

Nexus 9500-R、Nexus 3000-R:排除输入丢弃故障

目录

[简介](#)

[背景信息](#)

[入口流量管理器\(ITM\)](#)

[入口VOQ缓冲区](#)

[计划和流量控制](#)

[常见原因](#)

[适用硬件](#)

[排除输入丢弃故障](#)

[常见场景 — 10G到1G流量 — 持续丢弃：](#)

[步骤1.检查哪个队列在输入丢弃接口上受影响。](#)

[步骤2.检查用于诊断的Broadcom计数器的图形表示：](#)

[步骤3.查找遇到输入丢弃的前面板端口的ASIC和杰里科端口属于哪些：](#)

[步骤4.了解入口端口具有的VOQ和VOQ连接器。](#)

[步骤5.从BCM的角度检查哪个队列具体为非空；即拥塞。](#)

[步骤6.从非空队列值中查找出口拥塞端口：](#)

[步骤7.根据您的之前的发现，检查ASIC 1中的前面板端口和到杰里科端口9的映射。](#)

[其它命令](#)

[其他实验测试：](#)

[步骤1.使用多个出口拥塞接口丢弃输入。](#)

[步骤2.由于SPAN而导致的输入丢弃。](#)

[步骤3.由于流量毛发引脚而导致的输入丢弃。](#)

[步骤4.发送目的IP为未知的数据包。](#)

[步骤5.当接入/中继端口转换到STP转发状态时，输入丢弃](#)

[步骤6.由于Eth1/9超出线速而导致的输入丢弃。](#)

简介

本文档介绍Cisco Nexus 9500-R EoR和Nexus 3000-R ToR的输入丢弃的原因和解决方案。输入丢弃表示由于拥塞而在输入队列中丢弃的数据包数。此数字包括由尾部丢弃和加权随机早期检测(WRED)引起的丢包。

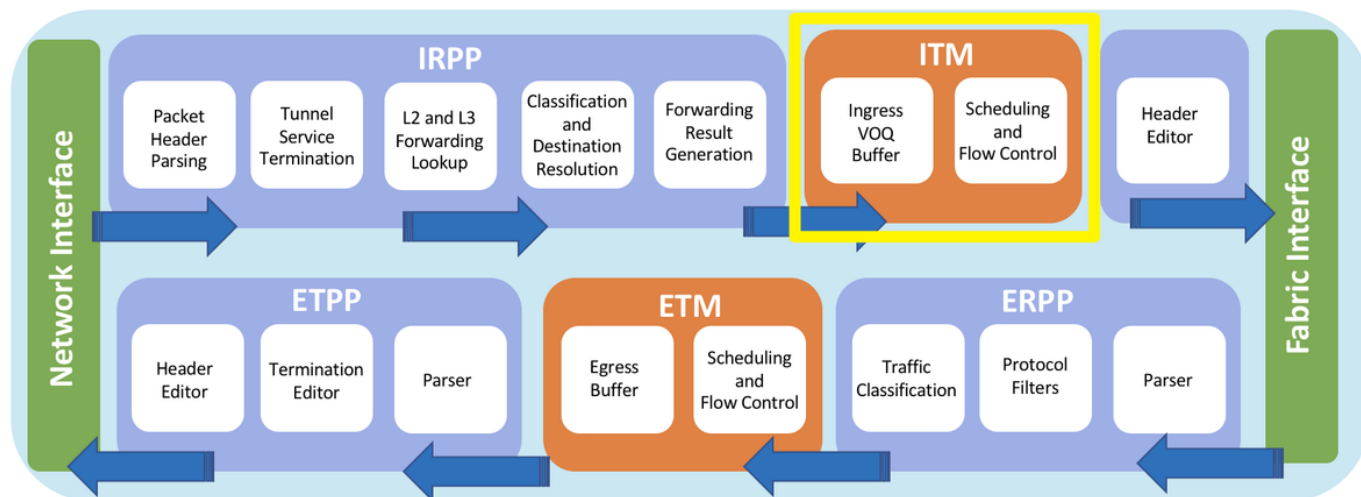
如果您遇到随机/偶发/历史（即不再出现）丢包，请联系思科TAC进一步调查。当输入丢弃频繁增加时，此逐步介绍功能非常有用。

背景信息

R系列使用入口VOQ架构。VOQ架构通过虚拟队列模拟入口缓冲区中的出口队列。每个出口端口有八个单播流量队列和八个组播流量队列。根据数据包中的服务类别(CoS)或差分服务代码点(DSCP)值，可将流量分类为流量类，然后在该流量类的相应虚拟队列中排队。

R系列使用分布式信用机制通过交换矩阵传输流量。在将数据包计划离开VOQ之前，入口缓冲区调度程序会请求出口缓冲区中特定端口的信用和优先级。向入口信用调度程序请求目标端口和优先级的信用。如果缓冲区空间可用，则出口调度程序将授予访问权限并将信用授予发送到入口缓冲区调度程序。如果出口缓冲区中没有可用的缓冲区空间，则出口计划不会授予信用，并且流量在VOQ中缓冲，直到下一个信用可用。

以下是 R平台的数据包转发管道。在本文中，您将重点介绍Ingress Traffic Manager组件。有关此链接上架构的详细信息[信息](#)



入口流量管理器(ITM)

入口流量管理器(ITM)是入口管道中的一个块。它执行与将流量排入VOQ、安排流量通过交换矩阵传输以及管理积分相关的步骤。

入口VOQ缓冲区

入口VOQ缓冲区块管理片内缓冲区和片外数据包缓冲区。两个缓冲区都使用VOQ架构，流量根据来自IRPP (入口接收方数据包处理器) 的信息排队。单播和组播流量总共可用96,000个VOQ。

计划和流量控制

在从入口管道传输数据包之前，需要安排数据包通过交换矩阵进行传输。入口调度程序向位于出口流量管理器块中的出口调度程序发送信用请求。当入口流量管理器收到信用时，它开始向入口传输数据包处理器发送流量。如果出口缓冲区已满，流量将缓冲在出口端口和流量类代表的专用队列中。

常见原因

通常，在各种Nexus硬件中，由于以下原因，可以看到输入丢弃

- 流量会阻塞出口接口 (例如10G入口和1G出口)
- 超订用SPAN目标端口 — 适用于特定硬件类型。

适用硬件

		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	
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            = 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
FACE                               |
|
|\|/
|

```

如果QUEUE_DELETED_PACKET_COUNTER大于零，则表示数据包被IQM（入口队列管理器）在队列后删除。这可能是由于活动队列没有收到任何可能表明计划方案配置错误的积分。您可以通过**bcm-shell mod X "getReg IQM_QUEUE_DELETED_PACKET_COUNTER"**检查此项

ENQ_DISCARDED_PACKET_COUNTER表示数据包在入队前被丢弃。您还可以在BCM中看到此计数器集（读取时清除命令）：

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

通过show hardware internal errors module X(读取时命令清除),您始终可以快速注意到以下情况：

```
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
```

步骤3.查找遇到输入丢弃的前面板端口的ASIC和杰里科端口属于哪些：

```
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
```

显示此示例的Eth1/33。在实际网络中，您还不知道拥塞的出口端口。

步骤4.了解入口端口具有的VOQ和VOQ连接器。

```
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 .

此命令显示特定端口的入口VoQ流的详细信息。此外，它还显示了VoQ的当前信用余额。

端口的VOQ的派生方式如下：

LC基于0 — 模块1为0，模块2为1，等等
每个LC有256个系统端口ID

$ID = (LC * \text{系统端口ID}) + \text{FP编号}$

$\text{Eth1/9} = (0 * 256) + 9 = 9$

$\text{VOQ ID} = 32 + (\text{系统端口ID} * 8)$

$\text{Eth1/9} = 32 + (9 * 8) = 104$

因此，Eth1/9的VOQ将为104，与之前收集的输出匹配

```
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
shared buffers occupied	127792 <<<<
granted buffers occupied	3280
shared buffer max occupancy	127792 <<<<

步骤5.从BCM的角度检查哪个队列具体为非空；即拥塞。

```
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>

步骤6.从非空队列值中查找出口拥塞端口：

如果队列是303，请回想一下，这些队列实际上是一个范围，因此可以是303 + 7或303-7 — 问题是，哪个端口的VOQ在296-303范围内匹配，或者是303-310？

众所周知，Eth1/9上的队列7拥塞，因此303实际上是其范围中的最高值，因此296-303的范围是一个有充分根据的猜测。

```
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>

显示asic 0的相同值 — 此处不显示为简单；您会注意到，在Voq列下，您感兴趣的范围不在该ASIC中

请注意上述输出中的几点：

- 出口拥塞端口位于ASIC 1上。
- 出口拥塞端口的VOQ为296,303等于该端口上的队列7。
- 注意Credit Balance列 — 此接口上剩余的可授予的信用非常少，这就是我们的入口Eth1/9开始缓冲的原因。

步骤7.根据您之前的发现，检查ASIC 1中的前面板端口和到杰里科端口9的映射。

```
Nexus-R# show interface hardware-mappings | i i Eth1/9|--|Name|Eth1/33
```

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

此时，您已发现出口拥塞端口 — 确定是否有错误地突入网络，您已配置SPAN，且目标端口为1G，同时采购一个或多个10G接口，或者这是瓶颈/设计问题。

其它命令

这些功能更先进 — 在正常情况下，无需查找出口拥塞端口。


```

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-vsq 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> ] [ ]

```

```

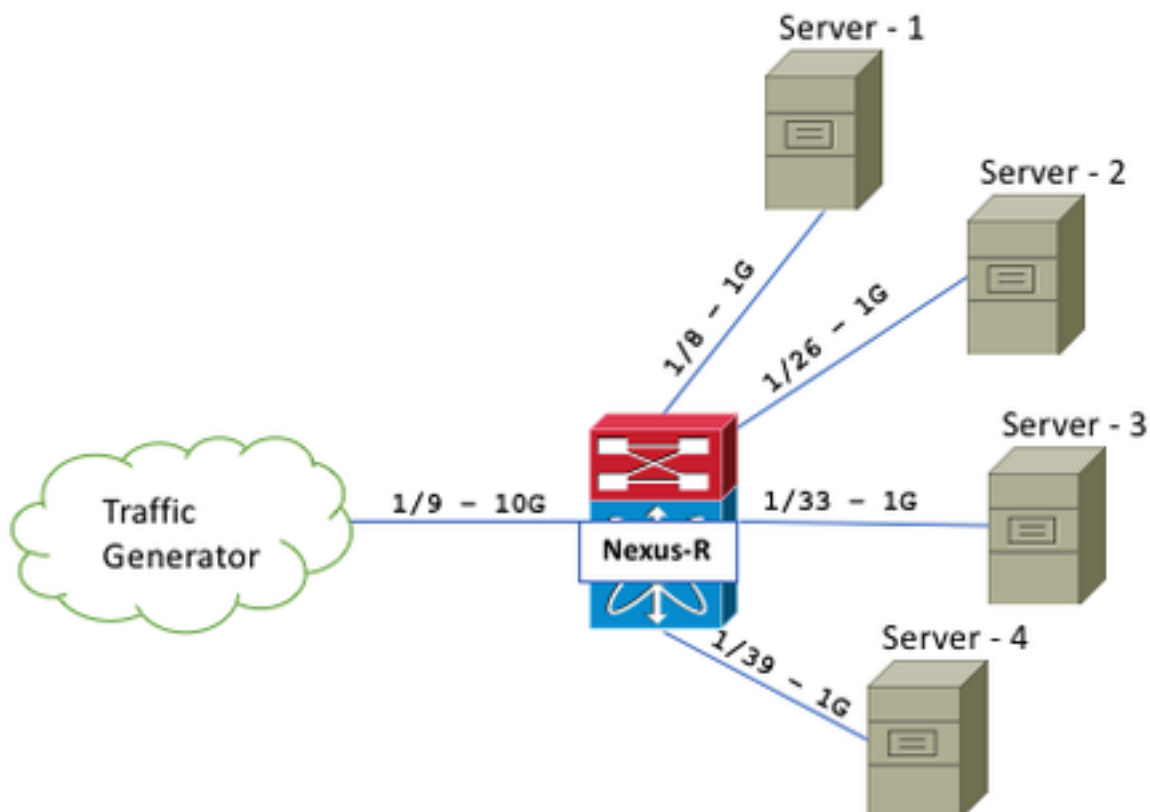
bcm-shell mod X
diag cosq print_flow_and_up dest_id=<flow_id>
diag cosq voq id=<voqid> detailed=1
diag cosq qpair e2e ps=<id>
cosq conn ing
cosq conn egr
dump IPS_CR_BAL_TABLE <voqID>
getReg IQM_QUEUE_MAXIMUM_OCCUPANCY_QUEUE_SIZE

```

其他实验测试：

步骤1.使用多个出口拥塞接口丢弃输入。

请考虑以下拓扑，其中流量生成器向每台服务器发送2G流量：



快速检查哪些队列不为空 — 注意有4个：


```

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

```

步骤4.发送目的IP为未知的数据包。

发送SRC 10.10.10.10和DEST 192.168.10.10的数据包，其中Eth1/9在10.10.10.1/24中，Eth1/33在172.16.0.1/30子网中是L3端口 — 无丢弃计数器，即使目的地在时也不丢弃输入未知。

步骤5.当接入/中继端口转换到STP转发状态时，输入丢弃

发送Eth1/9仅是宽中继（或接入端口）的数据包 — 当端口转换到STP转发状态时，这将注册为输入丢弃。

```

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是由于数据包处理器(PP)决定丢弃或从数据包处理器(PP)块接收的无效目标而导致的丢弃。

步骤6.由于Eth1/9超出线速而导致的输入丢弃。

将10G+发送到Eth1/9将导致不同类型的丢包，因为您首先从Eth1/9最大化 — 仍计为输入丢弃：

```
bcm-shell.0> 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 |
\|/

```

```

|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|
|
| 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 |
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 = 0 |
| RCY_PACKET_COUNTER = 16,837 |
|
| IRE_FDT_INTRFACE_CNT = 0 |
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|
| IDR |
EGQ |
|
| 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 |
| ENQ_DISCARDED_PACKET_COUNTER = 48,716,055 |
DELETED_PKT_CNT = 40 |
| Rejects: VOQ_MX_QSZ_STATUS |
|
<snip>

```