

Slow Drain Device Detection and Congestion Avoidance

White Paper

November 2013

<u>Scope</u>	<u>3</u>
<u>Introduction</u>	<u>3</u>
<u>Cisco Solution</u>	<u>3</u>
<u>Background</u>	<u>4</u>
<u>Flow Control in Fibre Channel</u>	<u>4</u>
<u>Types of Flow Control</u>	<u>4</u>
<u>Fibre Channel Class of Service</u>	<u>5</u>
<u>Virtual Output Queue</u>	<u>6</u>
<u>Slow Drain Example</u>	<u>6</u>
<u>Slow Drain Detection and Congestion Avoidance</u>	<u>7</u>
<u>Slow Ports</u>	<u>7</u>
<u>Stuck Ports</u>	<u>8</u>
<u>Port Monitor</u>	<u>9</u>
<u>Detection of Credit Not Available Event on Cisco MDS 9500 Series</u>	<u>11</u>
<u>Detection of Credit Not Available Event on Cisco MDS 9700 Series</u>	<u>11</u>
<u>Identifying a Slow Drain Port</u>	<u>12</u>
<u>Levels of Performance Degradation</u>	<u>12</u>
<u>Level 1: Latency</u>	<u>12</u>
<u>Level 2: Retransmission</u>	<u>12</u>
<u>Level 3: Extreme Delay</u>	<u>12</u>
<u>Link Event "LR Rcvd B2B" on Fibre Channel Ports</u>	<u>13</u>
<u>Troubleshooting Slow Drain</u>	<u>14</u>
<u>Level-3 Troubleshooting: Extreme Delay</u>	<u>14</u>
<u>Check for Link Failure with "LR Rcvd B2B" Message</u>	<u>14</u>
<u>Check for Credit Loss on an Egress Port</u>	<u>15</u>
<u>Level-2 Troubleshooting: Retransmission</u>	<u>17</u>
<u>Check for Transmit Frame Drops</u>	<u>17</u>
<u>Check for Frame Transmit Timeout History on Egress Port</u>	<u>18</u>
<u>Level-1 Troubleshooting: Latency</u>	<u>19</u>
<u>Check for Credit Not Available</u>	<u>19</u>
<u>Lack of B2B Credits for the Distance the ISL is Traversing</u>	<u>19</u>
<u>Check ISLs for Lack of Transmit Credits</u>	<u>20</u>
<u>Check for Frame Queuing on Ingress Ports</u>	<u>20</u>
<u>Check for Arbitration Timeouts</u>	<u>22</u>
<u>Monitoring Tx and Rx B2B Transitions to Zero</u>	<u>23</u>
<u>Recovery</u>	<u>24</u>
<u>Error Disable or Flap the Port Through Port-Monitor Port Guard</u>	<u>24</u>
<u>Dropping Packets When the Congestion-Drop Threshold Is Reached</u>	<u>25</u>
<u>Setting the No-Credit-Drop Timeout Value</u>	<u>26</u>
<u>Automatic Credit Loss Recovery</u>	<u>27</u>
<u>Appendix A: Slow Drain Port-Monitor Detection Configuration</u>	<u>28</u>
<u>Configuring the Tx-Credit-Not-Available Counter</u>	<u>28</u>
<u>Appendix B: Cisco MDS 9000 Family Counter Names and Descriptions</u>	<u>29</u>

Scope

This white paper discusses the Cisco® MDS 9500 Series Multilayer Directors with Generation 2 (DS-X9112, DS-X9124, DS-X9148, and DS-X9304-18K9), Generation 3 (DS-X9248-48K9 and DS-X92xx-96K9), and Generation 4 (DS-X92xx-256K9) line cards, and the Cisco MDS 9700 Series Multilayer Directors (DS-X9448-768K9). Although some of the details are different, many of the principles and functions are broadly applicable to the entire Cisco MDS 9000 Family of switches. The appendices are provided for reference.

Introduction

A Slow Drain Device is a device that does not accept frames at the rate generated by the source. In the presence of slow devices, Fibre Channel networks are likely to lack frame buffers, resulting in switch port credit starvation and potentially choking Inter-Switch Links (ISLs). Frames destined for slow devices need to be carefully isolated in separate queues and switched to egress ports without congesting the backplane. A decision then needs to be made about whether the frames are considered stuck and when to drop them.

Cisco Solution

Cisco provides a Slow Drain Device Detection and Congestion Avoidance (henceforth referred to Slow Drain) feature that helps detect, identify, and resolve the condition exhibited by slow devices. Reasons for Slow Drain include:

Edge devices

An edge device can be slow to respond for a variety of reasons:

- Server performance problems: application or OS
- Host bus adapter (HBA) problems: driver or physical failure
- Speed mismatches: one fast device and one slow device
- Nongraceful virtual machine exit on a virtualized server, resulting in packets held in HBA buffers
- Storage subsystem performance problems, including overload
- Poorly performing tape drives

Inter Switch Links (ISL)

- Lack of B2B credits for the distance the ISL is traversing
- The existence of slow drain edge devices

Any device exhibiting such behavior is called a Slow Drain Device. Cisco MDS 9000 Family switches constantly monitor the network for symptoms of slow drain and can send alerts and takes automatic actions to mitigate the situation. These actions include:

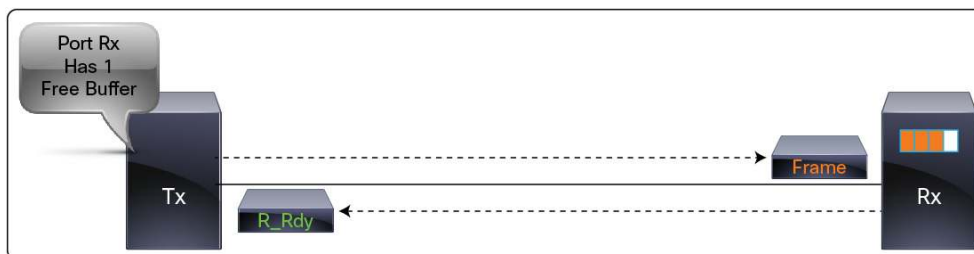
- Drain the buffers
- Perform link reset on the affected interfaces
- Flap the affected interfaces
- Error-disable the port
- Drop new packets arriving at line rate

Background

Flow Control in Fibre Channel

Fibre Channel is designed as a no-frame-drop network. To achieve this, Fibre Channel implements a credit-based flow-control mechanism (Figure 1).

Figure 1. Flow Control in Fibre Channel



Here, the transmitter (Tx) can send frames in the amount of the buffer-to-buffer (B2B) credits reported by the receiver (Rx). For each packet sent, the Rx port needs to send an R_Rdy (Receiver_Ready, Fibre Channel Primitive). If the Rx port withholds this R_Rdy then the Tx port's remaining B2B credits will stay at zero. At this point no further packets can be sent by the Tx port. Only upon receiving an R_Rdy can the Tx port resume sending an additional packet. This strategy prevents frames from getting lost when the Rx port runs out of buffers and helps ensure that the receiver is always in control.

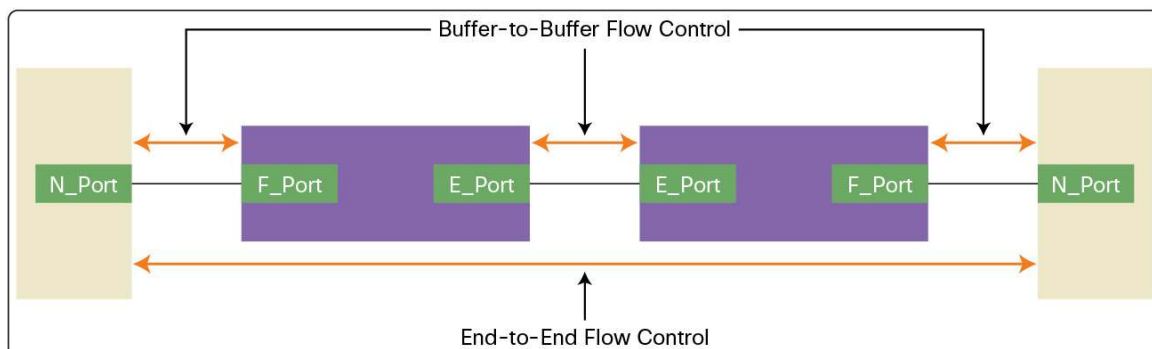
Whenever a Fibre Channel (FC) device logs into the fabric, it informs the switch of the available B2B credits that the switch has to transmit. Likewise, the switch informs the Fibre Channel device of the available B2B credits that the Fibre Channel device has to transmit. This helps ensure that neither side can overrun the other side.

Types of Flow Control

Fibre Channel defines two types of flow control (Figure 2):

- Buffer-to-Buffer (port to port)
- End-to-End (source to destination)

Figure 2. Types of Flow Control in Fibre Channel



With Buffer-to-Buffer (B2B) flow control, the transmitting port must assume that each frame has been delivered correctly. The receiving port on the end device is responsible for detecting that an error has occurred and reporting the error to the transmitting port on the other end device (this process is handled at the Upper-Layer Protocol [ULP] layer using the Small Computer System Interface [SCSI] Check_Condition command status code). Unfortunately, if the error occurs at the end of a sequence of frames, the transmitting port may have already emptied its ULP data buffers by the time it finds out about the error, in which case it cannot immediately retry the transmission, thus delaying successful delivery of the data.

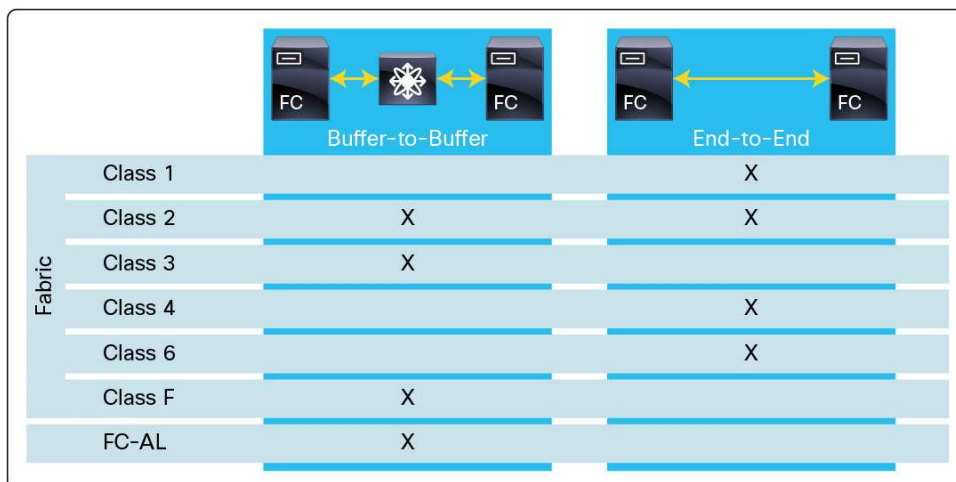
More important, B2B flow control does not relieve the congestion that can occur when a transmitter injects frames into the fabric faster than the receiver can receive those frames. When that happens, the fabric must either hold or drop the frames.

End-to-End flow control involves only the port at which a frame originates and the ultimate destination port, regardless of the number of Fibre Channel switches that are in the data path. When end-to-end flow control is used, the transmitting port is responsible for ensuring that all frames are delivered. Only when the transmitting Node port (N_Port) receives the last Acknowledgment (ACK) frame in response to a sequence of frames sent does it know that all frames have been delivered correctly, and only then will it empty its ULP data buffers. If a returning ACK frame indicates that the receiving port has detected an error, the transmitting N_Port has access to the ULP data buffers and can resend all the frames in the sequence. Thus, end-to-end flow control trades a degree of efficiency for increased reliability.

Fibre Channel Class of Service

Fibre Channel defines several Classes of Service (CoS), which can be used by different applications to choose the optimal type of delivery priority and flow control depending on the type of application data. Each CoS uses a connection-oriented, packet-switched, or quality-of-service (QoS) communication strategy (Figure 3).

Figure 3. Fibre Channel Class of Service



As of today, all data traffic is carried over Fibre Channel using Class 3. Inter switch control traffic such as Exchange Link Parameter (ELP), and Exchange Switch Capabilities (ESC) use class F. Fabric Login (FLOGI) and Port Login (PLOGI) uses class 2 or 3.

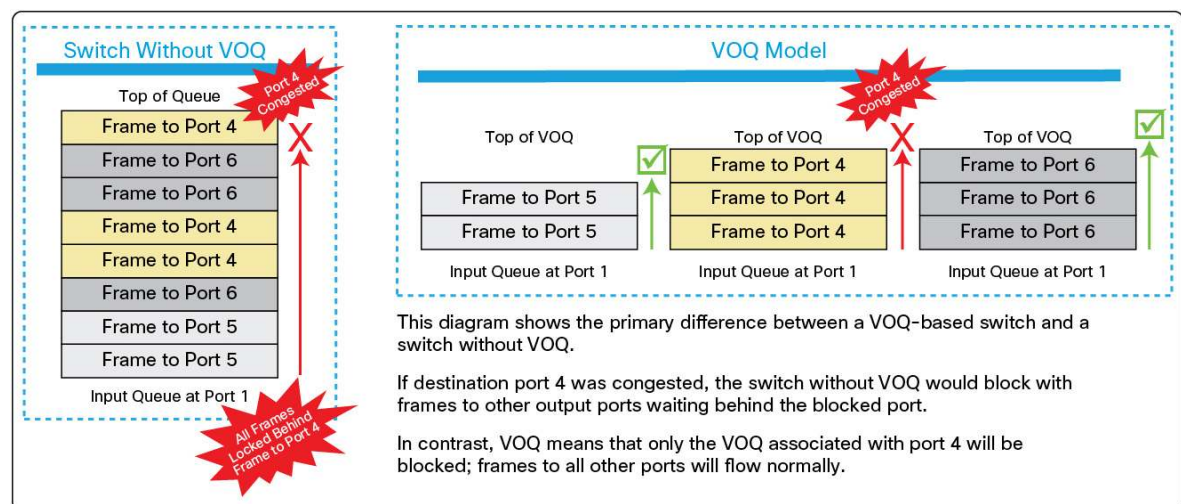
Class 3 is a best-effort packetized service:

- The receiving port does not acknowledge receipt of frames. If the fabric cannot deliver the frame for any reason, the frame can be discarded without notifying the sending port. However, Class 3 is not really unreliable, because it relies on ULP to help ensure that frames are delivered, by detecting and recovering from lost frames
- Class 3 does not guarantee fixed latency because data paths are variable
- Class 3 does not guarantee in-order delivery. For most Fibre Channel applications, including storage applications, the ULP is responsible for guaranteeing in-order delivery

Virtual Output Queue

Head-of-line blocking occurs when the frame at the head of the queue cannot be sent because of congestion at its output port. In this situation, frames behind this frame are blocked from being sent to their destination, even though their respective output ports are not congested. A virtual output queue (VOQ) exists at the ingress port and prevents head-of-line blocking through the use of multiple VOQs. Individual VOQs may be blocked, but traffic queued for different (nonblocked) destinations can continue to flow without getting delayed behind frames waiting for the blocking to clear on a congested output port (Figure 4).

Figure 4. Cisco MDS 9000 Family Virtual Output Queue



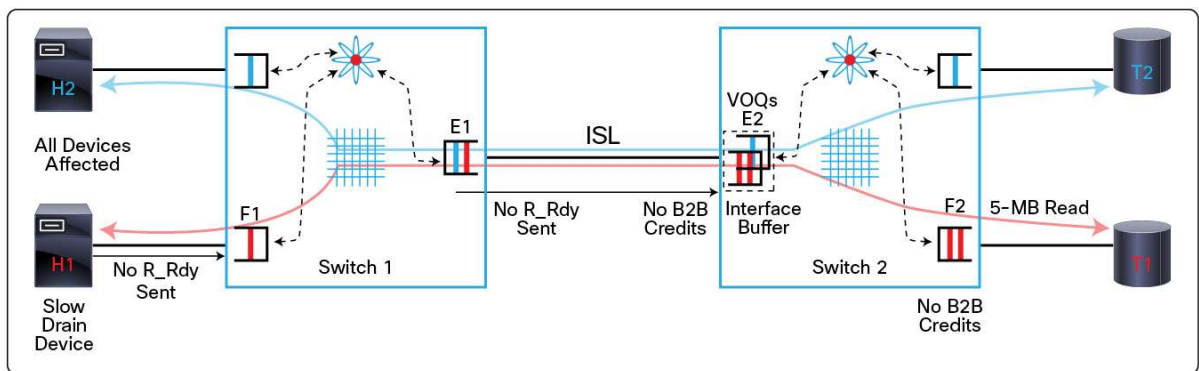
Cisco MDS 9000 Family line cards support up to 4096 VOQs, allowing these line cards to address up to 1024 destination ports per chassis, with four QoS levels per port.

Slow Drain Example

Consider the situation in Figure 5. Host H1 sends a large 5MB read to Target T1. A SCSI packet is 2148 bytes with 36 bytes of start-of-frame (SoF), Fibre Channel header, end-of-frame (EoF), and cyclic redundancy check (CRC) data and a payload of 2112 bytes.

Therefore, the response from the target will be about 2500 SCSI data packets. Host H1 cannot process all the data packets and begins to delay the sending of an R_Rdy message to the port F1 on Switch 1, creating a slow drain situation. On Switch 2, target T1 continues to send packets to port F2, filling its VOQ. However, because port F1 cannot send packets to host H1 fast enough, the packet buffers start to become full, and this congestion then cascades upstream to port E1 and eventually to port E2. At this point, the congestion has spread in the network such that the only link between the two switches is not able to send data, and the single slow device (host H1) has now adversely affected the performance of the entire network. Given the seriousness of this situation, it is extremely important to be able to monitor the network for slow drain devices and detect their presence at the earliest possible time so as to be able to take preventive measures.

Figure 5. Slow Drain Scenario



Slow Drain Detection and Congestion Avoidance

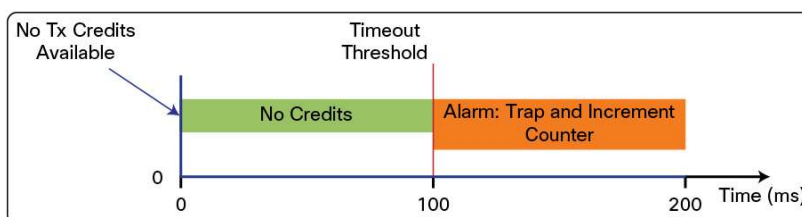
The Cisco slow drain detection and congestion avoidance solution is based on the capability to identify the following and take remedial action:

- Slow ports
- Stuck ports

Slow Ports

A Slow port is a port that receives Tx Credits slowly: that is, the receiver of the Fibre Channel packet does not immediately return an R_Rdy message to the sender (Figure 6). This causes the Tx B2B credits to drop to zero for a period of time. If the Cisco MDS 9000 Family device has packets to send but has zero Tx B2B credits, then those packets must wait. If the Tx B2B credits are at zero for 100 milliseconds (ms) or longer, then the supervisor will detect this and increment the FCP_SW_CNTR_TX_WT_AVG_B2B_ZERO counter and generate a trap (if enabled). Even if a port has zero B2B credits for just a few milliseconds, significant congestion exists and can cause problems in the rest of the fabric.

Figure 6. Slow Port Detection



Cisco MDS 9000 Family switches constantly monitor for symptoms of slow ports using the following counters:

- Tx Credit Not Available: When Tx B2B credits are at 0 for 100 ms, the switch increments the FCP_SW_CNTR_TX_WT_AVG_B2B_ZERO counter
- Rx Credit Not Available: When Rx B2B credits are at 0 for 100 ms, the switch increments the FCP_SW_CNTR_RX_WT_AVG_B2B_ZERO counter
- Transmit B2B credit transitions to zero: Each time the “transmit B2B credit remaining” drops to zero, the THB_TMM_PORT_TBBZ_CH0 counter is incremented along with the “transmit B2B credit transitions from zero” value in the show interface counters command
- Receive B2B credit transitions to zero: Each time the “receive B2B credit remaining” drops to zero, the THB_RCM_RCPx_RBBZ_CH0 counter is incremented along with the “receive B2B credit transitions from zero” value in the show interface counters command

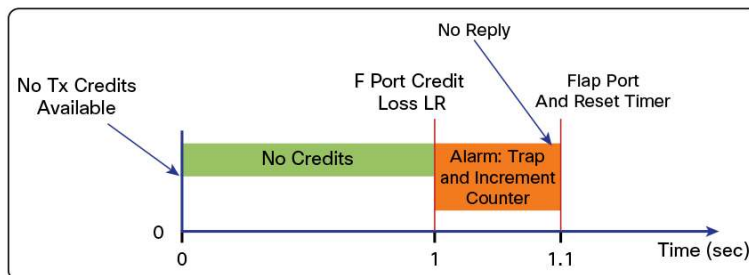
Note: All these events will occur without any configuration by the user.

Stuck Ports

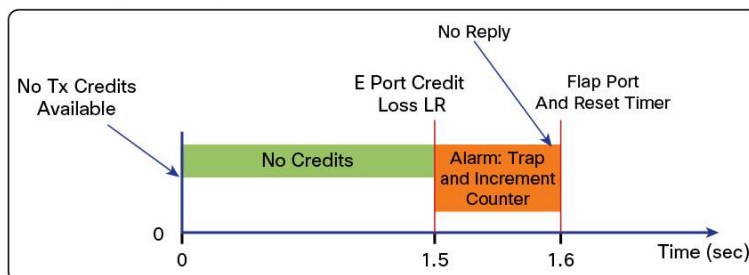
A Stuck port is a port that is continuously out of Tx credits for a given time interval that is longer than the slow port threshold (this threshold is set at 1 second for F ports and 1.5 seconds for E ports and is not configurable). A stuck port usually results in no traffic flow for an extended period of time (Figure 7).

Figure 7. Stuck Port Detection

Stuck port processing on F ports



Stuck port processing on E ports



Cisco MDS 9000 Family switches constantly monitor for symptoms of stuck ports using the following counters:

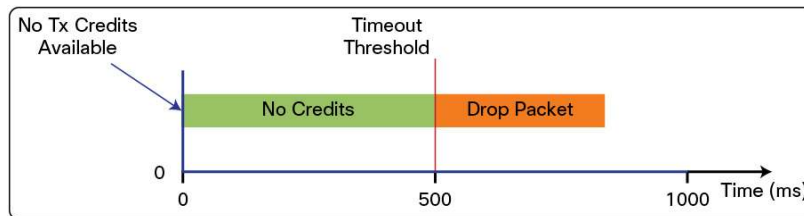
- Credit Loss Recovery Event: When the Tx B2B credits are at 0 for 1 second (F ports) or 1.5 seconds (E ports), then the switch will increment the FCP_SW_CNTR_CREDIT_LOSS counter. It will also invoke the credit loss recovery mechanism by initiating a link reset (LR)

- LR Tx (Link Reset Transmitted): When the switch sends an LR primitive, it is not counted anywhere. However, Link Reset Response (LRRs) are counted for the interface. Also, input LR and LRR primitives are counted through IP_FCMAC_INTR_PRIM_RX_SEQ_LR and IP_FCMAC_INTR_PRIM_RX_SEQ_LRR
- LR Rx (Link Reset Received): Input LR primitives are counted through IP_FCMAC_INTR_PRIM_RX_SEQ_LR

Note: All these events will occur without any configuration by the user.

In addition, frames may be dropped inside the switch and the Timeout Discards counter incremented if frames are received at a rate at which they cannot be transmitted (Figure 8).

Figure 8. Timeout Discards



- Timeout Discards (frames discarded because timeout occurred within the switch): When a packet is discarded as a result of a timeout, the switch increments various timeout drop counters. On Generation-4 line cards, this counter is the THB_TMM_TOLB_TIMEOUT_DROP_CNT counter. Please see [Appendix B](#) for information about other-generation line cards. This value is counted against the potential egress port. By default, the timeout value is 500 ms, but this value can be modified. Please see [Dropping Packets When the Congestion-Drop Threshold Is Reached](#) in the Recovery section of this document for more information.

Port Monitor

The Port Monitor process allows the monitoring of various conditions in the switch and can alert as well as take automatic action when the condition occurs. The port monitor process provides more than 15 different counters that can be monitored. Both rising and falling threshold values can be configured to generate actions when they occur. The following counters apply to the slow drain solution:

- Tx (Transmit) Credit Not Available
- Credit Loss Recovery Event
- Timeout Discards (frames discarded because a timeout occurred within the switch)
- LR Tx (Link Reset transmitted)
- LR Rx (Link Reset received)

These counters can also be monitored using Simple Network Management Protocol (SNMP) with the following object identifiers (OIDs):

- fclfTxWtAvgBBCreditTransitionToZero: 1.3.6.1.4.1.9.9.289.1.2.1.1.38
Note: There is no OID in the Rx direction.
- fclfCreditLoss: 1.3.6.1.4.1.9.9.289.1.2.1.1.37
- fclfLinkResetOut: 1.3.6.1.4.1.9.9.289.1.2.1.1.10
- fclfTimeOutDiscards: 1.3.6.1.4.1.9.9.289.1.2.1.1.35
- fclfOutDiscards: 1.3.6.1.4.1.9.9.289.1.2.1.1.36

For more information about each OID, such as the object name and MIB, use the [Cisco SNMP Object Navigator](#).

If no other user-defined port-monitor policy is explicitly enabled, the policy named “slowdrain” will be active on all Cisco MDS 9000 Family switches. This policy contains only two monitored counters, as shown below, and cannot be changed. If more counters need to be monitored then a new policy should be created, the “slowdrain” policy deactivated and the new policy activated.

```
MDS# sh port-monitor
```

```
-----  
Port Monitor   : enabled  
-----
```

```
Policy Name    : slowdrain  
Admin status   : Active  
oper status    : Active  
Port type      : All Access Ports  
-----
```

Counter	Threshold	Interval	Rising Threshold	event	Falling Threshold	event	PMON	Portguard
-----	-----	-----	-----	-----	-----	-----	-----	-----
Credit Loss Reco	Delta	1	1	4	0	4	Not enabled	
TX Credit Not Available	Delta	1	10	4	0	4	Not enabled	

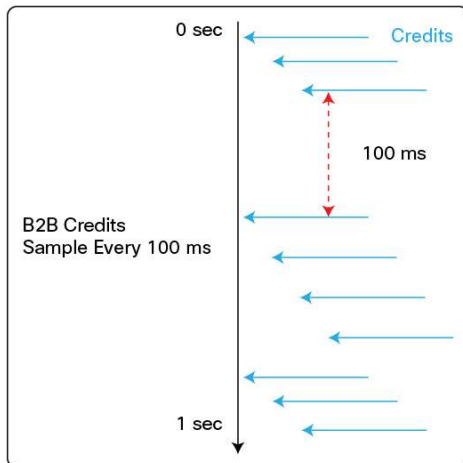
```
-----
```

See [Appendix A](#) for a Port-Monitor Configuration example.

Detection of Credit Not Available Event on Cisco MDS 9500 Series

Cisco MDS 9500 Series switches implement a software-based slow drain detection algorithm. The linecard supervisor continually polls the hardware port application-specific integrated circuit (ASIC) to determine the health of the system. Specifically, the B2B credits are polled every 100 ms (which is a very long time in Fibre Channel); see Figure 9.

Figure 9. B2B Credit Sampling on Cisco MDS 9500 Series



This approach gives a good snapshot of what the system is currently experiencing. This approach does use additional system resources and imposes a limit on the frequency of problem detection:

- The supervisor needs to constantly dedicate the processor to polling the hardware
- The Supervisor needs to constantly make a decision about whether or not to trigger an action on the basis of a predefined policy
- Because this feature is a snapshot mechanism, the software may not be able to detect that a credit is not available if the credit happens to be available just at the time that the software polls the hardware, potentially resulting in an inaccurate depiction of the hardware resources

Detection of Credit Not Available Event on Cisco MDS 9700 Series

The Cisco MDS 9700 Series switch uses a hardware-based slow drain detection algorithm. In this approach, slow drain detection and monitoring is built into the port ASIC, and instead of relying on the software for polling, the hardware can automatically detect every time a credit is not available and take appropriate action without any intervention from the supervisor.

The benefits of hardware-based slow drain detection algorithm are as follows:

1. The slow port detection threshold can be made configurable and it can be made less than the current value of 100ms. Currently, this is not user configurable but it will be in a future release of NX-OS.
2. The slow port detection timer is set when the Tx credits go to zero. If the timer expires while the port is still at zero Tx credits then it is counted immediately not waiting for a polling interval.
3. Reduced load on the linecard supervisor since no slow port polling is taking place.

Identifying a Slow Drain Port

Levels of Performance Degradation

Problems can essentially be divided into three levels of service degradation (Table 1).

Table 1. Levels of Performance Degradation

Level	Host Symptoms	Default Switch Behavior
1	Latency	Frame queuing
2	SCSI retransmission	Frame dropping
3	Extreme delay	Link reset

Level 1: Latency

The most subtle indicator of a slow drain device is use of all data frame credits on ISLs. If this is the only symptom, then data is delayed but not dropped, and the device is reporting latency. In the case of congestion without drops or link resets, ISLs should be checked for links that most frequently have zero low priority Tx credits available. These links can be suspected to be upstream of the slow drain device.

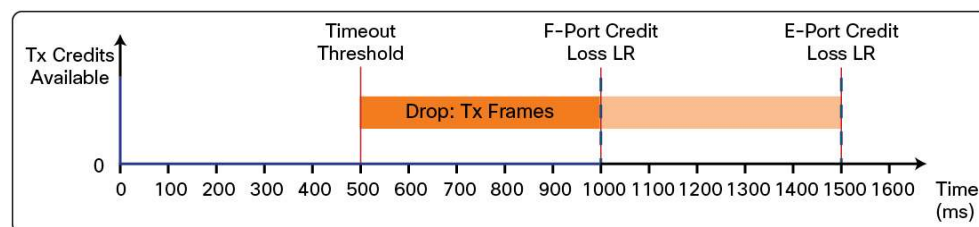
Level 2: Retransmission

Any frame that cannot be delivered to its destination port is held for a maximum of 500 ms (default). If that value is reached, the frame is dropped (Tx timeout drop). This behavior is independent of port recovery, and the timeout value can be tuned to a lower value in Cisco NX-OS Software Release 4.2(7) and later.

Level 3: Extreme Delay

If a port stays at zero Tx B2B credits for a long period of time, the credit monitor process tries to reset the peer of the local port to bring it to a known good state. It does this by sending logical and physical resets. Figure 10 shows the timing.

Figure 10. Credit Monitor Process



F Ports

```
timer = 0      Port Tx B2B credits go down to 0 (TBBZ or lack_of_transmit_credit);
                no credits received (considered to be credit loss)
timer = 1      Send LR, wait 100 ms for reply
timer = 1.1    If no reply, flap port, reset timer
```

E Ports

```
timer = 0      Port tx B2B credits go down to 0 (TBBZ or lack_of_transmit_credit);
                no credits received (considered to be credit loss)
timer = 1.5    Send LR, wait 100 ms for reply
timer = 1.6    If no reply, flap port, reset timer
```

Here:

- Timer values 0, 1, 1.1, 1.5, and 1.6 are measured in seconds. This is performed by the creditmon (system) process
- TBBZ refers to the total count of Tx B2B credits transitions to zero and is applicable to Generation 4 and Cisco MDS 9700 Series modules. The count is a representation over a period of time (since the module is up or the counter wraps around) of the number of times that a port has no credits to transmit or receive
- Lack of transmit credit is applicable to Generation 2 and Generation 3 modules

Link Event “LR Rcvd B2B” on Fibre Channel Ports

If a port stays at zero Rx B2B credits for a long period of time, a link reset may be initiated by the adjacent Fibre Channel device (because it presumably has zero Tx credits). When this reset occurs, the Cisco MDS 9000 Family switch port receives an LR primitive. The port checks its ingress buffers and determines whether at least one packet is still queued. If no packets are queued (that is, if all received packets have been delivered to their respective destination egress ports), then an LRR primitive is returned. Both the adjacent Fibre Channel device and the Cisco MDS 9000 Family switch port are now back at their full complement of B2B buffers. The link continues to function uninterrupted.

If at least one packet is still queued, the Cisco MDS 9000 Family switch starts a 90 ms LR Rcvd B2B timer. If the Fibre Channel packets can be transmitted to the egress port, then the LR Rcvd B2B timer is canceled and an LRR message is sent back to the adjacent Fibre Channel device. However, if the egress port remains congested and Fibre Channel frames are still queued at the ingress port, the LR Rcvd B2B timer will expire. No LRR is transmitted back to the adjacent Fibre Channel device, and both the ingress port and the adjacent Fibre Channel device initiate a link failure by transmitting a Not Operational Sequence (NOS).

The link will go down with the following message:

```
%PORT-5-IF_DOWN_LINK_FAILURE: %$VSAN 100%$ Interface fc5/32 is down (Link failure
Link Reset failed nonempty rcv queue)
```

The link event log looks like this:

```
module-11# show port-config internal link-events
***** Port Config Link Events Log *****
-----
Time                PortNo          Speed  Event  Reason
-----
...
Jul 28 00:46:39 2012 00670297 fc11/25    ---   DOWN   LR Rcvd B2B
```

Typically, the FCP_SW_CNTR_RX_WT_AVG_B2B_ZERO counter will also increment during this time. Consequently, the preceding message is an indication of severe slow drain congestion, but the cause is not the port that failed. The slow or stuck port that caused the failure is the one with which the adjacent Fibre Channel device is attempting to communicate. It could be an F port or an E port in a multiswitch environment.

Refer to the [Level-3 Troubleshooting: Extreme Delay](#) section later in this document for methods for identifying the actual slow or stuck port.

In identifying the slow drain device, be aware of the following:

- Logs are detailed and can roll over on an active port. Troubleshooting should begin quickly when slow drain problems are detected
- If credit recovery and/or transmit frame drops occur on an ISL, then traffic destined to any egress port on the switch may be affected and so multiple edge devices may report errors. If either condition is seen on ISLs, then the investigation should continue on the ISL peer switch. If an edge switch shows signs of frame drops, then each edge port on the edge switch should be checked
- If credit recovery and/or transmit frame drops occur on an edge port, traffic to or from that port is affected as well as the fabric path to that affected port. Therefore, edge ports that show some or all the signs of congestion may be the source of the fabric congestion
- If a port has serious congestion, the port will show timeout frame drops, credit loss, and/or link reset events

Troubleshooting Slow Drain

Cisco suggests looking for the most severe problems first (Level 3); only when those problems are resolved should one start looking for less severe problems (Level 2 and then Level 1).

Level-3 Troubleshooting: Extreme Delay

The creditmon process on the switch monitors each port for this condition and attempts to correct it by logically resetting the link (LR) to reset the state with its peer. A port without any Tx credits (Tx B2B = 0) for an extended period of time (which is an abnormal condition) initiates credit loss recovery. This scenario can occur if R_Rdy frames are lost because of corruption or if the adjacent device is withholding R_Rdy frames because of slow drain. If the peer port does not reply, then the link is physically reset (flapped). The time limit for F ports is 1 second, and for E ports the limit is 1.5 seconds. Troubleshooting should begin on switches exhibiting credit loss recovery syslog entries and traps.

Note: A Link Reset (LR) is only a credit reset and does not actually flap the port if it is successful.

Check for Link Failure with “LR Rcvd B2B” Message

If the link failed with “LR Rcvd B2B” or “Link failure Link Reset failed nonempty Recv queue” message, then the port that failed is not the cause of the slow drain but is only a port that was affected by the slow or stuck port. To identify the slow or stuck port that caused the link failure, use following steps:

1. Determine whether more than one link is failing
2. Check the VSAN zoning database to see with which devices the adjacent Fibre Channel device is zoned. Map these to egress E ports or F ports. To map to egress E ports, use the “show fspf internal route vsan <vsan> domain <dom>” command. To map to local F ports, use the “show flogi database vsan <vsan>” command. If more than one is link failing with “LR Rcvd B2B,” then combine the egress E ports or F ports found and check for overlap. Overlapping ports are likely candidates for slow or stuck ports
3. Check the ports found in step 2 for indications of slow drain. Examples are credit loss (FCP_SW_CNTR_CREDIT_LOSS), 100 ms Tx B2B zero (FCP_SW_CNTR_TX_WT_AVG_B2B_ZERO), and timeout discard (THB_TMM_TOLB_TIMEOUT_DROP_CNT)
4. If the slow port is determined to be an E port, then continue slow drain troubleshooting on the adjacent switch indicated by the Fabric Shortest Path First (FSPF) next-hop interface

5. If the port is determined to be a Fibre Channel over IP (FCIP) link or PortChannel, then check the FCIP links for signs of IP retransmission or other problems such as link failures. "show ips stats all" can be used to check for problems

Check for Credit Loss on an Egress Port

Credit loss events should be investigated. This problem is reported only through a counter with no indication as to when these events occurred. Repeat the command to see if the counter is moving.

On the Cisco MDS 9700 Series DS-X9448-768-K9 module:

```
module-3# show process creditmon statistics
```

```
*****Credit Monitor Info*****
Port      Mode      Monitor  Cr Loss      Slow-Port-Detection
              on?      Detected?
-----
   1         E        yes         0         yes         0
   2         E        yes         0         yes         0
...
  13         F        yes         0         yes         2
  14         F        yes         0         yes         0
    ^
    |
egress port                                count
```

The command shows that port fc3/13 in F port mode has experienced credit-loss recovery twice. Debugging should continue on the edge device to which this port is connected.

On the Cisco MDS 9500 Series DS-X92xx-256K9, DS-X92xx-96K9, and DS-X92xx-48K9 modules:

```
switch# attach module <module_number>
module-x# show process creditmon statistics
*****      Credit Monitor Info      *****
Port  Mode  Monitor VL Cr Thresh      Cr Loss
              Exceeded  Event
-----
   1     E    yes   NA    0         7
   ^
   |
egress port                                count
```

The following command shows the affected interface and the timestamp when credit loss recovery occurred:

```
switch# show process creditmon credit-loss-events
      Module: 01      Credit Loss Events: NO
CLI is not supported on module 3
      Module: 04      Credit Loss Events: YES
-----
| Interface | Total |          Timestamp          |
|           | Events |                             |
-----
| fc4/3     |      2 | 1. Fri Jun 28 20:42:12 2013 |
|           |      | 2. Thu Jun 27 10:24:17 2013 |
-----
| fc4/4     |      1 | 1. Wed Jun 26 22:32:55 2013 |
-----
| fc4/20    |      2 | 1. Thu Jun 27 10:24:17 2013 |
|           |      | 2. Wed Jun 26 22:32:55 2013 |
-----
| fc4/21    |      1 | 1. Thu Jun 27 10:06:54 2013 |
-----
      Module: 06      Credit Loss Events: NO
      Module: 09      Credit Loss Events: NO
```

Note: The previous command will be available for the Cisco MDS 9710 Multilayer Director in a future Cisco NX-OS release.

Each credit recovery event is logged and should be investigated. The log is cleared if the module is reset for any reason. The following is the port link event finite state machine (FSM) showing all events and states pertaining to Link Reset (LR) for the port undergoing credit loss recovery:

```
module-x# show hardware internal fcmac port <port_number> link-event | include
LINK_LR_TX      ===== FCP Port#x Link State Machine Event Log =====
MMDDYY HHMMSS usecs          Event                      Current State
-----
032910 171903 987601 (0000)  E_LINK_LRR                      LINK_LR_TX
032910 171900 787616 (0000)  E_LINK_LRR                      LINK_LR_TX
032910 171857 587598 (0000)  E_LINK_LRR                      LINK_LR_TX
032910 171854 387591 (0000)  E_LINK_LRR                      LINK_LR_TX
032910 171851 187606 (0000)  E_LINK_LRR                      LINK_LR_TX
-----
^                               ^
|                               |
event timestamp                time in 10 ms increments between events
```

Note: Each time an LRR message is received on a port, it is recorded in the respective port's FSM log.

Level-2 Troubleshooting: Retransmission

Frames destined for a heavily used egress port will be dropped if the egress port is too slow. By default, frames are dropped after being held for 500 ms.

Check for Transmit Frame Drops

Frames can be dropped in a switch for multiple reasons. For transmit timeout drops, troubleshooting should begin at the edge switches connected to the end devices reporting problems. If a slow drain device is causing frame drops, they will be identified as transmit drops by the MAC ASIC on the egress port. Possible egress port MAC ASIC names are F16, Thunderbird (Tbird), Stratosphere, Aakash, Atlantis-fcp, and Taquile-fcp. This problem is reported only through a counter with no indication as to when these events occurred. Repeat the command to see if the transmit drop counter is moving, or use "show logging onboard <module x> error-stats" to see time-stamped entries.

On Cisco MDS 9500 Series DS-X92xx-256K9 and MDS 9710 DS-X9448-768-K9 modules:

```
switch# show hardware internal statistics module <x> pktflow dropped
Hardware statistics on module 01:
```

```
...
|-----|
| Device:F16 Mac Driver           Role:FCMAC           Mod: 1 |
| Last cleared @ Wed Jun  5 17:43:58 2013 |
|-----|
Instance:1
ID   Name                               Value              Ports
--   ---
286  F16_TMM_TIMEOUT_STATS_DROP          0000000013174890   13-16 -
1825 F16_TMM_PORT_FRM_DROP_CNT           0000000013419266   13 -
1954 F16_TMM_TOLB_TIMEOUT_DROP_CNT      0000000013419266   13 -
```

In the preceding example, ports fc1/13-16 are experiencing timeout drops and should be investigated.

On Cisco MDS 9500 Series DS-X92xx-96K9 and DS-X92xx-48K9 modules:

```
switch# show hardware internal packet-flow drop
```

```
Module: 03      Dropped Packets: YES
```

DEVICE NAME		PORTS	DROPPED COUNT	
			RX (Hex)	TX (Hex)
Aakash	3		0	1c
Aakash	4		0	10
^	^			^
MAC device name		egress	count	

Check for Frame Transmit Timeout History on Egress Port

On Cisco MDS 9500 Series DS-X92xx-256K9 and MDS 9710 DS-X9448-768-K9 modules:

On Cisco MDS 9500 Series DS-X92xx-96K9 and DS-X92xx-48K9 modules:

ERROR STATISTICS INFORMATION FOR DEVICE ID 118 DEVICE Aakash

Interface Range	Error Stat Counter Name	Count	Time Stamp MM/DD/YY HH:MM:SS
fc5/1	AK_FCP_CNTR_RXF_TIMEOUT_FRAMES	12378	06/06/13 09:57:46
fc5/1	AK_FCP_CNTR_LAF_C3_TIMEOUT_FRAME	2900	06/06/13 09:57:46
fc5/1	AK_FCP_CNTR_LAF_TOTAL_TIMEOUT_FR	12378	06/06/13 09:57:46
fc5/1	AK_FCP_CNTR_RXF_TIMEOUT_FRAMES	7968	06/06/13 09:57:36
fc5/1	AK_FCP_CNTR_LAF_C3_TIMEOUT_FRAME	1794	06/06/13 09:57:36

Level-1 Troubleshooting: Latency

A slow draining egress port may not be slow enough to cause frame drops, but still slow enough to exert pressure in the switch to ingress ports. Such cases of slight congestion are not indicated directly by any counters or logs. Such problems can be inferred from instantaneous outputs and by observing any trends. By default, data frames are buffered for up to 500 ms in a switch. This congestion may show up as poor host-disk response time.

Check for Credit Not Available

Switch credit nonavailability can be monitored using the command shown here:

```
switch# show system internal snmp credit-not-available
```

```
Module: 1      Number of events logged: 0
Module: 3      Number of events logged: 0
Module: 4      Number of events logged: 3
```

Port	Threshold	Rising/Falling	Interval(s)	Event Time	Type	Duration of time not available
fc4/3	10/0(%)		1	Fri Jun 28 20:42:13 2013	Rising	50%
fc4/3	10/0(%)		1	Fri Jun 28 20:42:14 2013	Falling	0%
	0/0(%)		0	Wed Dec 31 19:00:00 1969	Falling	0%

```
Module: 6      Number of events logged: 0
Module: 9      Number of events logged: 0
```

Lack of B2B Credits for the Distance the ISL is Traversing

Each fibre channel frame is a fixed length on the actual fibre. As an example, a 2KB fibre channel frame at 4Gbps is approximately 1km in length. As the speed doubles the length of the frame halves so the same 2KB frame that was 1km at 4Gbps is 0.5km at 8Gbps and only 0.25km at 16Gbps. So, the minimum number of B2B credits needed to utilize the full ISL bandwidth can be obtained by taking the round trip distance and dividing it by the average frame size.

Check ISLs for Lack of Transmit Credits

Data frames on ISLs are sent using low priority transmit credits. By monitoring ISL low priority credits, instances of zero credits should be noted. Every time the number of credits drops to zero, all devices using that ISL experience a definite delay, and the problem should be traced to the next downstream hop. This process also applies to ISLs in PortChannels.

In the following example, fc2/1 has no Tx credits available, and fc3/1 has available Tx credits but may be below its full complement. This scenario may indicate congestion in the fabric:

```
switch# show interface | include "fc|Belong|low priority" | exclude "description" |  
exclude "Peer" | include "trunking" next 2  
fc2/1 is trunking  
    Belongs to port-channel 20  
    0 low priority transmit B2B credit remaining  
fc3/1 is trunking  
    Belongs to port-channel 20  
    185 low priority transmit B2B credit remaining
```

Do Not Monitor Tx B2B Transitions to Zero on ISLs in a PortChannel

Few credits are reserved for absolute priority (or control) traffic on an ISL. Essentially, 1 in every 8 credits (with a maximum of 15) is reserved for absolute priority traffic only. All the user and data traffic (which could be high, medium, or low priority) is sent using the remaining shared credits. For example, if an ISL has 64 credits, then 8 of them are reserved for control traffic, and the data traffic cannot use more than 56 credits. However, if there is more control data to be sent on the same ISL, then the control traffic, having absolute priority, can use the remaining 56 credits, leaving no credits for regular data traffic.

Following the design just described, if congestion occurs in a PortChannel, all ISLs should be experiencing the flow control, but the transitions to zero counter may reach zero only on one of them. This result occurs because the control traffic consumes the high priority credits. The other interfaces do not carry control frames, and so the high priority Tx B2B credits will always be available, and the total Tx B2B credits always stay greater than 0.

Check for Frame Queuing on Ingress Ports

The status of port egress queues may be displayed to show which destinations have frames waiting to be transmitted. Because frames that are destined for an egress port are queued at the ingress port, this information can be found on the ingress module on which the frame arrives. All possible ingress sources must be checked to build a complete picture of traffic flows to a given destination port. A destination port index that appears occasionally in the command output likely indicates a normal device. Port indices that appear regularly are likely to indicate slow drain devices. If the same destination device shows up regularly on both fabrics, then it most likely is a slow drain device.

The following command output shows egress queues for all ingress ports on a given port ASIC. When this output indicates that ISLs are the ingress ports, the previous check should be used to confirm such upstream congestion. If the two outputs correlate, then a delay will occur for any downstream device. Any ports that appear regularly in the output shown here should be investigated as the cause of congestion in the fabric. Note that this information is only an instantaneous snapshot of the queue; the command should be repeated to see whether the queue is moving.

On Cisco MDS 9500 Series DS-X92xx-256K9 modules:

```
9513# slot <x> show hardware internal que inst <#> memory iqm-statusmem<0/1>
```

```
+-----+
| IQM: PG0 Status Memory for Tbird Que Driver
```

```
| Inst 1; port(s) 9-16
```

```
|
```

Note: Only non-zero entries are displayed

Each non-zero bit indicates pending frame in VOQ for that IB

```
+-----+-----+-----+-----+-----+
| GI (Hex) | Prio 0 | Prio 1 | Prio 2 | Prio 3 |
```

```
+-----+-----+-----+-----+-----+
|          c | 000000 | 000000 | 000000 | 000001 |
```

```
+-----+-----+-----+-----+-----+
```

^

|

egress port (slow)

^

|

ingress port

This command-line interface (CLI) captures the mapping of the destination index to the global index (DI to GI mapping), revealing where the frames are destined if they are queued for an extended time. This mapping indicates that one or more packets are queued to destination index 0xC, which is associated with port fc4/13. This mapping can be determined using the command shown here:

```
9513# show system internal fcfwd idxmap port-to-interface
```

Port to Interface Table:(All values in hex)

```
+-----+
glob|                               |VL|lcl| if |slot|port| mts | port| flags
idx |          if_index            | |idx|type|   |   | node| mode|
-----|-----|-----|-----|-----|-----|-----|-----|-----
...  |
```

```
    b| 0100b000 fc1/12            | 0| 0b| 01 | 00 | 0b | 0102| 00 | 00
```

```
  194| 01613000 fc13/20           | 0| 13| 01 | 0c | 13 | 0d02| 00 | 00
```

```
    c| 0100c000 fc4/13          | 0| 0c| 01 | 00 | 0c | 0102| 00 | 00
```

```
  195| 01614000 fc13/21           | 0| 14| 01 | 0c | 14 | 0d02| 00 | 00
```

Also, this mapping can be verified with this command:

```
9513# show port internal info interface fc4/13
```

```
fc4/13 - if_index: 0x0118C000, phy_port_index: 0xc
```

```
      local_index: 0xc
```

Ingress ports that have one or more queued packets are indicated by a 24 bit hexadecimal map(3 bytes). In the above example this is 000001(hexadecimal). This indicates the first port in Thunderbird ASIC instance 1. Instance 1 is for ports 9 through 16 as shown by "Inst 1; port(s) 9-16". This is specific to this MDS 9500 Series DS-X9232-256K9 module. Other module types will have different layouts. Note that since the instance only has 8 ports only the rightmost byte(8 bits) is applicable. Hence 000001 is really 01 which when broken down to bits is 00000001. Each 1 bit represents a port that has one or more queued packets.

Since, in this example, the first bit is on, that indicates the first interface in instance 1. That works out to interface fc4/9 where the packets are ingressing. In this example they are destined to fc4/13(GI - Global Index 0xc). For a Cisco MDS 9500 Series DS-X9248-256K9 module, each Thunderbird instance handles 12 ports, so, instance 1 would start with port 13.

Similar information can be obtained on other modules:

- Cisco MDS 9710 DS-X9448-768K9 module:
show hardware internal f16_que inst 1 table iqm-statusmem0
- Cisco MDS 9500 Series DS-X92xx-96K9 and DS-X92xx-48K9 modules:
show hardware internal up-xbar 0 queued-packet-info

Note: The information displayed is real-time data, not historical data. Consequently, it should be done while the slow drain event is occurring.

Check for Arbitration Timeouts

When an ingress port needs to send a frame to an egress port, the frame is first put in the VOQ for the egress port. When the frame arrives at the head of the VOQ, a request is made to the central arbiter to transmit the frame across the xbar to the egress port. When the egress port has no transmit buffers free, the arbiter will not send a grant for the ingress request. The ingress port considers the request timed out after 90 ms.

The egress port should be investigated because this behavior is an indication of transmission delays of 90 ms or greater. These credit request timeout events can be viewed on a per egress port basis along with the ingress port numbers and a timestamp of the earliest and latest events.

On the Cisco MDS 9500 Series DS-X92xx-256K9 and MDS 9710 DS-X9448-768-K9 modules:

```
module# show logging onboard flow-control request-timeout
```

```
-----
Module: 1
-----
```

Dest	Source	Events	Timestamp		Timestamp
Intf	Intf	Count	Latest		Earliest
fc3/1	fc1/10,	8	Tue Jun 12 22:32:11 2012		Tue Jun 12 22:32:12 2012
	fc1/11,				
	fc1/12,				
	fc1/22,				
	fc1/23,				

On the Cisco MDS 9500 DS-X92xx-96K9 and DS-X92xx-48K9 modules, enter the same command at the switch prompt instead.

Note: Arbitration timeouts are not packet drops. Arbitration requests are retried and, if the request is granted, the packet can be transmitted to the egress port successfully. If a packet never receives a grant, it will eventually be dropped at the congestion-drop interval and be counted as a timeout drop.

Monitoring Tx and Rx B2B Transitions to Zero

Each time the remaining B2B credits drop to zero, the switch increments the “B2B credit transitions to zero” counter:

```
9513# show interface fc10/5 counters
fc10/5
...
 0 input OLS, 0 LRR, 0 NOS, 211 loop inits
111 output OLS, 0 LRR, 57 NOS, 104 loop inits
0 link failures, 0 sync losses, 4 signal losses
58 transmit B2B credit transitions from zero
115 receive B2B credit transitions from zero
 32 receive B2B credit remaining
 0 transmit B2B credit remaining
 0 low priority transmit B2B credit remaining
```

The counters actually increment when the credits drop to zero. The actual “credit transitions...” text in the “show interface n counters” command has had several changes. The latest change was made through the following bugid:

CSCug35184 show interface counters - transitions of rx BB credit to zero state

This fix changed the text to the following:

```
n Transmit B2B credit transitions to zero
n Receive B2B credit transitions to zero
```

A variety of text may be seen based on different releases of NX-OS, but it all indicates credit transitions to zero. The transitions to zero are also maintained in the hardware statistics through the counters shown here:

- Receive counters: FCP_CNTR_RCM_CH0_LACK_OF_CREDIT, AK_FCP_CNTR_RCM_CH0_LACK_OF_CREDIT, THB_RCM_RCP0_RBBZ_CH0, and F16_RCM_RCP0_RBBZ_CH0
- Transmit counters: FCP_CNTR_QMM_CH0_LACK_OF_TRANSMIT_CREDIT, AK_FCP_CNTR_QMM_CH0_LACK_OF_TRANSMIT_CREDIT, THB_TMM_PORT_TBBZ_CH0, and F16_RCM_RCP0_TBBZ_CH0

See [Appendix B](#) for an explanation of the counters.

Note: It is normal for the “transitions to zero” and associated hardware statistics counters to increment occasionally. However, when the transitions to zero are excessive, performance will be affected. Identifying what is excessive is difficult to do, and there are no strict rules. This should only be considered if all the troubleshooting steps (above) have been performed.

Recovery

Cisco MDS 9000 Family switches can automatically take several recovery actions to mitigate the effects of slow drain devices. These actions include the following:

- Error Disable or flap the port through port-monitor port guard
- Dropping packets when the congestion-drop threshold is reached
- Setting the no-credit-drop timeout value
- Performing credit-loss recovery

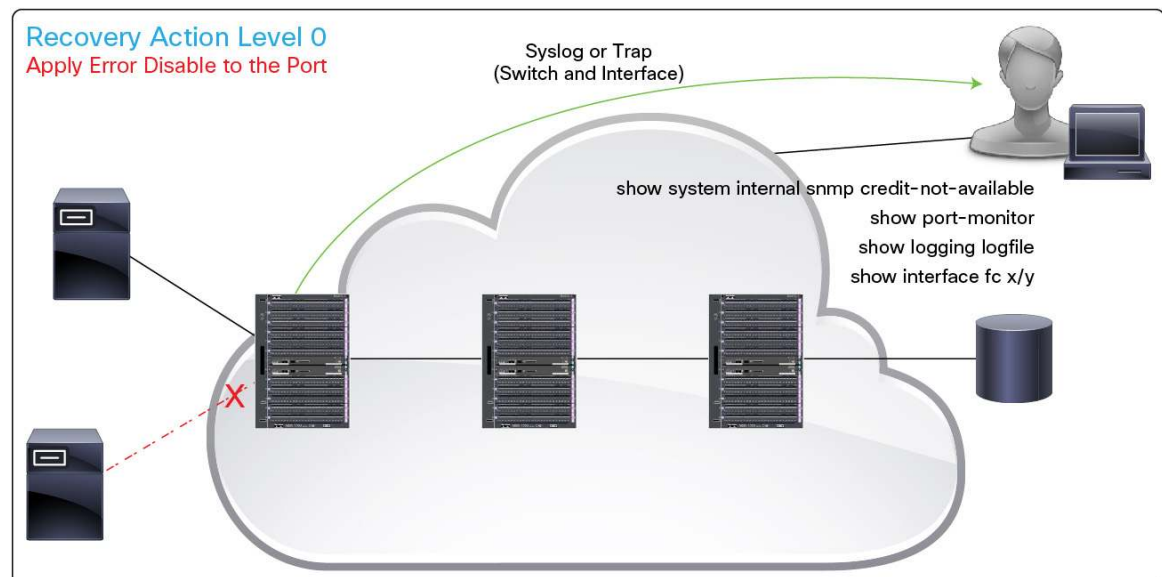
Some of these actions are enabled by default, and some need to be configured manually.

Error Disable or Flap the Port Through Port-Monitor Port Guard

As mentioned earlier, the port monitor has several counters that can be monitored, and when the configured rising threshold is reached, port guard can take one of two actions:

- Flap the port
- Error disable the port (Figure 11)

Figure 11. Error Disable the Port



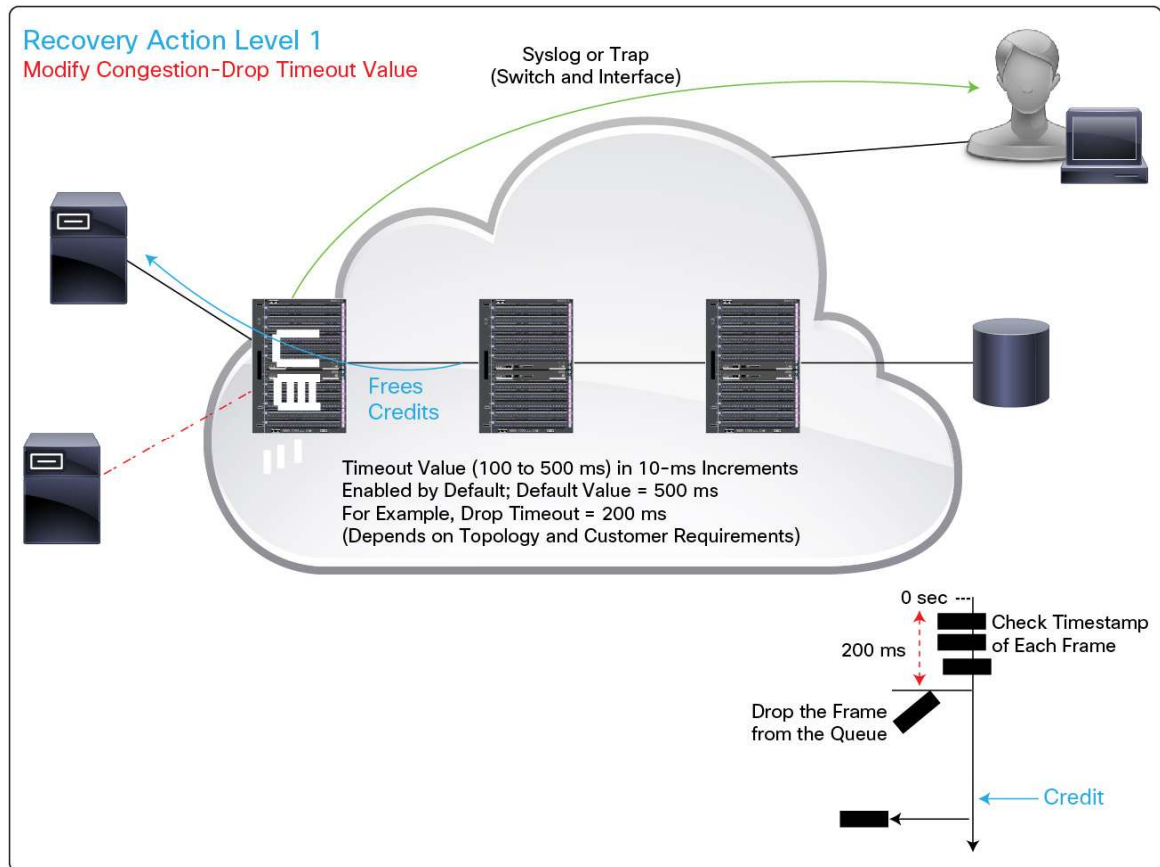
The action (Error Disable and link flap) and threshold depend on the topology and customer requirements and are customer driven configurations. Error Disabled ports can be recovered by using the “shut” and “no shut” commands.

Note: Error Disabling a port is a very disruptive event and is not recommended.

Dropping Packets When the Congestion-Drop Threshold Is Reached

By default, Cisco MDS 9000 Family switches drop packets that have not been delivered to their egress ports at 500 ms. Changing the congestion-drop timeout value to a lower value can help drop packets that have been stuck in the system more quickly, freeing up the buffers faster in the presence of slow drain (Figure 12). By default, this value is set to 500 ms.

Figure 12. Modifying Congestion-Drop Timeout Value



This value can be set at the switch level for each port type as:

```
switch(config)# system timeout congestion-drop <value> mode (F) / (E)
switch(config)# system timeout congestion-drop default mode (F) / (E)
```

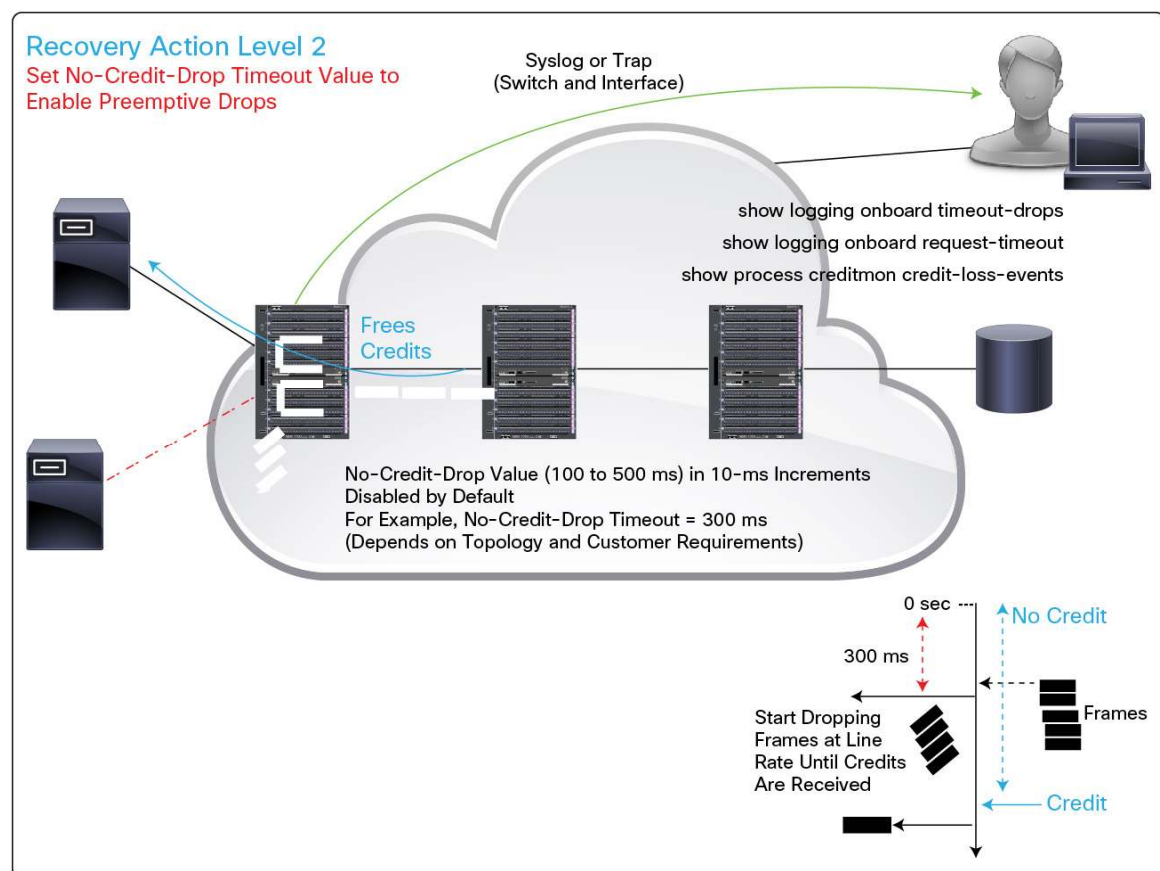
Note: Reduce the congestion-drop timeout to 200ms or lower (depending upon topology or requirements). This will timeout frames and speed up the freeing of buffers within the switch.

Setting the No-Credit-Drop Timeout Value

When a port has zero Tx B2B credits for the value configured, new packets that arrive will immediately be dropped (they are not subject to the current congestion-drop timeout value). This behavior frees buffer resources more quickly than in the normal congestion-drop timeout scenario and can alleviate much of the problem on an ISL in the presence of a slow drain device. By default, this option is off. It can be configured at the switch level for the F port type only (Figure 13):

```
switch(config)# system timeout no-credit-drop <value> mode F
switch(config)# system timeout no-credit-drop default mode F
```

Figure 13. Set No-Credit-Drop Timeout Value



Even if there is little slow drain behavior in the fabric currently, and if hosts and storage are not generating any delays, configuring the congestion-drop and no-credit-drop values as 200 ms can reduce the effects of problems that occur in the future.

If however, there is evidence of existing slow drain in the fabric then that needs to be cleaned up prior to implementation. Specifically, check the fabric logging onboard for FCP_CNTR_RX_WT_AVG_B2B_ZERO, AK_FCP_CNTR_RX_WT_AVG_B2B_ZERO, FCP_SW_CNTR_RX_WT_AVG_B2B_ZERO, FCP_CNTR_TX_WT_AVG_B2B_ZERO, AK_FCP_CNTR_TX_WT_AVG_B2B_ZERO, and FCP_SW_CNTR_TX_WT_AVG_B2B_ZERO counters, using the following CLI command:

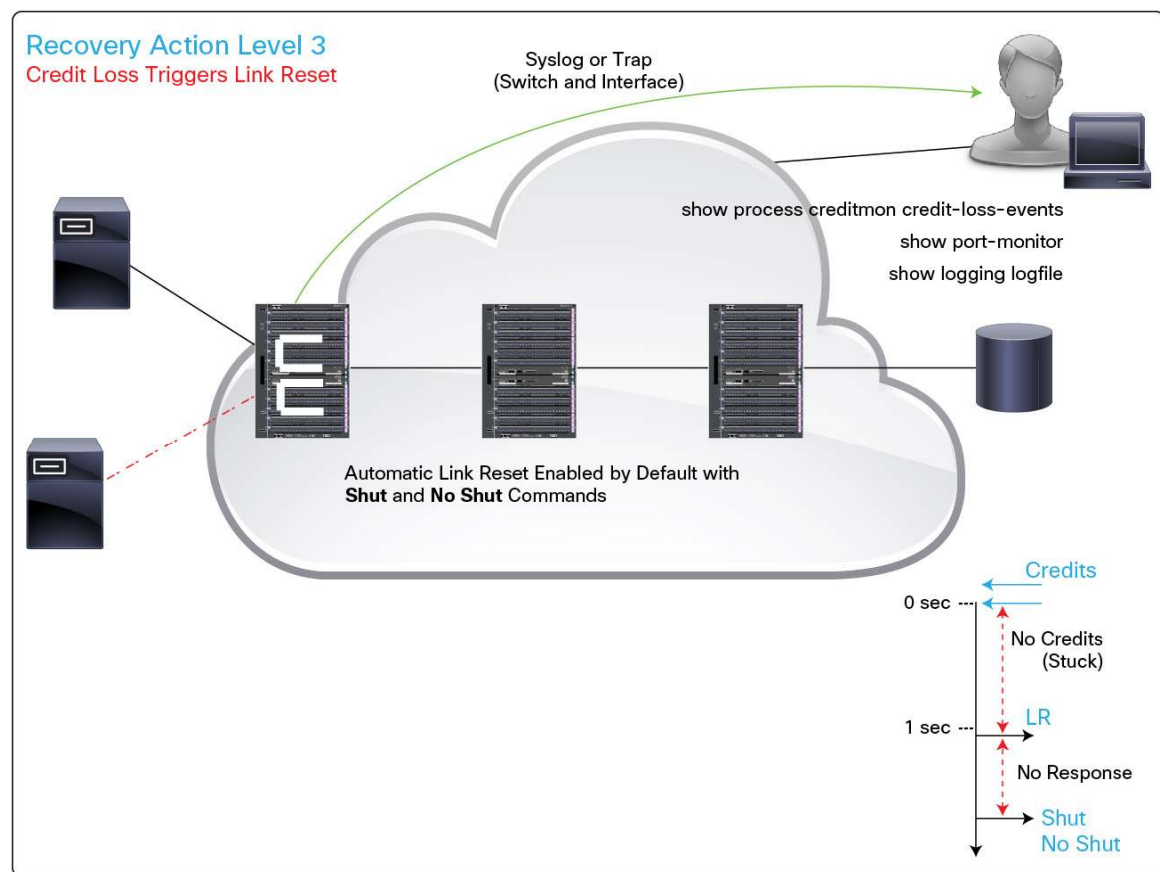
```
show logging onboard starttime MM/DD/YY-00:00:00 error-stats
```

Because these values represent 100 ms intervals at zero Rx or Tx credits, if these messages are not seen, then the fabric is probably performing okay, and setting the no-credit-drop timeout value should be considered. This value can be configured using the Cisco Fabric Manager or Cisco Data Center Network Manager (DCNM) and choosing Tools > Run CLI Commands.

Automatic Credit Loss Recovery

Credit loss triggers a Link Reset, which attempts to recover the B2B credits in both directions (Figure 14). If successful, this reset does not cause a link flap or any link disruption.

Figure 14. Credit Loss Triggers Link Reset



Timers are set to 1 second for F ports and 1.5 seconds for E ports and are not configurable. Link Reset (LR) occurs if the interface runs out of credits for 1 second (F port) or 1.5 seconds (E port) continuously.

Note: This option is enabled by default. Every time a Link Reset Response is lost or not returned, the system automatically triggers a link flap.

Appendix A: Slow Drain Port-Monitor Detection Configuration

Slow Drain Port-Monitor policy can be created for access ports, trunk ports, or all ports. Only one policy can be active for each port type at a time. If the port type is all ports, then there can be only one active policy. This appendix presents a sample configuration.

Configuring the Tx-Credit-Not-Available Counter

The tx-credit-not-available counter can be configured at the switch level or port level (F port, E port, or all ports):

```
switch(config)# port-monitor name Cisco
switch(config-port-monitor)# port-type <access/all/trunk>
    access-port    Configure port-monitoring for access ports
    all            Configure port-monitoring for all ports
    trunks        Configure port-monitoring for trunk ports
switch(config-port-monitor)# counter tx-credit-not-available poll-interval <1>
<delta> rising-threshold <10> event <4> falling-threshold <0> event <4> portguard
errordisable
```

- **port-type:** This setting allows the user to customize the specific policy to access or trunk ports or all ports
- **poll-interval:** This counter indicates the polling interval within which slow drain statistics will be collected. It is measured in seconds (it is configured to 1 second in this example)
- **Threshold type:** This setting determines the method for comparing the counter with the threshold. If the type is set to **absolute**, the value is compared to the threshold at the end of the interval. If the type is set to **delta**, the change in value of the counter during the polling interval is compared to the threshold
- **rising-threshold:** An alert is generated if the counter value was lower than the **rising-threshold** value in the last polling interval and is greater than or equal to this threshold at this interval. Another alert will not be generated until the counter is less than or equal to a falling threshold at the end of another polling interval
- **event:** This setting indicates the event to be triggered when the **rising-threshold** value is reached. This event could be syslog or a SNMP trap
- **falling-threshold:** An alert is generated if the counter was higher than the **rising-threshold** value prior in a last polling interval and lower than or equal to the **falling-threshold** value at this interval
- **portguard:** This option is an advanced feature and can be set to apply error disable or flap the affected port

For example, in the sample command, the poll interval is 1 second, and the rising threshold is set to 10 percent, or 100 ms. So if there are no Tx credits for 100 ms, then the rising-threshold criterion is met. If a credit is indeed returned, then the falling-threshold criterion is met, and the appropriate counter is incremented. In addition, when the rising-threshold criterion is met, the port is put into the error disable state. It will remain in that state until someone manually issues a **shut** or **no shut** command on that port.

For additional information about the slow drain counters, please refer to the [Port-Monitor Configuration Guide](#).

Appendix B: Cisco MDS 9000 Family Counter Names and Descriptions

Table 2 lists the counter names referenced throughout this document and describes their meaning.

Table 2. Counters

Counter Name	Counter Description
FCP_CNTR_RCM_CH0_LACK_OF_CREDIT ² AK_FCP_CNTR_RCM_CH0_LACK_OF_CREDIT ³ THB_RCM_RCP0_RBBZ_CH0 ⁴ F16_RCM_RCP0_RBBZ_CH0 ⁵	Total count of transitions to zero for Rx B2B credits on ch0; these transitions typically indicate that the switch is applying back pressure to the attached device because of perceived congestion, and this perceived congestion can be the result of a lack of Tx B2B credits being returned on an interface over which this device is communicating
FCP_CNTR_LAF_TOTAL_TIMEOUT_FRAMES ² AK_FCP_CNTR_LAF_TOTAL_TIMEOUT_FRAMES ³ THB_TMM_TOLB_TIMEOUT_DROP_CNT ⁴ F16_TMM_TOLB_TIMEOUT_DROP_CNT ⁵	Timeout drops at egress
FCP_CNTR_QMM_CH0_LACK_OF_TRANSMIT_CREDIT ² AK_FCP_CNTR_QMM_CH0_LACK_OF_TRANSMIT_CREDIT ³ THB_TMM_PORT_TBBZ_CH0 ⁴ F16_RCM_RCP0_TBBZ_CH0 ⁵	Total count of transitions to zero for Tx B2B credits on ch0; these transitions are typically the result of the attached device's withholding of R_Rdy primitive from the switch due to congestion in that device
None ² None ³ THB_TMM_PORT_FRM_DROP_CNT ⁴ F16_TMM_PORT_FRM_DROP_CNT ⁵	Number of frames dropped in tolb_path or np_path ; these drops include all types of packet drops: timeout, offline, abort drops at egress, etc.
None ² None ³ THB_TMM_PORT_TWAIT_CNT ⁴ F16_TMM_PORT_TWAIT_CNT ⁵	Packet is available to send, but no credit is available; increments every cycle (cycle = 2.35 nanoseconds)
FCP_CNTR_LAF_C3_TIMEOUT_FRAMES_DISCARD ² AK_FCP_CNTR_LAF_C3_TIMEOUT_FRAMES_DISCARD ³ THB_TMM_TO_CNT_CLASS_3 ⁴ F16_TMM_TO_CNT_CLASS_3 ⁵	Count of class-3 Fibre Channel frames dropped as a result of congestion-drop timeout
FCP_CNTR_RX_WT_AVG_B2B_ZERO ² AK_FCP_CNTR_RX_WT_AVG_B2B_ZERO ³ FCP_SW_CNTR_RX_WT_AVG_B2B_ZERO ⁴ FCP_SW_CNTR_RX_WT_AVG_B2B_ZERO ⁵ (unable to generate)	Count of the number of times an interface was at zero Rx B2B credits for 100 ms; this status typically indicates that the switch is withholding R_Rdy primitive to the device attached on that interface due to congestion in the path to devices with which it is communicating
FCP_CNTR_TX_WT_AVG_B2B_ZERO ² AK_FCP_CNTR_TX_WT_AVG_B2B_ZERO ³ FCP_SW_CNTR_TX_WT_AVG_B2B_ZERO ^{4,5}	Count of the number of times that an interface was at zero Tx B2B credits for 100 ms; this status typically indicates congestion at the device attached on that interface
FCP_CNTR_FORCE_TIMEOUT_ON ² AK_FCP_CNTR_FORCE_TIMEOUT_ON ³ FCP_SW_CNTR_FORCE_TIMEOUT_ON ^{4,5}	Count of the number of times the "system timeout no-credit-drop threshold" has been reached by this port; when a port is at zero Tx B2B credits for the time specified, the port starts to drop packets at line rate
FCP_CNTR_FORCE_TIMEOUT_OFF ² AK_FCP_CNTR_FORCE_TIMEOUT_OFF ³ FCP_SW_CNTR_FORCE_TIMEOUT_OFF ^{4,5}	Count of the number of times that the port has recovered from the system timeout no-credit-drop condition; this status typically means that R_Rdy primitive has been returned or possibly that an LR and LRR even has occurred
FCP_CNTR_LAF_CF_TIMEOUT_FRAMES_DISCARD ² AK_FCP_CNTR_LAF_CF_TIMEOUT_FRAMES_DISCARD ³ THB_TMM_TO_CNT_CLASS_F ⁴ F16_TMM_TO_CNT_CLASS_F ⁵	Count of class-F Fibre Channel frames dropped due to congestion-drop timeout
FCP_CNTR_CREDIT_LOSS ² AK_FCP_CNTR_CREDIT_LOSS ³ FCP_SW_CNTR_CREDIT_LOSS ^{4,5}	Count of the number of times that creditmon credit loss recovery has been invoked on a port

Notes:

- ¹: Generation 1 modules are no longer supported by NX-OS 5.0 (and later releases) and are not covered by this white paper
- ²: Generation 2 DS-X9112, DS-X9124, and DS-X9148 and DS-X9304-18K9 modules
- ³: Generation 3 DS-X9248-48K9 and DS-X92xx-96K9 modules
- ⁴: Generation 4 DS-X92xx-256K9 module
- ⁵: Cisco MDS 9710 DS-X9448-768K9 module
- AK: Aakash (Generation 2 or Generation 3 line card MAC address ASIC)
- THB: Thunderbird (Generation 4 MAC address ASIC)
- F16: F16 (Generation 5 MAC address ASIC)



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)