Nexus 3500 Output Drops and Buffer QoS

Contents

Introduction

This document describes commands used in order to troubleshoot the type of traffic dropped on the Nexus 3500 platform and the output buffer (OB) in which this traffic is dropped.

Methodology

1. [Check for Output Drops](#)
2. [Determine if the Drops are Unicast or Multicast](#)
3. [Determine Which Output Buffer is Used](#)
4. [Check the Active Buffer Monitoring](#)

Check for Output Drops

Check the physical interface statistics in order to determine if the traffic is dropped in the egress direction. Determine if the "output discard" counter in the TX direction increments and/or is non-zero.

```
Nexus3548# show interfce Eth1/7
Ethernet1/7 is up
```

```
Dedicated Interface
Hardware: 100/1000/10000 Ethernet, address: a44c.116a.913c (bia a44c.116a.91ee)
Description: Unicast Only
Internet Address is 1.2.1.13/30
MTU 1500 bytes, BW 1000000 Kbit, DLY 10 usec
reliability 255/255, txload 35/255, rxload 1/255
Encapsulation ARPA
full-duplex, 1000 Mb/s, media type is 1G
Beacon is turned off
Input flow-control is off, output flow-control is off
Rate mode is dedicated
Switchport monitor is off
EtherType is 0x8100
Last link flapped 00:03:48
Last clearing of "show interface" counters 00:03:55
1 interface resets
30 seconds input rate 200 bits/sec, 0 packets/sec
30 seconds output rate 0 bits/sec, 0 packets/sec
Load-Interval #2: 5 minute (300 seconds)
  input rate 40 bps, 0 pps; output rate 139.46 Mbps, 136.16 Kpps
RX
  1 unicast packets  118 multicast packets  0 broadcast packets
  119 input packets  9830 bytes
  0 jumbo packets  0 storm suppression bytes
  0 runts  0 giants  0 CRC  0 no buffer
  0 input error  0 short frame  0 overrun  0 underrun  0 ignored
  0 watchdog  0 bad etype drop  0 bad proto drop  0 if down drop
  0 input with dribble  0 input discard
  0 Rx pause
TX
  23605277 unicast packets  0 multicast packets  0 broadcast packets
  23605277 output packets  3038908385 bytes
  0 jumbo packets
  0 output errors  0 collision  0 deferred  0 late collision
  0 lost carrier  0 no carrier  0 babble 11712542 output discard
  0 Tx pause
```

Determine if the Drops are Unicast or Multicast

Once it is determined that the interface drops the traffic, enter the **show queuing interface <x/y>** command in order to find out if the traffic dropped is multicast or unicast. In releases earlier than 6.0(2)A3(1), the output looks like:

```
Nexus3548# show queuing interface Eth1/7
Ethernet1/7 queuing information:
 TX Queuing
   qos-group  sched-type  oper-bandwidth
       0         WRR            100

 RX Queuing
   Multicast statistics:
       Mcast pkts dropped                    : 0
   Unicast statistics:
   qos-group 0
   HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
       Ucast pkts dropped                    : 11712542
```

In Release 6.0(2)A3(1) and later, the output looks like:

```
Nexus3548# show queuing interface Eth1/7
Ethernet1/7 queuing information:
   qos-group  sched-type  oper-bandwidth
       0        WRR            100
   Multicast statistics:
       Mcast pkts dropped                       : 0
   Unicast statistics:
   qos-group 0
   HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
Ucast pkts dropped                          : 11712542
```

> **Note:** If the multicast slow receiver is configured for the port, see for feature information, drops are not tracked with the **show queuing interface Eth<x/y>** command due to a hardware limitation. See Cisco bug ID CSCuj21006.

Determine Which Output Buffer is Used

In Nexus 3500, there are three buffer pools used in the egress direction. The output of the **show hardware internal mtc-usd info port-mapping** command provides the mapping information.

```
Nexus3548# show hardware internal mtc-usd info port-mapping
OB Ports to Front Ports:
========= OB0 =========    ========= OB1 =========    ========= OB2 =========
45 47 21 23 09 11 33 35    17 19 05 07 41 43 29 31    13 15 37 39 25 27 01 03
46 48 22 24 10 12 34 36    18 20 06 08 42 44 30 32    14 16 38 40 26 28 02 04


Front Ports to OB Ports:
=OB2= =OB1= =OB0= =OB2=    =OB1= =OB0= =OB2= =OB1=    =OB0= =OB2= =OB1= =OB0=
12 14 04 06 08 10 00 02    00 02 04 06 08 10 12 14    12 14 04 06 08 10 00 02
13 15 05 07 09 11 01 03    01 03 05 07 09 11 13 15    13 15 05 07 09 11 01 03

Front port numbering (i.e. "01" here is e1/1):
=OB2= =OB1= =OB0= =OB2=    =OB1= =OB0= =OB2= =OB1=    =OB0= =OB2= =OB1= =OB0= 01 03 05 07 09 11
13 15    17 19 21 23 25 27 29 31    33 35 37 39 41 43 45 47 02 04 06 08 10 12 14 16    18 20 22
24 26 28 30 32    34 36 38 40 42 44 46 48
```

Note: Text in Red font is _not_ CLI output, it's purely to help those reading
the document faster match the actual front port instead of having to manually
count up.

The first part of the results indicate that the OB pool 0 is used by front-ports such as 45, 46, 47, 48, and so on and OB1 is used by front-ports 17, 18, and so on.

The second part of the results indicate that Eth1/1 is mapped to OB2 port 12, Eth1/2 is mapped to OB2 port 13, and so on.

The port in discussion, Eth1/7, is mapped to OB1.

See the Buffer Management section in this document for further information.

Check the Active Buffer Monitoring

See the Cisco Nexus 3548 Active Buffer Monitoring whitepaper and the section in this document for further information about this feature.

Counters Actively Increment

If the output discards actively increment, enable Active Buffer Monitoring (ABM) with this command. Note that the command allows you to monitor unicast or multicast, but not both. Also, it lets you configure the sampling interval and threshold values.

```
hardware profile buffer monitor [unicast|multicast] {[sampling <interval>] |
[threshold <Kbytes>]}
```

Brief Output

Once the ABM is enabled, you can view the results with this command.

```
Nexus3500# show hardware profile buffer monitor interface e1/7 brief
Brief CLI issued at: 09/30/2013 19:43:50

                 Maximum buffer utilization detected
               1sec    5sec    60sec    5min    1hr
               ------  ------  ------   ------  ------
Ethernet1/7    5376KB  5376KB  5376KB    N/A     N/A
```

These results indicate that 5.376 MB out of 6 MB of the OB1 buffer has been used by unicast traffic that left Eth1/7 for the past 60 seconds.

Detailed Output

```
Nexus3500# show hardware profile buffer monitor interface Eth1/7 detail
Detail CLI issued at: 09/30/2013 19:47:01

Legend -
384KB  - between   1 and 384KB of shared buffer consumed by port
768KB  - between 385 and 768KB of shared buffer consumed by port
307us  - estimated max time to drain the buffer at 10Gbps

Active Buffer Monitoring for Ethernet1/7 is: Active
KBytes                384  768 1152 1536 1920 2304 2688 3072 3456 3840 4224 4608 4992 5376 5760
6144
us @ 10Gbps           307  614  921 1228 1535 1842 2149 2456 2763 3070 3377 3684 3991 4298 4605
4912

                      ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
----
09/30/2013 19:47:01     0    0    0    0    0    0    0    0    0    0    0    0    0  250
0    0
09/30/2013 19:47:00     0    0    0    0    0    0    0    0    0    0    0    0    0  252
0    0
09/30/2013 19:46:59     0    0    0    0    0    0    0    0    0    0    0    0    0  253
0    0
09/30/2013 19:46:58     0    0    0    0    0    0    0    0    0    0    0    0    0  250
0    0
09/30/2013 19:46:57     0    0    0    0    0    0    0    0    0    0    0    0    0  250
0    0
09/30/2013 19:46:56     0    0    0    0    0    0    0    0    0    0    0    0    0  250
0    0
09/30/2013 19:46:55     0    0    0    0    0    0    0    0    0    0    0    0    0  251
0    0
09/30/2013 19:46:54     0    0    0    0    0    0    0    0    0    0    0    0    0  251
0    0
```

```
09/30/2013 19:46:53      0    0    0    0    0    0    0    0    0    0    0    0    0 250
0    0
09/30/2013 19:46:52      0    0    0    0    0    0    0    0    0    0    0    0    0 253
0    0
09/30/2013 19:46:51      0    0    0    0    0    0    0    0    0    0    0    0    0 249
0    0
...
```

Information in each row is logged at a second interval. Each column represents the buffer usage. As mentioned in the command results, if there is a non-zero value reported for column "384", it means that buffer usage was between 0-384 KBytes when the ABM polled the OB usage. The non-zero number is the number of times the usage was reported.

These results indicate that OB1 averaged 5.376 MB of usage between 249 - 253 times for each second in the last 10 seconds for Eth1/7. It takes 4298 microseconds (us) in order to clear the buffer of this traffic.

Generate a Log When a Threshold Is Crossed

If the drop counter and buffer usage increments periodically, then it is possible to set a threshold and generate a log message when the threshold is crossed.

```
logging level mtc-usd 5
hardware profile buffer monitor unicast sampling 10 threshold 4608
```

The command is set to monitor unicast traffic at a 10 nanoseconds interval and when it goes above 75% of the buffer it generates a log.

You can also create a scheduler in order to collect ABM statistics and interface counter output every hour and append it to bootflash files. This example is for multicast traffic:

```
hardware profile buffer monitor multicast

feature scheduler
scheduler job name ABM
show hardware profile buffer monitor detail >> ABMDetail.txt
show clock >> ABMBrief.txt
show hardware profile buffer monitor brief >> ABMBrief.txt
show clock >> InterfaceCounters.txt
show interface counters errors >> InterfaceCounters.txt
scheduler schedule name ABM
time start now repeat 1:0
job name ABM
```

Notable Cisco Bug IDs

- Cisco bug ID CSCum21350: Rapid port flaps cause all ports in same QoS buffer to drop all TX Multicast/Broadcast traffic. This is fixed in Release 6.0(2)A1(1d) and later.
- Cisco bug ID CSCuq96923: The multicast buffer block is stuck, which results in egress multicast/broadcast drops. This issue is still under investigation.
- Cisco bug ID CSCva20344: Nexus 3500 Buffer block / lockup - no TX multicast or broadcast. Unreproducible issue, potentially got fixed in Releases 6.0(2)U6(7), 6.0(2)A6(8), and 6.0(2)A8(3).
- Cisco bug ID CSCvi93997: Cisco Nexus 3500 switches output buffer block stuck. This is fixed in Releases 7.0(3)I7(8) and 9.3(3).

Frequently Asked Questions

**Does ABM impact performance or latency?**

No, this feature does not impact latency or performance of the device.

What is the impact of the lower ABM hardware polling interval?

By default, the hardware polling interval is 4 milliseconds. You can configure this value as low as 10 nanoseconds. There is no performance or latency impact because of the lower hardware polling interval. The default hardware polling of 4 milliseconds is selected in order to make sure you do not overflow the histogram counters before the software polls every one second. If you lower the hardware polling interval then it might saturate the hardware counters at 255 samples. The device cannot handle a software polling lower than one second, in order to match the lower hardware polling due to CPU and memory restrictions. The whitepaper has the example of the lower hardware polling interval and its use case.

Appendix - Feature Information

Buffer Management
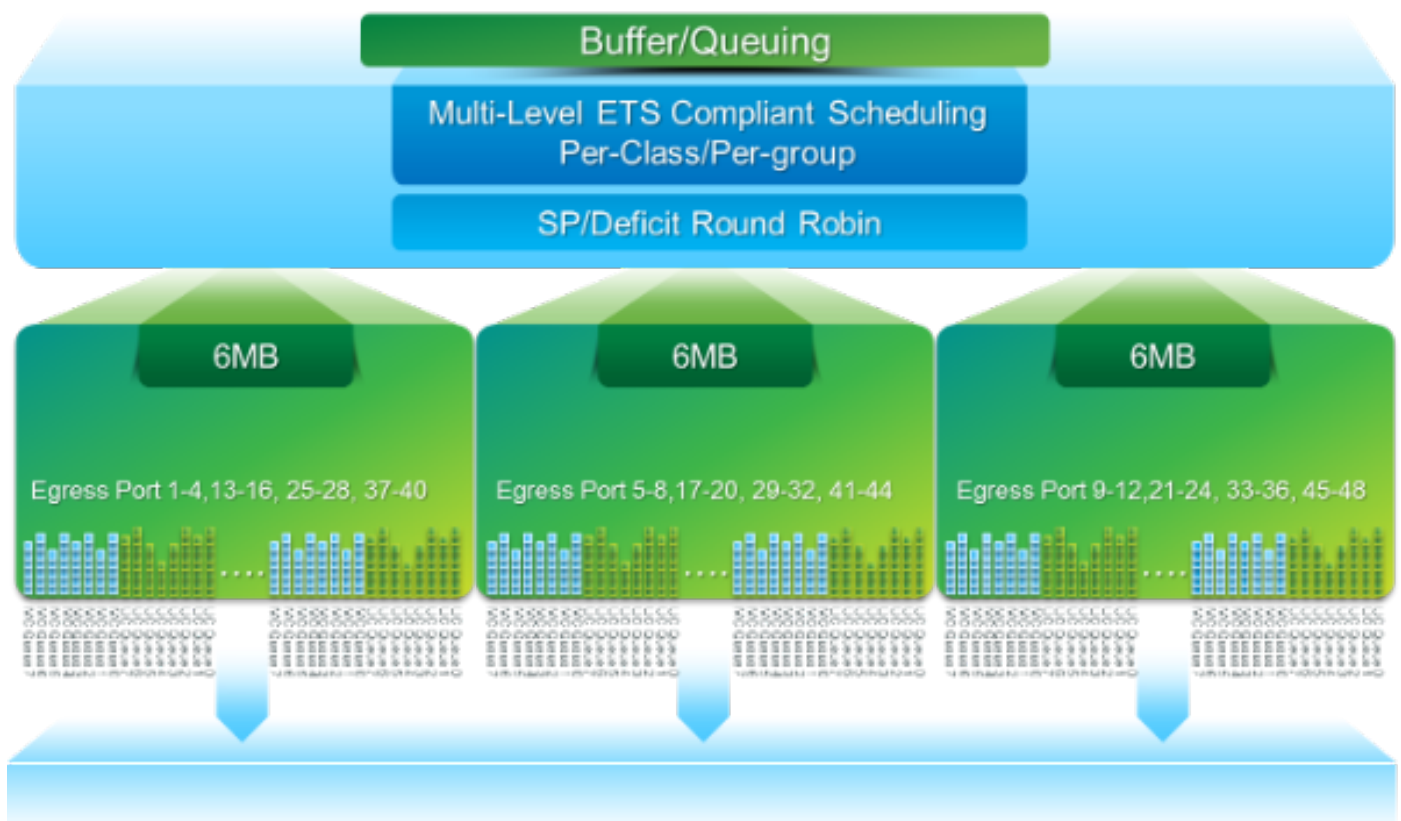
- 18 MB packet buffer shared by three OB blocks: ~4 MB Reserved: Size based on configured Maximum Transmission Unit (MTU) (Per port sum of 2 x MTU Size x # of enabled QoS groups)~14 MB Shared: Remainder of total buffer~767 KB of OB: 0 for CPU destined packets
- 6 MB for each OB is shared by a set of 16 ports (**show hardware internal mtc-usd info port-mapping** command)
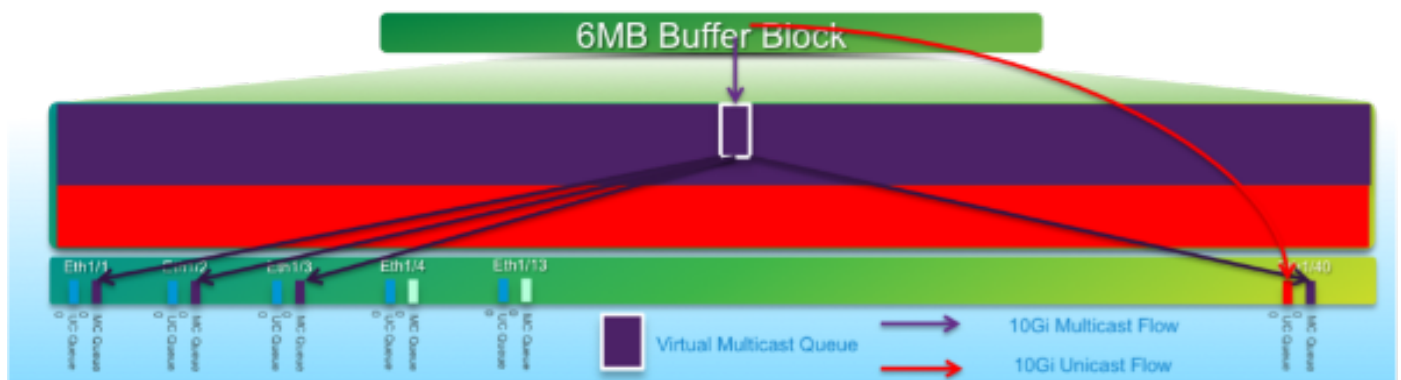
Scheduling

Three-layer scheduling:

- Unicast and multicast
- Traffic classes of the same scheduling scheme
- Traffic classes across the scheme
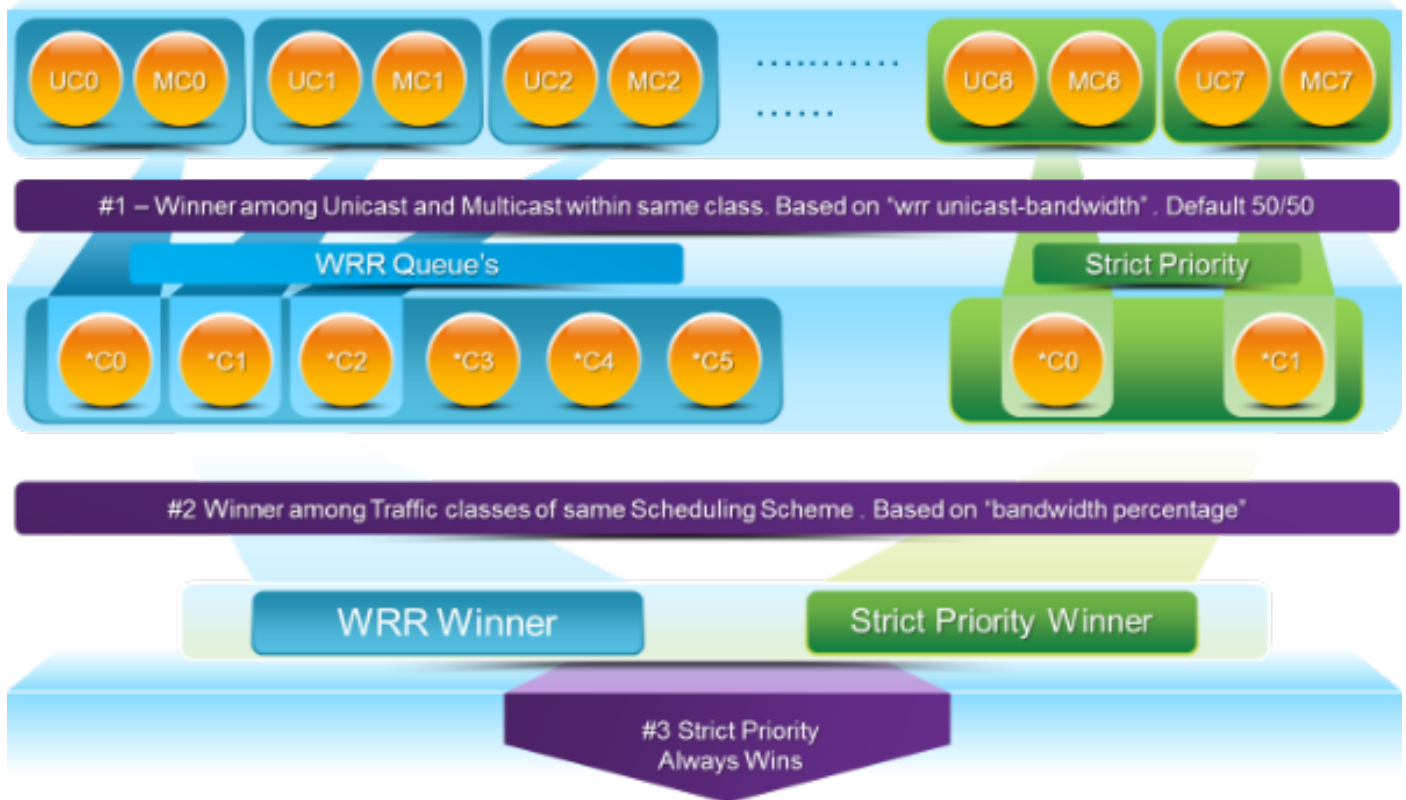
Multicast Slow-Receiver

In this diagram:

- Sustained congestion is introduced on 1 G Eth1/40.
- Other multicast receivers (Eth1/1 - 3) on the buffer block are impacted due to the multicast scheduling behavior. *Receivers on other buffer blocks remain unaffected.*
- "Multicast slow-receiver" can be applied to e1/40 in order to avoid traffic loss on non-congested ports.
- "Multicast slow-receiver" drains the multicast at a 10 G rate on Eth1/40. *Drops are still expected to occur on the congested port.*

- Configured with the **hardware profile multicast slow-receiver port <x>** command.

Active Buffer Monitoring

See the [Cisco Nexus 3548 Active Buffer Monitoring](#) whitepaper for an overview of this feature.

Hardware Implementation

- ASIC has 18 buckets and each bucket corresponds to a range of buffer utilization (for example, 0-384KB, 385-768KB, and so on).
- ASIC polls the buffer utilization for all ports every 4 milliseconds (default). This ASIC polling interval is configurable as low as 10 nanoseconds.
- Based on the buffer utilization for each hardware polling interval, the bucket counter for the corresponding range is incremented. That is, if port 25 consumes 500 KB of buffer, the bucket #2 (385-768KB) counter is incremented.
- This buffer utilization counter is maintained for each interface in histogram format.
- Each bucket is represented with 8 bits, so the counter maxes out at 255 and it is reset once the software reads the data.

Software Implementation

- Every one second, the software polls ASIC in order to download and clear all histogram counters.
- These histogram counters are maintained in the memory for 60 minutes with one second granularity.
- The software also makes sure it copies the buffer histogram to the bootflash every hour, which can be copied to the analyzer for further analysis.
- Effectively, this maintains two hours worth of buffer histogram data for all ports, the latest one hour in the memory and second hour in the bootflash.