

# Nexus 9500-R、Nexus 3000-R:排除輸入丟棄故障

## 目錄

[簡介](#)

[背景資訊](#)

[輸入流量管理員\(ITM\)](#)

[輸入VOQ緩衝區](#)

[計畫和流程控制](#)

[常見原因](#)

[適用硬體](#)

[排除輸入丟棄故障](#)

[常見場景 — 10G到1G流量 — 持續丟棄：](#)

[步驟1.檢查輸入丟棄介面上受影響的隊列。](#)

[步驟2.檢查Broadcom用於診斷的計數器的圖形表示：](#)

[步驟3.找出遇到輸入丟棄的前面板埠所屬的ASIC和Jericho埠：](#)

[步驟4.瞭解入口埠的VOQ和VOQ聯結器是什麼。](#)

[步驟5.從BCM的角度檢查Queue具體為Non-Empty;即擁塞。](#)

[步驟6.從非空隊列值中查詢出口擁塞埠：](#)

[步驟7.根據先前的查詢結果檢查哪個前面板埠位於ASIC 1中並對映到Jericho埠9。](#)

[其他命令](#)

[其它實驗室測試：](#)

[步驟1.輸入丟棄多個出口擁塞的介面。](#)

[步驟2.由於SPAN原因而丟棄輸入。](#)

[步驟3.由於交通髮夾導致輸入丟棄。](#)

[步驟4.傳送目的IP為Unknown的資料包。](#)

[步驟5.當接入/中繼埠轉變為STP轉發狀態時，輸入丟棄](#)

[步驟6.由於Eth1/9超過線路速率而丟棄輸入資料。](#)

## 簡介

本文檔介紹Cisco Nexus 9500-R EoR和Nexus 3000-R ToR輸入丟棄的原因和解決方案。輸入丟棄指示因擁塞而丟棄在輸入隊列中的資料包數。此數字包括尾部丟棄和加權隨機早期檢測(WRED)導致的丟包。

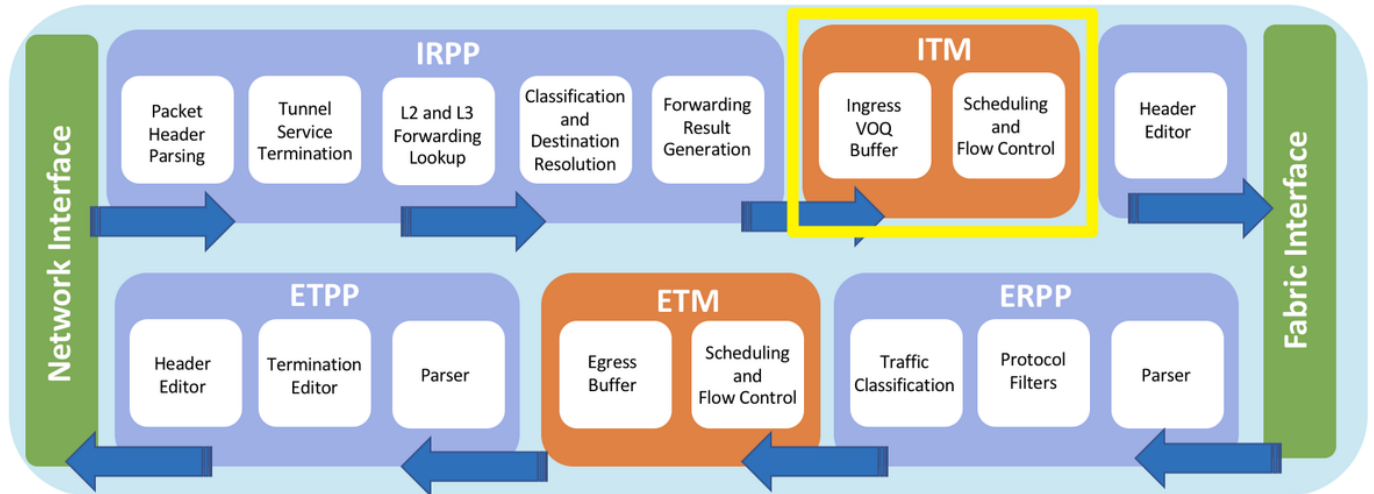
如果您遇到隨機/偶發/歷史（即不再發生）丟棄情況，請聯絡Cisco TAC進行進一步調查。當輸入丟棄頻繁遞增時，此瀏覽很有用。

## 背景資訊

R系列使用輸入VOQ架構。VOQ架構使用虛擬隊列模擬入口緩衝區中的出口隊列。每個出口埠有八個單播流量隊列和八個組播流量隊列。可以根據資料包中的服務等級(CoS)或差分服務代碼點(DSCP)值將流量分類為流量類別，然後在該流量類別的相應虛擬隊列中排隊。

R系列使用分散式信用機制通過交換矩陣傳輸流量。在安排封包離開VOQ之前，輸入緩衝區排程器會要求為輸出緩衝區中的特定連線埠和優先順序提供積分。從輸入信用排程程式為目標埠和優先順序請求信用。如果緩衝區空間可用，則出口排程程式將授予訪問許可權並將信用授予傳送到入口緩衝區排程程式。如果輸出緩衝區中沒有可用的緩衝區空間，則輸出排程不會授予積分，並且流量在VOQ中緩衝，直到下一個積分可用。

以下是一 R平台的資料包轉發管道。在本文章中，您將重點介紹Ingress Traffic Manager組件。關於此連結處的架構的更多詳細[資訊](#)



## 輸入流量管理員(ITM)

輸入流量管理器(ITM)是輸入管道中的一個塊。它執行與將流量排入VOQ相關的步驟，排程通過交換矩陣傳輸的流量，以及管理信用。

## 輸入VOQ緩衝區

輸入VOQ緩衝塊管理片上緩衝區和片外分組緩衝區。兩個緩衝區都使用VOQ架構，並且流量會根據來自IRPP (入口接收器資料包處理器) 的資訊排隊。單播和組播流量總計有96,000個VOQ。

## 計畫和流程控制

從入口管道傳輸資料包之前，需要計畫資料包以便通過交換矩陣傳輸。入口排程程式向位於出口流量管理器塊中的出口排程程式傳送信用請求。當輸入流量管理器收到信用時，它開始向輸入傳輸資料包處理器傳送流量。如果輸出緩衝區已滿，流量將在由輸出埠和流量類表示的專用隊列中緩衝。

## 常見原因

通常，由於以下原因，可以在各種Nexus硬體中看到輸入丟棄

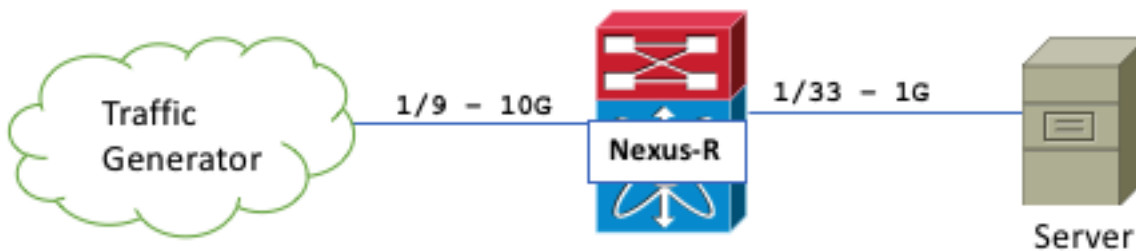
- 擁塞出口介面 (如10G ingress和1G egress) 的流量
- 超額訂用SPAN目的地連線埠 — 適用於特定硬體型別。

## 適用硬體

N9K-X9636C-R  
N9K-X9636Q-R  
N9K-X9636C-RX  
N9K-X96136YC-R  
N3K-C36180YC-R  
N3K-C3636C-R

## 排除輸入丟棄故障

常見場景 — 10G到1G流量 — 持續丟棄：



在本文中，「input discards」計數器的值和任何引用相同內容的HW內部計數器的值將隨著錯誤的增加而變化，同時必須即時捕獲相關命令。

步驟1.檢查輸入丟棄介面上受影響的隊列。

這一步以後會派上用場。

在本例中，它是Queue 7 (預設隊列) — 入口上共有8個隊列：

```
Nexus-R# show system internal qos queuing stats interface e1/9 | beg "QUEUE: 7"
QUEUE: 7
=====
ingress dropped packets: 113503981
ingress dropped bytes: 113503981000
enqueued packet count: 74115825
enqueued byte count: 74115825000
```

步驟2.檢查Broadcom用於診斷的計數器的圖形表示：

```
Nexus-R# bcm-shell mod 1 "diag counters g"
```

```

|
/|\
R F A C E | J E R I C H O N E T W O R K I N T 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	= 0	
-----+-----		
	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	= 0	
EPE_DSCRD_PKT_CNT	= 0	
RCY_PACKET_COUNTER	= 0	
IRE_FDT_INTRFACE_CNT	= 0	
-----+-----		
	IDR	
EGQ		
MMU_IDR_PACKET_COUNTER	= 10,659,302	
FQP_PACKET_COUNTER	= 606	
IDR_OCB_INTERFACE_COUNTER	= 0	
PQP_UNICAST_PKT_CNT	= 606	
PQP_DSCRD_UC_PKT_CNT	= 0	
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	
DEQUEUE_PKT_CNT	= 1,403,078	
EHP_MC_HIGH_PKT_CNT	= 0	
DELETED_PKT_CNT	= 0	
EHP_MC_LOW_PKT_CNT	= 0	
<b>ENQ_DISCARDED_PACKET_COUNTER</b>	<b>= 9,256,829</b>	
DELETED_PKT_CNT	= 0	
<b>Rejects: PORT_AND_PG_STATUS</b>		
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          = 0
|
EHP_MC_LOW_DSCRD_CNT           = 0
|
ERPP_LAG_PRUNING_DSCRD_CNT     = 0
|
ERPP_PMF_DISCARDS_CNT          = 0
|
ERPP_VLAN_MBR_DSCRD_CNT        = 0
+-----+-----+-----+-----+

```

```

|
FDA
|
CELLS_IN_CNT_P1      = 0      |   CELLS_OUT_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
|
CELLS_IN_MESHMC_CNT = 0      |   CELLS_OUT_MESHMC_CNT = 0
|   EGQ_PKT_CNT              = 606
CELLS_IN_IPT_CNT     = 606   |   CELLS_OUT_IPT_CNT     = 606
|   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
|   CFG_EVENT_CNT            = 606 *
|   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              = 0
|   TRANSMITTED_DATA_CELLS_COUNTER = 5,609,892
CELL_IN_CNT_TOTAL           = 0
+-----+-----+-----+-----+

```

```

/|\
|
                                J E R I C H O   F A B R I C   I N T E R
F A C E
|
                                \|/

```

大於零的QUEUE\_DELETED\_PACKET\_COUNTER表示資料包已被IQM ( 入口隊列管理器 ) 在隊列後刪除。這可能是由於活動隊列沒有收到任何信用額度，這表明計畫方案的配置錯誤。您可以通過 `bcm-shell mod X "getReg IQM_QUEUE_DELETED_PACKET_COUNTER"` 檢查此項

ENQ\_DISCARDED\_PACKET\_COUNTER表示在入隊之前丟棄資料包。您還可以在BCM中看到此

計數器集 ( 讀取時清除命令 ) :

```
Nexus-R# bcm-shell mod 1 "g iqm_reject_status_bmp" | i i PG|IQM0|IQM1
IQM_REJECT_STATUS_BMP.IQM0[0x1a7]=0x2000000: <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,
```

通過show hardware internal errors module X ( 命令在讀取時清除 ) , 您始終可以快速注意到這些錯誤 :

```
Nexus-R# show hardware internal errors module 1

slot 1
=====
|-----|
| Device:Forwarding ASIC Role:MAC Mod: 1 |
| Device Statistics Category :: ERROR
|-----|
Instance:0

IQM
-----
ENQ_DISCARDED_PACKET_COUNTER = 8,233,862
Rejects: PORT_AND_PG_STATUS

Instance:1
```

步驟3.找出遇到輸入丟棄的前面板埠所屬的ASIC和Jericho埠 :

```
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
```

顯示本示例的Eth1/33。在實際網路中, 您還不知道擁塞的出口埠。

步驟4.瞭解入口埠的VOQ和VOQ聯結器是什麼。

```
Nexus-R# attach module 1
module-1# show hardware internal jer-usd info voq asic 0 port 9
```

Unit	JerPort	Voq	VoqConn	SE	HR	CreditBal
0	9	104	176	82213	72	16a .

此命令向我們顯示特定埠的輸入VoQ流量的詳細資訊。此外，它還向我們顯示了VoQ的當前信用餘額。

連線埠的VOQ由此得出：

LC基於0 — 模組1為0，模組2為1等  
每個LC有256個系統埠ID

$ID = (LC * \text{系統埠ID}) + \text{FP編號}$   
 $\text{Eth1/9} = (0 * 256) + 9 = 9$   
 $\text{VOQ ID} = 32 + (\text{系統埠ID} * 8)$   
 $\text{Eth1/9} = 32 + (9 * 8) = 104$

因此，Eth1/9的VOQ為104，與之前收集的輸出相符

```
module-1# show hardware internal jer-usd ingress-vsqr buffer-occupancy front-port 9
```

VSQF BUFFER OCCUPANCY		
Front port 9		
max global shared		157286
max ocb buffer occupancy		0
COSQ 0		
rate class		4
granted buffers per port		3280
<b>shared buffers occupied</b>		<b>127792</b> <<<<
granted buffers occupied		3280
<b>shared buffer max occupancy</b>		<b>127792</b> <<<<

步驟5.從BCM的角度檢查Queue具體為Non-Empty;即擁塞。

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

Core 0:

```
Ingress VOQs Sizes (format: [queue_id(queue_size)]):
[303(191338496B)] << the Queue ID belongs to your Egress CONGESTED port!
```

Core 1:

```
<empty>
```

## 步驟6.從非空隊列值中查詢出口擁塞埠：

如果隊列是303，請記住，這些隊列實際上是一個範圍，因此可以是303 + 7或303-7 — 問題是，哪個埠的VOQ與296-303或303-310的範圍匹配？

眾所周知，Eth1/9上的隊列7是擁塞的，因此303實際上是其範圍內最高的，因此296-303的範圍是很有根據的猜測。

```
module-1# show hardware internal jer-usd info voq asic 1
```

Unit	JerPort	Voq	VoqConn	SE	HR	CreditBal
1	1	232	56	81957	8	3ffff
1	2	240	72	81989	16	3ffff
1	3	248	88	82021	24	3ffff
1	4	256	104	82053	32	3ffff
1	5	264	120	82085	40	3ffff
1	6	272	136	82117	48	3ffff
1	7	280	152	82149	56	3ffff
1	8	288	168	82181	64	3ffff
1	9	296	184	82213	72	3a5
1	10	304	200	82245	80	3ffff
1	11	312	216	82277	88	3ffff

<<< 296 + 7 would give us 303  
<< It cannot be this one as 303 is not included

```
<snip>
```

為asic 0顯示相同的 — 此處未顯示以簡潔明瞭；您會注意到Voq欄下方的關注範圍不在該ASIC中

請注意上述輸出中的幾點：

- 我們的出口擁塞埠位於ASIC 1上。
- 我們的出口擁塞埠的VOQ為296,303等於該埠上的隊列7。
- 注意Credit Balance列 — 此介面上剩餘的要授予的信用數量非常少，這就是我們的入口Eth1/9開始緩衝的原因。

## 步驟7.根據先前的查詢結果檢查哪個前面板埠位於ASIC 1中並對映到Jericho埠9。

```
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

這時，您發現了出口擁塞的埠 — 確定是否有異常情況突入網路，您已配置SPAN且目標埠為1G，同時尋找一個或多個10G介面，或者是否存在瓶頸/設計問題。

## 其他命令

這些技術更先進 — 在正常情況下無需查詢出口擁塞埠。

```
attach module X
show hardware internal jer-usd tm_debug asic <slot> module <module>
show hardware internal jer-usd info voq [ asic <instance> ] [ port <port> ] [ ]
```



```

show hardware internal jer-usd info non-empty voq asic [ <instance> ] [ ]
show hardware internal jer-usd info voq-profile { QueueThreshold drop_p <dp> | OCBThreshold } [
asic <instance> ] [ port<port> ] [ ]
show hardware internal jer-usd info voq-connector front-port <port> [ ]
show hardware internal jer-usd stats vsq { front-port <port> | inband asic <slot> | recycle-port
<port> asic <slot> }
show hardware internal jer-usd ingress-vsqs buffer-occupancy front-port <port>
show hardware internal jer-usd info IQM { counter | rate } asic <instance> dst-port <port> [
interval <int> ] [ ]
show hardware internal jer-usd info SCH { counter | rate } asic <instance> dst-port <port> [
interval <int> ] [ ]

```

```

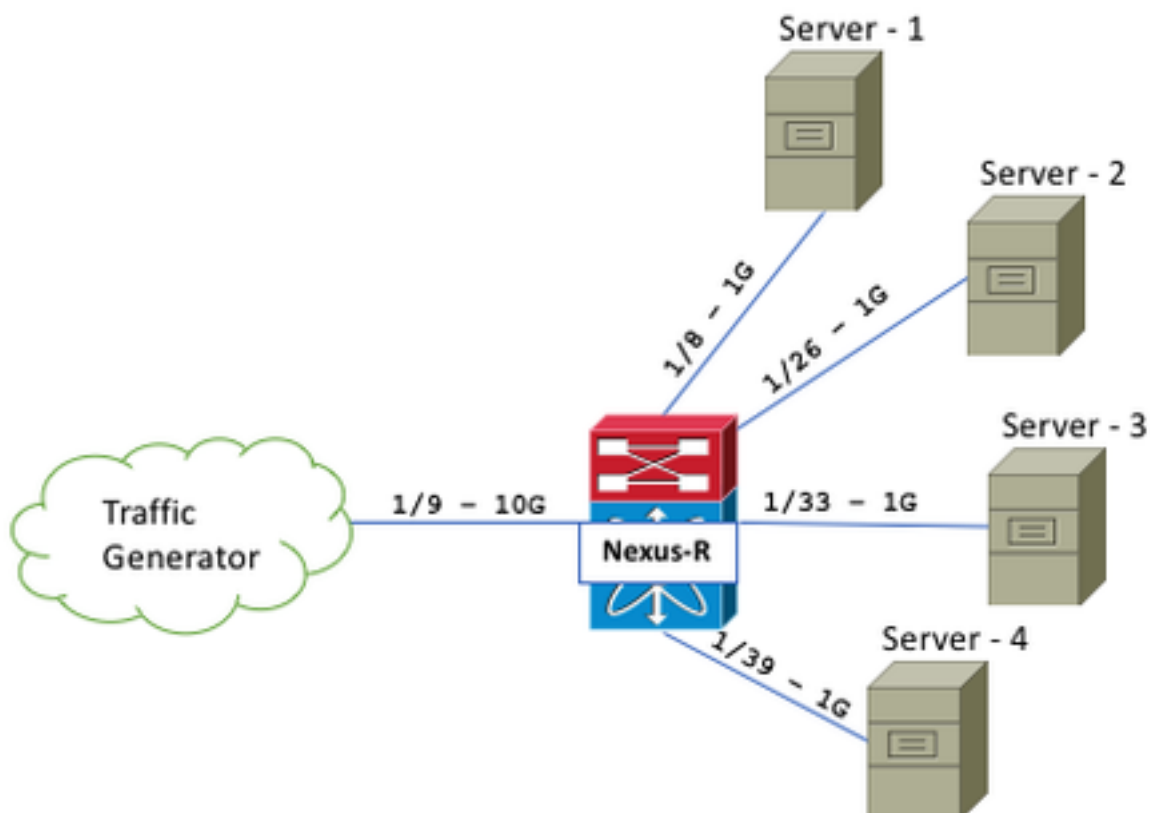
bcm-shell mod X
diag cosq print_flow_and_up dest_id=<flow_id>
diag cosq voq id=<voqid> detailed=1
diag cosq qpair e2e ps=<id>
cosq conn ing
cosq conn egr
dump IPS_CR_BAL_TABLE <voqID>
getReg IQM_QUEUE_MAXIMUM_OCCUPANCY_QUEUE_SIZE

```

## 其它實驗室測試：

步驟1.輸入丟棄多個出口擁塞的介面。

請考慮以下拓撲，其中流量生成器向每台伺服器傳送2G流量：



快速檢查哪些隊列不為空 — 注意有4個隊列：

```

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)]

```

確定這些隊列屬於哪些介面 — 首先檢查ASIC 0 ( 它只演示了一個介面 ) :

```

module-1# show hardware internal jer-usd info voq asic 0
-----+-----
|Unit|JerPort| Voq| VoqConn| SE| HR|CreditBal|
-----+-----
| 0| 1| 40| 48| 81957| 8| 3ffff|
| 0| 2| 48| 64| 81989| 16| 3ffff|
| 0| 3| 56| 80| 82021| 24| 3ffff|
...
| 0| 8| 96| 160| 82181| 64| 7b| << 96 + 7 = 103, this is port Eth1/8
<snip>

`show interface hardware-mappings`
-----
Name          Ifindex  Smod Unit  HPort HName NPort VPort SrcId
-----
Eth1/8        1a000e00 0  0  8  xe8  7  -1  0

```

對其他三個「隊列」值重複相同的過程：247、303和351。

## 步驟2.由於SPAN原因而丟棄輸入。

設定Eth1/33為SPAN目的地連線埠，同時設定Eth1/9為RX方向的SPAN來源連線埠

```

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

```

## 步驟3.由於交通髮夾導致輸入丟棄。

使用SRC 10.10.10.10和DEST 192.168.10.10傳送資料包，其中Eth1/9位於10.10.10.1/24中 — 這不會導致輸入丟棄；但是，您確實會看到以下計數器：

```

Nexus-R# bcm-shell mod 1 "diag counters g"

```

```

| /|\
| J E R I C H O N E T W O R K I N T E R F A C E |
\|/ |
-----+-----

```

```

-----+-----
_PACKET_COUNTER = 0 | DELETED_PKT_CNT = 12,027,201 |
| | Discards: INVALID_OTM SRC_EQUAL_DEST
-----+-----
-----+-----

```

**步驟4.傳送目的IP為Unknown的資料包。**

使用SRC 10.10.10.10和DEST 192.168.10.10傳送資料包，其中Eth1/9位於10.10.10.1/24中，Eth1/33是172.16.0.1/30子網中的L3埠 — 無丟棄計數器，即使目標未知，也不會丟棄任何輸入卡。

**步驟5.當接入/中繼埠轉變為STP轉發狀態時，輸入丟棄**

傳送資料包，其中Eth1/9隻是一個寬中繼埠（或接入埠） — 當埠轉變為STP轉發狀態時，此埠註冊為輸入丟棄。

```

Nexus-R(config)# int e1/9
Nexus-R(config-if)# switchport mode trunk
Nexus-R# bcm-shell mod 1 "diag counters g" | i i --|IQM|ENQ_DISCARD|Rejects
-----+-----
-----+-----
-----+-----
-----+-----
-----+-----
-----+-----
-----+-----
PQP_MC_PKT_CNT = 1,678,949 |
| IQM |
PQP_DSCRD_MC_PKT_CNT = 11,369,033 |
| ENQ_DISCARDED_PACKET_COUNTER = 1,289,182 |
DELETED_PKT_CNT = 11,369,081 |
| Rejects: QUEUE_NOT_VALID_STATUS |
Discards: SRC_EQUAL_DEST |
-----+-----
-----+-----

```

```

Nexus-R# show span int e1/9

Vlan Role Sts Cost Prio.Nbr Type
-----
VLAN0001 Desg BLK 2 128.9 P2p
VLAN0010 Desg BLK 2 128.9 P2p
<snip>

```

QUEUE\_NOT\_VALID\_STATUS是由於資料包處理器(PP)決定丟棄或從資料包處理器(PP)塊接收的無效目標而丟棄的。

**步驟6.由於Eth1/9超過線路速率而丟棄輸入資料。**

將10G+傳送到Eth1/9將導致不同的丟棄型別，因為您首先要將Eth1/9最大化 — 仍會計為「輸入丟棄」：

```

bcm-shell.0> diag counters g

```

R F A C E

\|/

		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	= 0	
-----		
		IRE
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	= 0	
RCY_PACKET_COUNTER	= 16,837	
IRE_FDT_INTRFACE_CNT	= 0	
-----		
		IDR
MMU_IDR_PACKET_COUNTER	= 54,128,577	
FQP_PACKET_COUNTER	= 50,703	
IDR_OCB_INTERFACE_COUNTER	= 0	
PQP_UNICAST_PKT_CNT	= 50,683	
PQP_DSCRD_UC_PKT_CNT	= 0	
PQP_UC_BYTES_CNT	= 5,216,716	
-----		
PQP_MC_PKT_CNT	= 20	
		IQM
PQP_DSCRD_MC_PKT_CNT	= 20	
PQP_MC_BYTES_CNT	= 2,079	
ENQUEUE_PKT_CNT	= 5,463,323	
EHP_UNICAST_PKT_CNT	= 50,683	
DEQUEUE_PKT_CNT	= 5,594,400	
EHP_MC_HIGH_PKT_CNT	= 20	
DELETED_PKT_CNT	= 0	
EHP_MC_LOW_PKT_CNT	= 0	
<b>ENQ_DISCARDED_PACKET_COUNTER</b>	<b>= 48,716,055</b>	
DELETED_PKT_CNT	= 40	
<b>Rejects: VOQ_MX_QSZ_STATUS</b>		

<snip>