BROCADE
A Broadcom Inc. Company

USER GUIDE

Brocade Fabric Congestion Troubleshooting Guide

Supporting Fabric OS

Fabric-Con-Trbl-UG100
2 July 2018

Contents

About this Document

This guide is intended to help you identify the causes of congestion that are impacting your fabric and application performance. It provides guidelines for troubleshooting the source of the congestion to a fabric port, a problem device, or both.

A well-designed Fibre Channel (FC) SAN should not cause application performance issues. Unless you are receiving fabric-performance related alerts from the FC SAN fabric management tools, first determine if your issues are one of the following more typical causes:

- CPU or memory exhaustion—For example, extended periods of 100% utilization of CPU or memory resources allocated to the application.
- Data network congestion—For example, TCP/IP packet loss due to congestion on the front-end network connecting application and database servers.
- Storage target I/O latency—For example, an abnormally long delay from the target when responding to an I/O request.

Requirements for using this guide

The following are requirements for using this guide:

- A technical understanding of Fibre Channel networks and the Fibre Channel Flow Control mechanism. If you do not have this understanding, we recommend that you read through the "Fibre Channel flow control" appendix before proceeding with this guide.
- Access to RASLog messages that provide Monitoring and Alerting Policy Suite (MAPS) alerts.
- Access to and knowledge of MAPS to display alerts and dashboards.

Prerequisites for congestion detection and troubleshooting

The following are prerequisites for using MAPS to access fabric performance alerts and dashboards discussed in this guide for troubleshooting causes of congestion. Refer to the *Brocade Monitoring and Alerting Policy Suite Configuration Guide* for details.

- Confirm that a Fabric Vison license is installed on the switch. Although limited MAPS functionality is available in Brocade Fabric OS without the Fabric Vision license, full MAPS functionality is required for the procedures in this guide.
- Enable a MAPS monitoring policy that sets the thresholds for generating alerts. Aggressive, moderate, and conservative predefined policies are available.
- Set an action for where alerts are generated, such as RASLog, email, or SNMP.

How to use this guide

To use this guide:

1. Use Congestion Troubleshooting Overview on page 7 chapter to understand, determine, and mitigate the causes of congestion
   a. Read through What is congestion? on page 7, Symptoms of congestion on page 7, and Causes of congestion on page 8.
   b. Use the flowchart and steps in Determining the cause of congestion on page 8 as a guide to help determine if the cause of your fabric congestion is lost credits, a credit-stalled device, or oversubscription.
   c. Use the steps in Lost credits on page 10 to verify that the cause of your congestion is lost credits and mitigate the congestion.
   d. Use the steps in Credit-stalled devices on page 12 to verify that the cause of your congestion is a credit-stalled device and mitigate the congestion.
   e. Use the steps in Oversubscription on page 14 to verify that the cause of your congestion is oversubscription and to mitigate the congestion.
2. Use the Appendixes for reference material on topics such as a description of congestion causes using an analogy, Fibre Channel Flow Control, more detailed information on congestion causes and their detection, metrics used to identify congestion and its causes, severity levels used for congestion, and a glossary of congestion terms.

Document feedback

Quality is our first concern at Broadcom, and we have made every effort to ensure the accuracy and completeness of this document. However, if you find an error or an omission, or you think that a topic needs further development, we want to hear from you.

Send your feedback to documentation.pdl@broadcom.com.

Provide the publication title, part number, and as much detail as possible, including the topic heading and page number if applicable, as well as your suggestions for improvement.

Congestion Troubleshooting Overview

What is congestion?

Congestion occurs when frames enter the fabric faster than they exit the fabric. As a result, frames build up in the fabric's switches while waiting for transmission. This causes traffic moving through the fabric to slow down or become "congested." Congestion can occur on device links and inter-switch links (ISLs). Back pressure from a congested port in the fabric can cause traffic to slow down on upstream ISLs. This "congestion spreading" can cause traffic from unrelated flows that use the same ISL to slow down "victim" flows.

The performance capacity in the fabric is impacted because links carry data slower than they should. For example, moderate congestion could cause a 16 Gbps link to function at 4 Gbps. Severe congestion could cause that link to function at a few Kbps or over a million times slower than its capacity. Quality of service impacts due to mild-to-moderate congestion can include transmission delays, resulting in performance degradation due to increased I/O latency. Severe congestion may result in frame loss and eventually link resets. When fabric congestion reaches a level that causes frame loss, the storage protocol I/O timeouts, such as those for SCSI or NVMe over Fabrics (NVMeOF), can take up to 60 seconds. This congestion results in severe performance degradation or application failure.

Congestion becomes apparent when an application is underperforming or, in severe cases, when it fails.

NOTE

An understanding of Fibre Channel flow control is essential for a technical understanding of congestion. Refer to Fibre Channel flow control on page 27 for more information.

Symptoms of congestion

Symptoms to note in a mildly, moderately, or severely congested fabric are the following.

- Mild congestion—The traffic load is approaching the effective bandwidth of the link or device, but credit and queue latency remains low. There is typically no impact to the application performance that reaches a level that the application or user reports, but Brocade's Fabric Performance Impact (FPI) alerts may be triggered. However, these alerts could indicate that the application traffic load is increasing and may require a fabric or HBA upgrade to increase bandwidth. Action should be taken, based on the FPI alerts, before the traffic load or errors impact application performance to a reportable level.
- Moderate congestion—Your application is sluggish. While "sluggish" is subjective to the user, this may mean a noticeable lag in response time ranging from momentary to seconds. Congestion in the fabric is impacting the application traffic flows. This indicates that a corrective action or an upgrade to the fabric or HBA is needed to increase bandwidth capacity. As congestion increases from moderate to severe, users may notice decreases in application performance.
- Severe congestion—Your application has stopped functioning at an acceptable level.

NOTE
This classification of congestion symptoms is based on Brocade's expertise in Fibre Channel SAN and reflect Brocade's view of proper Fibre Channel SAN performance. These symptoms manifest differently depending on the normal operational status of the SAN deployment and the application's requirements.

Causes of congestion

Congestion can be caused by lost buffer credits, credit-stalled devices, and oversubscription.

- **Lost credit**—Lost credits are caused by links with physical layer errors. The effects of lost credits increase in severity as more credits are depleted. Lost credit can be initially identified by credit latency at a port or by queue latency upstream from the port. As the problem persists, frame loss due to timeouts occurs at the port or at upstream ports and can eventually lead to link resets, which occur after 2 seconds if all buffer credits are lost. Lost credits may affect unrelated flows in the fabric due to the congestion-spreading effect.

  Lost credits occur when the link between a sender and a receiver experiences errors that corrupt the receiver ready signal (R_RDY), causing the credit to be permanently lost. The R_RDY signal is sent from the receiver to the sender to indicate that it has processed a received frame and that the buffer is now available to receive a new frame. Lost credits can occur on device-to-switch links, inter-switch links (ISLs), and back-end switch ports. Lost credits prevent a transmitter from sending frames as fast as possible. This results in degraded throughput or, if all credits are lost, zero throughput on the link.
- **Credit-stalled device**—A credit-stalled device is a misbehaving device that stops returning R_RDY signals (buffer credits) to the switch. This causes the switch to stop sending frames to the device. A credit-stalled device that stops returning credits to the switch for hundreds of milliseconds or more causes frame drops and link resets. A credit-stalled device is sometimes referred to as a "slow-drain device."

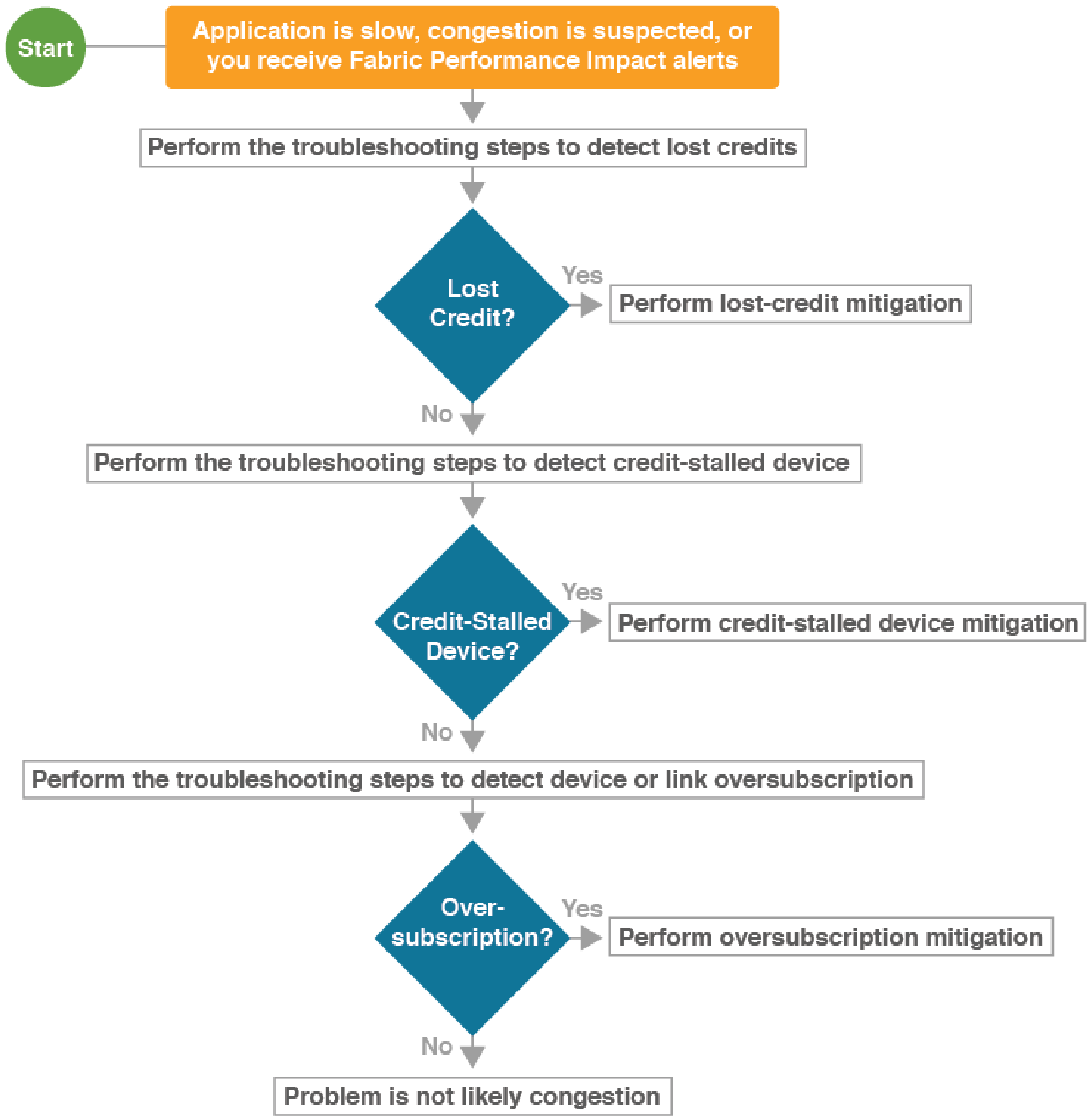
  Credit-stalled devices can be identified by credit latency or frame loss at a port. In the case of frame loss, the credit stall is long enough to cause queue latencies greater than 220 ms to 500 ms. Once frame loss occurs, application performance suffers severely and can be detected by users. If the credit-stalling behavior degrades significantly, link resets occur, indicating a credit stall for more than 2 seconds. Credit-stalled devices may affect flows unrelated to the misbehaving, credit-stalled device due to congestion spreading in the fabric.
- **Oversubscription**—Congestion due to oversubscription occurs when a fabric device or link is asked to handle more data than it can efficiently process, causing the surplus data to back up into the fabric. Congestion may also occur on ISLs when data flows from multiple devices exceed the capacity of the ISL, even though no individual device has requested more data than it can handle.

  Oversubscription is identified by queue latency on upstream port(s) and high bandwidth utilization at a downstream port. Congestion from oversubscription is typically caused by a bandwidth mismatch between the source and destination ports, such as a speed mismatch when 16 Gbps device is sending to a 4 Gbps device. This may affect flows that share the same path through the fabric.

Determining the cause of congestion

It is important to determine the cause of congestion in your fabric and perform proper mitigation before congestion becomes severe and leads to sluggish or stalled applications. The following flowchart and steps provide an overview of how to determine if your fabric congestion is caused by lost credits, credit-stalled devices, or oversubscription. Use them to guide troubleshooting; refer to Congestion Troubleshooting Examples on page 17 for additional details regarding each potential cause of congestion. Once the cause is determined, you can perform proper mitigation procedures as detailed in Lost credits on page 10, Credit-stalled devices on page 12, or Oversubscription on page 14.

FIGURE 1 Determining the cause of fabric congestion



Use the following steps to determine the cause of congestion using MAPS alerts and the MAPS dashboard "3.1 Summary Report" and "3.2 Rules Affecting Health" sections to display current port health and counts of triggered MAPS rules.

1. Determine if credit loss due to link errors is occurring on fabric ports on the flow path between a host and a target using the MAPS dashboard.
    a. Identify link errors on switch ports. Refer to Identifying lost credits on page 10.
    b. If link errors persist on ports over time, perform lost-credit mitigation by referring to Mitigating lost credits on page 11.

2. If link errors are not occurring, determine if F_Ports on the flow path between a host and a target are experiencing credit latency using the MAPS dashboard.
   a. Identify C3 frame timeouts, IO_PERF_IMPACT, and IO_FRAME_LOSS on switch ports. Refer to Identifying credit-stalled devices on page 12.
   b. If F_Ports are experiencing credit latency, perform credit-stalled device mitigation by referring to Mitigating credit-stalled devices on page 13.
3. If link errors or credit latency are not occurring on ports, the likely cause of congestion is oversubscription. Use the MAPS dashboard to determine if queue and credit latency is occurring on E_Ports and if high bandwidth utilization is occurring on F_Ports with flows traversing the impacted E_Port's ISL.
   a. Identify high bandwidth utilization, C3 frame timeouts, IO_PERF_IMPACT, and IO_FRAME_LOSS on switch ports. Refer to Identifying oversubscription on page 14.
   b. If oversubscription is verified, perform oversubscription mitigation by referring to Mitigating oversubscription on page 15.

   NOTE
   If your evaluation eliminates lost credit, credit-stalled devices, or oversubscription as the causes of application or fabric performance issues, then congestion is not likely the cause.

Lost credits

Link errors, such as CRC (cyclic redundancy check) and ITW (invalid transmission word) errors, can affect frame integrity and cause valid R_RDYs to stop being returned to the sending switch, which ultimately causes lost credits on device-to-switch links, inter-switch links (ISLs), and back-end switch ports. Lost credits are caused by the signal being degraded by the transmitter, the cable system, or the receiver, such that the transmitted data cannot be successfully recovered.

This section provides information on using MAPS to identify lost credits and provides troubleshooting and mitigation instructions.

Identifying lost credits

MAPS monitors and provides alerts for various link errors and statistics through the Port Health category to notify you about the following link errors occurring on switch ports:

- Loss of sync
- ITW
- CRC
- Loss of signal

You can use the MAPS dashboard "3.1 Summary Report" and "3.2 Rules Affecting Health" sections to display current port health and counts of triggered MAPS rules on all switch ports.

To display the MAPS dashboard, use the **mapsdb --show** command. For more information on the MAPS dashboard and creating MAPS rules and policies, refer to the "Port Health" section and "MAPS dashboard overview" sections of the *Brocade Monitoring and Alerting Policy Suite Configuration Guide*.

Troubleshooting lost credits

If you are experiencing performance problems with an application running on a host, perform the following steps along the flow path between the host and storage device.

NOTE
For examples of using MAPS tools to identify lost credit, refer to Lost Credit on page 17.

1. The most common link errors that occur with lost credits are CRC or ITW errors. Other link errors can include link reset, loss of sync, and loss of signal. Check for MAPS alerts warning that these errors have exceeded thresholds set in MAPS rules, such as the number of errors occurring per minute. MAPS alerts display through RASLog messages, email, SNMP, or whatever you set as an action through MAPS for generating alerts.
2. Use the **mapsdb --show** command to display the MAPS dashboard to find more details on these alerts.
   a) Check the dashboard "3.1 Summary Report" section to see if the Port Health category is out of operating range.
   b) Check the dashboard "3.2 Rules Affecting Health" section for incidents of MAPS rules exceeding thresholds for link errors on the same F_Ports identified in the MAPS alerts that you have received.
3. Determine if link errors are indicated on host HBA or storage device ports using tools available for those devices.
4. Perform a fault isolation analysis using prescribed Brocade tools and procedures to isolate the cause of errors to a switch port, host device, storage device, or cable and link components between switches and devices. If problems originate at the host or storage device, use tools and documentation available for the device for further troubleshooting and problem resolution.

Mitigating lost credits

Link errors that lead to lost credits are generally caused by faulty cables, faulty or improperly seated transceivers, and faulty connections. Connections between internal back-end links between ASICs can also suffer from physical errors.

In summary, to mitigate the effects of congestion due to lost credits, perform the following tasks:

- Maintain cables, transceivers, connections, and the network infrastructure.
- Enable credit recovery.
- Launch ClearLink (D_Port) Diagnostics.

The following best practices are recommended for mitigating lost credits:

- Deploy credit recovery:
  - Make sure that you upgrade to Fabric OS v7.2.1 or later on Gen 4 platforms to enable lost credit recovery for front-end ports.
  - Enable lost credit recovery on front-end and back-end links.
  - Deploy Gen 5 and Gen 6 switches and devices that support buffer credit recovery.
- Regularly maintain cables, connections, transceivers, and other network components. Reseat or replace transceivers for ports with significant physical link errors and lost credits. Clean connectors or replace the cable between ports with significant physical link errors and lost credit.
- Run ClearLink (D_Port) Diagnostic tests on the supported switches and devices to validate the link connections before application deployments and to isolate faulty components, such as a small form-factor pluggable (SFP) transceiver or a cable.
- For Gen 5 and Gen 6 devices, always employ at least two member trunks that use Brocade Trunking on front-end 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 event).

Although the conditions leading to lost credit are not controllable by Fabric OS, Fabric OS can detect the condition and recover in some fashion. Some recovery methods are more rudimentary such as a link reset (LR), while others are more sophisticated such as credit recovery.

- Permanent loss of all credits on a port can be handled through either a manual or automatic link reset (LR) on the port. A manual link reset can ensure that partial lost credit is recovered.

- For Gen 5 and later platforms, the Brocade buffer credit recovery feature allows links to automatically recover credits after buffer credits are lost. The feature also maintains performance. If a credit is lost, a recovery attempt is initiated.

  The credit recovery feature monitors and detect lost credit in traffic flow on front-end F_Ports, E_Ports, EX_Ports and back-end links (internal links between ASICs). Alerts are generated when lost credit and link resets occur. You can use this information to determine if credit starvation is due to downstream congestion or actual permanent lost credit. If the number of available credits returns to the maximum assigned to a port or VC, credit starvation is caused by congestion. If credits return to a value lower than the assigned maximum by one or more credits, this indicates permanent lost credit.

Following are the credit-recovery mechanisms supported by product generation:

- Buffer credit recovery with end devices—Gen 6
- Buffer credit recovery with switches—Gen 5 and Gen 6
- Hardware-assisted buffer credit recovery with end devices—Gen 5 and Gen 6

It should be noted that automatic recovery mechanisms, such as buffer credit recovery and link resets, mitigate the effects of lost credit, but should not be considered a solution to lost credits. If link resets and credit recovery are occurring, but performance issues do not indicate congestion, you should still investigate the cause and properly maintain cables, transceivers, connections, and the network infrastructure to reduce credit loss.

Credit-stalled devices

Credit-stalled devices have abnormal device behaviors that withholds credit from the fabric. Commonly, the problem is a defect in the device programming or operation. Therefore, issues should be escalated to the device vendor. Fabric Performance Impact (FPI) monitoring, a MAPS feature, uses frame delivery timeouts and credit latency to identify credit-stalled devices. MAPS generates FPI alerts when either the congestion levels or port latencies meet or exceed the thresholds specified in MAPS rules. Alerts may be generated through SNMP, RASLog, e-mail, or whatever you set as an action through MAPS for generating alerts.

This section provides information on using MAPS to identify credit-stalled devices and provides troubleshooting and mitigation instructions.

Identifying credit-stalled devices

The following types of FPI alerts are generated for F_Ports to indicate credit stall at a connected device:

- IO_PERF_IMPACT—This state can occur if the port does not have credit for a substantial period time or if frames are transmitted with delay. A port or device in this state can negatively impact overall network performance. You should take action to mitigate the effects of this moderate latency condition.
- IO_FRAME_LOSS—This state indicates a severe level of latency. In this state, frame timeouts either have already occurred or are very likely to occur. This is a severe issue and can impact the traffic due to backpressure. You should take immediate action to prevent application interruption.

You can use the MAPS dashboard "3.1 Summary Report" and "3.2 Rules Affecting Health" sections to display the current FPI status and counts of triggered MAPS rules for C3 frame timeouts, IO_PERF_IMPACT, and IO_FRAME_LOSS on switch ports.

To display the MAPS dashboard, use the **mapsdb --show** command. For more information on the MAPS dashboard and creating MAPS rules and policies, refer to the "Port Health" and "MAPS dashboard overview" sections of the *Brocade Monitoring and Alerting Policy Suite Configuration Guide*.

Troubleshooting credit-stalled devices

If you are experiencing performance issues, have eliminated lost credits as the cause, and are receiving IO_PERF_IMPACT and IO_FRAME_LOSS alerts at an F_Port, perform the following steps along the flow path between the host and the storage device.

> NOTE
> For examples of using MAPS tools to identify credit-stalled devices, refer to Credit-Stalled Devices on page 20.

1. Check for any of the following MAPS alerts occurring on an F_Port connected to a device, such as a host HBA. MAPS alerts display through RASLog messages, email, SNMP, or whatever you set as an action through MAPS for generating alerts:
   - DEVICE_LATENCY_IMPACT==IO_PERF_IMPACT conditions
   - DEVICE_LATENCY_IMPACT==IO_FRAME_LOSS conditions
   - Link resets
   - Class 3 frame timeouts (C3TXTO) that have exceeded thresholds
2. Use the **mapsdb --show** command to display the MAPS dashboard and for more details on these alerts.
   a) Check the dashboard "3.1 Summary Report" section to see if the FPI category is out of operating range.
   b) Check the dashboard "3.2 Rules Affecting Health" section for incidents when MAPS rules exceeded thresholds for C3 frame timeouts and to verify that FPI IO_LATENCY and IO_FRAME_LOSS are on the same F_Ports identified in the MAPS alerts that you have received.

Mitigating credit-stalled devices

After you determine that the cause of congestion is from credit-stalled devices, you can employ the mitigation techniques summarized in this section. However, to completely resolve the effect of congestion due to a credit-stalled device, the only solution is to stop credit-stalled behavior on all devices connected to the switch. *Therefore, it is important for you to escalate the issue of credit-stalled devices to the attention of the device's vendor to determine the root cause and for complete resolution.*

In summary, to reduce the effects of the credit-stalled device on fabric congestion and other device traffic until the issue with the credit-stalled device is resolved, perform the following tasks:

- Reset the credit-stalled device.
- Isolate the credit-stalled device.
- Remove the credit-stalled device.

Use the following Fabric OS features to mitigate the effects of the credit-stalled device until the problem can be remedied or the device is replaced:

- Isolate traffic from the credit-stalled device using the Brocade Slow-Drain Device Quarantine (SDDQ) feature. This action is automatically triggered when FPI detects an F_Port in either the IO_PERF_IMPACT state or the IO_FRAME_LOSS state.
- Isolate traffic from the credit-stalled device using QoS zoning. This requires that you manually configure QoS zones to manage traffic priority between specific host-target pairs.
- Enable the Brocade port toggle, fencing, or decommissioning features.

For more information on the Fabric OS SDDQ, QoS zoning, port toggling, port decommissioning, and port fencing features, refer to the *Brocade Monitoring and Alerting Policy Suite Configuration Guide*.

Oversubscription

Normal SAN configurations are oversubscribed by design under the assumption that hosts will reference storage independently rather than simultaneously. Use standard SAN design best practices, such as fan-in/fan-out and speed matching, to avoid congestion caused by oversubscription. Congestion caused by oversubscription is characterized by high queue or credit latency occurring at a switch E_Port and high bandwidth utilization (high rates of frame transmission) at a downstream F_Port in the same traffic flow path as the E_Port. In some fabric configurations, high bandwidth utilization, IO_PERF_IMPACT, and IO_FRAME_LOSS alerts may occur on multiple E_Ports. While the impact to the E_Port(s) is similar to the congestion spreading caused by a credit-stalled device, with device oversubscription, the downstream F_Port(s) do not exhibit abnormal credit latency. These ports will exhibit only high bandwidth utilization. Although periods of high bandwidth utilization on an F_Port are not necessarily abnormal, they can be used to identify devices suspected of being oversubscribed after verifying that lost credits and a credit-stalled device are not causing the upstream E_Port congestion.

This section provides information on using MAPS to identify oversubscription and provides troubleshooting and mitigation instructions.

Identifying oversubscription

To identify oversubscription, you can use the MAPS dashboard "3.1 Summary Report" and "3.2 Rules Affecting Health" sections to display the current Fabric Performance Impact status and counts of triggered MAPS rules for C3 frame discards, Fabric Performance Impact, and bandwidth utilization on switch ports.

NOTE
To display the MAPS dashboard, use the **mapsdb --show** command. On Gen 6 and later switches, use the **mapsdb --show congestion** command to display the Congestion dashboard. This dashboard displays a list of potentially congested ports with congestion summary information. For more information on MAPS dashboards and creating MAPS rules and policies, refer to the "Port Health" section and "MAPS dashboard overview" sections of the *Brocade Monitoring and Alerting Policy Suite Configuration Guide*.

Troubleshooting oversubscription

Use the following steps along the flow path between the host and storage device to confirm that oversubscription exists and whether the path of the oversubscription would impact the I/O for the application.

NOTE
For examples of using MAPS tools to identify oversubscription, refer to Oversubscription on page 22.

1. Check for any of the MAPS DEVICE_LATENCY_IMPACT alerts occurring on E_Ports, indicating high credit and transmit queue latency.
2. Check for Fabric Performance Impact alerts occurring on downstream F_Ports used by the same traffic flows as the E_Ports identified in Step 1 that indicate high transmit bandwidth utilization (high rates of frame transmission).
3. Use the **mapsdb --show** command to display the MAPS dashboard and more details on these alerts.
   a) Check the dashboard "3.1 Summary Report" section to see if the Fabric Performance Impact category is out of operating range.
   b) Check the dashboard "3.2 Rules Affecting Health" section for incidents of MAPS rules exceeding thresholds, indicating high bandwidth utilization and FPI alerts on the same F_Ports identified in the MAPS alerts that you have received.
4. Check the dashboard "3.2 Rules Affecting Health" section for incidents of FPI alerts that indicate IO_LATENCY or IO_FRAME_LOSS on the identified E_Ports.
5. Determine if the identified E_Ports are used by traffic flows that are related to an application that is experiencing poor or "sluggish" performance.

Mitigating oversubscription

Congestion due to oversubscription is primarily a result of improper SAN design. When oversubscription is identified as the source of congestion, the problem should be addressed by correcting the design.

To mitigate the effects of congestion due to oversubscription, perform the following tasks:

- Reduce link speed mismatch ratios between the target and the host end device.
- Reduce fan-in.
- Reduce queue depths.
- Add ISLs.

Common design mistakes include link-speed mismatches, a high storage target fan-in ratio, or 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, particularly when different generations of servers, storage, and SAN switches are deployed in the same network. Solutions might include reducing the number of hosts per target or upgrading the server HBA speed. Refer to *Brocade SAN Design and Best Practices Deployment Guide* for a detailed discussion of how to properly design a SAN.

Consider the following practices for mitigating oversubscription:

- For ISL oversubscription:
  - Follow recommendations in the "Data Flow Considerations" section of the *Brocade SAN Design and Best Practices Deployment Guide* for fan-in ratios, oversubscription, and congestion.
  - If possible, connect server and storage devices with a high traffic load between them to the same switch. This will localize traffic and reduce ISL oversubscription.
  - Implement traffic isolation (TI) zones to isolate critical traffic. Refer to the *Brocade Fabric OS Administration Guide* for more information.
  - Isolate legacy 2 Gbps devices with traffic localization on the same switch ASIC ports or with a dedicated ISL.
  - Avoid locating devices with more than two different speed generations on ISLs in a single fabric.
- For device port oversubscription:
  - Follow recommendations in the "Data Flow Considerations" section of the *Brocade SAN Design and Best Practices Deployment Guide* for fan-in ratios, oversubscription, and congestion.
  - If the oversubscribed device is a server, consider reducing the queue depth on the server. The queue depth is the number of pending I/O requests being sent by the server to multiple targets to improve efficiency and avoid delays.
  - Decrease the number of devices sending traffic to the oversubscribed device.
  - Upgrade the device to a faster port speed.

Congestion Troubleshooting Examples

This chapter contains examples of congestion caused by lost credit, a credit-stalled device, and oversubscription and examples of using MAPS features to detect these congestion causes.

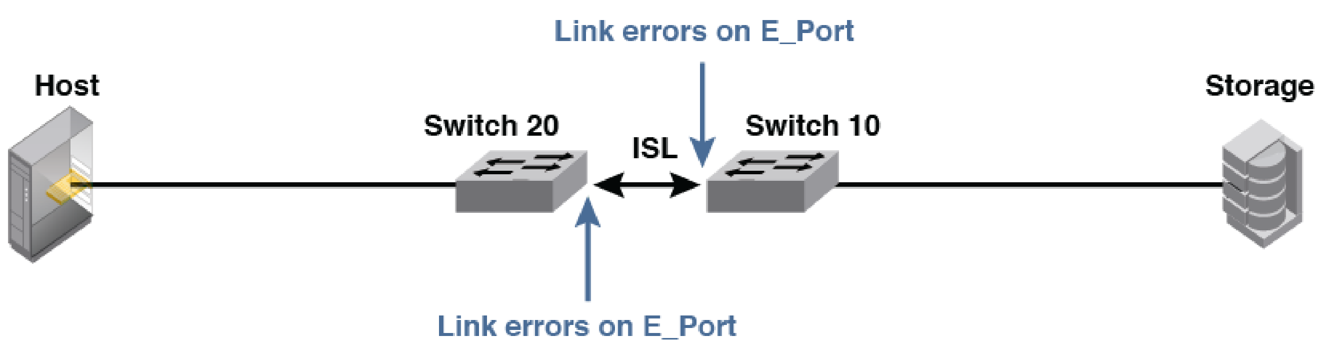
Lost Credit

Sluggish application performance can be due to a lost-credits condition on the link between the host and the switch, the ISL between two switches, or the link between a switch and a storage device. Following are examples of congestion due to lost-credit conditions and using MAPS to detect the link errors in the host application's path that cause lost credit.

Lost credits on an ISL and MAPS detection

The following illustration depicts a lost-credit condition occurring on the ISL between two switches, 10 and 20, which can be determined by the link errors occurring on a switch E_Ports. The port with errors can help isolate the port at fault. For example, errors on switch 10 can be caused by faulty transmission from the E_Port on switch 20. Link errors may also be caused by faulty transmission medium, which includes cable and connections between the ports, or by faulty transmitter or receiver hardware in either switch. Use the Brocade MAPS dashboard to identify the fabric E_Ports with link errors.

FIGURE 2 Lost credits between switch E_Ports



Following are examples of RASLog alerts received for E_Port 2 on switch 10 that warn of CRC and ITW errors exceeding error-per-minute thresholds set in MAPS rules.

```
2018/03/19-16:21:11, [MAPS-1003], 19516, FID 101, WARNING,
sw10______FID_101______auto__, port2, E-Port 2, Condition=ALL_E_PORTS(CRC/min>0),
Current Value:[CRC, 142 CRCs], RuleName=defALL_E_PORTSCRC_0, Dashboard Category=Port Health.

2018/03/19-16:31:47, [MAPS-1003], 2, FID 101, WARNING,
sw10______FID_101______auto__, port2, E-Port 2, Condition=ALL_E_PORTS(ITW/min>15),
Current Value:[ITW, 212 ITWs], RuleName=defALL_E_PORTSITW_15, Dashboard Category=Port Health.
```

Following is an example of the "3.1 Summary" and "3.2 Rules Affecting Health" sections of the MAPS dashboard. Note the following in the examples that indicate lost credits on the ISL:

- The "3.1 Summary Report" indicates that Port Health is out of operating range for the day. When a category contains an "out-of-range" error, the dashboard displays the "3.2 Rules Affecting Health" section showing the rules triggered in that category since midnight.
- The "3.2 Rules Affecting Health" section shows the number of times a MAPS rule was violated in the past hour in the Repeat Counts column and the five most recent counts of triggered rules in the Triggered Value (Units) column. Note the rules for ITW and CRC exceeding thresholds for Port Health occurring on E_Port 2 , which is the same port identified in the MAPS RASLog alert message. These errors caused Port Health to be "out of operating range."

  NOTE
  The column headings in the following example have been edited slightly and output is truncated to allow the example to display clearly.

```
switch10:admin> mapsdb --show

-------------------------------------------Output truncated-----------------------------------------------

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

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


3.2 Rules Affecting Health:
===========================

Category         |RepeatCount|Rule Name           |Execution Time   |Object    |Triggered   |
(Violation Count)|           |                    |                 |          |Value(Units)|
--------------------------------------------------------------------------------------------
Port Health(24)  |1          |defALL_E_PORTSITW_20|03/19/18 16:31:47|E-Port 2  |212 ITWs    |
                 |1          |defALL_E_PORTSITW_15|03/19/18 16:31:47|E-Port 2  |212 ITWs    |
                 |1          |defALL_E_PORTSCRC_2 |03/19/18 16:21:11|E-Port 2  |142 CRCs    |
                 |1          |defALL_E_PORTSCRC_0 |03/19/18 16:21:11|E-Port 2  |142 CRCs    |
```

Lost credits on the switch-to-device link and MAPS detection

The following illustrations depict a lost-credit condition occurring on the link between a switch and a device. The first illustration shows link errors between an F_Port on switch 20 and a host's HBA port. The second illustration depicts link errors between an F_Port on switch 10 and a storage device. Link errors, which indicate a lost-credit condition, may be detected at the switch port, the device port, or both ports. Errors on the switch F_Port can be caused by faulty transmission from the device port. Link errors on the device may be caused by faulty transmission from the switch port. Link errors can be caused by faulty transmission medium, which includes cable and connections between the ports, or faulty transmitter or receiver hardware. Use tools available for your host or storage device to determine if link errors are indicated on the device. Use the Brocade MAPS dashboard to determine if errors are indicated on the switch port.

FIGURE 3 Lost credits on link between switch F_Port and HBA

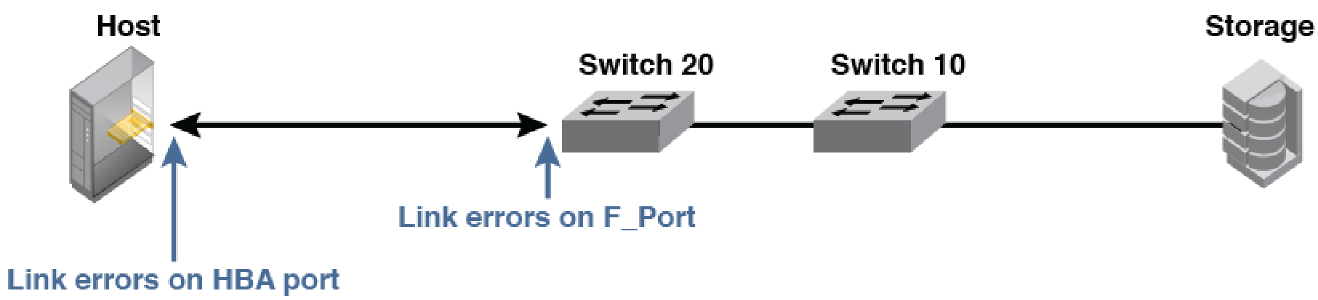
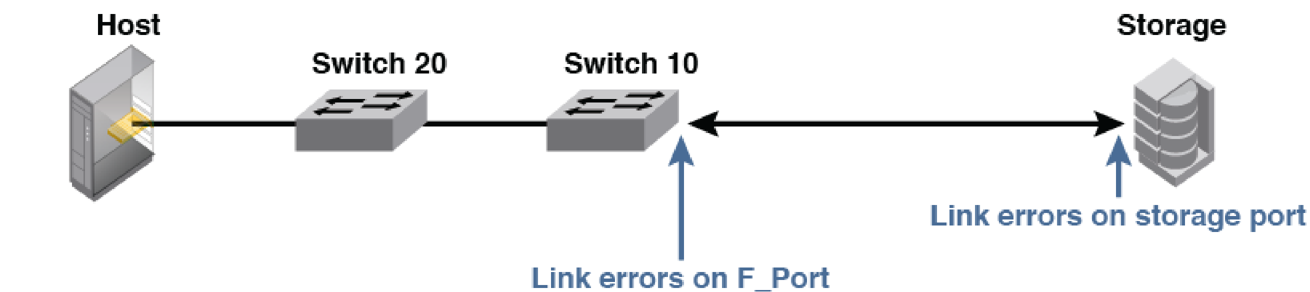


FIGURE 4 Lost credits between switch F_Port and storage device



Following are examples of RASLog alerts received for F_Port 35 on switch 20 that warn of CRC and ITW errors exceeding error-per-minute thresholds set in MAPS rules. The first alert shows 142 CRC errors, which exceeds the rule set for CRC errors per minute at greater than zero. The second alert shows 212 ITW errors, which exceeds the rule set for ITW errors per minute at greater than 15.

```
2018/03/19-17:20:26, [MAPS-1003], 102, FID 101, WARNING,
sw20___SDD_____fid_101________, port35, F-Port 35, Condition=ALL_HOST_PORTS(CRC/min>0),
Current Value:[CRC, 142 CRCs], RuleName=defALL_HOST_PORTSCRC_0, Dashboard Category=Port Health.

2018/03/19-17:23:08, [MAPS-1003], 105, FID 101, WARNING,
sw20___SDD_____fid_101________, port35, F-Port 35, Condition=ALL_HOST_PORTS(ITW/min>15),
Current Value:[ITW, 212 ITWs], RuleName=defALL_HOST_PORTSITW_15, Dashboard Category=Port Health.
```

Following is an example of a RASLog alert for a link reset initiated by lost credits on F_Port 35. Note that this error will not be included in the following MAPS dashboard example.

```
2018/03/23-19:07:47, [C4-1014], 2880, SLOT 2 | CHASSIS, WARNING, stats_switch20,  Link Reset on Port
S3,P-1(35) vc_no=5 crd(s)lost=3
auto trigger. Link Reset performed on slot 3 blade port 35
```

Following is an example of the "3.1 Summary" and "3.2 Rules Affecting Health" sections of the MAPS dashboard. Note the following in the examples that indicate lost credits on the link between the F_Port and the host:

- The "3.1 Summary Report" indicates that Port Health is out of operating range for the day. When a category contains an "out-of-range" error, the dashboard displays the "3.2 Rules Affecting Health" section showing the rules triggered in that category since midnight.

- The "3.2 Rules Affecting Health" section shows the number of times a rule was violated in the past hour in the Repeat Counts column and the five most recent counts of errors that triggered MAPS rules in the Triggered Value (Units) column. The incidents of ITW and CRC rules exceeding thresholds on F_Port 35 caused Port Health to be "out of operating range."

  NOTE
  The column headings in the following example have been edited slightly and output is truncated to allow the example to display clearly.

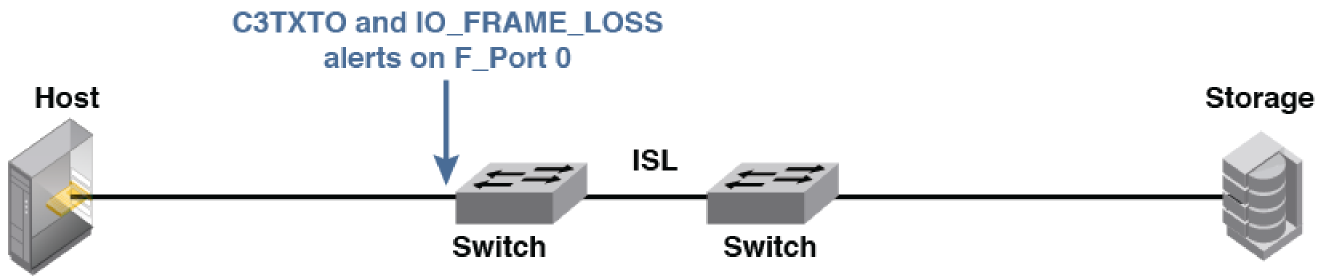
```
switch20:admin> mapsdb --show

-------------------------------------------Output truncated------------------------------------------------

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

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


3.2 Rules Affecting Health:
===========================

Category         |RepeatCount|Rule Name               |Execution Time   |Object    |Triggered   |
(Violation Count)|           |                        |                 |          |Value(Units)|
-----------------------------------------------------------------------------------------------
Port Health(4)   |1          |defALL_HOST_PORTSITW_20|03/19/18 17:23:08|F-Port 35|212 ITWs    |
                 |1          |defALL_HOST_PORTSITW_15|03/19/18 17:23:08|F-Port 35|212 ITWs    |
                 |1          |defALL_HOST_PORTSCRC_2 |03/19/18 17:20:26|F-Port 35|142 CRCs    |
                 |1          |defALL_HOST_PORTSCRC_0 |03/19/18 17:20:26|F-Port 35|142 CRCs    |
```

Credit-Stalled Devices

A credit-stalled device is indicated by Fabric Performance Impact alerts on the switch F_Port connected to the device, indicating severe credit latency. Following is an example of congestion due to a credit-stalled device and the use of MAPS to detect increased credit latency on the device port.

The following illustration depicts a credit-stalled device (host HBA) connected to a switch at F_Port 0. MAPS has triggered C3 transmit timeout and IO_FRAME_LOSS alerts, indicating severe credit latency.

FIGURE 5 Credit-stalled device example



Following are a series of RASLog alerts received over time for F_Port 0 on switch 062165 that indicate credit latency and a credit-stalled device attached to the port. Note the alerts for C3 timeouts, the link reset alert for lost frames, and the critical IO_FRAME_LOSS alert.

The following alert shows that 18 C3 transmit timeouts occurred in a minute, which violates the MAPS rule set for 11 timeouts per minute.

```
2018/03/15-12:21:52, [MAPS-1002], 122, FID 128, ERROR, sw062165, port0, F-Port 0,
Condition=ALL_HOST_PORTS(C3TXTO/min>11), Current Value:[C3TXTO, 18 Timeouts],
RuleName=defALL_HOST_PORTSC3TXTO_11, Dashboard Category=Port Health.
```

The following alert shows that an IO_FRAME_LOSS condition has occurred due to the large number of C3 transmit timeouts.

```
2018/03/15-12:22:04, [MAPS-1001], 124, FID 128, CRITICAL, sw062165, port0, F-Port 0,
Condition=ALL_PORTS(DEV_LATENCY_IMPACT==IO_FRAME_LOSS), Current Value:[DEV_LATENCY_IMPACT,
IO_FRAME_LOSS,  (64 C3TX Timeouts) ], RuleName=defALL_PORTS_IO_FRAME_LOSS_UNQUAR, Dashboard
Category=Fabric Performance Impact.
```

The following alert shows that 46 C3 transmit timeouts occurred in one minute, which violates the MAPS rule set for 11 timeouts per minute.

```
2018/03/15-12:22:52, [MAPS-1002], 125, FID 128, ERROR, sw062165, port0, F-Port 0,
Condition=ALL_HOST_PORTS(C3TXTO/min>11), Current Value:[C3TXTO, 46 Timeouts],
RuleName=defALL_HOST_PORTSC3TXTO_11, Dashboard Category=Port Health.
```

Following is an example of the "3.1 Summary Report" and "3.2 Rules Affecting Health" sections of the MAPS dashboard. Note the following in the example that indicate severe credit latency at F_Port 0 and a credit-stalled device connected to the port:

- The "3.1 Summary Report" indicates that Fabric Performance Impact is out of operating range for the day. When a category contains an "out-of-range" error, the dashboard displays the "3.2 Rules Affecting Health" section, showing the rules triggered in that category since midnight.
- The "3.2 Rules Affecting Health" section shows the number of times that a rule threshold was exceeded in the past hour in the Repeat Counts column and the five most recent counts of rule thresholds being exceeded in the Triggered Value (Units) column. Note the following progression of incidents in the example:
  - Class 3 frame transmit timeouts (C3TXTO) incidents
  - Fabric Performance Impact (FPI) incidents, which includes IO_FRAME_LOSS

  NOTE
  The column headings in the following example have been edited slightly and output is truncated to allow the example to display clearly.

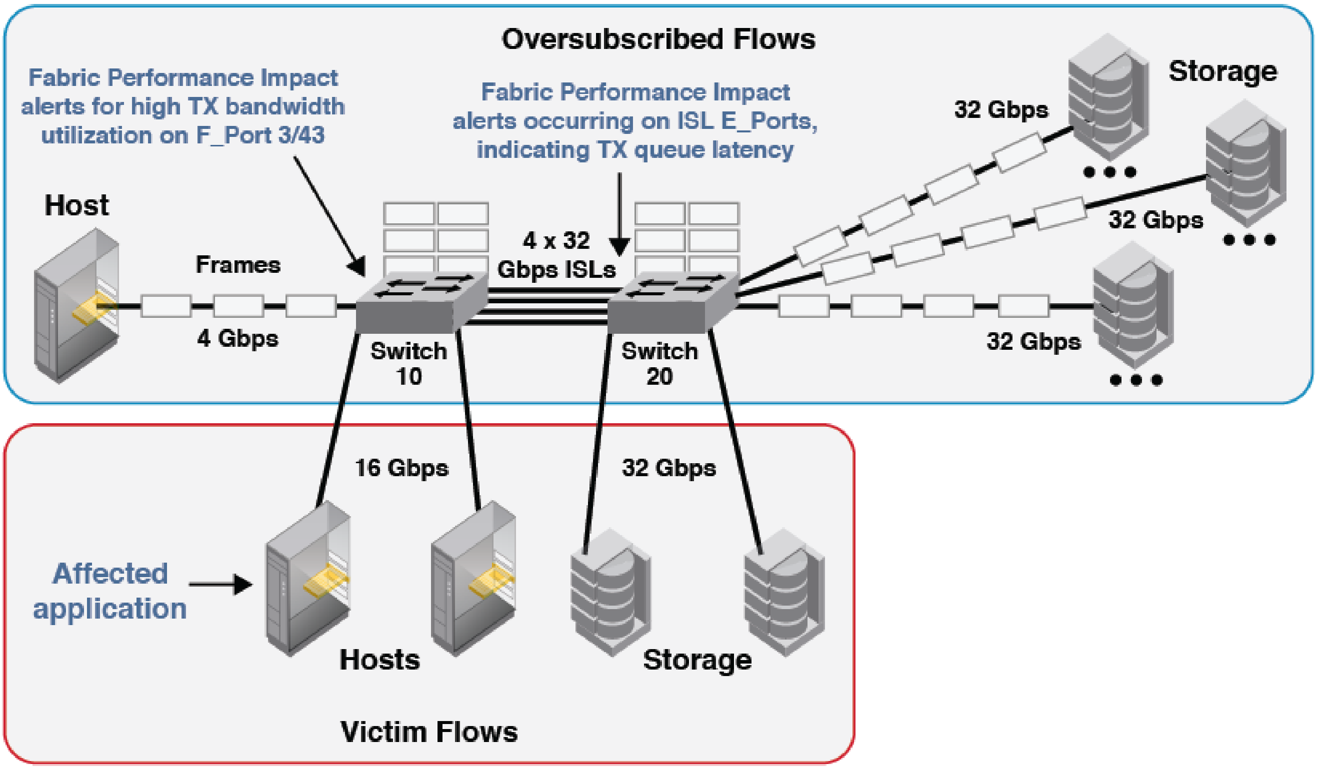
```
sw062165:admin> mapsdb --show

--------------------------------Output truncated-----------------------------------------------
```

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

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


3.2 Rules Affecting Health:
===========================

Category          |Repeat|Rule Name                   |Execution Time    |Object    |Triggered     |
(Violation Count) |Count |                            |                  |          |Value (Units)|
-------------------------------------------------------------------------------------------------
Port Health(102)  |1     |defALL_HOST_PORTSC3TXTO_20  |03/15/18 12:21:58|F-Port 0  |64 Timeouts   |
                  |2     |defALL_HOST_PORTSC3TXTO_11  |03/15/18 12:22:52|F-Port 0  |46 Timeouts   |
                  |      |                            |                  |F-Port 0  |18 Timeouts   |
                  |      |                            |                  |          |              |
                  |31    |defALL_HOST_PORTSC3TXTO_20  |03/15/18 11:56:58|F-Port 0  |64 Timeouts   |
                  |      |                            |                  |F-Port 0  |122 Timeouts  |
                  |      |                            |                  |F-Port 0  |64 Timeouts   |
                  |      |                            |                  |F-Port 0  |65 Timeouts   |
                  |      |                            |                  |F-Port 0  |64 Timeouts   |
                  |32    |defALL_HOST_PORTSC3TXTO_11  |03/15/18 11:57:22|F-Port 0  |64 Timeouts   |
                  |      |                            |                  |F-Port 0  |58 Timeouts   |
                  |      |                            |                  |F-Port 0  |64 Timeouts   |
                  |      |                            |                  |F-Port 0  |64 Timeouts   |
                  |      |                            |                  |F-Port 0  |129 Timeouts  |
Fabric            |1     |defALL_PORTS_IO_LATENCY_    |03/15/18 12:23:04|F-Port 0  |IO_LATENCY_   |
Performance       |      |CLEAR                       |                  |          |CLEAR         |
Impact (8)        |1     |defALL_PORTS_IO_FRAME_LOSS_|03/15/18 12:22:04|F-Port 0  |IO_FRAME_LOSS|
                  |      |UNQUAR                      |                  |          |              |
                  |2     |defALL_PORTS_IO_FRAME_LOSS_|03/15/18 11:34:04|F-Port 0  |IO_FRAME_LOSS|
                  |      |UNQUAR                      |                  |          |              |
                  |3     |defALL_PORTS_IO_LATENCY_    |03/15/18 11:58:04|F-Port 0  |IO_LATENCY_   |
                  |      |CLEAR                       |                  |          |CLEAR         |
                  |      |                            |                  |F-Port 0  |IO_LATENCY_   |
                  |      |                            |                  |          |CLEAR         |
                  |      |                            |                  |F-Port 0  |IO_LATENCY_   |
                  |      |                            |                  |          |CLEAR         |
                  |1     |defALL_PORTS_IO_FRAME_LOSS  |03/15/18 11:24:04|F-Port 0  |IO_FRAME_LOSS|
```

Oversubscription

The following illustration depicts an example of congestion caused by oversubscription occurring in a fabric consisting of multiple storage targets communicating with multiple hosts. Hosts on 4 Gbps and 16 Gbps links are connected to switch 10 and multiple storage devices on 32 Gbps links are connected to switch 20. All hosts are performing read requests to multiple storage devices on 32 Gbps links connected to switch 20, so traffic is flowing from storage to the hosts.

FIGURE 6 Oversubscription example—multiple targets to multiple hosts



In this example, MAPS alerts indicate that congestion is occurring because the 4 Gbps host link from switch 10 is oversubscribed:

- Fabric Performance Impact alerts for high transmit (TX) bandwidth utilization are occurring on F_Port 3/43, which connects the 4 Gbps host connected to switch 10.
- Fabric Performance Impact alerts are occurring on ISL E_Ports on switch 20 indicating transmit queue latency.

The following actions are causing congestion due to oversubscription in this example:

- A host connected to switch 10 through a 4 Gbps HBA has requested large read I/O from storage.
- Traffic from the faster storage devices overwhelms the slower 4 Gbps link, causing frames to back-up in switch 10 buffers.
- Switch 10 slows down incoming traffic on the ISLs from switch 20 using FC flow control to avoid frame loss due to buffer overflow. As a result, frames back up on switch 20 buffers.
- The two 16 Gbps hosts connected to switch 10 and storage devices connected to switch 20 become "victims" of congestion occurring in the fabric due to the oversubscribe 4 Gbps host link and resulting switch congestion.
  - Frame flow to the 16 Gbps hosts decreases to a point where application performance is impacted.
  - Frame flow also decreases from 32 Gbps storage devices sharing the same ISLs from switch 20 to switch 10.

High bandwidth utilization on the F_Port and MAPS detection

Following are RASLog Fabric Performance Impact alerts received on switch 10 that indicate high transmit bandwidth utilization on F_Port 3/43, which is connected to the 4 Gbps host.

The following Fabric Performance Impact warning shows that the current value of 98.82 percent transmit utilization triggered the MAPS rule of 95 percent utilization per minute 38 times in the past hour on F_Port 3/43.

```
2018/05/10-20:56:00:619991, [MAPS-1005], 2/2, SLOT 2 | FID 14, WARNING, SWITCH 10, slot3 port43, F-Port 
3/43,
Condition=ALL_HOST_PORTS(TX/min>95.00), Current Value:[TX, 98.82 %], Rule defALL_HOST_PORTSTX_95 triggered 
38 times in
1 hour and last trigger time Thu May 10 20:53:30 2018, Dashboard Category=Fabric Performance Impact., 
raslogAction.c,
line: 187, comp:raslog, ltime:2018/05/10-20:56:00:619724
```

The following Fabric Performance Impact warning shows that the current value of 98.79 percent transmit utilization triggered the MAPS rule of 95 percent on F_Port port 3/43.

```
2018/05/10-20:56:12:634962, [MAPS-2046], 3/3, SLOT 2 | FID 14, WARNING, SWITCH 10, slot3 port43, F-Port 
3/43,
Condition=ALL_HOST_PORTS(TX/min>95.00), Current Value:[TX, 98.79 %], RuleName=defALL_HOST_PORTSTX_95,
Dashboard Category=Fabric Performance Impact., raslogAction.c, line: 170, comp:raslog,
ltime:2018/05/10-20:56:12:634705
```

The following alert shows that the current value of 67.67 percent transmit utilization triggered the MAPS rule of 60 percent utilization per hour on F_Port 3/43.

```
2018/05/10-21:03:00:618264, [MAPS-2046], 4/4, SLOT 2 | FID 14, WARNING, SWITCH 10, slot3 port43, F-Port 
3/43,
Condition=ALL_HOST_PORTS(TX/hour>60.00), Current Value:[TX, 67.67 %], RuleName=defALL_HOST_PORTSTX_60,
Dashboard Category=Fabric Performance Impact., raslogAction.c, line: 170, comp:raslog,
ltime:2018/05/10-21:03:00:618097
```

The MAPS dashboard output for switch10 in this example indicates high transmit bandwidth utilization for F_Port 3/43:

- The "3.1 Summary Report" indicates that Fabric Performance Impact is out of operating range for the day.
- The "3.2 Rules Affecting Health" section shows the number of times a rule was violated in the past hour in the Repeat Counts column and the five most recent transmit bandwidth utilization percentages that triggered rules in the Triggered Value (Units) column. Note the following in the example for port 3/43, connected to the 4 Gbps link to the host:
  - The repeat counts and measured values that triggered the rule for transmit utilization bandwidth of greater than 95 percent per minute (defALL_HOST_PORTSTX_95). Note that there were 25 incidents of thresholds exceeding the 95-percent-per-minute rule at 21:30:06 in the past hour. Also, bandwidth utilization for the 4 Gbps link consistently triggered the 95-precent-per-minute rule with values above 97 percent.
  - The repeat counts and measured values that triggered the rule for transmit utilization bandwidth of greater than 60 percent per hour (defALL_HOST_PORTSTX_60).

  NOTE
  The column headings in the following example have been edited slightly and output is truncated to allow the example to display clearly.

```
sw10:admin> mapsdb --show

----------------------------------------Output truncated----------------

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

Category                 |Today                      |Last 7 days         |
--------------------------------------------------------------------------
Port Health              |In operating range         |No Errors           |
BE Port Health           |No Errors                  |No Errors           |
Extension GE Port Health |No Errors                  |No Errors           |
Fru Health               |In operating range         |In operating range  |
Security Violations      |No Errors                  |No Errors           |
Fabric State Changes     |In operating range         |No Errors           |
Switch Resource          |In operating range         |In operating range  |
Traffic Performance      |In operating range         |In operating range  |
```

```
Extension Health         |In operating range        |No Errors          |
Fabric Performance Impact|Out of operating range    |In operating range |


3.2 Rules Affecting Health:
===========================

Category             |Repeat|Rule Name                |Execution Time    |Object       |Triggered    |
(Violation Count)    |Count |                         |                  |             |Value (Units)|
-------------------------------------------------------------------------------------------------------
Fabric Performance   |1     |defALL_HOST_PORTSTX_60   |05/10/18 21:03:00|F-Port 3/43  |67.67 %      |
Impact (30)          |      |                         |                  |             |             |
                     |25    |defALL_HOST_PORTSTX_95   |05/10/18 21:30:06|F-Port 3/43  |98.93 %      |
                     |      |                         |                  |F-Port 3/43  |98.78 %      |
                     |      |                         |                  |F-Port 3/43  |98.82 %      |
                     |      |                         |                  |F-Port 3/43  |97.54 %      |
                     |      |                         |                  |F-Port 3/43  |98.82 %      |
                     |4     |defALL_HOST_PORTSTX_95   |05/10/18 20:59:12|F-Port 3/43  |98.81 %      |
                     |      |                         |                  |F-Port 3/43  |99.72 %      |
                     |      |                         |                  |F-Port 3/43  |98.07 %      |
                     |      |                         |                  |F-Port 3/43  |98.79 %      |
```

Transmit queue latency on E_Ports and MAPS detection

Following is the the RASLog alert received for E_Port 11/8 on switch 20. Note that although only the alert for port 11/8 is shown, alerts for the other ISL ports (11/9, 11/10, and 11/11) are exactly the same.

```
2018/05/10-20:41:00, [MAPS-2070], 1693, SLOT 1 | FID 14, WARNING, switch_20, slot11 port8, E-Port 11/8,
Condition=ALL_PORTS (DEV_LATENCY_IMPACT==IO_PERF_IMPACT), Current Value:[DEV_LATENCY_IMPACT,
IO_PERF_IMPACT,
(10 ms Frame Delay)], RuleName=defALL_PORTS_IO_PERF_IMPACT_UNQUAR, Dashboard Category=Fabric Performance
Impact.
```

Each E_Port on switch 20 is connected to a 32 Gbps ISL between switch 20 and 10 (refer to Oversubscription on page 22). Alerts are occurring on the switch 20 E_Ports since traffic is flowing from storage to hosts during read operations. These alerts provide DEV_LATENCY_IMPACT==IO_PERF_IMPACT warnings indicating severe latency on port transmit queues. The alerts indicate that defALL_PORTS_IO_PERF_IMPACT_UNQUAR rule thresholds have exceeded. When rules thresholds are exceeded, host traffic through the ports is placed into a quarantined state and is moved to lower-priority QoS virtual channels. A MAPS SDDQ action is required to move the traffic and an UNQUAR action is required if you want MAPS to remove the quarantine state.

The following example output from the **mapsdb --show congestion** command, which summarizes congestion conditions for switch 20. Note the Fabric Performance Impact incidents that have occurred for the switch 20 E_Ports. The output indicates four congested ports and provides a table summarizing Fabric Performance Impact alerts that have occurred from time 20:40 to 21:17

```
sw20:admin> mapsdb --show congestion
-------------------------------------------------------------
DB start time:  Thu May 10 20:40:10 2018
Total congested ports for time window (20:40 - 21:17): 4
-------------------------------------------------------------

                                 |Frequency details for time window (20:40 - 21:17)|
Port        |Current Min State   |Frame Loss  |Perf Impact |Medium |Low     |Info    |
---------------------------------------------------------------------------------
E-Port 11/11|Perf Impact         |0           |22          |15     |0       |0       |
E-Port 11/9 |Perf Impact         |0           |18          |19     |0       |0       |
E-Port 11/8 |Perf Impact         |0           |18          |19     |0       |0       |
E-Port 11/10|Medium St           |0           |7           |30     |0       |0       |
```

The MAPS dashboard output on switch 20 in this example indicates severe latency on E_Port transmit queues:

- The "3.1 Summary Report" indicates that Fabric Performance Impact is out of operating range for the day.

- The "3.2 Rules Affecting Health" section shows the number of times a rule was violated in the past hour in the Repeat Counts column and the five most recent counts of rule thresholds exceeding in the Triggered Value (Units) column. Note the following incidents of rule thresholds that have exceeded in the example for the switch 20 E_Ports:
  - There was a total of 18 Fabric Performance Impact incidents.
  - There were nine incidents of IO_PERF_IMPACT_UNQUAR in the past hour, indicating incidents when ports were removed from the quarantined state. These incidents indicate that host traffic through the ports is quarantined to a lower-priority QoS vc due to increased latency conditions at the port transmit queues.
  - There were nine incidents of the IO_LATENCY_CLEAR in the past hour, indicating that latency at the port transmit queues decreased to a point where the port quarantined states were cleared.

  NOTE
  The column headings in the following example have been edited slightly and output is truncated to allow the example to display clearly.

```
sw20:admin> mapsdb --show

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

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


 3.2 Rules Affecting Health:
 ===========================

 Category           |Repeat|Rule Name                   |Execution Time    |Object        |Triggered      |
 (Violation Count)  |Count |                            |                  |              |Value(Units)   |
 --------------------------------------------------------------------------------------------------------
 Fabric Performance |9     |defALL_PORTS_IO_LATENCY_CLE|05/10/18 20:55:00|E-Port 11/11 |IO_LATENCY_CLEAR|
 Impact(18)         |      |AR                          |                  |              |                |
                    |      |                            |                  |E-Port 11/8  |IO_LATENCY_CLEAR|
                    |      |                            |                  |E-Port 11/9  |IO_LATENCY_CLEAR|
                    |      |                            |                  |E-Port 11/10 |IO_LATENCY_CLEAR|
                    |9     |defALL_PORTS_IO_PERF_IMPACT|05/10/18 20:57:00|E-Port 11/11 |IO_PERF_IMPACT  |
                    |      |_UNQUAR                     |                  |              |                |
                    |      |                            |                  |E-Port 11/10 |IO_PERF_IMPACT  |
                    |      |                            |                  |E-Port 11/11 |IO_PERF_IMPACT  |
                    |      |                            |                  |E-Port 11/9  |IO_PERF_IMPACT  |
                    |      |                            |                  |E-Port 11/8  |IO_PERF_IMPACT  |
```

Appendix

Fibre Channel flow control

NOTE

An understanding of Fibre Channel flow control is essential for gaining a technical understanding of congestion.

To prevent a device (either host or storage) from being overwhelmed with frames, the Fibre Channel architecture provides a flow control mechanism based on buffer credits. A buffer credit is a value representing the number of frame buffers available at a receiving device port.

To prevent frame loss due a lack of frame buffer, Fibre Channel buffer-to-buffer flow control is implemented on every link. Briefly, buffer-to-buffer flow control functions as follows between end devices on a link:

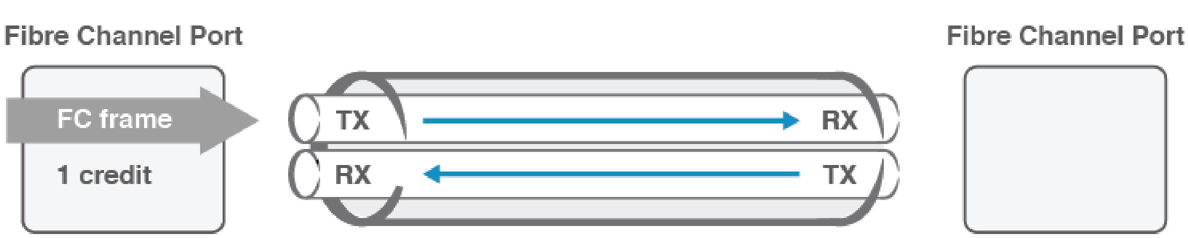
1. During link initialization, the number of buffer credits allocated for each device established.
2. During fabric login, each side of the link tells the other side of the link how many buffer credits it has.
3. After receiving information on the number of credits available on the remote peer, the local device initializes its transmit credit count (BBC_TRC) to that number. The BBC_TRC represents the number of frames that the port can transmit without overrunning its peer's buffers. When the device transmits a frame, it decrements the BBC_TRC count by one.
4. The remote port sends a receiver ready (R_RDY) back to the local device once it has received a frame and processed it (freeing up the buffer).
5. The local device increments its transmit credit count by one when it receives the R_RDY signal. If the transmit credit count reaches zero, the port cannot transmit and must wait until it receives an R_RDY to replenish its BBC_TRC.

If a recipient does not return credits to the transmitting device, no frames can be sent. If frames arrive at a device faster than they can be processed, the flow control mechanism will pace the credit response in order to slow the ingress frame rate. A credit-based flow control mechanism prevents frame loss and eliminates the need for frame retransmission due to buffer overflow at the receiving device. Unless there is severe congestion resulting in frame time outs, frames are not dropped.

The following are steps for the frame transmission process using the buffer-to-buffer flow control mechanism:

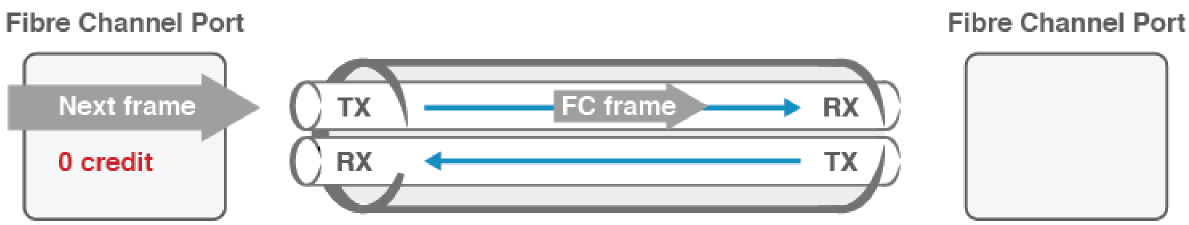
1. A frame is ready to be transmitted to the receiving port and a credit is available.

FIGURE 7 Frame waiting for transmission



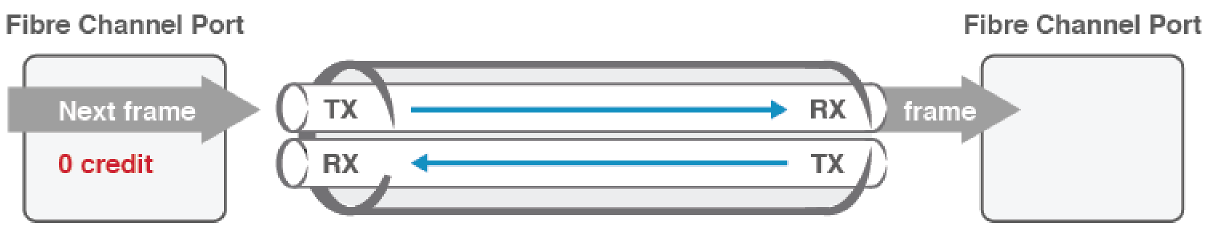
2. The frame is transmitted and the credit count of the transmitting port is decremented. The next frame is waiting in queue for a credit from the receiving port before it can be transmitted.

FIGURE 8 Frame transmitted and credit count decremented on transmitting port



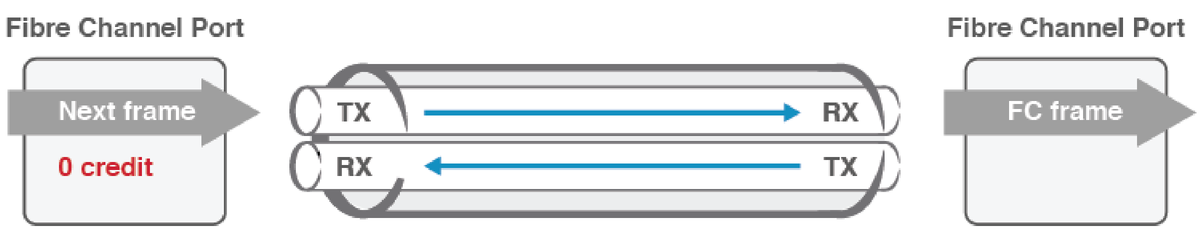
3. The frame is received.

FIGURE 9 Frame received



4. The frame is processed.

FIGURE 10 Frame processed by receiving device

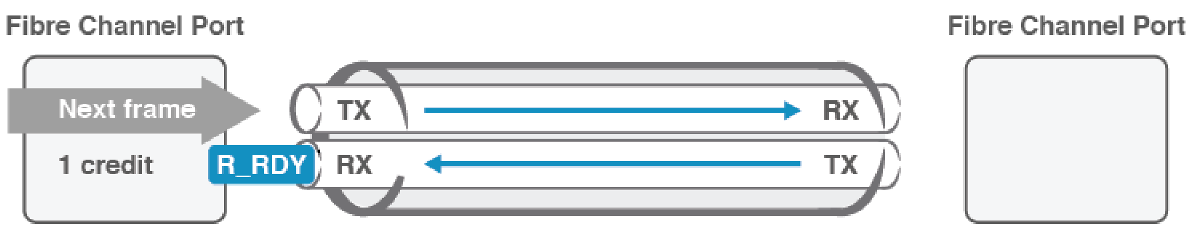


5. Receiver Ready (R_RDY) is returned.

FIGURE 11 R_RDY returned to transmitting device

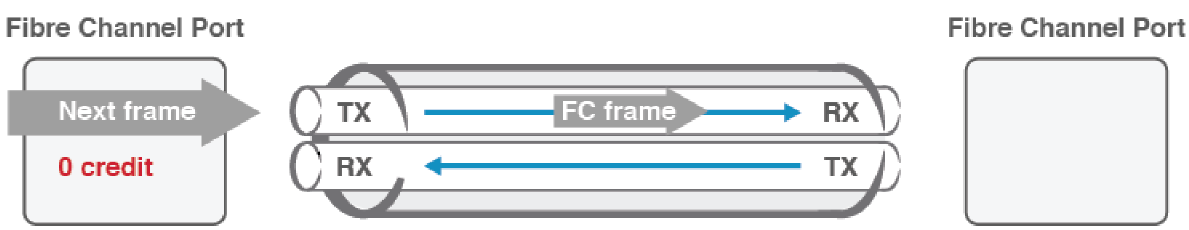
6. The credit count is incremented, and the next frame can be transmitted.

FIGURE 12 Credit count incremented on transmitting device



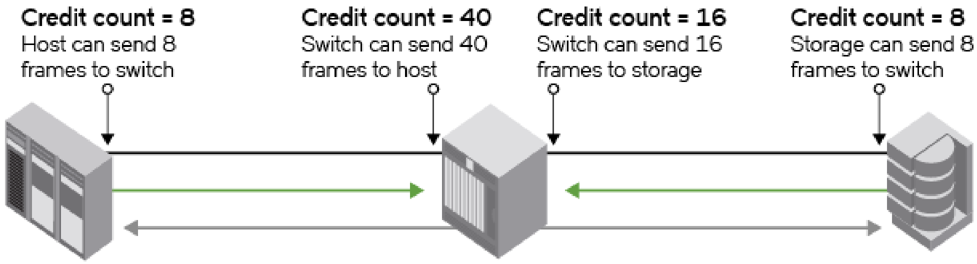
7. The frame is transmitted, and the credit count of the transmitting port is decremented. The next frame is waiting in queue for a credit from the receiving port before it can be transmitted.

FIGURE 13 Frame transmitted and credit count decremented on transmitting port



Adequate flow control requires enough capacity in the hardware to allow continuous transmission of frames on the link, while waiting for the acknowledgment to be sent by the receiver at the other end. The following illustration shows an example of credit accounting after devices exchange credit counts at fabric login.

FIGURE 14 Buffer credit exchange at fabric login



The optimal number of buffer credits required for a device is determined by the distance (frame delivery time), the processing time at the receiving port, the size of the frames being transmitted, and the speed of the link. For example, as the link speed increases, the frame delivery time is reduced and the number of buffer credits must consequently be increased to maintain full link utilization, even when there is a short distance between communicating devices.

Buffer-to-buffer credit management affects performance over distances. Allocating a sufficient number of buffer credits for long-distance traffic is essential to performance. As the distance between switches and the link speed increases, additional buffer credits are required for the ports used for long-distance connections.

Brocade ISL credit management

On Brocade inter-switch links (ISLs), buffer-to-buffer flow control works somewhat differently than it does between device F_Ports. A physical ISL is logically subdivided into virtual channels (VCs), each with independent buffers and VC_RDY mechanism to manage the buffers. Each VC has its own transmit credit counter (BBC_TRC) that is initialized when the E_Port comes online. The number of credits for each VC are exchanged and values are negotiated between ISL endpoints. VC_RDY is used instead of R_RDY to return credits, since the VC_RDY ordered sets allow identification of the VC that returns the credit.

Since VCs create multiple logical data paths across a single physical link or connection and since flow control is implemented independently on each of these paths, traffic in one VC can be isolated from the congestion affecting another VC. Any congestion issues can be isolated within a virtual channel so that they do not impact other virtual channels. In addition, the Quality of Service (QoS) feature (enabled by default) allows traffic to be distributed by priority level across multiple VCs of an ISL based on the SID and DID of each frame. For more information on virtual channels and an illustration of virtual channel priority levels on a QoS-enabled ISL, refer to the "Virtual Channels" section of the *Brocade Fabric OS Administration Guide*.

For more information on buffer credit management, flow control, QoS and VC priority levels, adaptive networking services, and allocating buffer credits for long-distance applications, different Brocade devices, VCs, and varied environments, refer to the "Buffer Credit Management" section of the *Brocade Fabric OS Administration Guide*.

Congestion and its Causes

This appendix provides more detailed information on congestion and its causes in the following sections:

Congestion and Its Causes—An Analogy

This section further explains congestion and its causes using an analogy.

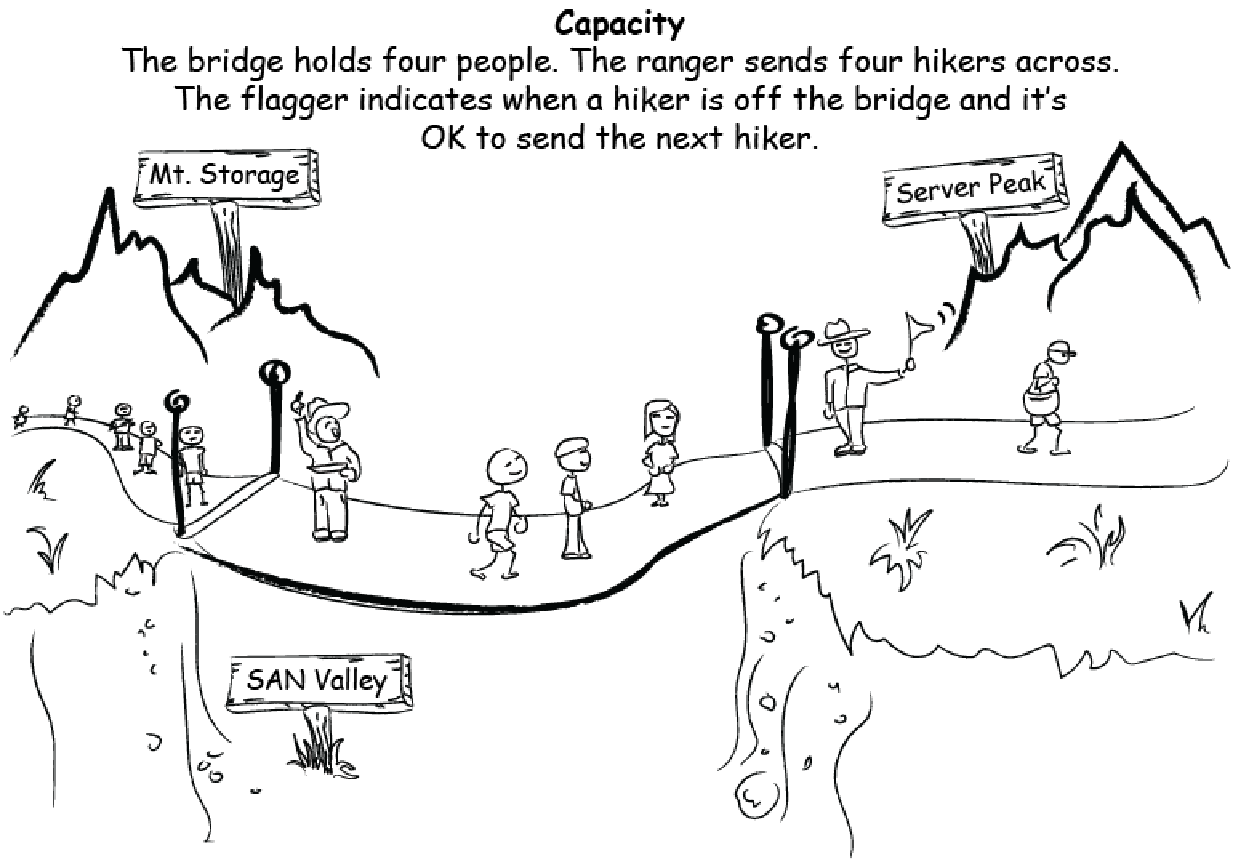
Fibre Channel is a lossless, credit-based network. To prevent a device (host or storage) from being overwhelmed with frames, the Fibre Channel architecture provides a flow control mechanism based on buffer credits. When two devices connect, they tell each other how many buffer credits, or the number of "chunks" of data or frames, that they can store. Each sender keeps track of this number, so that it knows how many more “frames” the receiver can accept. Each receiver uses a signal called “Receiver Ready” to tell the sender that it has room for the next frame. When the sender’s count reaches zero, it stops sending frames until the receiver says, “I’m ready!” which means it has made room for more frames. Congestion occurs when frames are entering the fabric faster than they are existing the fabric. Frames build up in the fabric, causing the movement of frames to slow down or become congested.

Imagine that the link between a sending and receiving device is a suspension bridge between two peaks called Mt. Storage and Server Peak. The steady stream of hikers coming up the trail between the peaks and traversing the bridge represents the frames being sent by a sender. Two park rangers regulate the number of hikers crossing the bridge. The Mt. Storage Ranger is at the entrance to the bridge and is the sender. The Server Peak Ranger is at the exit from the bridge and is the receiver. At the start of the day, the rangers agree on the maximum number of hikers that can be on the bridge at any time. The Mt. Storage Ranger keeps track of the number of hikers on the bridge and the Server Peak Ranger tells the Mt. Storage Ranger when a hiker gets off the bridge by waving the “Receiver Ready” flag. The flag waves represent R_RDY signals sent by a receiver

Let’s assume that the bridge can only hold four hikers at a time and it takes a hiker 1 minute to cross the bridge. The Mt. Storage ranger sends four hikers over the bridge and the Server Peak ranger lets the Mt. Storage Ranger know when a hiker crosses the bridge by waving the "Receiver Ready" flag. Upon seeing the flag, the Mt. Storage ranger knows that it is o.k. to send another hiker over the bridge. Every time a hiker exits the bridge the Server Peak ranger waves his flag and the Mt. Storage ranger allows another hiker onto the bridge. On a clear day there are always 4 hikers on the bridge and the bridge is being utilized at a maximum efficiency of 4 hikers per minute. However, if problems occur, the steady stream of hikers on the trail can slow and may even cease crossing the bridge, much like frames congesting in a fabric.

The following illustration indicates how the Server Peak ranger in our analogy is limiting the number of hikers on the bridge.

FIGURE 15 Regulating capacity of hikers on bridge



Lost credit

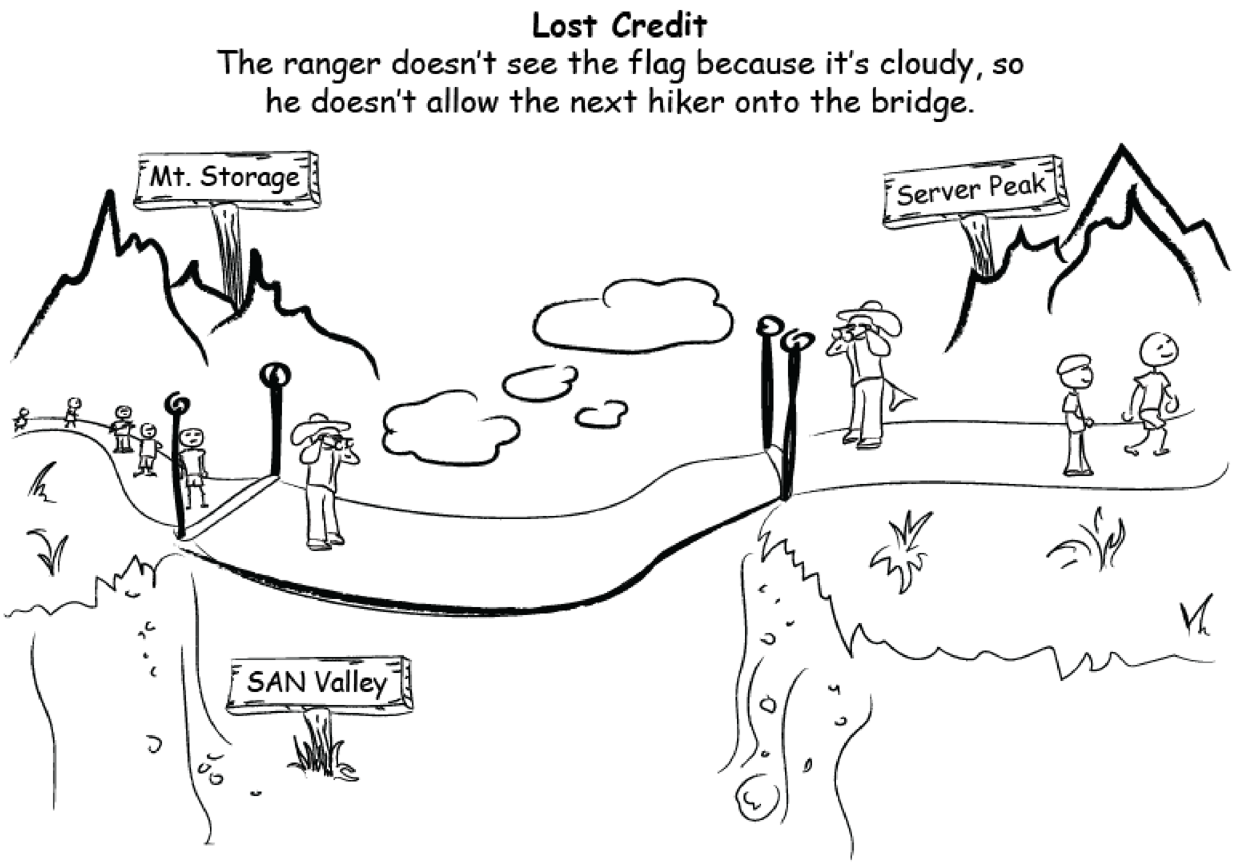
Lost credits occur when there are errors on the link that affect the signal that returns a credit from the receiver to the sender.

For our bridge and ranger analogy, the clouds in the following figure represent link errors. If a cloud moves in and obstructs the view between the rangers, and the Mt. Storage ranger misses a flag wave by the Server Peak ranger, the ranger would not know to let another hiker onto the bridge until another hiker exits the bridge and the Server Peak ranger again waves the flag. Now there would only be three hikers on the bridge at one time and only three hikers crossing the bridge every minute. For every cloud obstructed flag wave, one less person would be allowed on the bridge at a time. The bridge would be used less efficiently until eventually there are no hikers on the bridge. Eventually, the Server Peak ranger realizes something is wrong and notifies the Mt. Storage ranger to restart the process.

Assume that we have a 16-Gbps link that is long enough so that four frames can be transiting it at any time. The receiver initially gives the sender four buffer credits and sends an R_RDY for every frame it receives and processes. For every R_RDY lost due to error, the effective throughput of the link is reduced. One lost R_RDY reduces the throughput to 12-Gbps, two lost R_RDYs reduce throughput to 8-Gbps, and so on, until the throughput drops to zero. After 2 seconds of no credits, the sender realizes something is wrong and resets the link, thereby reinitializing the buffer credits. However, during those 2 seconds with no traffic, the 16-Gbps link could have handled over 1,600,000 frames. That is equivalent to our rangers not realizing something was wrong for 400,000 minutes or almost 40 weeks.

The following illustration depicts the Mt. Storage ranger in our analogy missing the signal flag so doesn't allow another hiker on the bridge. This "lost signal" reduces the capacity of hikers using the bridge.

FIGURE 16 Lost signal from Server Peak ranger reduces capacity of hikers on bridge

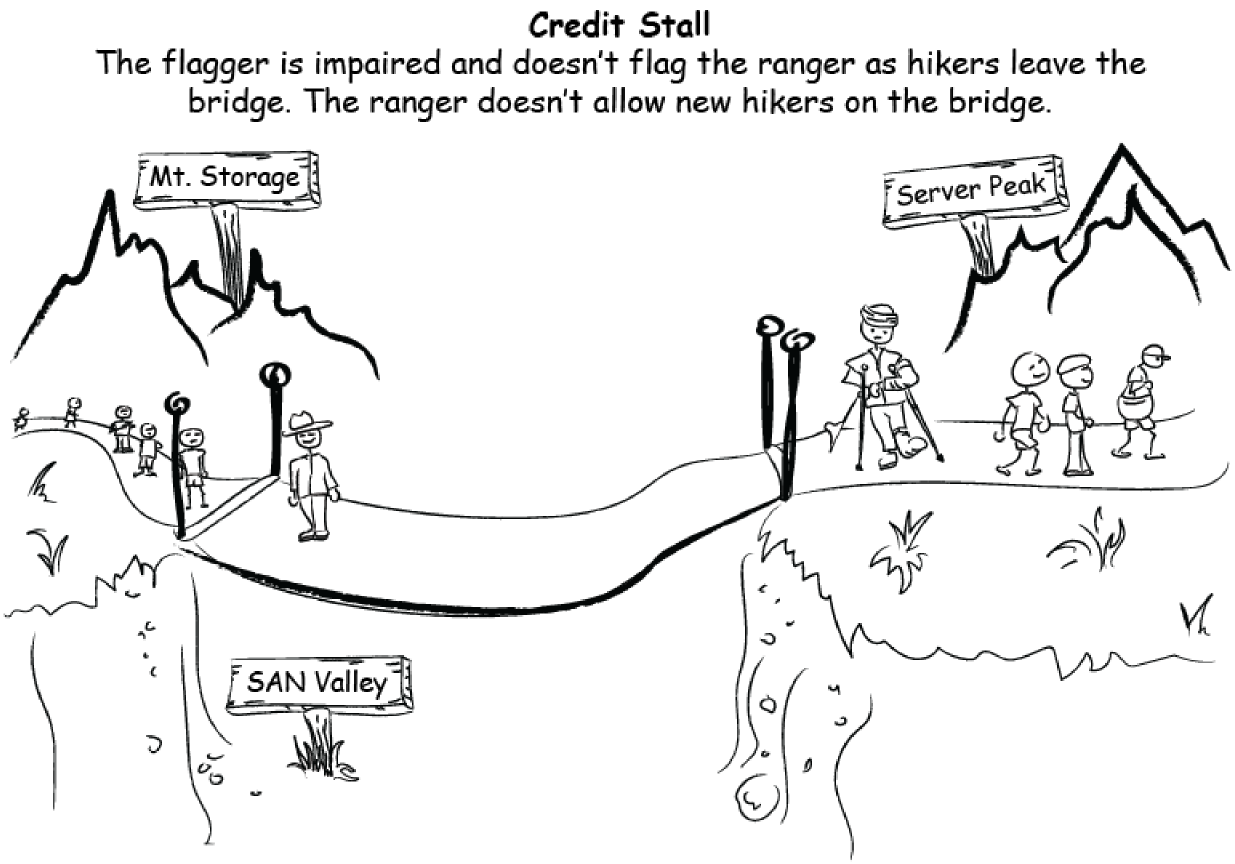


Credit-stalled devices

Often called slow-draining devices, these are devices that fail to send the "Receiver Ready" signal to return credits. Without credit, the sender cannot transmit frames, which causes them to back-up in the fabric. In our park ranger analogy, a credit-stalled device is like the Server Peak ranger becoming impaired by injury and cannot wave the "Receiver Ready" flag when a hiker gets off the bridge. As a result, the Mt. Storage ranger stops allowing new hikers on the bridge. A line of hikers forms at the bridge entrance, and soon, no one is on the bridge! This analogy mimics congestion in a fabric. The line of hikers represents frames waiting to be transmitted. The impaired Server Peak ranger represents a credit-stalled receiving device that is not working properly by not sending the "Receiver Ready" signal. To put our bridge analogy in perspective for SAN links, consider that over 800,000 frames can pass over a 16-Gbps FC link every second. So a misbehaving, credit-stalled server that stops returning credits for 1/10 of a second is equivalent to a ranger being unable to raise his flag for weeks.

The following illustration depicts the Server Peak ranger in our analogy failing to signal the Mt. Storage ranger to send another hiker across the bridge, effectively shutting the bridge down to foot traffic. As a result, hikers back up down the trail waiting to cross, until park management addresses the issue of the injured Server Peak ranger.

FIGURE 17 Server Peak flagger stalls when sending signal for another hiker

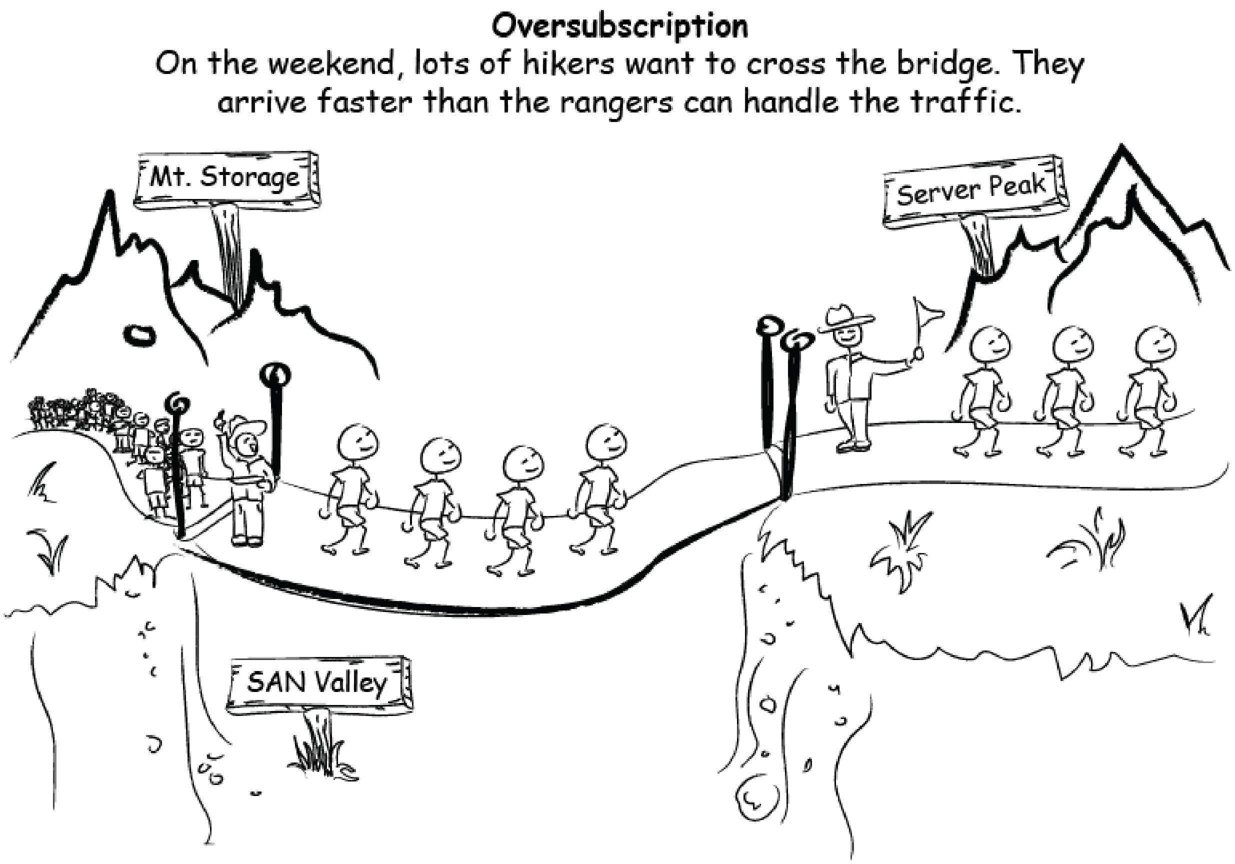


Oversubscription

This is a condition that occurs when a device requests more data than it can consume efficiently, and the extra data backs up into the fabric. Let's say in our suspension bridge analogy that flyers have been sent encouraging all hikers to see the beauty of Server Peak. As a result, lots of hikers would like to make their way across the suspension bridge toward Server Peak. Both rangers work as quickly as possible to keep hikers moving across the bridge, but the bridge holds only so many hikers at a time. Soon a long line forms at the entrance of the bridge. Eventually, some of the hikers get tired of waiting and go home! In this analogy, the flyer encouraging all hikers to come to Server Peak is like a device asking for more data than it can consume at full speed. As a result, the data arrives so quickly that it soon backs up in the fabric and causes congestion. In severe cases, frames are discarded by the fabric because they cannot be delivered in time, much like the hikers that got tired of waiting and went home!

The following illustration depicts that so many hikers are arriving at the bridge so fast that the rangers cannot regulate them on the bridge. As a result, hikers back up down the trail—some get tired of waiting and go home.

FIGURE 18 Hikers arrive at bridge faster than rangers can handle them



Description of congestion causes

This section provides more detailed information and illustrations to help you understand the causes of congestion defined in Causes of congestion on page 8.

Lost credits

The advantage of buffer-to-buffer flow control in a high-performing network is that it provides consistent and reliable frame delivery from sender to receiver. However, signal corruption due to physical layer issues on the link can interfere with the flow control mechanism. If corruption occurs, valid R_RDY or VC_RDY signals may not be returned after a frame is received and processed at the other end of a link, resulting in the BBC transmit credit count not being replenished. The result is that a credit is not available to enable transmission of a frame to the receiving device. A lost credit at the switch can result in degraded fabric performance.

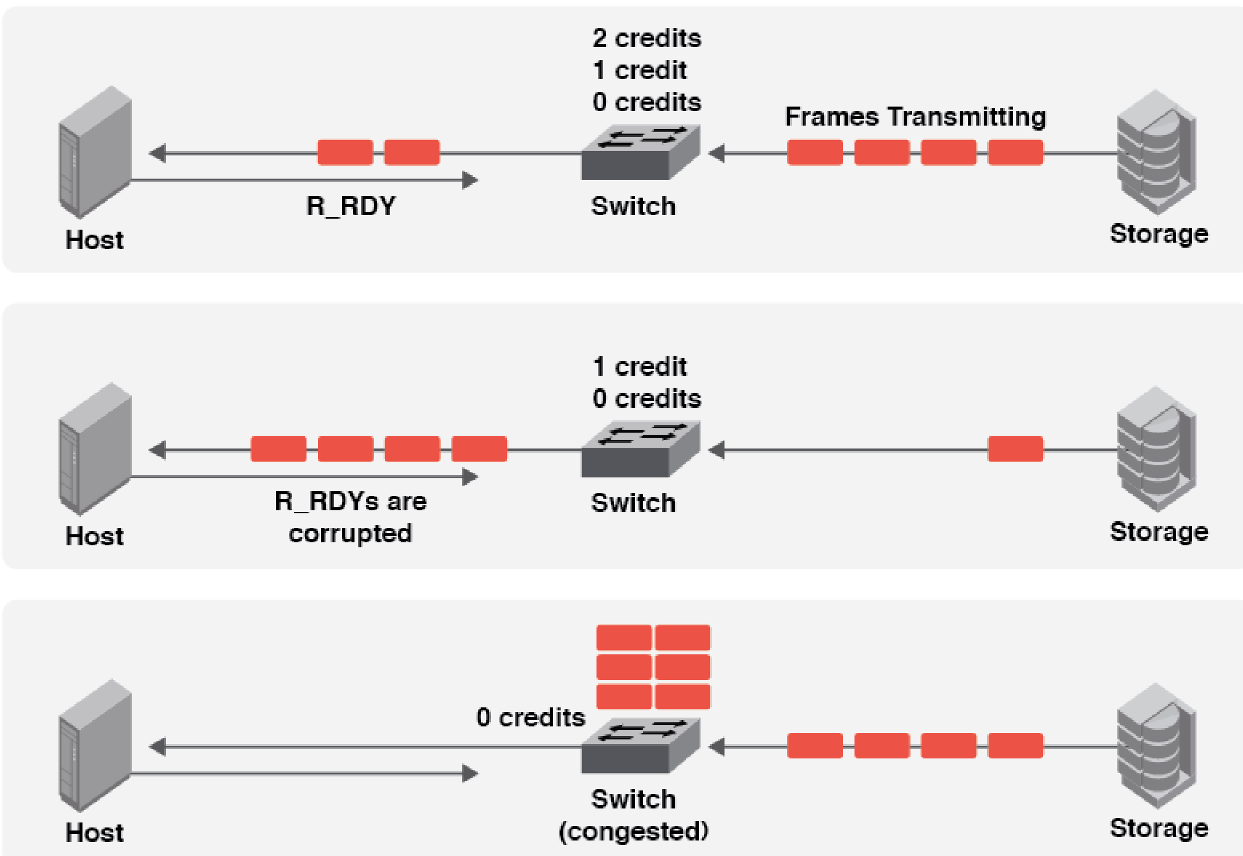
The following figure illustrates an example of a lost-credit condition causing congestion in the fabric.

1. In the first panel, the switch is transmitting frames to the host and R_RDY signals are returning to the switch to maintain credits for continued transmission of frames from the switch.
2. In the second panel, physical link errors, such as invalid transmission words (ITWs), are preventing R_RDYs from returning to the switch. The number of buffer credits available to the switch depletes by one credit each time it sends a frame and it does not receive an R_RDY in return. As the switch loses credits, link performance degrades. For example, if the switch requires at least

five credits for adequate performance on a 500-m link. and the credits deplete to two credits, then the link will have one half of its maximum capacity because of the round trip required for the frame to travel to the host and an R_RDY to return from the host.

3. As illustrated in the third panel, at some point, credits will deplete to zero and the switch port cannot send frames at all. When frames back up in the switch, frame delivery timeouts occur and frames drop, affecting I/O.

FIGURE 19 Lost credits



Permanent lost credit is usually caused by an external condition that corrupts the R_RDY primitive signal, such as electrical noise, faulty transceivers, dirty transceiver and cable connectors, poorly seated blades and transceivers, and bad cables/fibers. Corrupt primitives are dropped by the receiver as malformed frames. Although permanent lost credit is very infrequent, if lost credit occurs over time, all available buffer credits over a link may eventually become depleted, and traffic between the two endpoints ceases until credit is recovered by a link reset. Permanent lost credit can occur on a port's VC ISLs, back-end (BE) ports, or another link where VCs are supported.

Credit-stalled devices

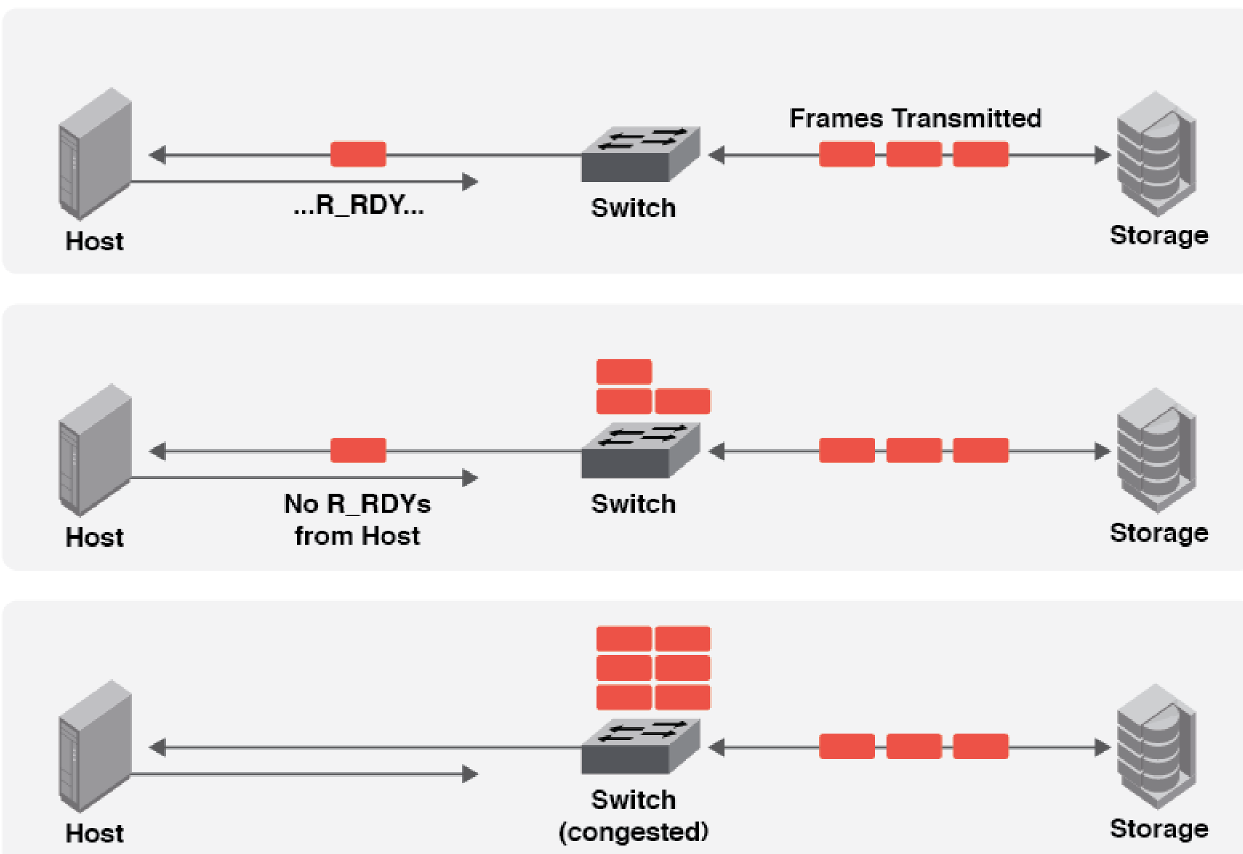
Credit-stalled devices are a major source of congestion in today's fabrics. These devices can cause the fabric to hold frames for excessive periods of time, which can result in application performance degradation or, in extreme cases, I/O failure. A credit-stalled device is often referred to as a "slow-drain" device. The device stops returning R_RDYs (buffer credits) to the transmitting switch for tens or hundreds of milliseconds or even seconds.

A classic example of a credit-stalled device is an HBA that stops returning credit when an out-of-order frame condition is encountered. The HBA detects the missing frame and waits for it to arrive. While it is waiting, it stops returning credits. This "stalling" behavior creates back pressure in the fabric, which results in congestion. Effectively, "credit stalling" reduces the capacity of the link. A link that could be running at 16 Gbps may actually be running at 2 Gbps, 4 Gbps, or 8 Gbps rate (or slower).

The following figure illustrates how a credit-stalled host causes congestion:

1. The first panel illustrates normal frame transmission from the switch to the host with no congestion at the switch.
2. The second panel shows valid R_RDY signals are not being returned to the switch. Eventually, buffer credits at the switch are depleted, and the switch cannot return frames.
3. Frames then back up at the host and in switch queues. If frames are held longer than the edge hold time set on the switch, frames are dropped. Dropped frames are logged as C3 discards.

FIGURE 20 Credit-stalled device



Stalled credits on a device can be caused by the following:

- Misbehaving device drivers.
- Incorrectly configured or misbehaving device firmware.
- Faulty hardware.
- Insufficient cache memory.

- Unsolicited command overflow. This occurs when a target device receives more commands than it can process. The incoming commands are queued and. if the queue reaches capacity, the device stops returning credits. As a result, all traffic flow will cease from the fabric to the device.

When a device goes into credit-stall behavior, 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 that is connected to the credit-stalled device) and ultimately affects the fabric.

Some known credit-stall issues have been remedied by HBA firmware upgrades. Consult with your supplier for availability of qualified firmware versions. For example, Emulex addresses credit-stalled or slow draining device behavior with following firmware updates:

- Version 2.82a4 for 4Gbit/sec LPe11000-series adapters. This was originally released in June 2015 as Version 2.82x14
- Version 2.02a1 for 8Gbit/sec LPe12000-series adapters. This was originally released in June 2015 as Version 2.01a13

To download Emulex Version 2.02a1 firmware, use the following steps:

1. Go to www.broadcom.com.
2. Select Support>Support Documents and Downloads to display the "Support Documents and Downloads" page.
3. On the "Support Documents and Downloads" page, select the following:
   a. Product Group — Storage Adapters, Controllers, and ICs
   b. Product Family — Fibre Channel Host Bus Adapters>OEM>Emulex
   c. Product — LPe12000 FC Host Bus Adapter
   d. Asset Type — Firmware
4. Select the Search button.
5. Select the firmware link in the search results to display downloads.

To download the Emulex LPe12000 Family Firmware Release Notes, follow the preceding steps, except select Release Notes for the Asset Type in step 3d.

Oversubscription

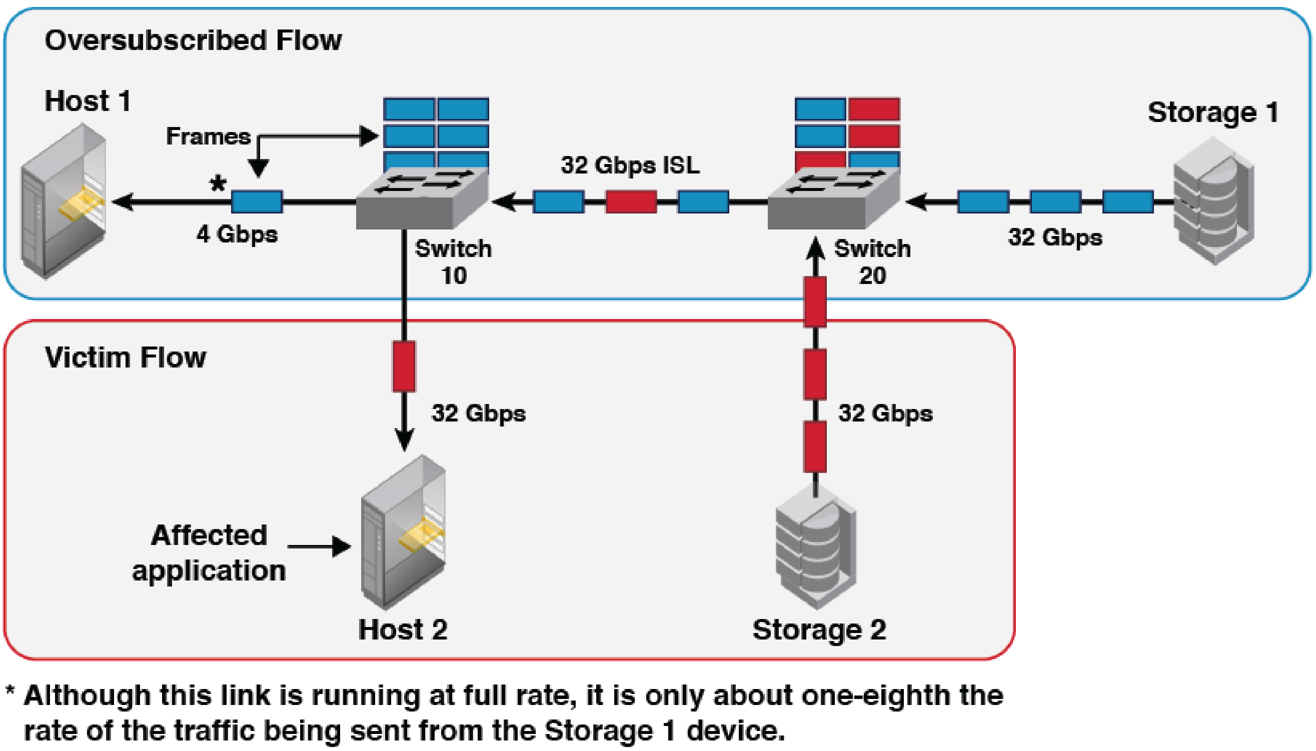
Fabrics are oversubscribed by design based on the assumption that the I/O behavior of the devices attached to the fabric is never 100% committed at any one time. Thus, it is a common practice to share fabric resources by assigning a total theoretical traffic rate that is greater than the total capacity. For example, multiple 4-Gbps host devices may be assigned a single 8-Gbps storage port to take advantage of unused 8-Gbps capacity as the devices will rarely use the storage port at the same time. Therefore, oversubscription to some extent always exists in a fabric by design. Oversubscription becomes a problem only when the design assumptions do not correspond to the actual application and device behavior. In these instances, oversubscription leads to congestion and related problems, such as underperforming applications, erosion of quality of service (queuing delays, frame loss, or adverse effects on shared resources), and possibly I/O failure.

Oversubscription, in terms of the number of source ports accessing target ports and devices accessing ISLs, can cause instances of insufficient link capacity, which can lead to congestion. The most common cause of oversubscription is too many storage targets accessing too few servers. This causes oversubscribed links between the servers and storage. As Fibre Channel link bandwidth has increased to 16-Gbps or 32-Gbps, 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 can cause serious congestion when servers issue many data requests, but cannot process the data sent at the rate returned by storage, effectively limiting throughput of the storage port.

Oversubscription is identified by queue latency on upstream port(s) and high bandwidth utilization at a device port. Oversubscription congestion is typically caused by a bandwidth mismatch between the source and destination ports, and it may affect flows that share the same path through the fabric.

The following figure illustrates congestion from read oversubscription caused by a speed mismatch between a 32-Gbps storage device sending to a 4-Gbps host.

FIGURE 21 Read oversubscription



The following actions are causing oversubscription and congestion in this example.

1. Host 1, connected to the fabric through a 4 Gbps HBA, has requested a large read I/O from Storage 1, a 32 Gbps storage device.
2. Host 2, connected to the fabric through a 32 Gbps HBA, has requested a large read I/O from Storage 2, also a 32 Gbps storage target.
3. Traffic from the faster Storage 1 device overwhelms the slower 4 Gbps link from Switch 10 to Host 1 causing frames destined to Host 1 to back-up in Switch 10 buffers.
4. Switch 10 slows downs incoming traffic on the ISL from Switch 20 using FC flow control to avoid frame loss due to buffer overflow.
5. Frames from both Storage 1 and Storage 2 back-up in Switch 20 buffers, as the throughput of the ISL has been throttled by Switch 10.
6. The traffic from Storage 2 to Host 2 has been slowed down (victimized) by the oversubscription congestion caused by the speed mismatch between Host 1 and Storage 1.
7. The result is poor performance of an application on the 32-Gbps host.

Metrics

This section describes the metrics used to determine when congestion is occurring in a fabric, the causes of congestion, and the severity of congestion.

Identifying Congestion

The following table provides an overview of the metrics used to determine if congestion is occurring in a fabric.

TABLE 1 Metrics that Identify Congestion

Metric	Metric definition	Hardware statistic counters or algorithms used for the metric
Credit latency	Time at zero credit. Length of time that a frame waits at the top of the queue until a credit is available for frame transmission.	The following switch counters indicate the number of times transmit credits are at zero for ports and VCs: • tim_txcrd_z • tim_txcrd_z_vc
Queue latency	Transmit queue latency. Length of time that a frame is queued for transmission until it transmits. This is the time that a frame takes to move from the bottom of a transmit queue to the top, plus transmission time.	Switch counters include the following: • tim_latency_vc (Gen 5 platforms) • max_latency_vc (Gen 6 platforms) • lat_tot_pkt_vc (Gen 6 platforms) • lat_low_time_vc (Gen 6 platforms) • lat_hi_time_vc (Gen 6 platforms)
Bandwidth utilization	Percent utilization. Average amount of link capacity used.	MAPS monitors bandwidth utilization on each port in the switch every second. High bandwidth alerts occur when port bandwidth utilization exceeds 60%/hr, 75%/hr, 90%/hr, or 95%/min. These alerts indicate possible sources of oversubscription that could lead to congestion.
Frame loss	Discarded frame count. Number of frames discarded due to timeout of the transmission hold time (HT). Default HT for E_Ports is 500 ms and edge HT (EHT) set for F_Ports is 220 ms.	Switch counters used include the following: • er_rx_c3_timeout • er_tx_c3_timeout **NOTE** Receive (rx) timeouts on an F_Port indicates that frames received on the port are being discarded because of backpressure from upstream ports (ISLs or other devices). Transmit (tx) timeouts on an F_Port indicates that the F_Port the is the source of congestion and is causing backpressure.
Link reset	Link reset count. Number of times that a credit has been delayed for 2 seconds.	Switch counters used include the following: • Lr_out • Lr_in

Determining Congestion Causes

The following table compares details of observed fabric metrics to causes of congestion.

TABLE 2 Observed metrics compared to congestion causes

Metric	Causes	Metric observations
Bandwidth utilization	Oversubscription	High bandwidth utilization at a device port may be an indication of oversubscription, especially if it accompanies queue latency at an associated ISL port.

TABLE 2 Observed metrics compared to congestion causes (continued)

Metric	Causes	Metric observations
Queue latency	Oversubscription Credit-stalled device	When a device port stops issuing credit, this indicates a credit-stalled device. Queue latency at an ISL port is an indication of downstream congestion caused by oversubscription or a credit-stalled device.
Credit latency	Oversubscription Credit-stalled device	Credit latency at a device port is an indication of a credit-stalled device. Credit latency at an ISL port is an indication of a downstream congestion caused by oversubscription or a credit-stalled device.
Frame loss	Credit-stalled device Lost credit	Frame loss at a device port indicates that the device is credit-stalled and has withheld credit for a period between 220 ms and 500 ms. Frame loss at an ISL port indicates downstream congestion caused by oversubscription, credit-stalled devices, or lost credit.
Link reset	Credit-stalled device Lost credit	A link reset at a device port indicates that the device is credit-stalled and has withheld credit for over 2 seconds. A link reset at an ISL port indicates that the peer switch is unable to return credit, which could be caused by downstream oversubscription or credit-stalled devices. The link reset may also be caused by complete credit depletion on the ISL link, which indicates physical errors.

Determining congestion severity

The impact to the fabric (and other traffic flows) by congestion varies based on the severity of the latency and frame loss exhibited in the fabric. The following summarizes congestion severity levels based on credit latency, queue latency, and frame loss that are used by Brocade tools, such as the MAPS feature, that detect congestion conditions. The different severity levels result in application symptoms ranging from poor performance to timeout and errors.

The following table defines congestion severity based on various combinations of oversubscription, credit-stalled device, and lost credit metrics.

TABLE 3 Severity levels for congestion causes

Cause	Mild	Moderate	Severe
Oversubscription	1. High bandwidth at the device port. 2. Low credit latency at the ISL port. 3. Less than 10-ms queue latency at the ISL port. 4. No frame loss or link resets.	1. High bandwidth at the device port. 2. Moderate credit latency at the ISL port. 3. 10-ms to 80-ms queue latency at the ISL port. 4. No frame loss or link resets.	1. High bandwidth at the device port. 2. High credit latency at the ISL port. 3. Greater than 80-ms queue latency at the ISL port. 4. Frame loss at an upstream (ISL) port indicates queue latency greater than the hold time (HT) set for the switch. The default HT set for E_Ports is 500 ms and edge HT (EHT) set for F_Ports is 220 ms. 5. No link resets.
Credit-stalled (slow-drain) device	1. A small period of credit stall at the device port and upstream ISL port. 2. Less than 10-ms queue latency at the device port and upstream ISL port. 3. No frame loss or link resets.	1. A medium period of credit stall at the device port and upstream ISL port. 2. 10-ms to 80-ms queue latency at the device port and upstream ISL port. 3. No frame loss or link resets.	1. A large period of credit stall at the device port and upstream ISL port. 2. Greater than 80-ms queue latency at the device port and upstream ISL port. 3. Frame loss at an upstream (ISL) port indicates queue latency greater than the hold time (HT) set for the switch.

TABLE 3 Severity levels for congestion causes (continued)

Cause	Mild	Moderate	Severe
			The default HT set for E_Ports is 500 ms and edge HT (EHT) set for F_Ports is 220 ms. 4. Link reset at an ISL port (indicates device credit stall for more than 2 seconds).
Lost credit	1. Low credit latency at the port. 2. Less than 10-ms queue latency at the port or upstream from the port. 3. No frame loss or link resets.	1. Moderate credit latency at the port. 2. Between 10-ms to 80-ms queue latency at the port or upstream from the port. 3. No frame loss or link resets.	1. High credit latency at the port. 2. Greater than 80-ms queue latency at the port or upstream from the port. 3. Frame loss at the port or upstream from the port, which indicates device credit stall greater than the HT set for the switch. The default HT set for E_Ports is 500 ms and edge HT (EHT) set for F_Ports is 220 ms. 4. Link reset at the port or upstream from the port (indicates device credit stall for more than 2 seconds).

Glossary of congestion terms

Following are definitions of terms used in this publication to describe congestion and congestion troubleshooting.

- Back-end ports—A back-end port connects a core switching blade to a port blade or application blade in a chassis-type switch.
- Back pressure—The accumulation of frames within the fabric (frame queueing).
- Bandwidth utilization—The percentage of available port (or trunk) bandwidth being used for transmitted or received traffic.
- Bottleneck—A bottleneck is caused by congestion that occurs when R_RDY buffer credits are not returned fast enough from a device to a switch. The switch can no longer transmit frames due to zero transmit credits, and R_RDYs are withheld to upstream devices preventing them from transmitting frames. Frames may be dropped, and timeouts may occur upstream from the point or "bottleneck" of congestion.
- C3 timeout—A Class 3 frame timeout.
- Congestion—Persistent and pervasive back pressure that affects traffic flow(s). In such cases, quality of service can erode with queuing delays, frame loss, or blocking of new connections.
- CRC—Cyclic redundancy check errors.
- Credit latency—A delay, in milliseconds, for R_RDY transmission between a device and a switch.
- Credit-stalled device—Also called a slow-drain device. A credit-stalled device is a cause of fabric congestion created when a receiving device stops issuing R_RDY buffer credits to the transmitting device for an abnormally long time and may not issue credit until a frame loss or link reset occurs. Because the transmitting device cannot increment its credit count without an R_RDY, it cannot move frames out the fabric. This immediately causes congestion as the frames destined for the credit-stalled device back up in the fabric.
- Discarded frame count—The number of frames discarded due to transmission hold time-out. The default HT set for E_Ports is 500 ms, and the edge HT (EHT) set for F_Ports is 220 ms.

- Fan-in—I/O requests from a server to multiple targets.
- Fan-in ratio—The ratio of server ports to storage ports.
- Fan-out—I/O requests from a single target to multiple servers.
- Fan-out ratio—The ratio of storage ports to server ports.
- Fibre Channel flow control—A method of controlling frame transmission between devices through I/O pacing based on credit response.
- FPI—Fabric Performance Impact (FPI), a MAPS feature, monitors the current condition of the latency detected on E_Ports and F_Ports over different time windows and uses this to determine the performance impact to the fabric and network.
- Frame delay—The time, in milliseconds (ms), for which the frame is held in a port's transmit queue before being sent.
- Frame loss—Frames discarded at a port during transmission if they remain in the transmission queue longer than the set hold time.
- Hold time—The amount of time that a Class 3 frame may remain in a queue before being discarded while waiting for credit to be given for transmission. Hold time (HT) can be configured for E_Ports and edge hold time (EHT) can be configured for F_Ports.
- IO_FRAME_LOSS—The state set by FPI for a port if frame timeouts either have already occurred or are very likely to occur. This state indicates a severe level of latency.
- ITW—Invalid transmission word error.
- Latency—The time it takes for a frame to get from source to destination. This time includes transmission time, propagation time, and queuing delays in switches.
- Link reset—Reinitializing the link between a switch and a device when transmission is delayed for 2 seconds due to a lack of transmit buffer credit. During link reset, the frame and credit loss counters are reset without performance degradation.
- Lost credit—A typical cause of fabric congestion that occurs when the link between the fabric and a device experiences errors that corrupt the buffer credit management system. Since credits are used in Fibre Channel fabrics to ensure lossless delivery of frames and to maintain the optimum link bandwidth, when a credit is lost due to transmission errors, the link bandwidth is effectively reduced. This leads to behaviors resembling oversubscription.
- Oversubscription—The assignment of a total traffic rate to a given port or link that is greater than the port or link can handle.
- Performance Impact (IO_PERF_IMPACT)—The state set for a port by FPI if the port does not have credit for a substantial time or if frames are transmitted with delay.
- Queue—A first-in-first-out list of items waiting for a service (such as frames waiting for transmission or I/O requests waiting for processing).
- Queue depth—For HBAs, the number of concurrent I/O requests allowed from a host or to a target in the SAN is referred to as queue depth. This number is configured on the HBA. For transmit queues, this is the number of frames in the queue awaiting transmission.
- Queue latency—The time, in milliseconds (ms), for which I/O requests are queued before being serviced. See also "Transmit queue latency."
- Transmit queue latency—The length of time for which a frame is queued before being transmitted. This is the time that a frame takes to move from the bottom of a transmit queue to the top, plus transmission time.