了解和解决 Nexus 5600/6000 的输入丢弃问题

目录

简介

本文档介绍如何解决思科 Nexus 5600/6000 系列交换机上的输入丢弃问题。

先决条件

要求

思科建议您对Cisco Nexus 6000系列配置有基本的了解。

使用的组件

本文档中的信息基于以下软件和硬件版本：

- 思科Nexus 6001
- 7.1(3)N1(1)

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

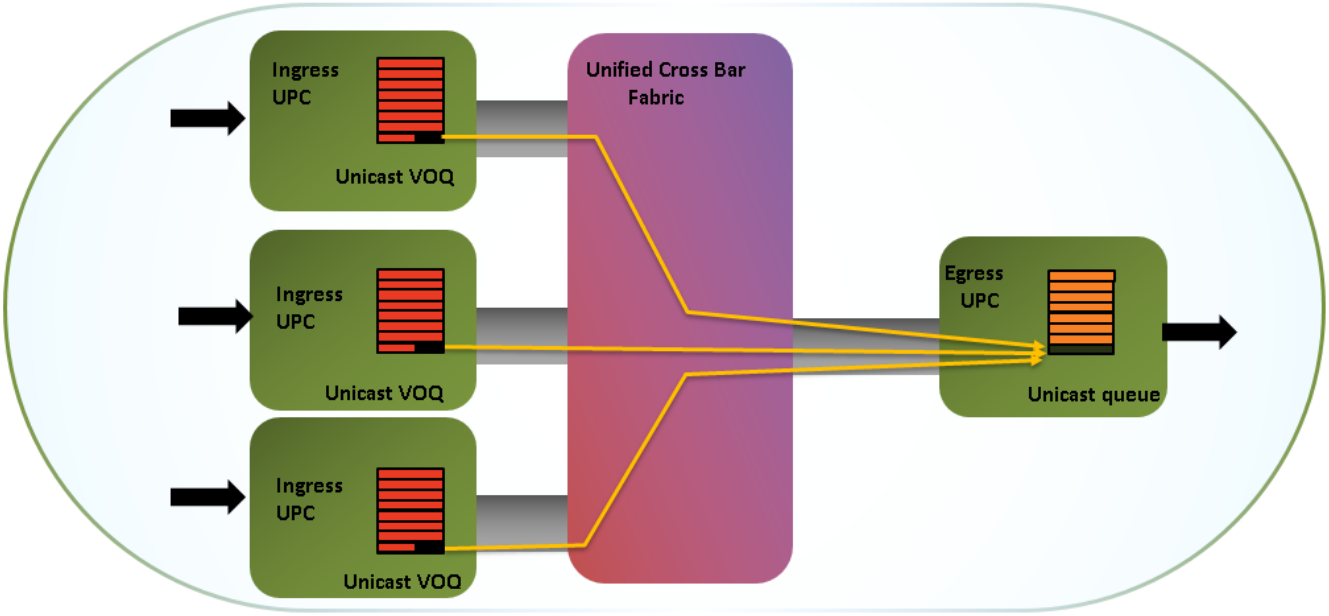
背景信息

输入丢弃是超订用出口端口的指示。这也意味着您可能会丢弃该特定端口上的单播流量。本文可帮

助您了解单播和组播流量如何在此平台上缓冲，以及输入丢弃如何随缓解步骤一起发生。

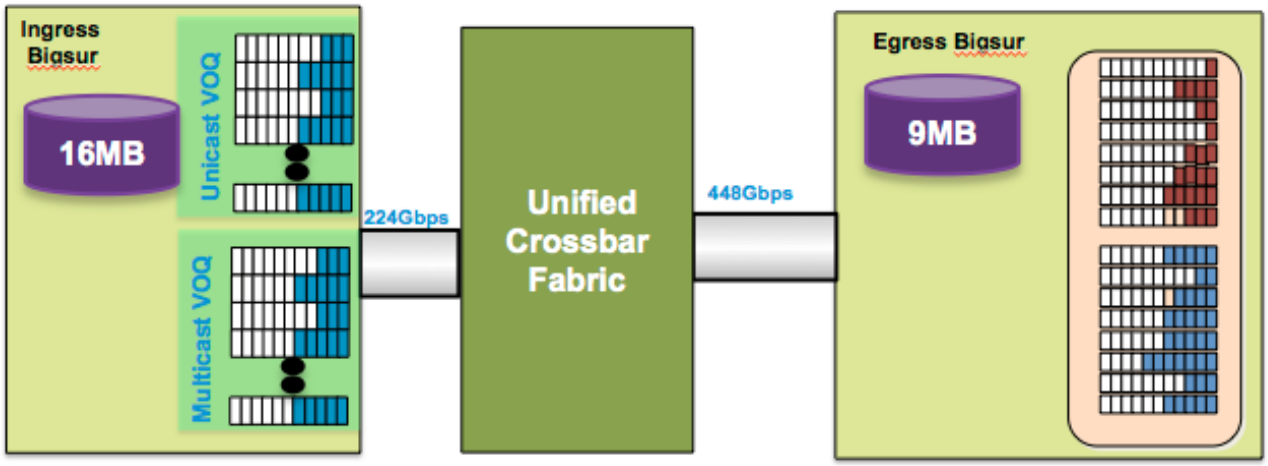
单播流量和缓冲

单播流量先在出口缓冲池排队，然后在出口队列已满后进入缓冲区，如图所示。



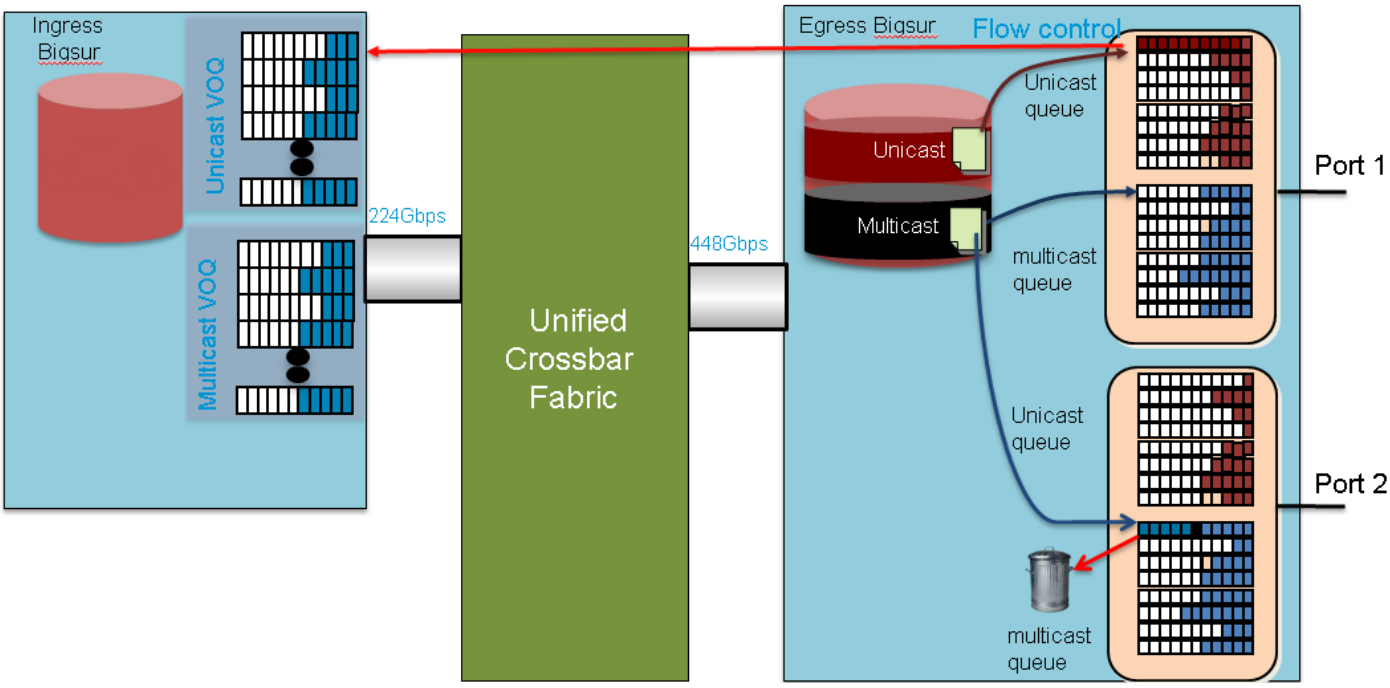
有16MB的入口共享缓冲区和9MB的出口共享缓冲区。缓冲区在12个10千兆端口或3个40千兆端口之间共享。共享缓冲区对突发吸收有利。

以下是供参考的内存分配的直观描述（Bigsur是ASIC/统一端口控制器的名称），如图所示。



组播流量流和缓冲

- 组播数据包在出口处被缓冲并丢弃
- 丢弃靠近拥塞点的组播数据包以避免线路头阻塞(HOLB)
- 维护单播的无损交换矩阵，如图所示。



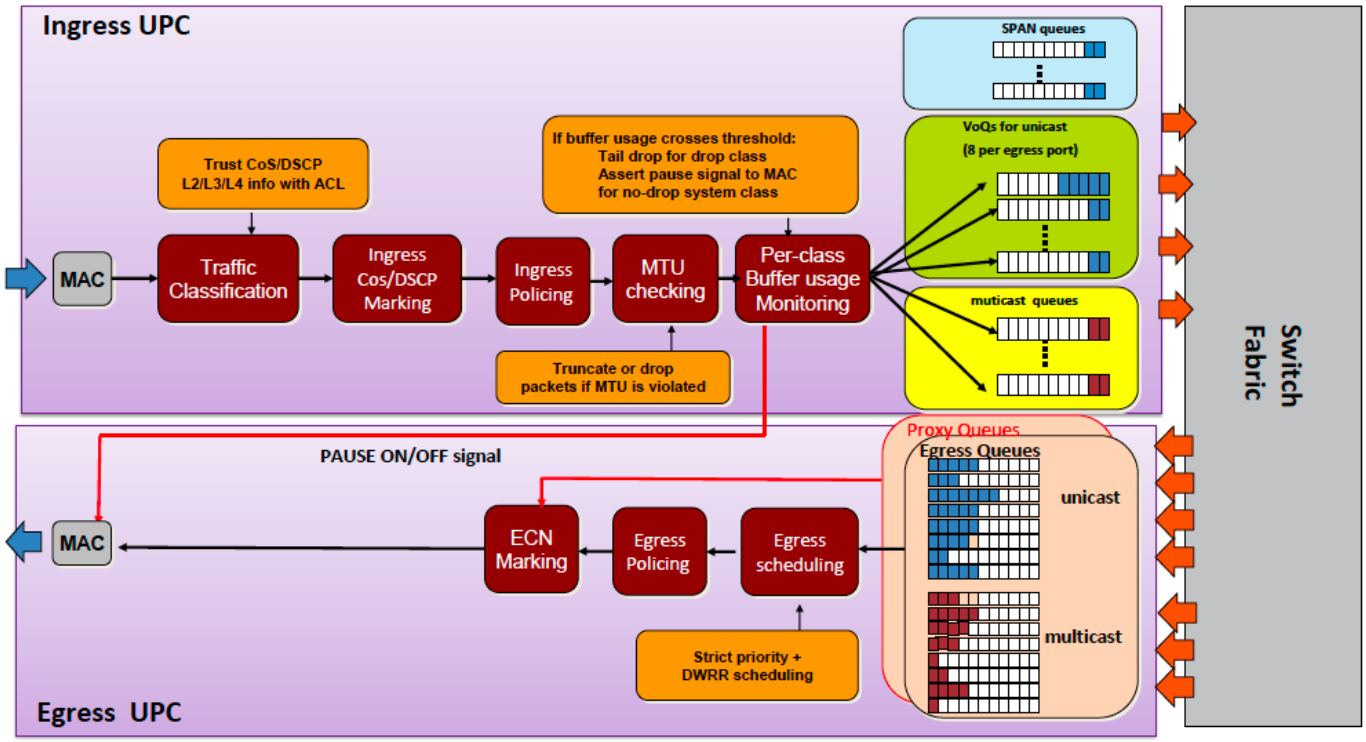
在大多数情况下，出口丢弃始终是由组播/广播/未知单播流量引起的。

什么原因导致输入丢弃？

拥塞的出口端口会导致出口缓冲区先填满，然后会对入口造成背压。这仅适用于单播流量。一旦入口缓冲区已满，您可能会丢弃入口上导致输入丢弃的流量。

此解释的级别非常高且易于摘要，但其内容更多，尤其是当您查看不同的流量类别、队列等时。虚拟输出队列(VOQ)的概念在Nexus平台上经常使用。VOQ是每个出口端口的每个IEEE 802.1p服务类别(CoS)的入口缓冲区分配。因此每个出口端口有8个VOQ。

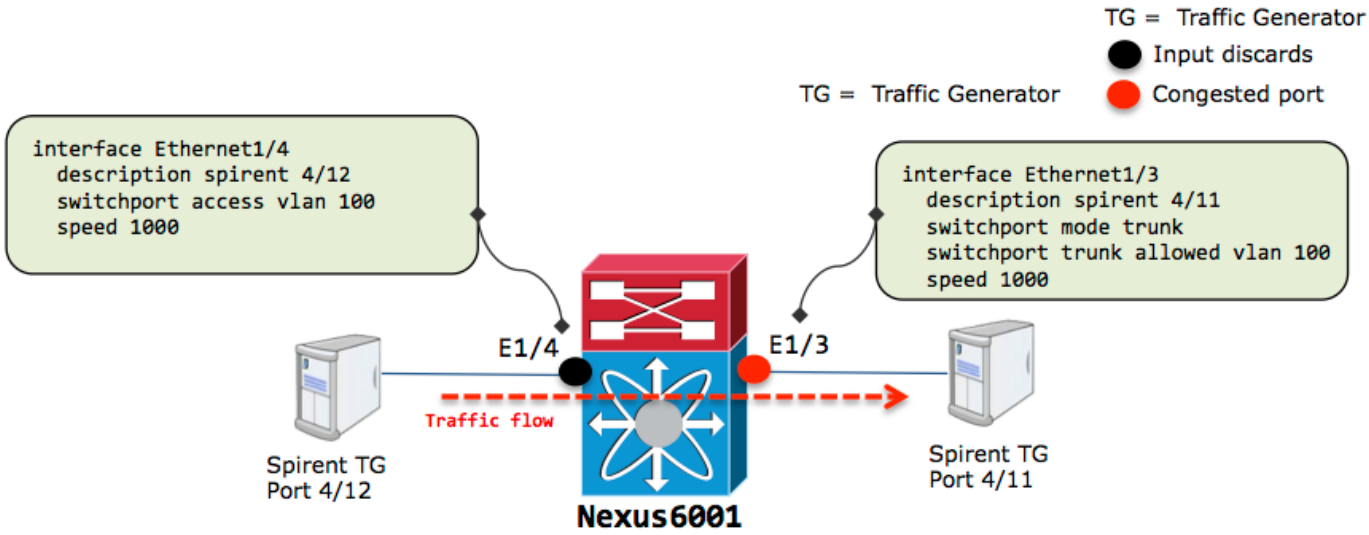
一个CoS中一个出口端口上的拥塞最终会渗透到入口端口上其相应VOQ的拥塞。达到限制后，流量将被丢弃。但是，它不会影响发往其他CoS或其他出口接口的流量，从而避免HOLB，否则会导致拥塞扩散。从入口到出口端口的流量以及正在播放的各种块如图所示。



故障排除情况

场景1.输入丢弃

实验设置:



线速流量流出e1/3并可能超订用：

```
nexus6001# sh int e1/3
Ethernet1/3 is up
 Dedicated Interface
 Hardware: 1000/10000 Ethernet, address: 002a.6a56.7a8a (bia 002a.6a56.7a8a)
 Description: spirent 4/11
 MTU 1500 bytes, BW 1000000 Kbit,, BW 1000000 Kbit, DLY 10 usec
 reliability 255/255, txload 251/255, rxload 25/255
```

```
 Encapsulation ARPA, medium is broadcast
 Port mode is trunk
 full-duplex, 1000 Mb/s
 Beacon is turned off
 Input flow-control is off, output flow-control is off
 Switchport monitor is off
 EtherType is 0x8100
 Last link flapped 11:39:20
 Last clearing of "show interface" counters 00:00:15
 0 interface resets
 30 seconds input rate 98683696 bits/sec, 8223 packets/sec
 30 seconds output rate 986853640 bits/sec, 82019 packets/sec
 Load-Interval #2: 5 minute (300 seconds)
   input rate 98.68 Mbps, 8.22 Kpps; output rate 986.85 Mbps, 82.01 Kpps
 RX
   124003 unicast packets  0 multicast packets  0 broadcast packets
   124003 input packets  186004500 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
   1236745 unicast packets  9 multicast packets  0 broadcast packets
   1236754 output packets  1860065401 bytes
   0 jumbo packets
   0 output error  0 collision  0 deferred  0 late collision
   0 lost carrier  0 no carrier  0 babble 0 output discard
   0 Tx pause



nexus6001# sh int e1/4
Ethernet1/4 is up
 Dedicated Interface

 Hardware: 1000/10000 Ethernet, address: 002a.6a56.7a8b (bia 002a.6a56.7a8b)
 Description: spirent 4/12
 MTU 1500 bytes, BW 1000000 Kbit,, BW 1000000 Kbit, DLY 10 usec
 reliability 255/255, txload 25/255, rxload 251/255
 Encapsulation ARPA, medium is broadcast
 Port mode is access
 full-duplex, 1000 Mb/s
 Beacon is turned off
 Input flow-control is off, output flow-control is off
 Switchport monitor is off
 EtherType is 0x8100
 Last link flapped 10:53:31
 Last clearing of "show interface" counters 00:00:04
 0 interface resets
 30 seconds input rate 986840376 bits/sec, 82236 packets/sec
 30 seconds output rate 98421072 bits/sec, 8223 packets/sec
 Load-Interval #2: 5 minute (300 seconds)
   input rate 986.84 Mbps, 82.23 Kpps; output rate 98.42 Mbps, 8.22 Kpps
 RX
   326332 unicast packets  0 multicast packets  0 broadcast packets
   326332 input packets  489496500 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  863 input discard >>>>>>
```

```
   0 Rx pause
 TX
   32633 unicast packets  2 multicast packets  0 broadcast packets
   32635 output packets  48819096 bytes
   0 jumbo packets
   0 output error  0 collision  0 deferred  0 late collision
   0 lost carrier  0 no carrier  0 babble 0 output discard
   0 Tx pause
```

在如下所示的模拟设置中，您知道超订用的原因，但在流量量变曲线突发的生产设置中，通过这些命令找出拥塞的出口端口可能是一个挑战。

此处列出的步骤可帮助您识别拥塞的出口端口。

步骤1.识别输入丢弃的端口

端口e1/4上看到的输入丢弃：

```
nexus6001# sh int e1/4 | in i disc
   0 input with dribble  3024 input discard
   0 lost carrier  0 no carrier  0 babble 0 output discard

nexus6001# sh queuing int e1/4
Ethernet1/4 queuing information:
 TX Queuing
   qos-group  sched-type  oper-bandwidth
       0       WRR            100

 RX Queuing
   qos-group 0 >>>> Drops in QOS 0
   q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
       Pkts received over the port             : 9612480
       Ucast pkts sent to the cross-bar        : 9587016
       Mcast pkts sent to the cross-bar        : 0
       Ucast pkts received from the cross-bar  : 961249
       Pkts sent to the port                   : 961261
       Pkts discarded on ingress               : 3024 >>>>>>
       Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

步骤2. ASIC标识

- 从此输出将接口映射到内部ASIC(UPC)编号。
- 从您注意到丢弃的入口端口ID中查找入口ASIC ID。

```
nexus6001# sh hard internal bigsur all-ports

Bigsur Port Info:
Port     |asic|inst|inst|
name     |idx |slot|asic|eport|logi|flag|adm|opr|if_index|diag|ucVer
---------+----+----+----+-----+----+----+---+---+--------+----+-----
sup1     |0   |0   |0   | 0 - |48  |b3  |en |dn |15010000|pass| 0.00
sup0     |0   |0   |0   | 1 - |49  |b3  |en |dn |15020000|pass| 0.00
1gb1/1   |1   |0   |1   | 2 - |0   |b3  |en |up |1a000000|pass| 0.00
1gb1/2   |1   |0   |1   | 3 - |1   |b3  |en |up |1a001000|pass| 0.00
1gb1/3   |1   |0   |1   | 0 - |2   |b3  |en |up |1a002000|pass| 0.00
```

```
1gb1/4   |1** |0   |1   | 1 - |3   |b3  |en |up |1a003000|pass| 0.00 >>>>** is the asic number
1gb1/5   |1   |0   |1   | 6 - |4   |b3  |en |up |1a004000|pass| 0.00
1gb1/6   |1   |0   |1   | 7 - |5   |b3  |en |up |1a005000|pass| 0.00
1gb1/7   |1   |0   |1   | 4 - |6   |b3  |en |up |1a006000|pass| 0.00
1gb1/8   |1   |0   |1   | 5 - |7   |b3  |en |up |1a007000|pass| 0.00
1gb1/9   |1   |0   |1   |10 - |8   |b3  |en |up |1a008000|pass| 0.00
1gb1/10  |1   |0   |1   |11 - |9   |b3  |en |up |1a009000|pass| 0.00
1gb1/11  |1   |0   |1   | 8 - |10  |b3  |en |up |1a00a000|pass| 0.00
xgb1/12  |1   |0   |1   | 9 - |11  |b3  |en |dn |1a00b000|pass| 0.00
xgb1/13  |2   |0   |2   | 2 - |12  |b3  |en |dn |1a00c000|pass| 0.00
xgb1/14  |2   |0   |2   | 3 - |13  |b3  |en |dn |1a00d000|pass| 0.00
xgb1/15  |2   |0   |2   | 0 - |14  |b3  |en |dn |1a00e000|pass| 0.00
xgb1/16  |2   |0   |2   | 1 - |15  |b3  |en |dn |1a00f000|pass| 0.00
xgb1/17  |2   |0   |2   | 6 - |16  |b3  |en |dn |1a010000|pass| 0.00
xgb1/18  |2   |0   |2   | 7 - |17  |b3  |en |dn |1a011000|pass| 0.00
xgb1/19  |2   |0   |2   | 4 - |18  |b3  |en |dn |1a012000|pass| 0.00
xgb1/20  |2   |0   |2   | 5 - |19  |b3  |en |dn |1a013000|pass| 0.00
xgb1/21  |2   |0   |2   |10 - |20  |b3  |en |dn |1a014000|pass| 0.00
xgb1/22  |2   |0   |2   |11 - |21  |b3  |en |dn |1a015000|pass| 0.00
xgb1/23  |2   |0   |2   | 8 - |22  |b3  |en |dn |1a016000|pass| 0.00
xgb1/24  |2   |0   |2   | 9 - |23  |b3  |en |dn |1a017000|pass| 0.00
xgb1/25  |3   |0   |3   | 2 - |24  |b3  |en |dn |1a018000|pass| 0.00
xgb1/26  |3   |0   |3   | 3 - |25  |b3  |en |dn |1a019000|pass| 0.00
xgb1/27  |3   |0   |3   | 0 - |26  |b3  |en |dn |1a01a000|pass| 0.00
xgb1/28  |3   |0   |3   | 1 - |27  |b3  |en |dn |1a01b000|pass| 0.00
xgb1/29  |3   |0   |3   | 6 - |28  |b3  |en |dn |1a01c000|pass| 0.00
xgb1/30  |3   |0   |3   | 7 - |29  |b3  |en |dn |1a01d000|pass| 0.00
xgb1/31  |3   |0   |3   | 4 - |30  |b3  |en |dn |1a01e000|pass| 0.00
xgb1/32  |3   |0   |3   | 5 - |31  |b3  |en |dn |1a01f000|pass| 0.00
xgb1/33  |3   |0   |3   |10 - |32  |b3  |en |dn |1a020000|pass| 0.00
xgb1/34  |3   |0   |3   |11 - |33  |b3  |en |dn |1a021000|pass| 0.00
xgb1/35  |3   |0   |3   | 8 - |34  |b3  |en |dn |1a022000|pass| 0.00
xgb1/36  |3   |0   |3   | 9 - |35  |b3  |en |dn |1a023000|pass| 0.00
xgb1/37  |4   |0   |4   | 2 - |36  |b3  |en |dn |1a024000|pass| 0.00
xgb1/38  |4   |0   |4   | 3 - |37  |b3  |en |dn |1a025000|pass| 0.00
xgb1/39  |4   |0   |4   | 0 - |38  |b3  |en |dn |1a026000|pass| 0.00
xgb1/40  |4   |0   |4   | 1 - |39  |b3  |en |dn |1a027000|pass| 0.00
xgb1/41  |4   |0   |4   | 6 - |40  |b3  |en |dn |1a028000|pass| 0.00
xgb1/42  |4   |0   |4   | 7 - |41  |b3  |en |dn |1a029000|pass| 0.00
xgb1/43  |4   |0   |4   | 4 - |42  |b3  |en |dn |1a02a000|pass| 0.00
xgb1/44  |4   |0   |4   | 5 - |43  |b3  |en |dn |1a02b000|pass| 0.00
xgb1/45  |4   |0   |4   |10 - |44  |b3  |en |dn |1a02c000|pass| 0.00
xgb1/46  |4   |0   |4   |11 - |45  |b3  |en |dn |1a02d000|pass| 0.00
xgb1/47  |4   |0   |4   | 8 - |46  |b3  |en |dn |1a02e000|pass| 0.00
xgb1/48  |4   |0   |4   | 9 - |47  |b3  |en |dn |1a02f000|pass| 0.00
40gb2/1  |5   |1   |0   | 2 - |0   |b3  |dis|dn |1a0f0000|pass| 0.00
40gb2/2  |5   |1   |0   | 1 - |1   |b3  |dis|dn |1a0f1000|pass| 0.00
40gb2/3  |6   |1   |1   | 2 - |2   |b3  |dis|dn |1a0f2000|pass| 0.00
40gb2/4  |6   |1   |1   | 1 - |3   |b3  |dis|dn |1a0f3000|pass| 0.00
Done.
```

步骤3.确定出口拥塞端口

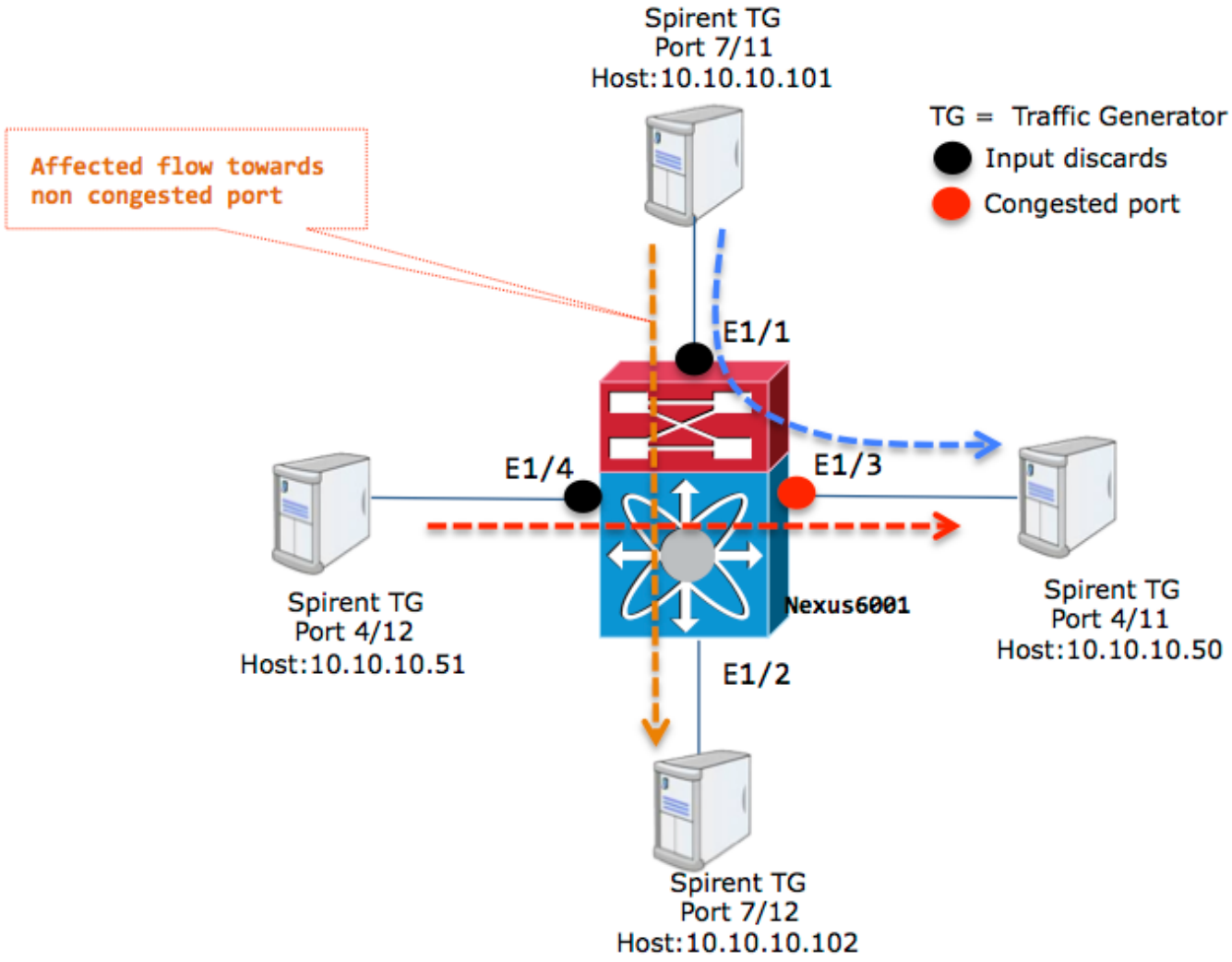
- 使用VOQ计数器识别拥塞出口端口。
- 使用计数器**voq asic-num**中的ASIC编**号，以便找出哪个出口端口对丢弃有贡献。**

```
nexus6001# sh plat soft qd info counters voq asic-num 1
+----------+------------------------------+------------------------+-----------+
|      port|                      TRANSMIT|               TAIL DROP|  HEAD DROP|
+----------+------------------------------+------------------------+-----------+
Eth1/3
 QUEUE-3                       3222876464                 8545008           0
```

```
Eth1/4
 QUEUE-3                          323451170                          0            0
Eth1/6
 QUEUE-3                             871362                          0            0
SUP_HI
 QUEUE-0                               2041                          0            0
+----------+------------------------------+------------------------+-----------+
```

场景2. HOLB的输入丢弃

实验设置:



所有端口都在VLAN 100中。

您可以看到e1/4和e1/1上的输入丢弃，这取决于通向se1/3的入口接口上的流量速率。

```
nexus6001# sh int e1/4 | in discard|rate
 30 seconds input rate 592103840 bits/sec, 49341 packets/sec
 30 seconds output rate 166412120 bits/sec, 13863 packets/sec
   input rate 592.10 Mbps, 49.34 Kpps; output rate 834.82 Mbps, 69.55 Kpps
   0 input with dribble  15245 input discard
   0 lost carrier  0 no carrier  0 babble 0 output discard

nexus6001# sh int e1/1 | in discard|rate
```

```
 30 seconds input rate 986839872 bits/sec, 82236 packets/sec
 30 seconds output rate 99790992 bits/sec, 8310 packets/sec
   input rate 986.84 Mbps, 82.23 Kpps; output rate 500.88 Mbps, 41.73 Kpps
   0 input with dribble  110632 input discard
   0 lost carrier  0 no carrier  0 babble 0 output discard
```

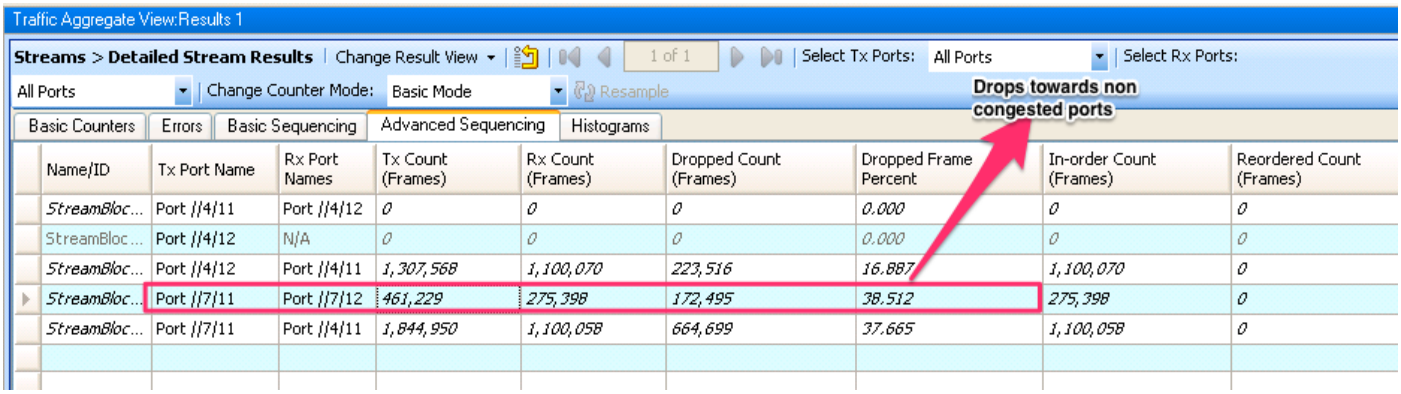
使用与场景1记录的相同流程。您可以找到出口拥塞端口。

```
nexus6001# sh plat so qd info counters voq asic-num 1 <snip>
+----------+-------------------------------+------------------------+-----------+
|      port|                       TRANSMIT|               TAIL DROP|  HEAD DROP|
+----------+-------------------------------+------------------------+-----------+
Eth1/3
 QUEUE-3                        3893719464               164782171           0
```

必须影响的流向10.10.10.50。10.10.101和10.10.102之间的流必须是干净的。

但情况并非如此。阻塞或缓慢排出的出口端口可能导致一个或多个入口端口上将流量发送到出口端口的所有缓冲区耗尽，从而影响这些入口端口上的所有流量。这是典型的HOLB问题。

Spirent流量生成器显示流被丢弃。端口号为Spirent端口号，如图所示。



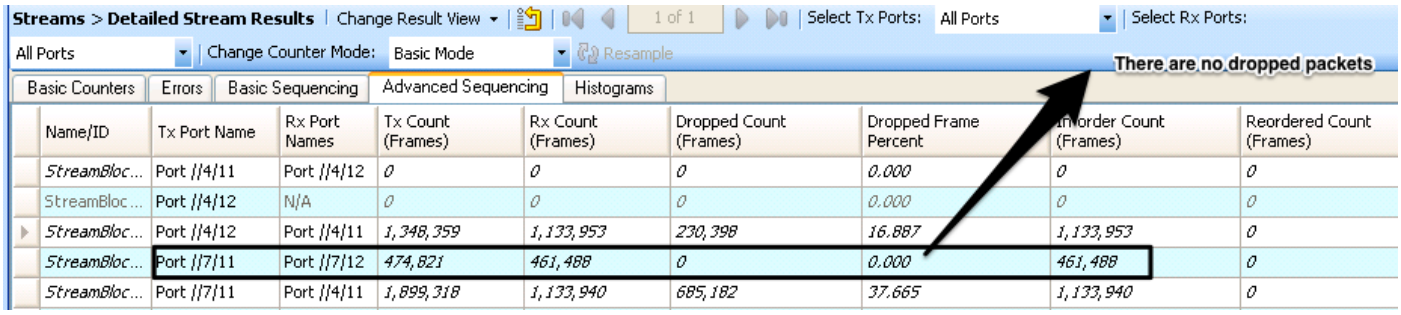
HOLB缓解：启用VOQ限制

为避免此场景，VOQ（仅用于单播流量）可以配置设置的阈值。

```
nexus6001(config)# hard unicast voq-limit
```

配置后，流向非拥塞端口的流量不受影响。

VOQ限制配置后的Spirent流量生成器视图如图所示。



虽然此配置显示了防止因HOLB而丢弃的明显优势。为什么这不是默认配置？

通常，生产环境中的流量在本质上可能会突发。通过禁用VOQ阈值，您允许入口缓冲区吸收流量微突发，而无需丢弃。

除非此情况需要启用VOQ限制，否则建议使用默认设置，即将其禁用。

HOLB缓解：通信分类

还有一种方法可通过使用QoS配置来缓解HOLB。由于入口丢弃仅影响特定VOQ，而该VOQ又是特定QOS类，因此您可以将受影响的流量映射到非拥塞端口到不同的QOS组。从此输出中，入口丢弃会影响QOS组0类。

```
nexus6001# sh queuing int e1/4
Ethernet1/4 queuing information:
 TX Queuing
   qos-group  sched-type  oper-bandwidth
       0       WRR            100

 RX Queuing
   qos-group 0 >>>> Drops in QOS 0
   q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
       Pkts received over the port             : 9612480
       Ucast pkts sent to the cross-bar        : 9587016
       Mcast pkts sent to the cross-bar        : 0
       Ucast pkts received from the cross-bar  : 961249
       Pkts sent to the port                   : 961261
       Pkts discarded on ingress               : 3024 >>>>>>
       Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

此处的配置将重要流量映射到QoS组2。

1.为不能丢弃的流量定义ACL。目标是将此流量分类到不同的QOS组，使其不受影响。

```
ip access-list SINGLEFLOW
 statistics per-entry
 10 permit ip 10.10.10.101/32 10.10.10.102/32
```

2. QoS分类：

```
class-map type qos match-all FIX_AFFECTED_FLOW
 match access-group name SINGLEFLOW
policy-map type qos QOS_POLICY_FIX_AFFECTED_FLOW
 class FIX_AFFECTED_FLOW
   set qos-group 2
```

3.网络QoS配置：

```
class-map type network-qos QOSGRP2
 match qos-group 2
policy-map type network-qos NQOS-GRP2
 class type network-qos QOSGRP2
 class type network-qos class-default
```

4.应用各种政策。网络QoS是系统范围，而分类策略可应用于单个接口。

```
system qos
 service-policy type network-qos NQOS-GRP2

interface Ethernet1/1
 service-policy type qos input QOS_POLICY_FIX_AFFECTED_FLOW
```

5. QoS组2类未看到丢弃：

```
nexus6001(config-if)# sh queuing int e1/1
Ethernet1/1 queuing information:
 TX Queuing
   qos-group  sched-type  oper-bandwidth
       0       WRR            100
       2       WRR              0
 RX Queuing
   qos-group 0
   q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
       Pkts received over the port             : 525111
       Ucast pkts sent to the cross-bar        : 327510
       Mcast pkts sent to the cross-bar        : 0
       Ucast pkts received from the cross-bar  : 0
       Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 197868 >>>>
       Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
   qos-group 2
   q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
   drop-type: drop, xon: 0, xoff: 0
   Statistics:
       Pkts received over the port             : 131413
       Ucast pkts sent to the cross-bar        : 132096
       Mcast pkts sent to the cross-bar        : 0
       Ucast pkts received from the cross-bar  : 0
       Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0 >>> No Drops
       Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

相关信息

- Nexus 6000系列交换机QoS配置示例
- 技术支持和文档 - Cisco Systems