Nexus 5000/6000シリーズのFEXパフォーマンス問題のトラブルシューティング

内容

概要

このドキュメントでは、Nexus 5000 もしくは 6000 シリーズ スイッチにアタッチするファブリック エクステンダ (FEX) のパフォーマンスをトラブルシューティングする方法について説明します。

注：このドキュメントで紹介されているコマンドはいずれも中断を伴うものではありません

。 Nexus 2000 スイッチを 5000 もしくは 6000 シリーズのスイッチに接続する必要があります。

背景説明

CLIの操作

FEXへの接続

FEX コマンド ラインで show コマンドを実行するため、FEX にアタッチします。

Nexus# attach fex *fex*
fex>

デバッグEXECモードに入る

FEXでデバッグモードに入り、高度なコマンドを実行してFEX ASIC名を指定します。FEXのASIC名については、表1を参照してください。

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

デバッグEXECモードの終了

デバッグEXECモードを終了するには、CTRL+Cキーボードシーケンスを使用します。

fex> [CTRL+C]

FEXの終了

FEXを終了するには、exitコマンドを使用します。

fex# exit

用語

ホスト インターフェイス (HI)

FEX上のサーバに面するポートは、一般に前面ポートと呼ばれます。FEX上のすべての前面ポートにはHI番号があります。通常、この番号はポート番号とは異なりますが、ポートを参照するコマンドのトラブルシューティングに使用されます。ASIC によって前面ポートの配置は異なります。

ネットワーク インターフェイス (NI)

NIは、親スイッチに接続するFEXのFEX制御ポートです。これらはネットワーク アップリンクと

も呼ばれます。これらにはモデルに依存する固有の NI 番号もあります。

FEX ファブリック ポート

これらのポートは、FEXへの固有リンクの親スイッチ側です。これらのポートは、**switchport mode fex-fabric**コマンドと**fex association**コマンドを使用して設定されます。

FEX ASIC名

各FEXは異なるASICで設計されています。ASIC名の省略形は、コマンドを実行するためにデバッグモードで使用されます。

FEXのほとんどのモデルには1つのASICがありますが、2148には6個のASICがあり、それぞれに8個の前面ポートがあります。これらは、トラブルシューティングコマンドでは**rmon**と呼ばれます。

ASIC名と関連付けられた略語が参照用にリストされています。

表 1.

FEX モデル	ASIC 名	略語
N2K-C2148T-1GE	redwood	rw
N2K-C2224TP-1GE N2K-C2248TP-1GE	portola	prt
N2K-C2232PP-10GE N2K-C2232TM-10GE	woodside	woo
N2K-C2248TP-E-1GE	princeton	pri
B22	woodside	woo
N2K-C2232TM-E-10GE	woodside	woo
N2K-C2248PQ-10GE	woodside/belmont	woo
N2K-C2348UPQ-10GE	tiburon	tib

前面ポート マッピング

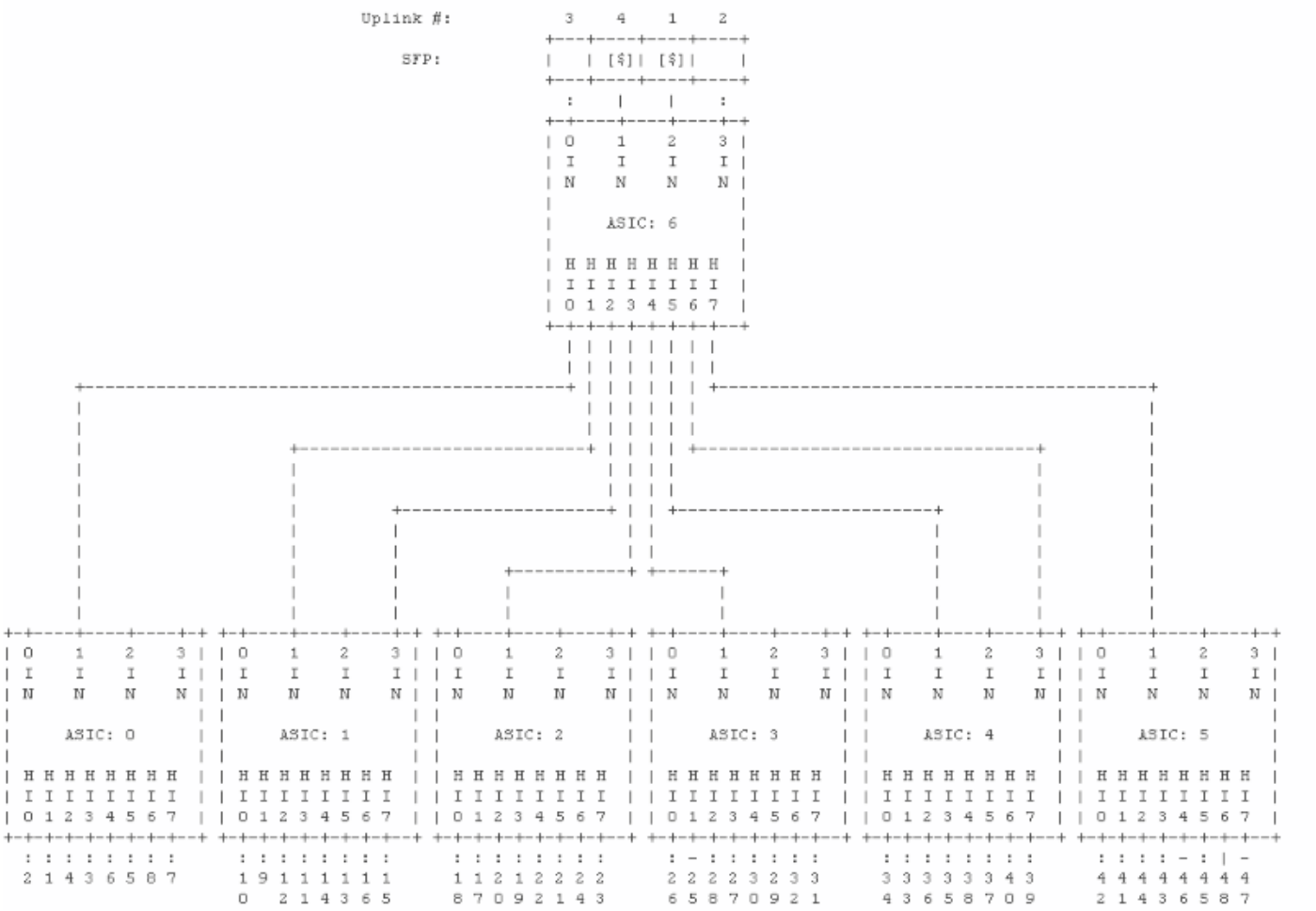
インターフェイスカウンタの出力を相互接続するには、前面ポート番号をHI番号に変換する必要があります。変換は FEX シャーシ モデルに依存します。

N2K-C2148T-1GE

この例では、前面ポート26（シャーシID/1/26）にrmon 3 HI 0が割り当てられています。

switch# attach fex *chassis_id*

fex-[chassis_id]# show platform software redwood sts



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

この例では、前面ポート10(135/1/10)にHI 9が割り当てられています。

switch# attach fex chassis_id

fex-[chassis_id]# dbgexec portola

prt> fp

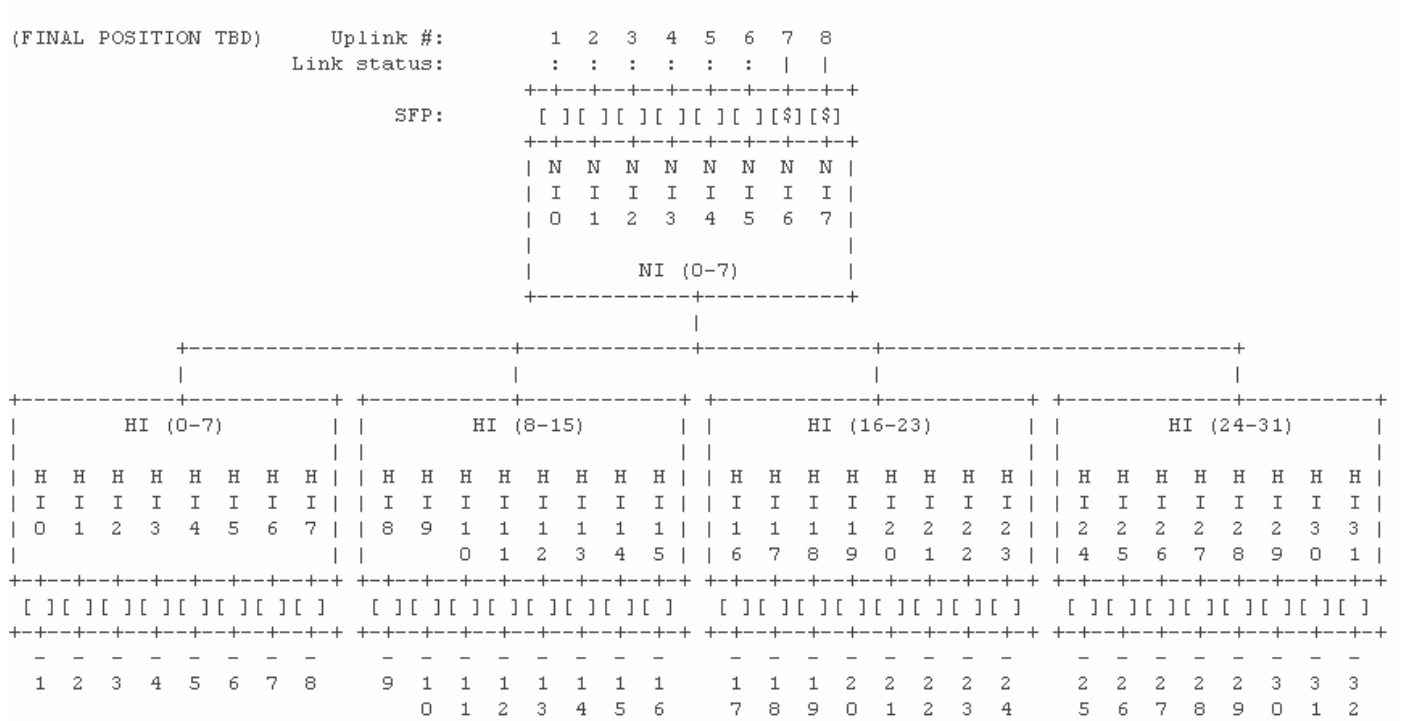
```
fex-135# dbgexec prt
prt> fp
Fabric port map:
Fabric port map:
       1     3
       |     :
    +-----+-----+
    | NI1 | NI0 |
    +-----+-----+
    | NI2 | NI3 |
    +-----+-----+
       :     |
       2     4
Front port map:
      1   3   5   7   9  11    13  15  17  19  21  23    25  27  29  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
prt>
```

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

この例では、前面ポート20(135/1/20)にHI 19が割り当てられています。

switch# attach fex *chassis_id*

fex-[chassis_id]# show platform software woodside sts

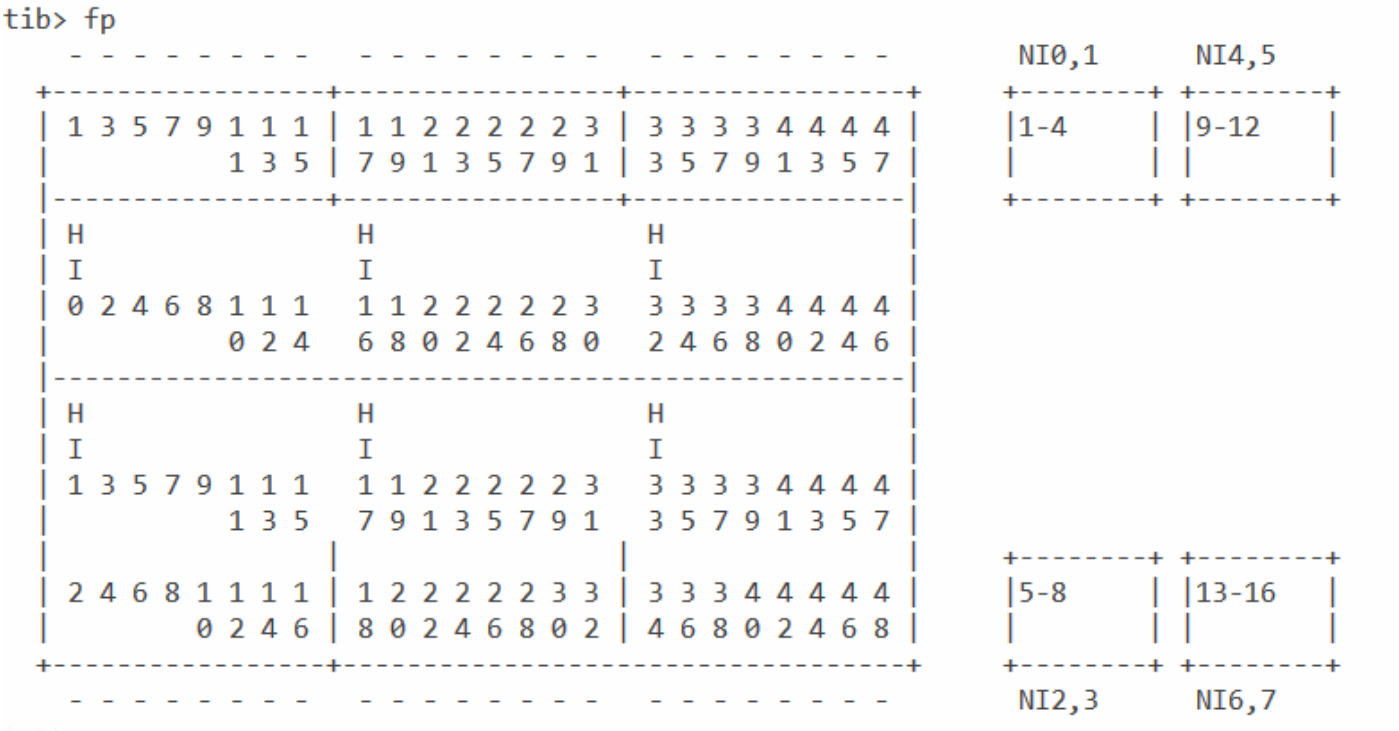


N2K-C2248TP-E-1G

```
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 & N2K-C2348UPQ-10GE

この例では、HI28が前面ポート29にマッピングされています。



SFPの確認

このコマンドは、ポートの着脱可能小型フォームファクタ(SFP)情報を表示します。

fex# show platform software woodside sfp rmon 0 HI5

この例では、HI5のSFPがCISCO-AVAGOによって作成された10G-Base-SR(LC)であることがわかります。

```
## 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： 銅線ポートを使用するFEXでこのコマンドを実行すると、コマンドエラーが表示されます。クエリー対象のSFPがないため、これは想定されています。ポートがファイバの場合はプロンプトが**no SFP found**に戻りますが、現在SFPは含まれていません。

損失の検索

showコマンドは、FEXファブリックポートリンクのFEX側のインターフェイスカウンタを表示するために、HIおよびNIポートのFEXプロンプトで実行できます。

HIポートカウンタの表示

次のコマンドは、**show int**と同様に、ポートカウンタの検証を**表示します。**

```
fex-128# show platform software woodside rmon 0 HI0
+---------------------+---------------------+----------------+-------------------------+----
-----------------+----------------+
| TX                        |         Current        |       Diff          |           RX
|       Current          |        Diff         |
+---------------------+---------------------+----------------+-------------------------+----
-----------------+----------------+
| TX_PKT_LT64     |                           0|                     0|    RX_PKT_LT64     |
0|                     0|
| TX_PKT_64         |                             0|                      0|    RX_PKT_64
|                           0|                     0|
| TX_PKT_65         |                             0|                      0|    RX_PKT_65
|                           0|                     0|
| TX_PKT_128       |                            0|                      0|    RX_PKT_128        |
0|                     0|
| TX_PKT_256       |                            0|                      0|    RX_PKT_256        |
0|                     0|
```

注：rmon 0は、FEXに1つのホストASICがある場合にのみ使用されます。2224、2248 および 2232 モデルには ASIC が 1 つだけ搭載されています。2148モデルには6つのasicがあるため、rmon 0 ~ 5が使用されます。詳細は、「前面ポートマッピング」セクションを参照してください。

NIポートカウンタの表示

このコマンドは、**show int**に似たネットワークアップリンクのポートカウンタを表示**します**。このコマンドは、リンクのFEX側を表示します。このコマンドでは、リンクの親スイッチ側は表示されません。

```
 fex-128# show platform software woodside rmon 0 NI0
+---------------------+---------------------+----------------+-------------------------+----
-----------------+----------------+
| TX                        |         Current        |       Diff          |           RX
|       Current          |        Diff         |
+---------------------+---------------------+----------------+-------------------------+----
-----------------+----------------+
| TX_PKT_LT64     |                           0|                     0|    RX_PKT_LT64     |
0|                     0|
| TX_PKT_64         |                             0|                      0|    RX_PKT_64
|                           0|                     0|
| TX_PKT_65         |                             0|                      0|    RX_PKT_65
|                           0|                     0|
| TX_PKT_128       |                            0|                      0|    RX_PKT_128        |
0|                     0|
| TX_PKT_256       |                            0|                      0|    RX_PKT_256        |
0|                     0|
```

ドロップ履歴の表示

ドロップ履歴は、dropsコマンドで表示**できます**。これにより、FEXがオンになってからのドロップがすべて表示されます。

このコマンドでは、DROP8カウンタを使用したデータトラフィックのドロップを表さないFEX CPUへのドロップも表示されます。これらは無視しても問題ありません。

注: tail drop [8]およびTAIL_DROP8は、FEX CPUへのテールドロップを表し、通常の条件下で発生する場合のパフォーマンスのトラブルシューティングには関係ありません。

```
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
```

最近のドロップと割り込みの表示

CPU に送信される割り込みには、輻輳やバッファ スペースの不足によるテール ドロップがあります。これらはshow new_intsコマンドで表示できます。

注: 6.0以降のコードではshow new_ints all

次の例は、フレームのテールドロップがSS1バッファにあることを示しています。

```
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.
```

次の例は、NI 3がシンボルエラーを受信していることを示しています。

```
 | 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.
|-----------------------------------------------|
```

次の例は、FEXテールがNI3に入るフレームをドロップすることを示しています。

```
 | 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.
|-----------------------------------------------|
```

リアルタイムでのポートトラフィックレートの表示

rate コマンドは、ポートのリアルタイムのトラフィック レートの統計情報を出力します。show intとは異なり、**平均ではな**く、現在の生のデータレートは2秒です。この例では、NI 3は現在、ネットワークからホストへの方向で2.96 kbpsを受信しています。対応する**親Nexusスイッチ**のshow intは、NI 3に接続されたFEXファブリックアップリンクのTX方向で2.96 Kbpsを示します。

```
prt> rate
+--------++------------+-----------+------------++------------+-----------+------------+-------+-------+---+
| Port   || Tx Packets |  Tx Rate  |   Tx Bit   || Rx Packets |  Rx Rate  |   Rx Bit   |Avg Pkt|Avg Pkt|   |
|        ||            |  (pkts/s) |    Rate    ||            |  (pkts/s) |    Rate    | (Tx)  | (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 |   |
+--------++------------+-----------+------------++------------+-----------+------------+-------+-------+---+
```

損失の軽減

テール ドロップはバッファ不足によって発生します。通常、複数のサーバが一度にHIFにバーストした場合、またはホストの出力バッファがNIFのクレジットを補充するのに十分な速さでアウトバウンドトラフィックを空にできない場合、バッファが枯渇します。

この損失を軽減するには、複数の選択肢があります。

サーバの再配置

ストレージアレイやビデオエンドポイントなどのバーストトラフィックフローを備えたサーバをFEXから移動し、親スイッチのベースポートに直接接続します。これはバーストの多いサーバがバッファを使い果たすことや、トラフィック フローの少ないホストから帯域を奪ってしまうことを防ぎます。

Nexus 5000および6000シリーズスイッチは、FEXモデルよりも大きなバッファを備えており、バーストサーバをベースポートに接続することで、ベースポートバッファが処理するバーストが大幅に増加するため、損失が軽減されます。

アップリンクの追加

FEX の一部のモデルでは、FEX から親スイッチへのアップリンクが追加されたときに、追加のバッファ スペースをアンロックできます。これにより、ネットワークアップリンクでのドロップが停止する可能性があります。

表 2

モデル	アップリンク追加時のバッファの増加
2148	none
2224	最大 2 個のアップリンクまでバッファ増加
2248TP	最大 4 個のアップリンクまでバッファ増加

2232　　最大 4 個のアップリンクまでバッファ増加
2248TP-E none
2248PQ　none

HIバッファの共有

FEXのほとんどのモデルは、すべてのホストポートでHIバッファを共有することでメリットを得ることができます。HIでドロップが見られる場合は、バッファを共有することでドロップを軽減できます。

FEXキュー制限をグローバルに変更します。

5k(config)# no fex queue-limit（その5k上のすべてのfexにグローバルに適用）

個々のFEXのFEXキュー制限を変更します。

FEX キュー

5k(config)# fex 100
5k(config-fex)# no hardware [model] queue-limit

Nexus 6000 FEXのロードバランシング強化

Nexus 6000 には、ロード バランシング アルゴリズムを HIF から NIF に変更する追加オプションがあります。デフォルトでは、パケットが異なるHIFポートに到着しても、同じNIFにキューイングされる可能性があります。uplink-load-balance-modeが有効な場合、複数のNIFに分散され、NIF出力バッファをより均等に使用できるようになります。

6k(config)# hardware N2248PQ uplink-load-balance-mode