Switch Congestion Management

Switch Configuration

Version 1.0

Author

Gairy Grannum
Sr. Principal Software Engineer
SANBlaze

Table of Contents

List of Figures

Introduction

Both Broadcom and Cisco implemented a method to manage congestion in an ISL environment. Both use the priority defined in the cs_ctl field of the header to direct traffic within the ISL configuration to Virtual Lanes for Cisco and Virtual Channels for Broadcom. Higher priority traffic is on one set of virtual lanes/channels while low priority traffic goes on another set of virtual lanes/channels.

To allow testing of these solutions, SANBlaze is providing an environment to allow priorities to be set on IO traffic generated to the switch. The SANBlaze solution supports SCSI traffic and will support NVME traffic if the port supports NVMe, and the system is licensed to enable NVMe. The SANBlaze solution only sets the priority for IO traffic; the priority is not set on ELS commands.

When the SANBlaze port logs into the switch, it will determine which implementation shall be supported (Cisco Virtual Lane feature or Broadcom's Virtual Channel feature). Once the switch manufacturer is determined, the appropriate settings needed to test the feature will be made available through the GUI.

Virtual Ports Priority

The virtual initiator or virtual target port can send IO traffic with a high, medium, or low priority. The actual value used for the cs_ctl field will be dependent on the switch implementation and the SANBlaze configuration pertaining to the switch.

The virtual port also allows for the priority to be set to high, medium, or low. It also allows for the priority to be disabled or a default priority to be sent. The behavior of the default priority is dependent on whether the virtual port is an initiator or a target.

In initiator mode, the default priority corresponds to the default priority configured for the port for the current protocol. A default priority can be set for SCSI and a separate priority can be set for NVMe (if NVME is supported for the port) for all the virtual ports on that port. The default port priority can be set to High, Medium, Low or Disabled.

In target mode, the default priority corresponds to the target returning whatever priority the initiator sent to it.

Virtual Ports Priority Override

The SANBlaze implementation also allows FPIN Peer Congestion and Congestion notifications to override the configured priority and instead send the Low priority to one or more nodes the virtual port is zoned with, depending on the notification. This behavior will continue until the notifications indicate the congestion condition has cleared or the timeout for the congestion notification expires.

In initiator mode, when Peer Congestion notifications are received for one or more of its targets it is logged into, the initiator will send a Low priority to the target(s) reported by the Peer Congestion. If the Peer Congestion notifications report the initiator as having congestion, the initiator will communicate with all its targets using a Low priority. This will also be true if the initiator receives a Congestion Notification. The sending of Low priorities will continue if the congestion condition is reported.

In target mode when Peer Congestion notifications are received for one or more of the initiators it is logged into, the target will send a Low priority to the initiator(s) reported by the Peer Congestion. If the

Peer Congestion notifications report the target as having congestion, the target will communicate with its initiators using a Low priority. This will also be true if the target receives a Congestion Notification. This sending of Low priorities will continue as long as the congestion condition is reported.

Configuring Switch Congestion Management Feature
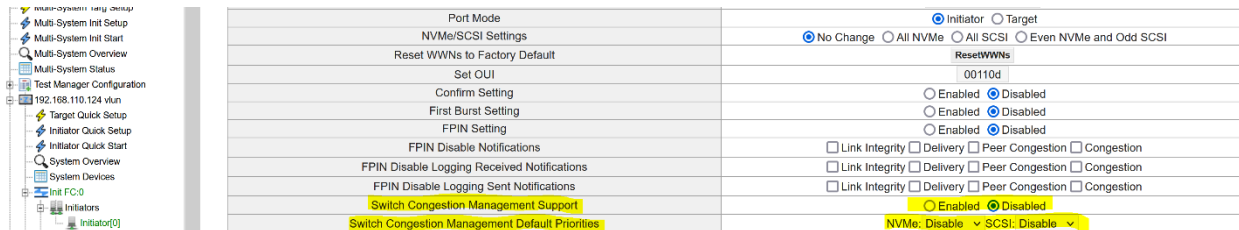
Initiator Port Configuration

Figure 1: Disabled Switch Congestion Management Settings on Configuration Page

On the configuration page of an initiator, the Switch Congestion Management fields are highlighted above. The settings will not be settable until the feature is enabled. This is done by changing the "Switch Congestion Management Support" to enabled and hitting apply. Once that is done, the "Switch Congestion Management Default Priorities" will have default values that can be changed.

Figure 2: Enabled Switch Congestion Management Settings on Configuration Page

Once the feature is enabled, the default priorities can be modified through the drop-down boxes. The NVMe option will only be displayed if the port supports NVMe and the system has an NVMe license.
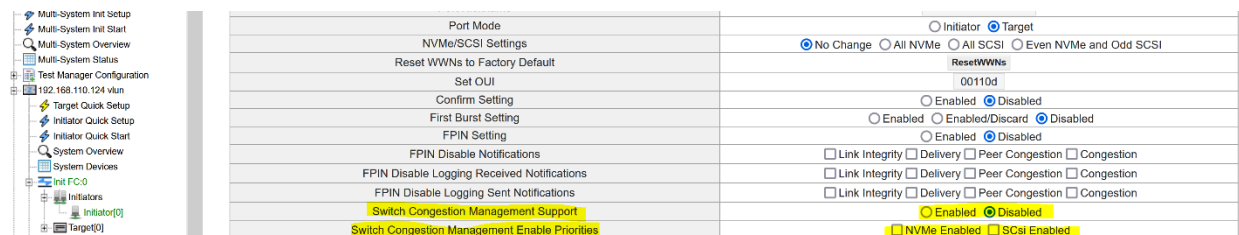
Target Port Configuration

Figure 3: Disabled Switch Congestion Management Settings on Configuration Page

On the configuration page of a target, the Switch Congestion Management fields are highlighted above. The settings will not be settable until the feature is enabled. This is done by changing the "Switch Congestion Management Support" to Enabled and hitting apply. Once that is done, the "Switch

Congestion Management Default Priorities" will be enabled and allow the default priorities to be enabled or disabled.
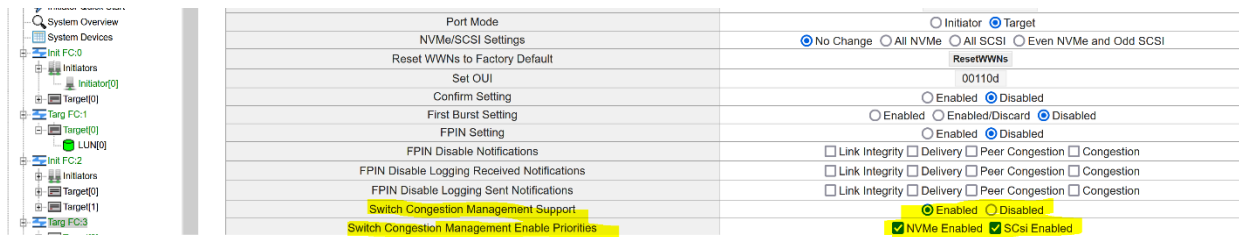

Figure 4: Enabled Switch Congestion Management Settings on Configuration Page

Configuration Support of Broadcom's QoS Class-Specific Control (CS_CTL)-based Frame Prioritization

Broadcom switches allow the prioritization of frames through its QoS Class-Specific Control (CS_CTL)-based frame prioritization. The feature is supported in two forms.

QoS Class-Specific Control (CS_CTL)-based Frame Prioritization Default Mode

The Default Mode of the feature allows a range of priorities to be specified for the high, medium and low priorities. Currently, the range for high priorities are 17 – 24, medium priorities are 9 – 16 and low priorities are 1 – 8.

QoS Class-Specific Control (CS_CTL)-based frame prioritization Auto Mode

The Auto Mode of the feature allows a logical value to be set for the high, medium, and low and the switch internally determines which Virtual Channel to use. Enabling this feature does require a reboot of the switch.

Configuration Setting for QoS Class-Specific Control (CS_CTL)-based Frame Prioritization

The figure below shows Broadcom's QoS Class-Specific Control (CS_CTL)-based frame prioritization settings on the Configuration page.
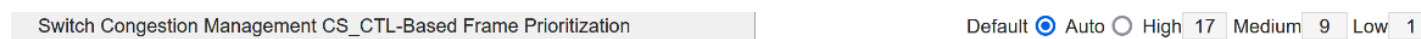

Figure 5: Broadcom's QoS Class-Specific Control (CS_CTL)-based Frame Prioritization Settings

The setting is common for both target and initiator page.

Default – if selected, the port will implement the Default Mode of the feature.

Auto – if selected, the port will implement the Auto Mode of the feature.

High - if default mode is selected, sets the high priority value in the cs_ctl field when the virtual port sends out a high priority (current range is 17 – 24) frame. If auto mode, it shows the value used in the cs_ctl field when the virtual port sends out a high priority (current value is 3) frame.

Medium - if default mode is selected, sets the medium priority value in the cs_ctl field when the virtual port sends out a medium priority (current range is 9 – 16) frame. If in auto mode, it shows the value used in the cs_ctl field when the virtual port sends out a medium priority (current value is 2) frame.

Low - if default mode is selected, sets the low priority value in the cs_ctl field when the virtual port sends out a low priority (current range is 1 – 8) frame. If in auto mode, it shows the value used in the cs_ctl field when the virtual port sends out a low priority (current value is 1) frame.

Enable Port Priority Override

Since the override priority mechanism is FPIN-based, FPIN needs to be enabled for the Port to receive and process the Peer Congestion and Congestion Notifications.
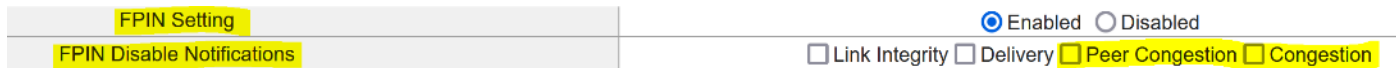

Figure 6: FPIN enable settings on Configuration Page

Ensure the FPIN Setting is enabled and the Congestion Notification you want to receive is unchecked in the "FPIN Disable Notifications" settings.
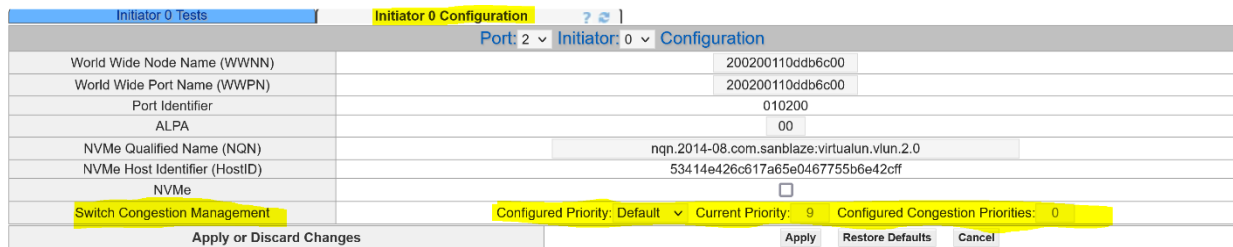
Virtual Initiator Port Configuration

Figure 7: Virtual Initiator Port Configuration Page Switch Congestion Management settings

Configured Priority – drop-down box that allows the priority sent by the virtual initiator to be set. If the Switch Congestion Management feature is not enabled, the drop-down box will still show the available settings but will not be settable.

Current Priority – shows the actual value that will be placed in the cs_ctl field when this initiator sends IO traffic. If the Switch Congestion Management feature is not enabled, the dialog box will show 0.

Configured Congestion Priorities - If FPIN is enabled with Congestion Notifications, this number will represent the number of targets the initiator is logged into that are receiving frames with a low priority value due to congestion being reported for these targets. The number will also include the initiator port if it is reported as having congestion.

If FPIN is enabled, the virtual port will also include the FPIN information showing which nodes (targets and initiators) have been reported as having congestion.

Figure 8: Virtual Initiator Port Configuration Page Switch Congestion Management settings wih FPIN

Port Congested – The dialog box shows if the initiator has been reported to have congestion or not. Yes indicates the initiator has been reported as having congestion.

Congested Targets - The drop-down box lists the targets the initiator is logged into that have been reported as having congestion.

When the congestion condition clears, the configured congestion priorities should go to 0, the Port Congested dialog box should show "No" and the Congested Targets drop-down list should be empty.

The virtual port's configuration page does not dynamically update; you must refresh the page to get the latest information.

Virtual Target Port Configuration

Figure 9: Virtual Target Port Configuration Page Switch Congestion Management Settings

Configured Priority – drop-down box that allows the priority sent by the virtual target to be set. If the Switch Congestion Management feature is not enabled, the drop-down box will still show the available settings but will not be settable.

Configured Congestion Priorities - If FPIN is enabled with Congestion Notifications, this number will represent the number of initiators the target is logged into that are receiving frames with a low priority value due to congestion being reported for these initiators. The number will also include the target port if it is reported as having congestion.

If FPIN is enabled, the virtual port will also include the FPIN information showing which nodes (targets and initiators) have been reported as having congestion.

Figure 10*: Virtual Target Port Configuration Page Switch Congestion Management Settings with FPIN*

Port Congested – The dialog box shows if the target has been reported to have congestion or not. Yes indicates the target has been reported as having congestion.

Congested Targets - The drop-down box lists the initiators the target is logged into that have been reported as having congestion.

When the congestion condition clears, the configured congestion priorities drop to 0, the Port Congested dialog box shows "No" and the Congested Initiators drop-down list shows as empty.

The virtual port's configuration page does not dynamically update; you mustrefresh the page to get the latest information.

Virtual Ports Priority Override Example

In this example, we have an initiator (WWPN 200200110ddb6c00, FCID 010200) who is logged into two targets. One target has a WWPN of 200300110ddb6b00 and an FCID of 010300 and the other target has a WWPN of 200500110de69f00 and an FCID of 010400.

When an FPIN is received with a Peer Congestion notification identifying that target 200300110ddb6b00 has congestion, the virtual port configuration page shows the following:

Figure 11*: Virtual Initiator Port Configuration Page Showing Target Congestion*

Switch Congestion Management *White Paper*

The current priority remains 9 to show the priority that will be used when congestion is not present. IO traffic sent to 200500110de69f00 will continue to use a value of 9. However, communication to target 200300110ddb6b00 will use a priority of 1.

Eventually, the reporting of congestion stops for target 200300110ddb6b00. However, congestion is later reported for both the initiator and target 00500110de69f00. A refresh of the configuration page shows:

Initiator 0 Tests	Initiator 0 Configuration	? ⟳
	Port: 2 ∨ **Initiator:** 0 ∨ **Configuration**	
World Wide Node Name (WWNN)	200200110ddb6c00	
World Wide Port Name (WWPN)	200200110ddb6c00	
Port Identifier	010200	
ALPA	00	
NVMe Qualified Name (NQN)	nqn.2014-08.com.sanblaze:virtualun.vlun.2.0	
NVMe Host Identifier (HostID)	53414e426c617a65e0467755b6e42cff	
NVMe	☐	
Switch Congestion Management	Configured Priority: Default ∨ Current Priority: 9 Configured Congestion Priorities: 2	
FPIN Identified Congested Nodes	Port Congested: Yes Congested Targets: 200500110de69f00 (fcid=010400) ∨	
Register Diagnostic Functions (RDF) registered	Link Integrity, Delivery, Peer Congestion, Congestion	
Apply or Discard Changes	**Apply** **Restore Defaults** **Cancel**	

Figure 12*: Virtual Initiator Port Configuration Page Showing Target and Initiator Congestion*

The current priority remains 9 to show the priority that will be used when congestion is not present. The Configured Congestion Priorities is set to 2 since the initiator port and the target will be configured for Low Priority traffic. Since the initiator is configured for congestion, all targets will receive low priority frames from the initiator.

Once the congestion condition clears, the display returns to the original display.

Initiator 0 Tests	Initiator 0 Configuration	? ⟳
	Port: 2 ∨ **Initiator:** 0 ∨ **Configuration**	
World Wide Node Name (WWNN)	200200110ddb6c00	
World Wide Port Name (WWPN)	200200110ddb6c00	
Port Identifier	010200	
ALPA	00	
NVMe Qualified Name (NQN)	nqn.2014-08.com.sanblaze:virtualun.vlun.2.0	
NVMe Host Identifier (HostID)	53414e426c617a65e0467755b6e42cff	
NVMe	☐	
Switch Congestion Management	Configured Priority: Default ∨ Current Priority: 9 Configured Congestion Priorities: 0	
FPIN Identified Congested Nodes	Port Congested: No Congested Targets: ∨	
Register Diagnostic Functions (RDF) registered	Link Integrity, Delivery, Peer Congestion, Congestion	
Apply or Discard Changes	**Apply** **Restore Defaults** **Cancel**	

Figure 13*: Virtual Initiator Port Configuration Page Once Target Congestion Clears*

Questions?

If you have questions, please contact us – we would be happy to hear from you.

support@sanblaze.com | sales@sanblaze.com | info@sanblaze.com

https://www.sanblaze.com/