Nexus 9500-R, Nexus 3000-R: Troubleshoot Input Discards

Contents

Introduction

Background Information

Ingress Traffic Manager (ITM)

Ingress VOQ Buffer

Schedule and Flow Control

Common Causes

Applicable Hardware

Troubleshoot Input Discards

Common Scenario - 10G to 1G Traffic Flow - Constant Drops:

Step 1. Check which Queue is Impacted on your Input Discards Interface.

Step 2. Check Broadcom's Graphical Representation of Counters used for Diagnostics:

Step 3. Find what ASIC and what Jericho Port your Front Panel Port experiencing Input Discards belongs to:

Step 4. Understand what VOQ and VOQ Connector your Ingress Port has.

Step 5. Check from BCM's perspective, which Queue specifically is Non-Empty; i.e Congested.

Step 6. Find your Egress Congested Port from the non-empty Queue value:

Step 7. Check which Front Panel Port is in ASIC 1 and Maps to Jericho Port 9 based on your Previous Finding.

Additional Commands

Additional Lab Tests:

Step 1. Input Discards with Multiple Egress Congested Interfaces.

Step 2. Input Discards due to SPAN.

Step 3. Input Discards due to Traffic Hair Pin.

Step 4. Send Packets with a Destination IP that is Unknown.

Step 5. Input Discards while an Access/Trunk port transitions into STP Forwarding State

Step 6. Input Discards due to Eth1/9 Exceeding Line Rate.

Introduction

This document describes the causes of and solutions for input discards for the Cisco Nexus 9500-R EoR and Nexus 3000-R ToR. An input discard indicates the number of packets dropped in the input queue because of congestion. This number includes drops that are caused by tail drop and Weighted Random Early Detection (WRED).

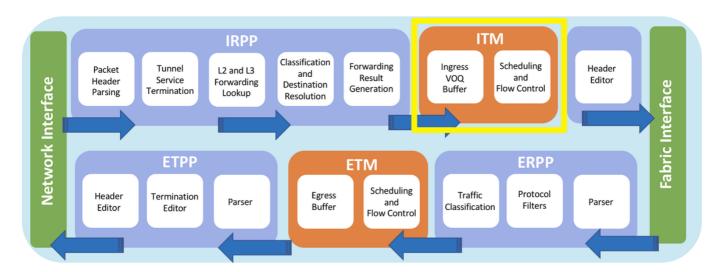
If you experience random/sporadic/historical (i.e no longer occurring) drops, please contact Cisco TAC for further investigation. This walk-through is useful when Input Discards are incremented frequently.

Background Information

The R-Series uses ingress VOQ architecture. VOQ architecture emulates egress queues in the ingress buffer with virtual queues. Each egress port has eight queues for unicast traffic and eight queues for multicast traffic. Traffic can be classified into traffic classes based on the Class-of-Service (CoS) or Differentiated Services Code Point (DSCP) value in the packets and then queued in the corresponding virtual queue for that traffic class.

The R-Series uses a distributed credit mechanism to transfer traffic over the fabric. Before a packet is scheduled to leave the VOQ, the ingress buffer scheduler requests a credit for the specific port and priority in the egress buffer. Credit is requested from an ingress credit scheduler for the destination port and priority. If buffer space is available, the egress scheduler grants access and sends the credit grant to the ingress buffer scheduler. If no buffer space is available in the egress buffer, the egress schedule does not grant a credit, and traffic is buffered in the VOQ until the next credit is available.

Below is the Packet Forwarding Pipeline for the -R platform. On this article, you focus on the **Ingress Traffic Manager** component. More details on the architecture at this link



Ingress Traffic Manager (ITM)

The ingress traffic manager (ITM) is a block in the ingress pipeline. It performs steps related to queue traffic into VOQ, schedule traffic for transmission over the fabric, and manage credits.

Ingress VOQ Buffer

The ingress VOQ buffer block manages both the on-chip buffer and the off-chip packet buffer. Both buffers use VOQ architecture, and traffic is queued based on the information from the IRPP (Ingress Receiver Packet Processor). A total of 96,000 VOQs are available for unicast and multicast traffic.

Schedule and Flow Control

Before a packet is transmitted from the ingress pipeline, the packet needs to be scheduled for transfer over the fabric. The ingress scheduler sends a credit request to the egress scheduler located in the egress traffic manager block. When the ingress traffic manager receives the credit, it starts sending traffic to the ingress transmit packet processor. If the egress buffer is full, traffic will be buffered in the dedicated queue represented by the egress port and traffic class.

Common Causes

Generally, input discards could be seen for below reasons across various Nexus hardware

- Traffic flows congesting egress interfaces (such as 10G ingress and 1G egress)
- Oversubscribed SPAN destination port- Applies to specific hardware types.

Applicable Hardware

PID N9K-X9636C-R N9K-X9636Q-R N9K-X9636C-RX

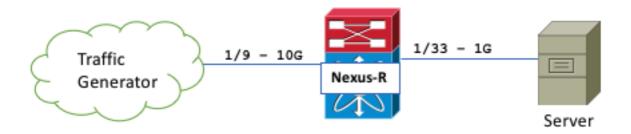
N9K-X96136YC-R

N3K-C36180YC-R

N3K-C3636C-R

Troubleshoot Input Discards

Common Scenario - 10G to 1G Traffic Flow - Constant Drops:



Throughout this article, the value for the counter of "input discards" and any HW internal counter that references the same will change as the errors were incrementing while testing and relevant commands must be grabbed live.

Step 1. Check which Queue is Impacted on your Input Discards Interface.

This step comes in handy later.

In our case, it is Queue 7, the default queue - There are 8 queues total on ingress:

Step 2. Check Broadcom's Graphical Representation of Counters used for Diagnostics:

```
Nexus-R# bcm-shell mod 1 "diag counters g"
/|\
                                               JERICHO NETWORK INTE
R F A C E
                                     \ | /
+-----
                                                                           NBI
RX_TOTAL_BYTE_COUNTER
                                          = 10,616,663,796
TX_TOTAL_BYTE_COUNTER
                                       = 41,136
RX_TOTAL_PKT_COUNTER
                                          = 10,659,301
TX_TOTAL_PKT_COUNTER
                                       = 606
 RX_TOTAL_DROPPED_EOPS
                                     IRE
EPNI
 CPU_PACKET_COUNTER
                                          = 606
NIF_PACKET_COUNTER
                                         = 10,659,302
EPE_BYTES_COUNTER
                                       = 41,136
OAMP_PACKET_COUNTER
                                          = 0
EPE_PKT_COUNTER
                                       = 606
OLP_PACKET_COUNTER
EPE_DSCRD_PKT_CNT
                                       = 0
RCY_PACKET_COUNTER
                                          = 0
 IRE_FDT_INTRFACE_CNT
                                     IDR
EGQ
                                      MMU_IDR_PACKET_COUNTER
                                         = 10,659,302
FQP_PACKET_COUNTER
                                       = 606
| IDR_OCB_INTERFACE_COUNTER
PQP_UNICAST_PKT_CNT
                                       = 606
PQP_DSCRD_UC_PKT_CNT
                                       = 48,408
PQP_UC_BYTES_CNT
PQP_MC_PKT_CNT
                                    IOM
PQP_DSCRD_MC_PKT_CNT
                                       = 0
PQP_MC_BYTES_CNT
ENQUEUE_PKT_CNT
                                          = 1,403,078
                                       = 606
EHP_UNICAST_PKT_CNT
DEQUEUE_PKT_CNT
                                          = 1,403,078
EHP_MC_HIGH_PKT_CNT
```

= 0

DELETED_PKT_CNT

```
EHP_MC_LOW_PKT_CNT
ENQ_DISCARDED_PACKET_COUNTER
                                         = 9,256,829
DELETED_PKT_CNT
                                       = 0
  Rejects: PORT_AND_PG_STATUS
RQP_PKT_CNT
                                       = 606
RQP_DSCRD_PKT_CNT
                                      = 0
PRP_PKT_DSCRD_TDM_CNT
                                       = 0
PRP_SOP_DSCRD_UC_CNT
                                       = 0
PRP_SOP_DSCRD_MC_CNT
                                       = 0
PRP_SOP_DSCRD_TDM_CNT
                                       = 0
EHP_MC_HIGH_DSCRD_CNT
EHP_MC_LOW_DSCRD_CNT
                                       = 0
ERPP_LAG_PRUNING_DSCRD_CNT
ERPP_PMF_DISCARDS_CNT
                                      = 0
ERPP_VLAN_MBR_DSCRD_CNT
FDA
CELLS_IN_CNT_P1 = 0
                                 CELLS_OUT_CNT_P1 = 0
CELLS_IN_CNT_P2 = 0
                            CELLS_OUT_CNT_P2 = 0
CELLS_OUT_CNT_P3 = 0
CELLS_IN_CNT_P3 = 0
                                   IPT
CELLS_IN_TDM_CNT = 0
                                  CELLS_OUT_TDM_CNT = 0
CELLS_IN_MESHMC_CNT
                                 CELLS_OUT_MESHMC_CNT = 0
                   = 0
EGQ_PKT_CNT
                                         = 606
                                  CELLS_OUT_IPT_CNT = 606
                   = 606
CELLS_IN_IPT_CNT
ENQ_PKT_CNT
                                         = 1,403,084
EGQ_DROP_CNT
                                      = 0
                                         = 1,402,472
FDT_PKT_CNT
EGQ_MESHMC_DROP_CNT
                                      = 0
CRC_ERROR_CNT
EGQ_TDM_OVF_DROP_CNT
                                      = 0
                                         = 606 *
CFG_EVENT_CNT
 CFG_BYTE_CNT
                                         = 48,408
                                    FDT
FDR
| IPT_DESC_CELL_COUNTER
                                         = 5,609,892
P1_CELL_IN_CNT
                                      = 0
| IRE_DESC_CELL_COUNTER
                                         = 0
P2_CELL_IN_CNT
                                      = 0
P3_CELL_IN_CNT
TRANSMITTED_DATA_CELLS_COUNTER
                                         = 5,609,892
```

A QUEUE_DELETED_PACKET_COUNTER being greater than zero would indicate that packets were DELETED by the IQM (Ingress Queueing Manager) afterenqueue. This would be due to an active queue not receiving any credits which would suggest a misconfiguration of the scheduling scheme. You would check this via bcm-shell mod X "getReg IQM QUEUE DELETED PACKET COUNTER"

ENQ_DISCARDED_PACKET_COUNTER means packets were discarded BEFORE enqueue. You can see this counter set in BCM as well (command is cleared on read):

```
Nexus-R# bcm-shell mod 1 "g iqm_reject_status_bmp" | i i PG|IQM0|IQM1
IQM_REJECT_STATUS_BMP.IQM0[0x1a7]=0x20000000: <VSQF_WRED_STATUS=0,
QNUM_OVF_STATUS=0,PORT_AND_PG_STATUS=1,OCCUPIED_BD_STATUS=0,
IQM_REJECT_STATUS_BMP.IQM1[0x1a7]=0: <VSQF_WRED_STATUS=0,VSQF_MX_SZ_STATUS=0,
PORT_AND_PG_STATUS=0,OCCUPIED_BD_STATUS=0,MULTICAST_ERROR_STATUS=0,
```

You can always notice these quickly with **show hardware internal errors module X** (command clears on read):

Step 3. Find what ASIC and what Jericho Port your Front Panel Port experiencing Input Discards belongs to:

```
Nexus-R# show interface hardware-mappings | i i Eth1/9|--|Name|Eth1/33
HName - Hardware port name. None means N/A

Name Ifindex Smod Unit HPort HName FPort NPort VPort SrcId

Eth1/9 1a001000 0 0 9 xe9 255 8 -1 0 << ASIC 0, Jericho Port 9
Eth1/33 1a004000 2 1 9 xe9 32 -1 0 << ASIC 1, Jericho Port 9
```

Displaying Eth1/33 for this example. In an actual network, you won't know the congested egress port yet.

Step 4. Understand what VOQ and VOQ Connector your Ingress Port has.

This command shows us details for the flow for ingress VoQ for a specific port. Additionally, it shows us the current credit balance of the VoQ.

The port's VOQ is derived in this way:

LCs are 0 based - Module 1 is 0, Module 2 is 1, etc There are 256 System Port IDs per LC

```
ID = (LC * System port ID) + FP number
Eth1/9 = (0 * 256) + 9 = 9
VOQ ID = 32 + (System Port ID * 8)
Eth1/9 = 32 + (9 * 8) = 104
```

Our VOQ for Eth1/9 will therefore be 104 which matches the output previously gathered

```
module-1# show hardware internal jer-usd ingress-vsq buffer-occupancy front-port 9
                   VSOF BUFFER OCCUPANCY
                    Front port 9
|max global shared
                                              157286
                                              0
max ocb buffer occupancy
                COSQ 0
rate class
granted buffers per port
                                                3280
|shared buffers occupied
                                               127792 | <<<<
granted buffers occupied
                                                3280
|shared buffer max occupancy |
                                              127792 <<<<
```

Step 5. Check from BCM's perspective, which Queue specifically is Non-Empty; i.e Congested.

Step 6. Find your Egress Congested Port from the non-empty Queue value:

If the Queue is 303, recall that these queues are actually a range so it can be 303 + 7 or 303-7 - The question is, which port has a VOQ that matches on a range of 296-303 or alternatively, 303-310?

It is known that Queue 7 on Eth1/9 is congested, so 303 actually is the highest in its range so the range of 296-303 is a well-educated guess.

Display the same for asic 0 - Not shown here for brevity; you would notice under the Voq column that your range of interest is not in that ASIC

Notice a few things on the above output:

- Our egress congested port is on ASIC 1.
- Our egress congested port has a VOQ of 296 and 303 would equate to Queue 7 on that port.
- Notice the Credit Balance column There are very few credits left on this interface to grant which is why our ingress Eth1/9 starts buffering.

Step 7. Check which Front Panel Port is in ASIC 1 and Maps to Jericho Port 9 based on your Previous Finding.

```
Nexus-R# show interface hardware-mappings | i i Eth1/9|--|Name|Eth1/33
HName - Hardware port name. None means N/A

Name Ifindex Smod Unit HPort HName FPort NPort VPort SrcId

Eth1/9 1a001000 0 0 9 xe9 255 8 -1 0 << ASIC 0, Jericho Port 9
Eth1/33 1a004000 2 1 9 xe9 32 -1 0 << ASIC 1, Jericho Port 9
```

At this point, you have found the egress congested port - Determine whether there's something wrongfully bursting into the network, you have configured SPAN and your destination port is 1G while sourcing one or more 10G interface or if this is a bottleneck/design issue.

Additional Commands

These are more advanced - Not needed to find Egress Congested port under normal scenarios.

```
Nexus-R# bcm-shell mod 1 "diag counters g"
                                       /|\
                                                 JERICHO NETWORK INTE
RFACE
                                       NBI
                                            = 10,616,663,796
RX_TOTAL_BYTE_COUNTER
TX_TOTAL_BYTE_COUNTER
                                         = 41.136
| RX_TOTAL_PKT_COUNTER
                                            = 10,659,301
TX_TOTAL_PKT_COUNTER
                                         = 606
  RX_TOTAL_DROPPED_EOPS
                                      IRE
EPNI
  CPU_PACKET_COUNTER
                                            = 606
 NIF PACKET COUNTER
                                            = 10,659,302
EPE_BYTES_COUNTER
                                         = 41,136
```

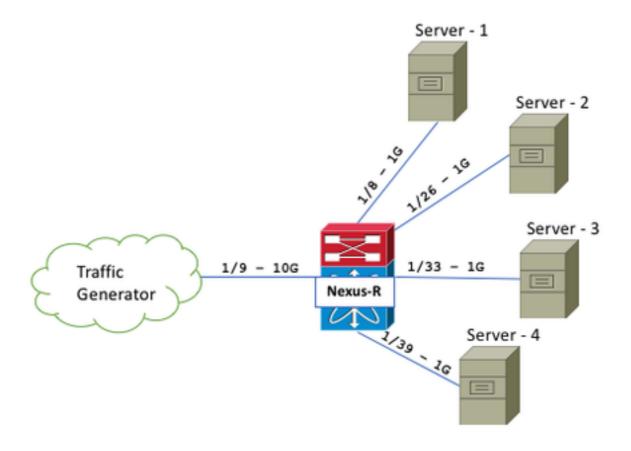
OAMP_PACKET_COUNTER EPE_PKT_COUNTER	= 0 = 606	1
OLP_PACKET_COUNTER	= 0	, l
EPE_DSCRD_PKT_CNT	= 0	
RCY_PACKET_COUNTER	= 0	
 IRE_FDT_INTRFACE_CNT	= 0	I
 	+	
+		+
	IDR	
EGQ	I	1
		I
 MMU_IDR_PACKET_COUNTER	= 10,659,302	
FQP_PACKET_COUNTER	= 606	
IDR_OCB_INTERFACE_COUNTER	= 0	
PQP_UNICAST_PKT_CNT	= 606	
DOD DOODD ITO DIVE ONE	- 0	
PQP_DSCRD_UC_PKT_CNT	= 0	l I
PQP_UC_BYTES_CNT	= 48,408	
+	·+	
PQP_MC_PKT_CNT	= 0	
	IQM	
PQP_DSCRD_MC_PKT_CNT	= 0	
 PQP_MC_BYTES_CNT	= 0	
ENQUEUE_PKT_CNT	= 1,403,078	
EHP_UNICAST_PKT_CNT	= 606	1
DEQUEUE_PKT_CNT	= 1,403,078	
EHP_MC_HIGH_PKT_CNT	= 0	
DELETED_PKT_CNT	= 0 = 0	
EHP_MC_LOW_PKT_CNT ENQ_DISCARDED_PACKET_COUNTER	= 9,256,829	
DELETED_PKT_CNT	= 0	
Rejects: PORT_AND_PG_STATUS		
		1
 ROP_PKT_CNT	= 606	
RQP_PRT_CNT	= 606	l I
RQP_DSCRD_PKT_CNT	= 0	
		· 1
PRP_PKT_DSCRD_TDM_CNT	= 0	
PRP_SOP_DSCRD_UC_CNT	= 0	l I
PRP_SOP_DSCRD_MC_CNT	= 0	
		, l
PRP_SOP_DSCRD_TDM_CNT	= 0	
EHP_MC_HIGH_DSCRD_CNT	= 0	
 EHP_MC_LOW_DSCRD_CNT	= 0	l I
	_ 0	
ERPP_LAG_PRUNING_DSCRD_CNT	= 0	
I		
ERPP_PMF_DISCARDS_CNT	= 0	
 ERPP_VLAN_MBR_DSCRD_CNT	= 0	l I
+	- ∨ ·+	 +
+	· 	+
		[
FDA	I	

```
CELLS_OUT_CNT_P1 = 0
CELLS_IN_CNT_P1 = 0
CELLS_IN_CNT_P2 = 0
                            CELLS_OUT_CNT_P2 = 0
CELLS_IN_CNT_P3 = 0
                            CELLS_OUT_CNT_P3 = 0
                              IPT
CELLS_IN_TDM_CNT = 0
                             CELLS_OUT_TDM_CNT = 0
EGQ_PKT_CNT
                                = 606
                            CELLS_OUT_IPT_CNT = 606
              = 606
CELLS_IN_IPT_CNT
ENQ_PKT_CNT
                                  = 1,403,084
EGQ_DROP_CNT
                                 = 0
FDT_PKT_CNT
                                   = 1,402,472
EGQ_MESHMC_DROP_CNT
                                 = 0
CRC_ERROR_CNT
                                   = 0
EGQ_TDM_OVF_DROP_CNT
                                 = 0
                                   = 606 *
CFG_EVENT_CNT
CFG_BYTE_CNT
                                   = 48,408
                               FDT
FDR
                               IPT_DESC_CELL_COUNTER
                                   = 5,609,892
                                 = 0
P1_CELL_IN_CNT
| IRE_DESC_CELL_COUNTER
P2_CELL_IN_CNT
P3_CELL_IN_CNT
TRANSMITTED_DATA_CELLS_COUNTER
                                 = 5,609,892
CELL_IN_CNT_TOTAL
/ | \
                                   JERICHO FABRIC INTER
F A C E
                               \ | /
```

Additional Lab Tests:

Step 1. Input Discards with Multiple Egress Congested Interfaces.

Consider this topology wherein the Traffic Generator is sending 2G of traffic towards each Server:



Quickly check which Queues are not empty - Notice there are 4:

```
Nexus-R# bcm-shell mod 1 "diag cosq non_empty_queue"

Core 0:
Ingress VOQs Sizes (format: [queue_id(queue_size)]):
[103(29475840B)] [247(29379584B)] [303(56452096B)] [351(76020736B)]
```

Determine what interfaces these Queues belong to - Check ASIC 0 first (it only demonstrates with one interface):

Repeat the same process for the other three Queue values: 247, 303 and 351.

Step 2. Input Discards due to SPAN.

Setting Eth1/33 as a SPAN destination port while setting Eth1/9 as a SPAN source port in the RX direction

```
Nexus-R# show run mon

monitor session 1
description SPAN TEST INPUT DISCARDS
source interface Ethernet1/9 rx
destination interface Ethernet1/33
no shut

Nexus-R# show int e1/9 | i i input.disc
0 input with dribble 9314306 input discard
```

Step 3. Input Discards due to Traffic Hair Pin.

Sending packets with SRC 10.10.10.10 and DEST 192.168.10.10 where Eth1/9 is in 10.10.10.1/24 - This does not result in an Input Discard; however, you do see this counter:

Step 4. Send Packets with a Destination IP that is Unknown.

Send packets with SRC 10.10.10.10 and DEST 192.168.10.10 where Eth1/9 is in 10.10.10.1/24 and Eth1/33 is an L3 port in the 172.16.0.1/30 subnet - No drop counter, no input discards even when the destination is **unknown**.

Step 5. Input Discards while an Access/Trunk port transitions into STP Forwarding State

Send packets where Eth1/9 is just a wide trunk (or access port) - This is registered as an Input Discard while the port transitions into an STP forwarding state.

QUEUE_NOT_VALID_STATUS is a drop due to the Packet Processor's (PP) decision to drop or an invalid destination received from the Packet Processor (PP) Blocks.

Step 6. Input Discards due to Eth1/9 Exceeding Line Rate.

Sends 10G+ into Eth1/9 would result in a different type of drop as you are maxing out Eth1/9 in the fist place - Does still count as an Input Discard:

```
bcm-shell.0> diag counters g
/ | \
                                                  JERICHO NETWORK INTE
RFACE
                                       \ | /
                                                                                NBI
RX_TOTAL_BYTE_COUNTER
                                            = 53,913,106,009
TX_TOTAL_BYTE_COUNTER
                                          = 1,164,231
RX_TOTAL_PKT_COUNTER
                                            = 54,145,395
TX_TOTAL_PKT_COUNTER
                                          = 17,029
 RX_TOTAL_DROPPED_EOPS
                                       TRE
EPNI
CPU_PACKET_COUNTER
                                             = 17,010
NIF_PACKET_COUNTER
                                            = 54,145,476
EPE_BYTES_COUNTER
                                          = 5,721,307
OAMP_PACKET_COUNTER
                                            = 0
EPE_PKT_COUNTER
                                          = 50,703
OLP_PACKET_COUNTER
                                            = 0
EPE_DSCRD_PKT_CNT
 RCY_PACKET_COUNTER
                                            = 16,837
 IRE_FDT_INTRFACE_CNT
                                            = 0
```

EGQ		
MMU_IDR_PACKET_COUNTER	= 54,128,577	
FQP_PACKET_COUNTER	= 50,703	
IDR_OCB_INTERFACE_COUNTER	= 0	
PQP_UNICAST_PKT_CNT	= 50,683	١.
		. 1
PQP_DSCRD_UC_PKT_CNT	= 0	١ .
POD WG DWEEG GWE	F 016 F16	
PQP_UC_BYTES_CNT	= 5,216,716	1
POP MC PKT CNT	= 20	
	IOM	' I
POP_DSCRD_MC_PKT_CNT	= 20	'
		' I
PQP_MC_BYTES_CNT	= 2,079	1
ENQUEUE_PKT_CNT	= 5,463,323	·
EHP_UNICAST_PKT_CNT	= 50,683	1
DEQUEUE_PKT_CNT	= 5,594,400	İ
EHP_MC_HIGH_PKT_CNT	= 20	
DELETED_PKT_CNT	= 0	
EHP_MC_LOW_PKT_CNT	= 0	
ENQ_DISCARDED_PACKET_COUNTER	= 48,716,055	
DELETED_PKT_CNT	= 40	
Rejects: VOQ_MX_QSZ_STATUS		-
<snip></snip>		