

Comprensione E Risoluzione Dei Problemi Relativi Ai Discard Di Ingresso Per Nexus 5600/6000

Sommario

[Introduzione](#)

[Prerequisiti](#)

[Requisiti](#)

[Componenti usati](#)

[Premesse](#)

[Flusso Del Traffico Unicast E Buffer](#)

[Flusso Del Traffico Multicast E Buffer](#)

[Cause dei rigetti di input](#)

[Risoluzione dei problemi](#)

[Scenario 1. Scartamenti input](#)

[Passaggio 1. Identificazione Delle Porte Con Input Ignorati](#)

[Passaggio 2. Identificazione ASIC](#)

[Passaggio 3. Identificazione della porta congestionata in uscita](#)

[Scenario 2. Input scartati con HOLB](#)

[Mitigazione HOLB: Abilita limite VOQ](#)

[Mitigazione HOLB: Classificazione traffico](#)

[Informazioni correlate](#)

Introduzione

In questo documento viene descritto come risolvere i problemi relativi agli scarti di input sugli switch Cisco Nexus serie 5600/6000.

Prerequisiti

Requisiti

Cisco raccomanda la conoscenza di base della configurazione di Cisco Nexus serie 6000.

Componenti usati

Le informazioni fornite in questo documento si basano sulle seguenti versioni software e hardware:

- Cisco Nexus 6001
- 7.1(3)N1(1)

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico

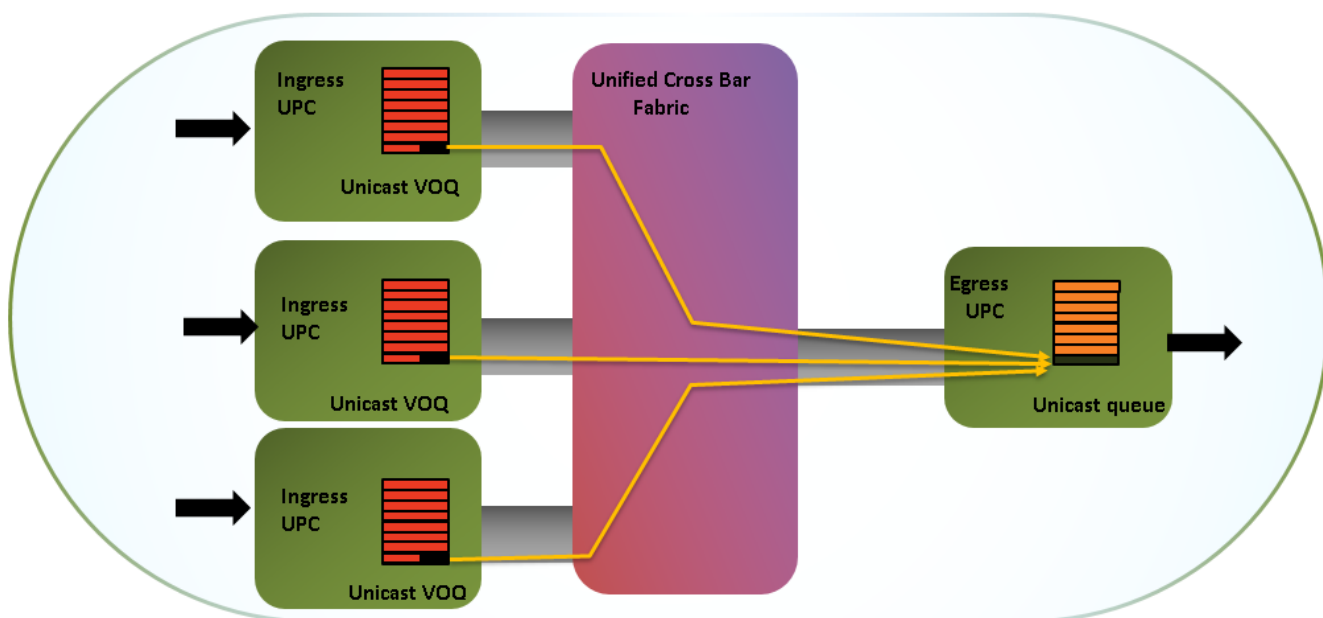
ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Premesse

I rigetti di input indicano una porta in uscita con sottoscrizione eccessiva. Significa anche che è probabile che si stia eliminando il traffico unicast su quella porta specifica. In questo articolo viene spiegato come il traffico unicast e multicast viene memorizzato nel buffer su questa piattaforma e come potrebbero verificarsi errori di input insieme alle operazioni di mitigazione.

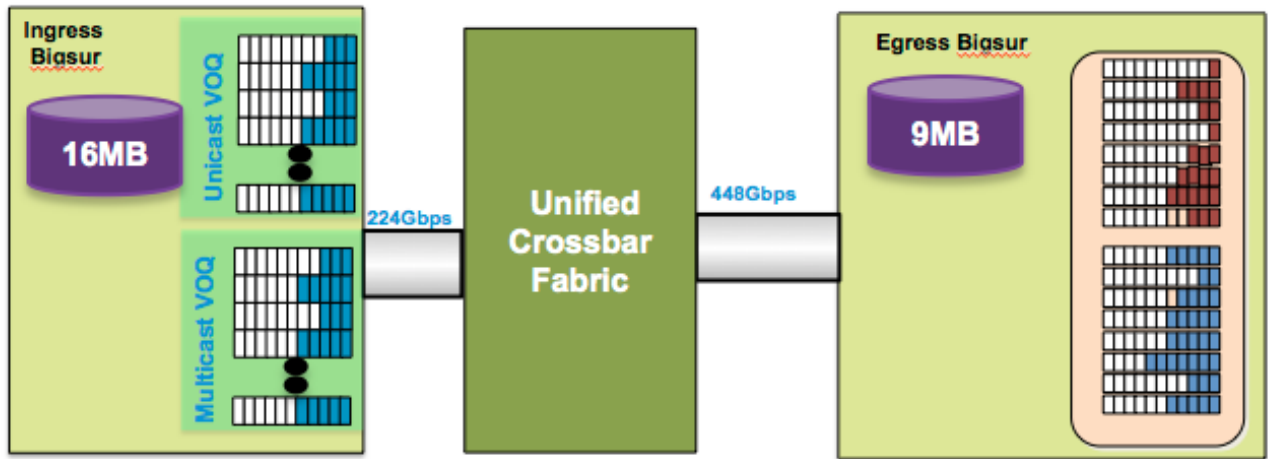
Flusso Del Traffico Unicast E Buffer

Il traffico unicast viene inserito prima nella coda del pool di buffer in uscita, quindi nel buffer in entrata quando la coda in uscita è piena, come mostrato nell'immagine.



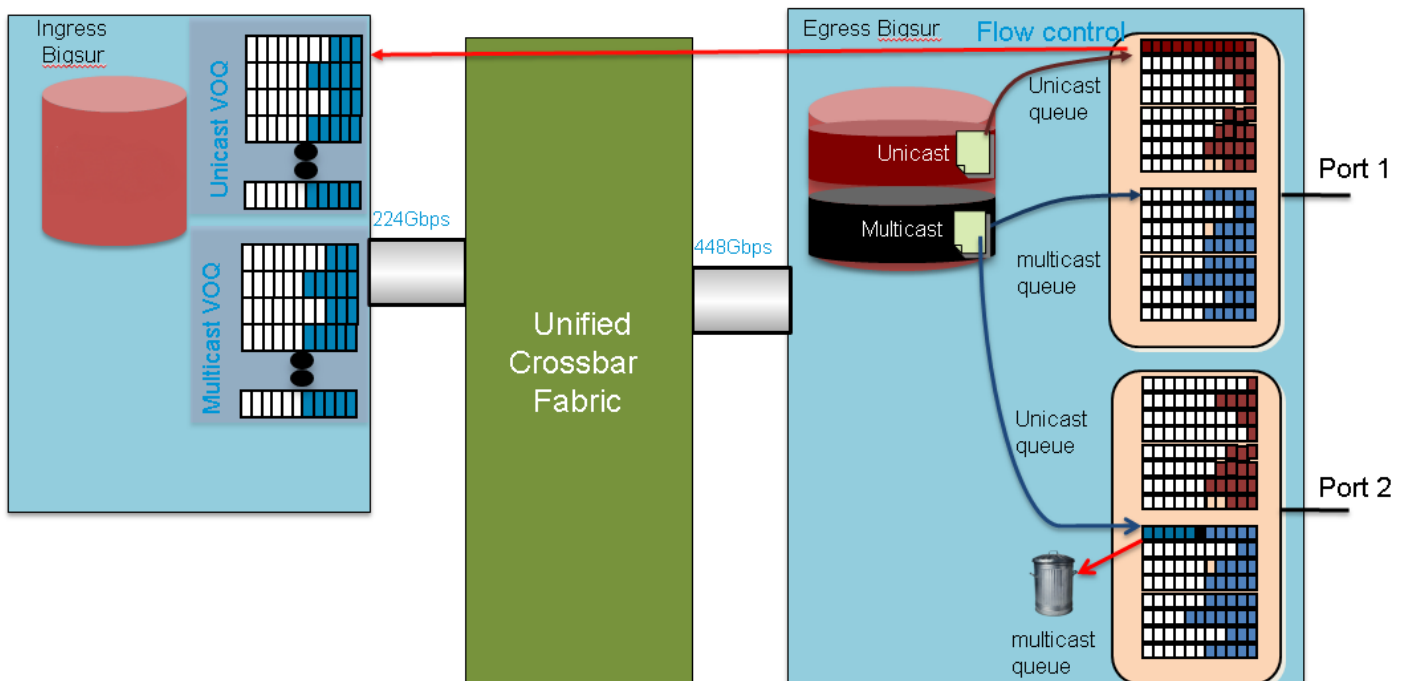
Sono disponibili 16 MB di buffer condiviso in entrata e 9 MB di buffer condiviso in uscita. I buffer sono condivisi tra 12 porte x 10 gig o 3 porte x 40 gig. Il buffer condiviso è utile per l'assorbimento della frammentazione.

Di seguito è riportata una rappresentazione visiva dell'allocazione di memoria per riferimento (BigSur è il nome dell'ASIC/Controller porta unificato), come mostrato nell'immagine.



Flusso Del Traffico Multicast E Buffer

- I pacchetti multicast vengono memorizzati nel buffer e scartati all'uscita
- Trascinare il pacchetto multicast vicino al punto di congestione per evitare il blocco Head of Line (HOLB)
- Mantenere il fabric senza perdite per unicast come mostrato nell'immagine.



Nella maggior parte dei casi, le cadute in uscita sono sempre dovute al traffico multicast/broadcast/unicast sconosciuto.

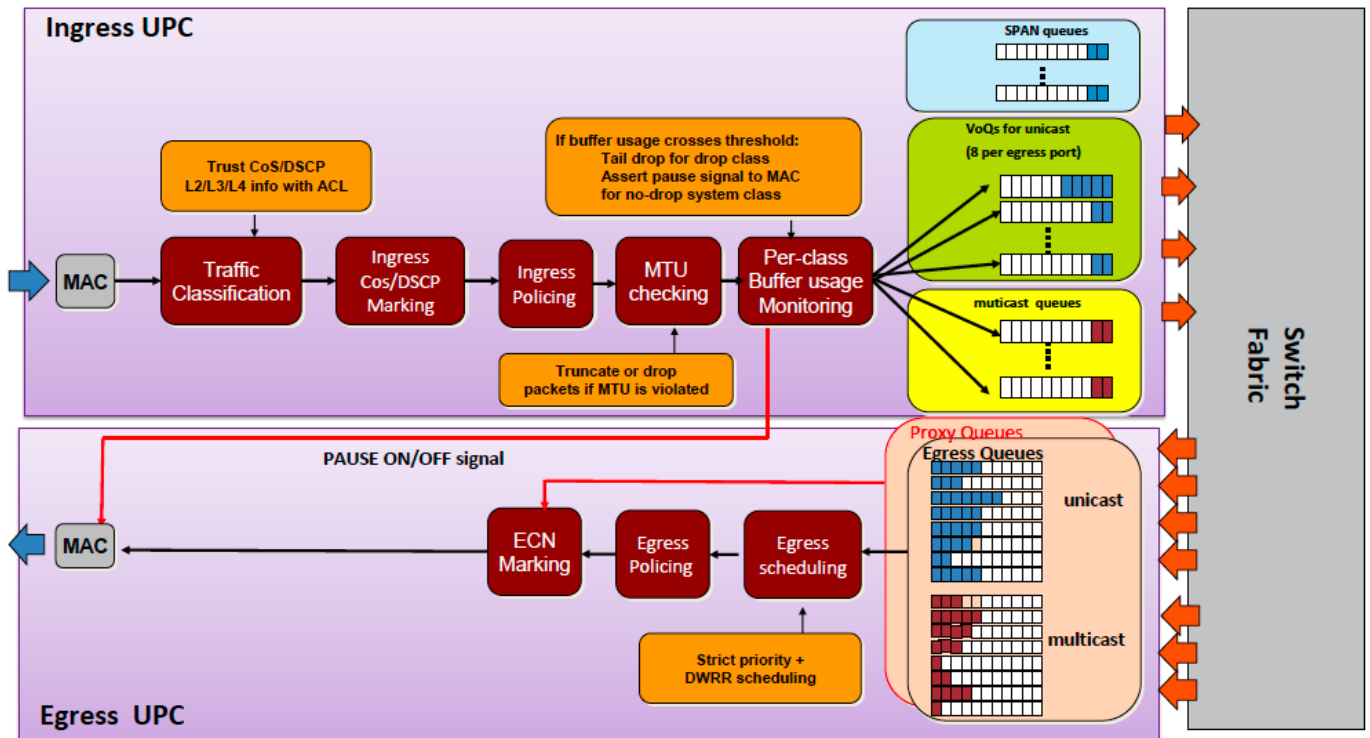
Cause dei rigetti di input

Una porta di uscita congestionata provoca il riempimento iniziale dei buffer di uscita e quindi la contropressione in entrata. Questa opzione è valida solo per il traffico unicast. Una volta che i buffer in entrata sono pieni, è possibile che il traffico in entrata causi la perdita di input.

Questa spiegazione è ad un livello molto alto e facile da digerire, ma c'è un po' di più in esso soprattutto quando si guarda a diverse classi di traffico, code ecc. Esiste un concetto di VOQ

(Virtual Output Queue) che viene spesso utilizzato sulla piattaforma Nexus. VOQ è un'allocazione di buffer in entrata per ogni classe di servizio (CoS) IEEE 802.1p per porta in uscita. Quindi ci sono 8 VOQ per porta di uscita.

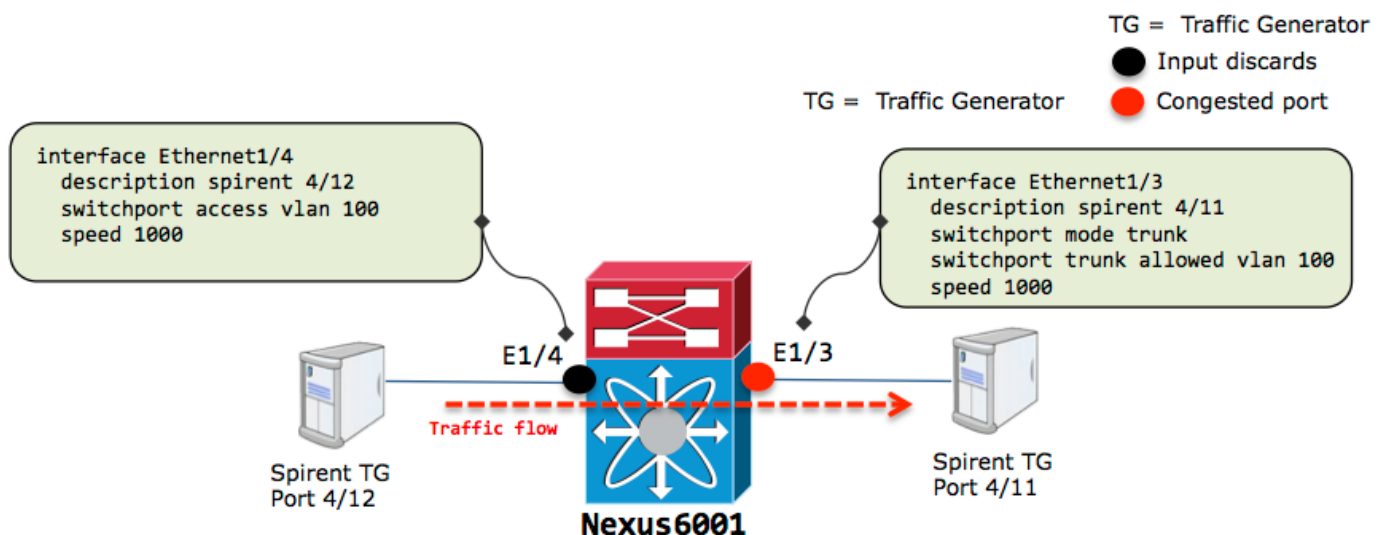
La congestione su una porta in uscita in un CoS finisce per interferire nella congestione del VOQ corrispondente sulla porta in entrata. Una volta raggiunto il limite, il traffico viene interrotto. Tuttavia, non influisce sul traffico destinato ad altri CoS o ad altre interfacce in uscita, evitando in tal modo l'HOLB, che altrimenti causerebbe la diffusione della congestione. Il flusso del traffico tra la porta in entrata e la porta in uscita e i vari blocchi in gioco è come mostrato nell'immagine.



Risoluzione dei problemi

Scenario 1. Scartamenti input

Configurazione lab:



Traffico di velocità della linea in uscita da e1/3 ed eventuale sovra-abbonamento:

```
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

```

In una configurazione simulata come questa, si conosce la causa della sovrassegnazione ma in una configurazione di produzione in cui il profilo del traffico aumenta e può essere difficile individuare le porte di uscita congestionate tramite questi comandi.

I passaggi elencati consentono di identificare le porte di uscita congestionate.

Passaggio 1. Identificazione Delle Porte Con Input Ignorati

Rifiuti di input rilevati sulla porta e1/4:

```

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)

```

Passaggio 2. Identificazione ASIC

- Mappare l'interfaccia al numero ASIC (UPC) interno da questo output.
- Individuare l'ID ASIC in entrata dall'ID della porta in entrata su cui si sono notate le perdite.

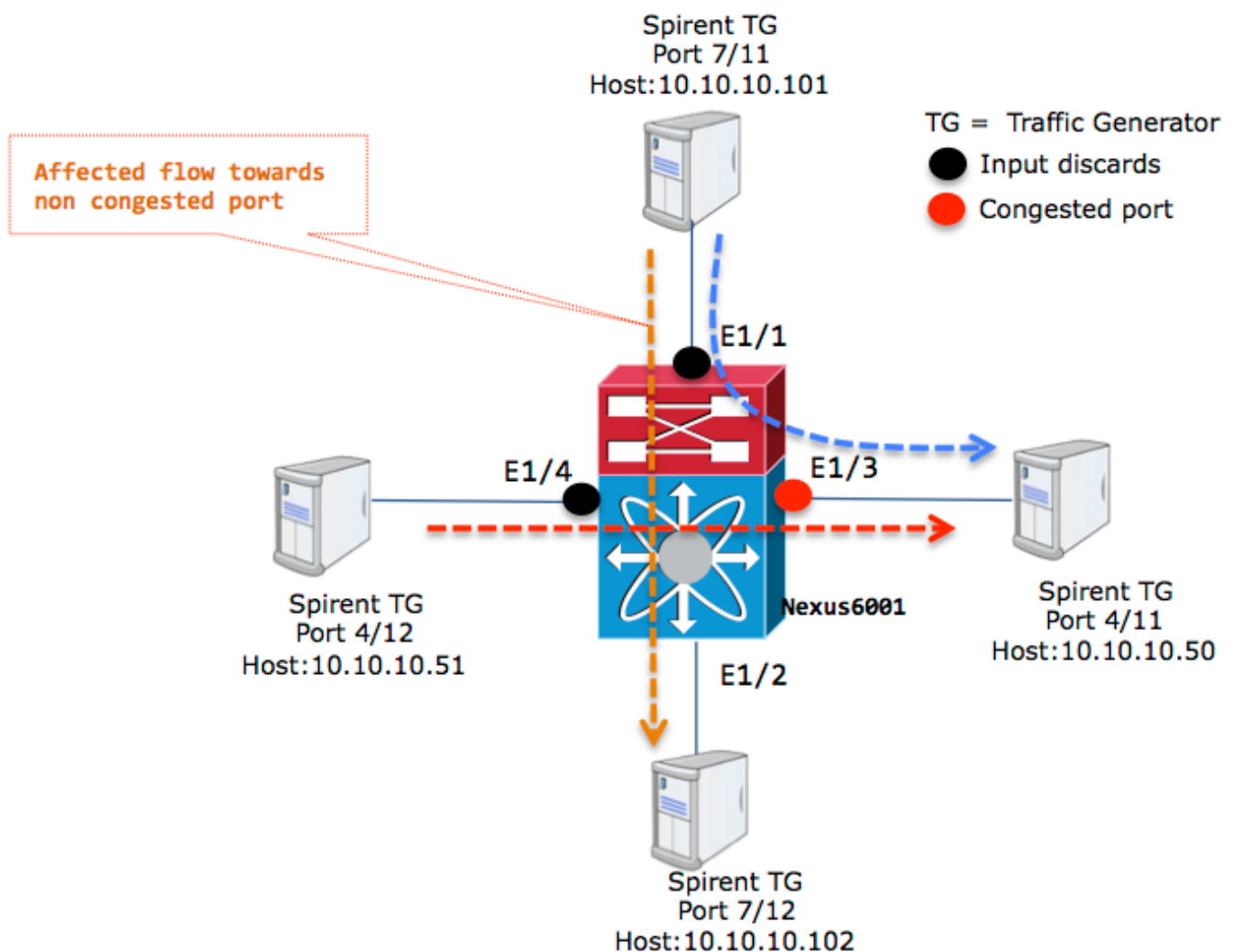
- Identificare la porta di uscita congestionata con i contatori VOQ.
- Usare il numero ASIC nei **contatori voq asic-num** per scoprire quale porta di uscita contribuisce alle cadute.

```
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

Scenario 2. Input scartati con HOLB

Configurazione lab:



Tutte le porte si trovano nella vlan 100.

Si possono vedere scarti di input sia su e1/4 che su e1/1 che dipendono dalla velocità del traffico in entrata sulle interfacce verso 1/3.

```
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
```

Utilizzando lo stesso processo documentato per lo scenario 1. È possibile trovare la porta congestionata in uscita.

```
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
```

Il flusso da modificare deve essere verso 10.10.10.50. Il flusso tra 10.10.10.101 e 10.10.10.102 deve essere pulito.

Ma non è così. Una porta in uscita bloccata o a basso drenaggio può causare l'esaurimento di tutti i buffer su una o più porte in entrata che inviano il traffico alla porta in uscita, con conseguente impatto su tutto il traffico su queste porte in entrata. Questo è il classico problema dell'HOLB.

I generatori di traffico a spirale mostrano che i flussi sono scesi. I numeri di porta sono numeri di porta spirale, come illustrato nell'immagine.

Name/ID	Tx Port Name	Rx Port Names	Tx Count (Frames)	Rx Count (Frames)	Dropped Count (Frames)	Dropped Frame Percent	In-order Count (Frames)	Reordered Count (Frames)
StreamBloc...	Port //4/11	Port //4/12	0	0	0	0.000	0	0
StreamBloc...	Port //4/12	N/A	0	0	0	0.000	0	0
StreamBloc...	Port //4/12	Port //4/11	1,307,568	1,100,070	223,516	16.887	1,100,070	0
StreamBloc...	Port //7/11	Port //7/12	461,229	275,398	172,495	38.512	275,398	0
StreamBloc...	Port //7/11	Port //4/11	1,844,950	1,100,058	664,699	37.665	1,100,058	0

Mitigazione HOLB: Abilita limite VOQ

Per evitare questo scenario, i VOQ (solo per il traffico unicast) possono essere configurati con una soglia impostata.

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

Dopo la configurazione, i flussi verso porte non congestionate non vengono influenzati.

La vista del generatore del traffico a spirale dopo la configurazione del limite VOQ è come mostrato nell'immagine.

Name/ID	Tx Port Name	Rx Port Names	Tx Count (Frames)	Rx Count (Frames)	Dropped Count (Frames)	Dropped Frame Percent	Reorder Count (Frames)	Reordered Count (Frames)
StreamBloc...	Port //4/11	Port //4/12	0	0	0	0.000	0	0
StreamBloc...	Port //4/12	N/A	0	0	0	0.000	0	0
StreamBloc...	Port //4/12	Port //4/11	1,348,359	1,133,953	230,398	16.887	1,133,953	0
StreamBloc...	Port //7/11	Port //7/12	474,821	461,488	0	0.000	461,488	0
StreamBloc...	Port //7/11	Port //4/11	1,899,318	1,133,940	685,182	37.665	1,133,940	0

Anche se questa configurazione mostra un chiaro vantaggio per evitare le cadute dovute a HOLB. Perché questa non è la configurazione predefinita?

In genere, il traffico in un ambiente di produzione potrebbe esplodere in natura. Disabilitando la soglia VOQ, si consente ai buffer in entrata di assorbire un micro burst del traffico senza che sia necessario cadere.

A meno che la situazione non renda necessario attivare il limite VOQ, si consiglia di utilizzare l'impostazione predefinita che lo lascia disattivato.

Mitigazione HOLB: Classificazione traffico

Esiste un altro metodo per ridurre l'utilizzo di HOLB con la configurazione QoS. Poiché gli scarti in entrata influiscono solo su una VOQ specifica che a sua volta è una classe QOS specifica, è possibile mappare il traffico interessato su una porta non congestionata a un gruppo QOS diverso. Da questo output, gli scarti in entrata influiscono sulla classe QOS Group 0.

```
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)
```

In questa configurazione viene eseguito il mapping del traffico importante a QoS-Group 2.

1. Definire l'ACL per il traffico che non deve essere eliminato. L'obiettivo è classificare il traffico in un gruppo QOS diverso in modo che non venga influenzato.

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

2. Classificazione QoS:

```
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. Configurazione QoS di rete:

```
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. Applicare le varie politiche. La QoS della rete è a livello di sistema, mentre il criterio di classificazione può essere applicato a una singola interfaccia.

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

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

5. Le cadute non sono visibili per il gruppo QoS classe 2:

```
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)

Informazioni correlate

- [Esempio di configurazione QoS degli switch Nexus serie 6000](#)
- [Documentazione e supporto tecnico – Cisco Systems](#)