be programmed on an individual port basis.

If 8G blades are installed into a Gen 5 platform (for example, an FC8-64 or FX8-24 blade in a DCX 8510), then the same EHT value will be programmed into all ports on the ASIC.

- If any single port on an ASIC is an F_Port, the alternate EHT value will be programmed into the ASIC, and all ports (E_Ports and F_Ports) will use this one value.
- If all ports on an ASIC are E_Ports, the entire ASIC will be programmed with the default hold time value (500 ms).

A unique EHT value can be independently configured for each logical switch for Gen 5 and later platforms. 8G blades installed in a Gen 5 platform will continue to use the default logical switch configured value for all ports on those blades regardless of which logical switches those ports are assigned to.

F.4 Default EHT Settings

The default setting used for edge hold time (EHT) is preloaded into the switch at the factory. Switch installed with FOS v8.x or later have the default EHT as 220ms:

The default setting can be changed using the `configure` command. The EHT can be changed without having to disable the switch, and it takes effect immediately after being set.

When you use the configure command to set the EHT, a suggested EHT value will be provided to you. If you accept this suggested setting by pressing `<enter>`, this suggested value will become the new value for EHT on the switch.

Once you set this value by running the `configure` command, the EHT value will be maintained across firmware upgrades, power-cycles, and HA fail-over operations. This is true for all versions of Fabric OS.

Example (Fabric OS 8.0):

Not all options are available on an enabled switch.

To disable the switch, use the `switchDisable` command.

```
sw0:FID128:admin> configure
Configure...
Fabric parameters (yes, y, no, n): [no] y
Configure edge hold time (yes, y, no, n): [no] y
Edge hold time in ms: (80(Low), 220(Medium), 500(High), 80-500(UserDefined)): (80..500) [220]
System services (yes, y, no, n): [no]
```

F.5 Recommended Settings

Edge hold time does not need to be set on core switches that are comprised of only ISLs; core switches, therefore, use only the standard hold time setting of 500 ms.Recommended values for platforms that contain initiators and targets are based on specific deployment strategies. End users typically either separate initiators and targets on separate switches or mix initiators and targets on the same switch.

A frame drop has more significance for a target than for an initiator because many initiators typically communicate with a single target port, whereas target ports typically communicate with multiple initiators. Frame drops on target ports usually result in error messages being generated in server logs that refer to a SCSI Transport. Multiple frame drops from the same target port can affect multiple servers in what appears to be a random fabric or storage problem. Since the source of the error is not obvious, you can waste time determining the source of the problem. Take extra care, therefore, when applying the EHT to switches where targets are deployed.

The most common value for EHT is 220 ms.The lowest EHT value of 80 ms should be configured only on edge switches that are comprised entirely of initiators. This lowest value is recommended for fabrics that are well maintained and when a more aggressive monitoring and protection strategy is being deployed.

Appendix G: Fabric Notification

G.1 Introduction to Fabric Notification

As described throughout this document, the resiliency of a SAN has many external conditions. To increase resiliency, it is imperative to have end devices (hosts and targets) to participate in mitigation. At the same time, it is important that a solution should be automated to the greatest possible extend. Fabric Notification is a Fibre Channel standard-based mechanism to facilitate the self-healing among fabric and device members to mitigate and recover from conditions that affect the performance of a SAN. Fabric Notification accomplish this by defining a set of protocols and communications between switch and device members.

G.2 Supported Release and License Requirement

Fabric Notification is supported by Brocade Fabric OS v9.0 and later. This feature is included in Fabric OS. It does not require a license.

G.3 Behavior

Fabric Notification is a broad term that covers a number of different notifications. These notifications are separated into two groups:

- Congestion Signals: Hardware generate primitive signals sent toward the point of congestion. These signals indicate that the number of frames in the transmit queue may impact fabric performance. The congestion signals are supported on Brocade Gen 7 platforms.
- Fabric Performance Impact Notifications (FPINs): Software generated notifications using ELS commands. The FPINs are supported on Brocade Gen 6 and Gen 7 platforms.

In order to receive Fabric Notification, a device must negotiate the capabilities and register to the connected fabric. When events occur that triggers Fabric Notification, Brocade Fabric OS will send appropriate notifications to registered devices.

FPINs are categorized according to the type of performance impacting issues. Following are set of FPIN descriptors:

- Congestion Descriptor: Send to a device connected on a congested port that was detected by FPI. The cause of congestion, credit-stalled device as compared to oversubscription, are part of the descriptor.
- Peer Congestion Descriptor: Send to devices that are in the same zone with the device on a congested port.
- Link Integrity Descriptor: Send to devices where CRC and ITW link errors are detected at the N_port. This notification is sent also to the peer devices in the same zone.
- Delivery Descriptor: Send to a device when an FCP command frame originated from that device has been dropped by the fabric.

Brocade MAPS includes FPINs as an alert action by default. The congestion descriptor, peer congestion descriptor, and link integrity descriptor are triggered by FPI, CRC, and ITW rules in MAPS. The delivery descriptor is configured by the `frameLog` command for the Frame Viewer feature. When devices receive Fabric Notification, they can choose from a range of responses to mitigate the issue. These include notifying the end user on the devices, reducing the I/O requests until the condition is cleared, or accelerating I/O failover to an alternative pass. The objective of these actions is to reduce response time and to improve overall resiliency.

G.4 Recommended Settings

Once you have upgraded a fabric to FOS v9.0, perform the following steps:

- Enable MAPS and FPI to detect issues related to congestion. Enable and include FPIN as an action.
- Configure Frame Viewer log to send FPINs

Revision History

53-1004609-04; July 8, 2021

- Updated the peer zoning description.

53-1004609-03; September 1, 2020

- Updated with Gen 7 and FOS v9.0 features.

53-1004609-02; April 30, 2020

- Provided a recommendation for Port Decommissioning with the impair option and updated the document to the latest FOS version.

53-1004609-01; December 7, 2016

- Initial release.