

排除Nexus 5000/6000系列上的FEX性能问题

目录

[简介](#)

[背景信息](#)

[导航CLI](#)

[连接到FEX](#)

[进入调试执行模式](#)

[退出调试执行模式](#)

[退出FEX](#)

[术语](#)

[主机接口\(HI\)](#)

[网络接口\(NI\)](#)

[FEX交换矩阵端口](#)

[FEX ASIC名称](#)

[前端口映射](#)

[N2K-C2148T-1GE](#)

[N2K-C2224TP-1GE/N2K-C2248TP-1GE](#)

[N2K-C2232PP-10GE/N2K-C2232TM-10GE](#)

[N2K-C2248TP-E-1G](#)

[N2K-C2248PQ-10GE和N2K-C2348UPQ-10GE](#)

[检验SFP](#)

[查找丢失](#)

[查看HI端口计数器](#)

[查看NI端口计数器](#)

[查看历史丢包](#)

[查看最近的丢包和中断](#)

[实时查看端口流量速率](#)

[减少损失](#)

[重新定位服务器](#)

[添加其他上行链路](#)

[共享HI缓冲区](#)

[Nexus 6000 FEX负载均衡增强](#)

简介

本文档介绍如何对可连接到Nexus 5000或6000系列交换机的交换矩阵扩展器(FEX)的性能进行故障排除。

注意：本文档中引入的任何命令都不具有破坏性。 您必须将Nexus 2000交换机连接到5000或6000系列交换机。

背景信息

导航CLI

连接到FEX

连接到FEX，在FEX命令行上运行show命令：

```
Nexus# attach fex FEX  
fex>
```

进入调试执行模式

进入FEX的调试模式以运行高级命令并指定FEX ASIC名称。有关FEX ASIC名称，请参阅表1。。

```
fex# dbgexec [prt/woo/red/pri]
```

退出调试执行模式

要退出调试执行模式，请使用CTRL+C键盘序列：

```
fex> [CTRL+C]
```

退出FEX

要退出fex，请使用命令exit:

```
fex# exit
```

术语

主机接口(HI)

Hi是FEX上面对服务器的端口。这些端口通常称为前端口。FEX上的每个前端端口都有HI编号。此编号通常与端口号不同，但用于排除指代端口的命令故障。每个基本设置前端口的表格不同。

网络接口(NI)

NI是FEX上连接回父交换机的FEX控制端口。这些也称为网络上行链路。这些NI编号也取决于模型。

FEX交换矩阵端口

这些端口是到FEX的唯一链路的父交换机端。这些端口配置有switchport mode fex-fabric和fex association命令。

FEX ASIC名称

每个FEX都使用不同的ASIC设计。ASIC名称的缩写是在调试模式下用于运行命令。

FEX的大多数型号都有一个ASIC，但2148有6个，每个都有8个前端端口。这些命令在故障排除命令中称为rmon。

列出ASIC名称和相关的恢复供参考：

表 1.

FEX型号	ASIC名称	放弃
N2K-C2148T-1GE	红木	rw
N2K-C2224TP-1GE	波托拉	端口
N2K-C2248TP-1GE		
N2K-C2232PP-10GE	伍德赛	吴
N2K-C2232TM-10GE		
N2K-C2248TP-E-1GE	普林斯顿	pri
B22	伍德赛	吴
N2K-C2232TM-E-10GE	伍德赛	吴
N2K-C2248PQ-10GE	伍德赛德/贝尔蒙	吴
N2K-C2348UPQ-10GE	蒂布龙	tib

前端口映射

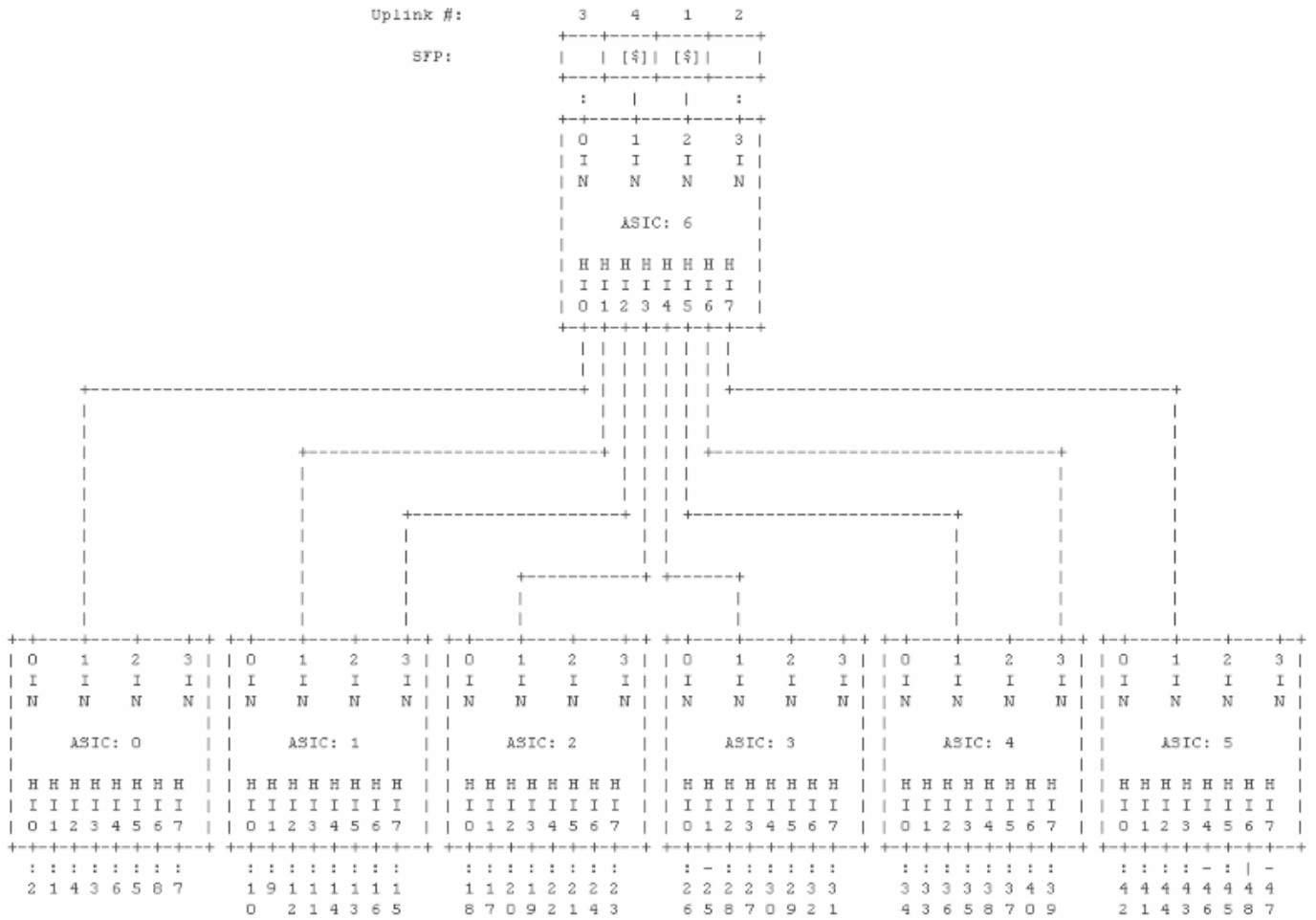
要交换接口计数器输出，可能需要将前端端口号转换为HI号。转换取决于FEX机箱型号。

N2K-C2148T-1GE

在本示例中，已为前端口26（机箱ID/1/26）分配rmon 3 HI 0:

```
switch# attach fex chassis_id
```

```
fex-[chassis]-id# show platform software redwood sts
```



N2K-C2224TP-1GE/N2K-C2248TP-1GE

在本例中，前端口10(135/1/10)已分配HI 9:

```

switch# attach fex chassis_id
fex-[chassis_id]# dbgexec portola
prt>fp

```

```

fex-135# dbgexec prt
prt> fp
Fabric port map:
Fabric port map:
  1   3
  |   :
+---+---+
| NI1 | NIO |
+---+---+
| NI2 | NI3 |
+---+---+
  :   |
  2   4
Front port map:
  1   3   5   7   9  11   13  15  17  19  21  23   25  27  29  31  33  35   37  39  41  43  45  47
  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
HIF | 3 | 7 | 2 | 6 | 11 | 16 | | 10 | 15 | 17 | 20 | 21 | 23 | | 26 | 30 | 27 | 31 | 35 | 39 | | 34 | 38 | 42 | 46 | 43 | 47 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
HIF | 1 | 5 | 0 | 4 | 9 | 13 | | 8 | 12 | 14 | 18 | 19 | 22 | | 24 | 28 | 25 | 29 | 32 | 37 | | 33 | 36 | 40 | 44 | 41 | 45 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
  2   4   6   8  10  12   14  16  18  20  22  24   26  28  30  32  34  36   38  40  42  44  46  48
prt> █

```

N2K-C2232PP-10GE/N2K-C2232TM-10GE

在本例中，前端口20(135/1/20)已分配HI 19:

```
switch# attach fex chassis_id
```

```
fex-[chassis_id]# show platform software woodside sts
```

```
(FINAL POSITION TBD)      Uplink #:      1  2  3  4  5  6  7  8
Link status:             :  :  :  :  :  :  :  |  |
SFP:                     [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ $ ] [ $ ]
                           | N  N  N  N  N  N  N  N  N  |
                           | I  I  I  I  I  I  I  I  I  |
                           | O  1  2  3  4  5  6  7  |
                           |
                           |              NI (0-7)              |
                           +-----+
                           |
                           +-----+-----+-----+-----+
                           |
+-----+-----+-----+-----+-----+-----+
|            HI (0-7)            | |            HI (8-15)            | |            HI (16-23)            | |            HI (24-31)            |
| H  H  H  H  H  H  H  H  | | H  H  H  H  H  H  H  H  | | H  H  H  H  H  H  H  H  | | H  H  H  H  H  H  H  H  |
| I  I  I  I  I  I  I  I  | | I  I  I  I  I  I  I  I  | | I  I  I  I  I  I  I  I  | | I  I  I  I  I  I  I  I  |
| O  1  2  3  4  5  6  7  | | 8  9  1  1  1  1  1  1  | | 1  1  1  1  2  2  2  2  | | 2  2  2  2  2  2  3  3  |
|                               | |                   0  1  2  3  4  5  | | 6  7  8  9  0  1  2  3  | | 4  5  6  7  8  9  0  1  |
+-----+-----+-----+-----+-----+-----+
[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]
+-----+-----+-----+-----+-----+-----+
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
  1  2  3  4  5  6  7  8      9  1  1  1  1  1  1  1      1  1  1  2  2  2  2  2      2  2  2  2  2  2  3  3  3
                               0  1  2  3  4  5  6      7  8  9  0  1  2  3  4      5  6  7  8  9  0  1  2
```

N2K-C2248TP-E-1G

```
fex-111# dbgexec pri
```

```
pri> fp
```

```
Fabric port map:
```

```
Fabric port map:
```

```
  1      3
  |      :
  +-----+
  | NI1 | NI0 |
  +-----+
  | NI2 | NI3 |
  +-----+
```

```
  2      4
```

```
Front port map:
```

```
  1  3  5  7  9  11  13  15  17  19  21  23  25  27  29  31  33  35  37  39  41  43  45  47
```

```
  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
```

```
HIF | 3 | 7 | 2 | 6 | 11 | 16 | 10 | 15 | 17 | 20 | 21 | 23 | 26 | 30 | 27 | 31 | 35 | 39 | 34 | 38 | 42 | 46 | 43 | 47 |
```

```
HIF | 1 | 5 | 0 | 4 | 9 | 13 | 8 | 12 | 14 | 18 | 19 | 22 | 24 | 28 | 25 | 29 | 32 | 37 | 33 | 36 | 40 | 44 | 41 | 45 |
```

```
  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
```

```
  2  4  6  8  10  12  14  16  18  20  22  24  26  28  30  32  34  36  38  40  42  44  46  48
```

N2K-C2248PQ-10GE和N2K-C2348UPQ-10GE

在本例中，HI28映射到前端口29:

tib> fp

-----			NI0,1	NI4,5
+-----+-----+-----+			+-----+	+-----+
1 3 5 7 9 1 1 1	1 1 2 2 2 2 2 3	3 3 3 3 4 4 4 4	1-4	9-12
	1 3 5	7 9 1 3 5 7 9 1		
		3 5 7 9 1 3 5 7		
+-----+-----+-----+			+-----+	+-----+
H	H	H		
I	I	I		
0 2 4 6 8 1 1 1	1 1 2 2 2 2 2 3	3 3 3 3 4 4 4 4		
	0 2 4	6 8 0 2 4 6 8 0		
	2 4 6	2 4 6 8 0 2 4 6		
+-----+-----+-----+				
H	H	H		
I	I	I		
1 3 5 7 9 1 1 1	1 1 2 2 2 2 2 3	3 3 3 3 4 4 4 4		
	1 3 5	7 9 1 3 5 7 9 1		
	7 9 1 3 5 7 9 1	3 5 7 9 1 3 5 7		
+-----+-----+-----+				
2 4 6 8 1 1 1 1	1 2 2 2 2 2 3 3	3 3 3 4 4 4 4 4	+-----+	+-----+
	8 0 2 4 6 8 0 2	4 6 8 0 2 4 6 8	5-8	13-16
	0 2 4 6	8 0 2 4 6 8 0 2		
+-----+-----+-----+			+-----+	+-----+
-----			NI2,3	NI6,7

检验SFP

此命令显示端口的小型可插拔(SFP)信息。

fex# show platform software woodside sfp rmon 0 HI5

在本例中，您看到HI5中的SFP是由CISCO-AVAGO制造的10G-Base-SR(LC):

```

## SFP Info:
  SFP FP-Port      : 0
  Fcot Num        : 0
  Fcot Type       : Not Found
10G-Base-SR      : Yes (Byte 3)
SONET            : No  (Bytes 4-5)
Ethernet         : No  (Byte 6)
FC               : No  (Bytes 7-10)
  SFP Type        : Gb Eth
  Min/Max Speeds  : [4294967295, 4294967295] Mbps

>> BASE ID FIELDS <<
Bytes  Name                Value
-----  ----
0       Identifier          : 0x03 (SFP Transceiver)
1       Ext. Identifier     : 0x04
2       Connector Type     : 0x07 (LC)
3-10    Transceiver        : 0x10 0x00 0x00 0x00 0x00 0x00 0x00 0x00
(4-5)   - SONET ComplCode  : 0x00 0x00 (None)
(6)     - Eth ComplCode    : 0x00 (Reserved)
(7)     - FC LinkLength   : 0x00 (None)
(7-8)   - FC TxType       : 0xFF (None)
(9)     - FC TxMedia      : 0x00 (None)
(10)    - FC Speed        : 0x00 (None)
11      Encoding           : 0x06 (64B/66B)
12      BR, Nominal       : 0x67
13      Reserved          : 0x00
14      Length(9m)-km    : 0x00
15      Length(9m)       : 0x00
16      Length(50m)      : 0x08
17      Length(62.5)     : 0x02
18      Length(Copper)   : 0x00
19      Reserved          : 0x1E
20-35   Vendor Name       : CISCO-AVAGO
36      Reserved          : 0x00
37-39   Vendor OUI        : 0x00 0x17 0x6A (0)
40-55   Vendor PN         : SFBR-7700SDZ
56-59   Vendor Rev       : 0x42 0x34 0x20 0x20 (B4 )
60-62   Reserved         : 0x03 0x52 0x00
63      CC_BASE          : 0x84

```

注意： 如果在使用铜缆端口的FEX上运行此命令，则会注意到命令错误。这是预期的，因为没有要查询的SFP。当该端口为光纤时，**提示符**将返回到未找到SFP，但当前不包含SFP。

查找丢失

在FEX提示符下为HI和NI端口运行Show命令，以便查看FEX交换矩阵端口链路FEX端的接口计数器。

查看HI端口计数器

此命令显示端口计数器验证，类似于show int:

```
fex-128# show platform software woodside rmon 0 HI0
```

TX	Current	Diff	Current	Diff	RX
TX_PKT_LT64	0	0	0	0	RX_PKT_LT64
TX_PKT_64	0	0	0	0	RX_PKT_64
TX_PKT_65	0	0	0	0	RX_PKT_65
TX_PKT_128	0	0	0	0	RX_PKT_128
TX_PKT_256	0	0	0	0	RX_PKT_256

注意：rmon 0仅在FEX具有一个主机ASIC时使用。2224、2248和2232型号只有一个基本设备。2148型号有6个星型，因此将使用rmon 0到5。有关详细信息，请参阅前端口映射部分。

查看NI端口计数器

此命令将显示与show int类似的网络上行链路的端口计数器。此命令显示链路的FEX端。此命令不会显示链路的父交换机端。

```
fex-128# show platform software woodside rmon 0 NI0
```

TX	Current	Diff	Current	Diff	RX
TX_PKT_LT64	0	0	0	0	RX_PKT_LT64
TX_PKT_64	0	0	0	0	RX_PKT_64
TX_PKT_65	0	0	0	0	RX_PKT_65
TX_PKT_128	0	0	0	0	RX_PKT_128
TX_PKT_256	0	0	0	0	RX_PKT_256

查看历史丢包

使用drops命令可以查看历史丢包。这显示FEX上自打开以来的所有丢包。

此命令还显示您丢弃到FEX CPU，该CPU不表示使用DROP8计数器的数据流量丢弃。这些都可以被安全地忽略。

注：尾部丢弃[8]和TAIL_DROP8表示FEX CPU的尾部丢弃，与排除性能故障无关，因为这种情况在正常情况下发生。


```
prt> drops
PRT_SS_CNT_TAIL_DROP1 : 3 SS0
PRT_SS_CNT_TAIL_DROP1 : 6 SS1
PRT_SS_CNT_TAIL_DROP1 : 1 SS2
PRT_SS_CNT_TAIL_DROP1 : 25 SS3
PRT_SS_CNT_TAIL_DROP1 : 2 SS5
PRT_SS_CNT_TAIL_DROP8 : 142 SS0
PRT_SS_CNT_TAIL_DROP8 : 73 SS1
PRT_SS_CNT_TAIL_DROP8 : 11 SS2
PRT_SS_CNT_TAIL_DROP8 : 62048 SS3
PRT_SS_CNT_TAIL_DROP8 : 4613 SS4
PRT_SS_CNT_TAIL_DROP8 : 552 SS5
```

查看最近的丢包和中断

发送到CPU的中断包括尾部丢弃，这些尾部丢弃由于拥塞和缺少缓冲空间而丢弃。可使用show new_ints命令查看：

注意：6.0及更高版本的代码使用show new_ints all

此示例显示SS1缓冲区中的帧尾丢弃：

```
prt> show new_ints
|-----|
| SS1 : ssx_int_norm_td
|-----+
| 1 | 00001c98 | tail drop[1] | frames are being tail dropped.
| 2 | 00005cac | tail drop[2] | frames are being tail dropped.
| 8 | 0000012e | tail drop[8] | frames are being tail dropped.
```

此示例显示NI 3收到符号错误：

```
| NI3 : nix_xe_INT_xg
|-----+
|2 |00000005 | rx_local_fault | Link is in local fault state
|3 |00000007 | rx_remote_fault | Link is in remote fault state
|4 |00000004 | rx_code_violation | MAC received unexpected XGMII control characters.
|5 |00000004 | rx_err_symbol | MAC received an XGMII error character.
|16|00000001 | rx_local_fault_edge | Local fault state has changed.
|17|00000001 | rx_remote_fault_edge | Remote fault state has changed.
|-----|
```

此示例显示FEX尾部丢弃入口NI3的帧：

```
| SS4 : ssx_int_err
|-----+
|0 |00031aa9 | wo_cr[0] | frames rcvd without credit for pausable classes. Pause
is missing.
|1 |00014e21 | wo_cr[1] | frames rcvd without credit for pausable classes. Pause
is missing.
|2 |00018a9f | wo_cr[2] | frames rcvd without credit for pausable classes. Pause
is missing.
|3 |00025efb | wo_cr[3] | frames rcvd without credit for pausable classes. Pause
is missing.
|-----|
```

实时查看端口流量速率

rate命令输出端口的实时流量速率统计信息。与show int不是平均值，它的原始当前数据速率是秒。在本例中，NI 3当前在网络到主机方向接收2.96kbps。相应父Nexus交换机上的show int显示，在连接到NI 3的FEX交换矩阵上行链路上，TX方向为2.96Kbps。

```
prt> rate
```

Port	Tx Packets	Tx Rate (pkts/s)	Tx Bit Rate	Rx Packets	Rx Rate (pkts/s)	Rx Bit Rate	Avg Pkt (Tx)	Avg Pkt (Rx)	Err
O-CI	11	2	4.80Kbps	12	2	8.64Kbps	252	430	
O-NI3	6	1	4.32Kbps	6	1	2.96Kbps	430	289	
O-NI1	6	1	4.32Kbps	5	1	1.89Kbps	430	217	

减少损失

尾部丢弃由缓冲区耗尽引起。通常，当多个服务器同时突发到HIF时，缓冲区会耗尽，或者主机出口缓冲区无法足够快地清空其出站流量以补充NIF上的信用。

有多种选项可用于减轻这种损失。

重新定位服务器

将任何具有突发流量（如存储阵列和视频终端）的服务器从FEX中移出，并直接将其连接到父交换机的基本端口。这将防止突发服务器耗尽缓冲区，并切断来自较不繁忙主机的流量。

Nexus 5000和6000系列交换机的缓冲区比FEX型号大，可将突发服务器连接到基本端口可减少损失，因为基本端口缓冲区能够处理大得多的突发。

添加其他上行链路

当从FEX到父交换机的更多上行链路被添加时，某些型号的FEX可以释放额外的缓冲空间。这可能会停止网络上行链路上的丢包。

表 2.

型号	添加上行链路时缓冲区增加
2148	none
2224	缓冲区增加多达2个上行链路
2248TP	缓冲区增加多达4个上行链路
2232	缓冲区增加多达4个上行链路
2248TP-E	none
2248PQ	none

共享HI缓冲区

大多数型号的FEX都可以通过跨所有主机端口共享HI缓冲区而受益。如果在HI上看到丢包，共享缓冲区可能会缓解这些丢包。

全局修改FEX队列限制：

5k(config)# no fex queue-limit (全局适用于该5k上的所有fex)

修改单个FEX的FEX队列限制：

FEX队列

5k(config)# fex 100

5k(config-fex)#无硬件[型号] queue-limit

Nexus 6000 FEX负载均衡增强

Nexus 6000还有一个附加选项，可将负载均衡算法从HIF更改为NIF。默认情况下，即使数据包到达不同的HIF端口，它们仍可能排入同一NIF的队列。启用上行链路负载均衡模式后，它们将分布在多个NIF上，并允许更基地使用NIF出口缓冲区。

6k(config)#硬件N2248PQ上行链路负载均衡模式