IB Fabric congestion Analysis v6.1

**HyungKwang Choi (2022.5)**

Table of Contents

[**1. Introduction** 3](#_Toc103066115)

[**1.1 Revision History** 3](#_Toc103066116)

[**1.2 Glossary** 3](#_Toc103066117)

[**1.3 Reference** 3](#_Toc103066118)

[**2. Script running** 4](#_Toc103066119)

[**3. Before running.** 6](#_Toc103066120)

[**3.1 What it does** 6](#_Toc103066121)

[**3.2 Usage.** 6](#_Toc103066122)

[**3.3 Python installation.** 6](#_Toc103066123)

[**4. Fabric Congestion key factors** 8](#_Toc103066124)

[**4.1 LinkDownCounter > 0** 8](#_Toc103066125)

[**4.2 PortXmitDiscards > 0** 8](#_Toc103066126)

[**4.3 port\_fec\_uncorrectable\_block\_counter > 0** 9](#_Toc103066127)

[**4.4 excessive\_buffer\_errors > 0** 10](#_Toc103066128)

[**4.5 CongestionIndex >= 10** 10](#_Toc103066129)

[**4.6 BER (Symbol/Effective)** 11](#_Toc103066130)

[**4.6 PortRcvErrors** 13](#_Toc103066131)

[**5. PM tips for switches “IB/Ethernet/SX6036” from sysdump** 14](#_Toc103066132)

[**5.1 From : Ethernet/IB switch** 14](#_Toc103066133)

[**5.1 From : SX6036** 15](#_Toc103066134)

[**6. Causing Fabric degradation** 20](#_Toc103066135)

[**6.1 Link speed downgrade** 20](#_Toc103066136)

# **1. Introduction**

## **1.1** **Revision History**

Table 1 tracks the revision history for this specification.

Table : Revision History

| Revision | Date | Author | Comments |
| --- | --- | --- | --- |
| 1.0 | Sep 28th, 2021 | Jie Wu | Initial Draft |
| 2.0 | Mar 15th, 2022 | HyungKwang Choi | Fixed script bug (FEC Uncorrectable) |
| 3.0 | Apr 14th, 2022 | HyungKwang Choi | Fixed script bug (Symbol BER/Err) |
| 4.0 | Apr 30th, 2022 | HyungKwang Choi | Modifying BER threshold from ‘1e-12’ to ‘1e-13’ |
| 5.0 | May 1th, 2022 | HyungKwang Choi | Added "Max Retransmission\_rate" |
| 5.1 | May 11th, 2022 | HyungKwang Choi | modifed "Running command" parsing code |
| 6.1 | May 13th, 2022 | HyungKwang Choi & Sam | In the Ibdiagnet "2.8.1", ibdiagnet2\_db.csv files , data written like "-1". Which triggers an exception |

## **1.2 Glossary**

Table 2 lists the terms and acronyms used in this Document.

Table 2: Terminology and Acronyms

| Term | Definition |
| --- | --- |
|  |  |
|  |  |

## **1.3 Reference**

Table 3 lists the reference referred at this Document.

Table 3: Reference

| Subject | Site |
| --- | --- |
| BER (Bit Error Rate) and Troubleshoot v1.0.pdf | My previous knowledge sharing for BER |
|  |  |
|  |  |

# **2. Script running**

**$ python check\_ib\_link\_errors\_v5.py -h**

usage: show\_ib\_inventory.py [-h] -i IBDIAGNET\_FOLDER [-o OUTPUT\_FILE] [-f] [-s] [-v]

optional arguments:

  -h, --help            show this help message and exit

  -i IBDIAGNET\_FOLDER, --ibdiagnet-folder IBDIAGNET\_FOLDER

                        read data from ibdiagnet2 output folder

  -o OUTPUT\_FILE, --output-file OUTPUT\_FILE

                        write IB SW and HCA inventory info to xlsx file

  -f, --show-ib-hosts   print detailed HCA inventory info

  -s, --show-ib-switches

                        print detailed IB switch inventory info

  -v, --version         print current script version

**$ python check\_ib\_link\_errors\_v5.py -i /var/tmp/ibdiagnet2 -o Saving\_data**

Running command: ibdiagnet --extended\_speeds all --pm\_per\_lane --get\_phy\_info --get\_cable\_info --pc --reset\_phy\_info -o ./ibdiagnet\_1st\_reset\_phy\_info/

Start to parse ibdiagnet2.net\_dump ...

Start to parse ibdiagnet2.db\_csv ...

Start to parse ibdiagnet2.net\_dump\_ext ...

INFO: The --top-n option is not set, only the first 10 records will be listed here.

LinkDowned Counters:

##################################################

                          SrcDevice SrcPort            SrcGUID  LinkDownedCounter                                DstDevice            DstGUID DstPort

                           SIB30183      74 0x0c42a103000ef7c6                  4 MT4123 ConnectX6   Mellanox Technologies 0x0c42a10300389dc8       1

                           SIB30237      37 0x0c42a103000ef206                  1                             n08509 HCA-1 0x0c42a10300407088       1

MF0;G1-IB-CORE-SW05A:MCS8500/S02/U1      15 0xb8599f0300f7d9e6                  1      MF0;G1-IB-CORE-SW05A:MCS8500/L15/U1 0x0c42a103000fd750      22

XmitDidscard Counters:

##################################################

                          SrcDevice SrcPort            SrcGUID  PortXmitDiscards                                DstDevice            DstGUID DstPort

                       n08509 HCA-1       1 0x0c42a10300407088               290                                 SIB30237 0x0c42a103000ef206      37

                           SIB30237      37 0x0c42a103000ef206               192                             n08509 HCA-1 0x0c42a10300407088       1

                           SIB30183      74 0x0c42a103000ef7c6                87 MT4123 ConnectX6   Mellanox Technologies 0x0c42a10300389dc8       1

MF0;G1-IB-CORE-SW05A:MCS8500/L15/U1      22 0x0c42a103000fd750                 4      MF0;G1-IB-CORE-SW05A:MCS8500/S02/U1 0xb8599f0300f7d9e6      15

Port FEC Uncorrectable Counters:

##################################################

                          SrcDevice SrcPort            SrcGUID  PortFECUncorrectableBlockCounter                          DstDevice            DstGUID DstPort

 MF0;G2-IB-CORE-SW6A:MCS8500/S13/U1      12 0x0c42a103000501aa                           4608243 MF0;G2-IB-CORE-SW6A:MCS8500/L12/U1 0x0c42a103000fdc30      33

MF0;G1-IB-CORE-SW04A:MCS8500/L03/U2       4 0x0c42a1030017d37c                           2402008                           SIB30064 0x0c42a103000eeca6      19

MF0;G1-IB-CORE-SW05A:MCS8500/L12/U1       2 0x0c42a103000fdbd0                           1558267                           SIB30022 0x0c42a103000be30e      16

INFO: ExcessiveBufferOverrunErrors counters are 0 on all links.

Congestion Indexes > 10:

##################################################

   SrcDevice SrcPort            SrcGUID  PortXmitWaitExt  PortXmitPktsExtended  CongestionIndexExt DstDevice            DstGUID DstPort

n00110 HCA-1       1 0x0c42a10300619450        561153409                     1         561153409.0  SIB30003 0x0c42a103001268d0      62

n00235 HCA-1       1 0x0c42a10300330432        557213155                     1         557213155.0  SIB30008 0x0c42a103000be12e       9

n00225 HCA-1       1 0x0c42a1030036bb58        463507299                     1         463507299.0  SIB30008 0x0c42a103000be12e      65

Effective BER Counters > 1e-12:

##################################################

                          SrcDevice SrcPort            SrcGUID  EffectiveBER     SymbolBER                          DstDevice            DstGUID DstPort

 MF0;G2-IB-CORE-SW6A:MCS8500/S13/U1      12 0x0c42a103000501aa  5.000000e-08 1.500000e-254 MF0;G2-IB-CORE-SW6A:MCS8500/L12/U1 0x0c42a103000fdc30      33

MF0;G1-IB-CORE-SW04A:MCS8500/L10/U2      15 0x0c42a103000fd200  3.000000e-08 1.500000e-254                           SIB30095 0x0c42a103000ef0a6      20

MF0;G1-IB-CORE-SW04A:MCS8500/L03/U2       4 0x0c42a1030017d37c  2.000000e-08  1.000000e-12                           SIB30064 0x0c42a103000eeca6      19

Symbol BER Counters > 1e-13:

##################################################

                          SrcDevice SrcPort            SrcGUID EffectiveBER    SymbolBER DstDevice            DstGUID DstPort

MF0;G1-IB-CORE-SW04A:MCS8500/L03/U2       4 0x0c42a1030017d37c 2.000000e-08 1.000000e-12  SIB30064 0x0c42a103000eeca6      19

# **3. Before running.**

## **3.1 What it does**

The script is used to collect & sort up which causes Fabric Congestion from ibdiagnet.

Parsing 4 files (ibdiagnet2.log, ibdiagnet2.net\_dump, ibdiagnet2.db\_csv,ibdiagnet2.net\_dump\_ext) from Ibdiagnet, then extract PM info which causes Fabric Congestion.

LinkDownedCounter > 0

PortXmitDiscards > 0

ExcessiveBufferOverrunErrors > 0

PortFECUncorrectableBlockCounter > 0

Effective BER > 1e-12

Symbol BER > 1e-13 or (Symbol Err > 0)

congestion indexes = PortXmitWaitExt / PortXmitPktsExtended > 10

## **3.2 Usage.**

First command, It clears all Fabric PM stats and port Phy info.

The second, it collects the count accumulated.

ibdiagnet -r --pc --pm\_pause\_time 600 -P all=1 --extended\_speeds all --pm\_per\_lane --reset\_phy\_info --get\_phy\_info --get\_cable\_info

ibdiagnet -r -P all=1 --extended\_speeds all --pm\_per\_lane --get\_phy\_info --get\_cable\_info

## **3.3 Python installation.**

[root@mtbc-r740-06 ~]# python3 --version

Python 3.6.8

Prerequisites:

Python 3.8+

Pandas library

OpenPyXL library

[root@mtbc-r740-06 ~]# pip3 install pandas

[root@mtbc-r740-06 ~]# pip3 install openpyxl

[root@mtbc-r740-06 ~]# pip3 list

DEPRECATION: The default format will switch to columns in the future. You can use --format=(legacy|columns) (or define a format=(legacy|columns) in your pip.conf under the [list] section) to disable this warning.

backports.entry-points-selectable (1.1.1)

certifi (2021.10.8)

chardet (3.0.4)

charset-normalizer (2.0.10)

distlib (0.3.4)

filelock (3.4.0)

idna (2.10)

importlib-metadata (4.8.2)

importlib-resources (5.4.0)

numpy (1.19.5)

pandas (1.1.5) <===========

pip (9.0.3)

platformdirs (2.4.0)

pynetbox (6.5.0)

PySocks (1.6.8)

python-dateutil (2.8.2)

pytz (2021.3)

requests (2.27.1)

setuptools (39.2.0)

six (1.16.0)

typing-extensions (4.0.1)

urllib3 (1.25.6)

virtualenv (20.10.0)

zipp (3.6.0)

# **4. Fabric Congestion key factors**

## **4.1 LinkDownCounter > 0**

* File location & Command

./ibdiagnet2.pm:link\_down\_counter=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0 🡸================

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................0

## **4.2 PortXmitDiscards > 0**

* File location & Command

./ibdiagnet2.pm:port\_xmit\_discard=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0 🡸================

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................0

ibportstate 5 79 reset

ibportstate 5 79 query

ibportstate 5 79 disable

ibportstate 5 79 enable

## **4.3 port\_fec\_uncorrectable\_block\_counter > 0**

* File location & Command

./ibdiagnet2.pm:port\_fec\_uncorrectable\_block\_counter=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery -T 5 13

# PortExtendedSpeedsCounters counters: Lid 5 port 13

PortSelect:......................13

CounterSelect:...................0x0000000000000000

SyncHeaderErrorCounter:..........0

UnknownBlockCounter:.............0

ErrorDetectionCounterLane0:......0

ErrorDetectionCounterLane1:......0

ErrorDetectionCounterLane2:......0

ErrorDetectionCounterLane3:......0

ErrorDetectionCounterLane4:......0

ErrorDetectionCounterLane5:......0

ErrorDetectionCounterLane6:......0

ErrorDetectionCounterLane7:......0

ErrorDetectionCounterLane8:......0

ErrorDetectionCounterLane9:......0

ErrorDetectionCounterLane10:.....0

ErrorDetectionCounterLane11:.....0

FECCorrectableBlockCtrLane0:.....0

FECCorrectableBlockCtrLane1:.....0

FECCorrectableBlockCtrLane2:.....0

FECCorrectableBlockCtrLane3:.....0

FECCorrectableBlockCtrLane4:.....0

FECCorrectableBlockCtrLane5:.....0

FECCorrectableBlockCtrLane6:.....0

FECCorrectableBlockCtrLane7:.....0

FECCorrectableBlockCtrLane8:.....0

FECCorrectableBlockCtrLane9:.....0

FECCorrectableBlockCtrLane10:....0

FECCorrectableBlockCtrLane11:....0

FECUncorrectableBlockCtrLane0:...0

FECUncorrectableBlockCtrLane1:...0

FECUncorrectableBlockCtrLane2:...0

FECUncorrectableBlockCtrLane3:...0

FECUncorrectableBlockCtrLane4:...0

FECUncorrectableBlockCtrLane5:...0

FECUncorrectableBlockCtrLane6:...0

FECUncorrectableBlockCtrLane7:...0

FECUncorrectableBlockCtrLane8:...0

FECUncorrectableBlockCtrLane9:...0

FECUncorrectableBlockCtrLane10:..0

FECUncorrectableBlockCtrLane11:..0

# mlxlink -d lid-3 -pc 36

# mlxlink -d lid-3 -m -c -e -p 36 --show\_device --show\_serdes\_tx --show\_fec --show\_ber\_monitor

* Tips

 > to get ‘fec\_uncorrectable\_block\_counter’, you have to add below options while running ibdiagnet

# --extended\_speeds all

> below extra fields are collected when you add the options “--extended\_speeds all”

sync\_header\_error\_counter=0x00000000

unknown\_block\_counter=0x00000000

fec\_corrected\_symbol\_counter\_total=0x0000000000000000

fec\_corrected\_symbol\_counter\_lane[0]=0x00000000

fec\_corrected\_symbol\_counter\_lane[1]=0x00000000

fec\_corrected\_symbol\_counter\_lane[2]=0x00000000

fec\_corrected\_symbol\_counter\_lane[3]=0x00000000

port\_fec\_correctable\_block\_counter=0x00008281

port\_fec\_uncorrectable\_block\_counter=0x00000000

port\_fec\_corrected\_symbol\_counter=0x00008292

> Clear FEC counters

#perfquery -x -R 0x20 1 (lid 20)

> with ‘Perfquery’ command, you will see  ‘FECUncorrectableBlockCtrLaneXXXX’. Those are sum of ..

FECCorrectableBlockCtrLaneX/FECUncorrectableBlockCtrLaneX vs. port\_fec\_correctable\_block\_counter / port\_fec\_uncorrectable\_block\_counter

## **4.4 excessive\_buffer\_errors > 0**

* File location & Command

./ibdiagnet2.pm: port\_fec\_uncorrectable\_block\_counter=0x0000044c

./ibdiagnet2.db\_csv

## **4.5 CongestionIndex >= 10**

* File location & Command

./ibdiagnet2.pm: port\_xmit\_wait, port\_xmit\_pkts

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................0

* Explanation on fabric ‘Congestion Indexes’

[PortXmitWait and Symbol Errors | Salesforce](https://mellanox.lightning.force.com/lightning/r/Knowledge_Article__kav/ka21T000000bpWSQAY/view?ws=%2Flightning%2Fr%2FCase%2F50050000016gclzAAA%2Fview)

**PortXmitWait** - This is not an error counter. It is a very high-resolution counter that is incremented every internal clock tick of the device whenever there is a packet queued on an output port, and this packet cannot be sent. This can happen whenever there is some temporary congestion in the fabric. For instance, if two nodes are sending to one node you can see this counter increment, or if there is a node mismatch in rates you can see this counter increment. When the counter increments it does not mean that any packets dropped. Packets are just stored in the buffer of the port until they can be sent. Therefore, it is natural to see these PortXmitWait counters increment under normal conditions.

* How to math PortXmitWait

CongestionIndex >= 10

   => PortXmitWaitExt/PortXmitPktsExtended

* How to troubleshoot

xmit\_waits in the cluster, but it seems like that may have been resolved by switch reboots and a switch replacement.

## **4.6 BER (Symbol/Effective)**

* File location & Command

ibdiagnet2.net\_dump\_ext

- Effective BER Counters > 1e-12, Symbol BER Counters > 1e-13 , Symbol Err > 0

ibdiagnet2.db\_csv

\*\* For details, please refer to my previous knowledge sharing (BER (Bit Error Rate) and Troubleshoot v1.0.pdf)

**> on a switch.**

TestLab [standalone: master] # show interfaces ib 10/1/14

IB15/1/4 state:

Logical port state : Active

Physical port state : LinkUp

Current line rate : 200.0 Gbps

Supported speeds : sdr, qdr, fdr, edr, hdr

Speed : hdr

Supported widths : 1X, 2X, 4X

Width : 4X

Max supported MTUs : 4096

MTU : 4096

VL capabilities : VL0 - VL7

Operational VLs : VL0 - VL3

Description :

IB Subnet : infiniband-default

Phy-profile : high-speed-ber

Width reduction mode : Not supported

Telemetry sampling : Disabled

Telemetry threshold : Disabled

Telemetry record : Disabled

Telemetry threshold level: N/A bytes

RX:

Bytes : 3480

Packets : 13

Errors : 7

Symbol errors : 735 <====

VL15 dropped packets: 0

TX:

Bytes : 11232

Packets : 39

Wait : 0

Discarded packets: 0

TestLab [standalone: master] # show interfaces ib 10/1/14 link-diagnostics

-----------------------------------------------------------------------

Interface Code Status

-----------------------------------------------------------------------

IB10/1/14 15 Bad signal integrity

[root@NVIDIA]# perfquery 199 14

# Port counters: Lid 199 port 14 (CapMask: 0x5300)

PortSelect:......................14

CounterSelect:...................0x0000

SymbolErrorCounter:..............38 <=======

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

[root@NVIDIA]# mlxlink -d lid-199 -p 14 -c -e -m

Operational Info

----------------

State : Active

Physical state : LinkUp

Speed : IB-HDR

Width : 4x

FEC : LL-FEC (271,257) + PLR

Loopback Mode : No Loopback

Auto Negotiation : ON

Supported Info

--------------

Enabled Link Speed : 0x00000061 (HDR,EDR,SDR)

Supported Cable Speed : 0x00000061 (HDR,EDR,SDR)

Troubleshooting Info

--------------------

Status Opcode : 15

Group Opcode : PHY FW

Recommendation : Bad signal integrity. <======

Physical Counters and BER Info

------------------------------

Time Since Last Clear [Min] : 143.1 <=== important to math BER rate

Effective Physical Errors : 122 <=========== This used for symbol error calculation

Raw Physical Errors Per Lane : 73506956266,43017712645,42717472449,58570852076

Effective Physical BER : 7E-8

Raw Physical BER : 1E-4

Link Down Counter : 0

Link Error Recovery Counter : 0

EYE Opening Info

----------------

Physical Grade : 1894, 2031, 2412, 2097

Height Eye Opening [mV] : N/A, N/A, N/A, N/A

Phase Eye Opening [psec] : N/A, N/A, N/A, N/A

Module Info

-----------

Identifier : QSFP28

Compliance : N/A

Cable Technology : 850 nm VCSEL

Cable Type : Active cable (active copper / optics)

OUI : Mellanox

Vendor Name : Mellanox

Vendor Part Number : P35346-001

Vendor Serial Number : THY1120057

Rev : A1

Wavelength [nm] : 850

Transfer Distance [m] : 30

Attenuation (5g,7g,12g) [dB] : N/A

FW Version : 38.100.59

Digital Diagnostic Monitoring : Yes

Power Class : 5.0 W max

CDR RX : ON,ON,ON,ON

CDR TX : ON,ON,ON,ON

LOS Alarm : N/A

Temperature [C] : 53 [-10..80]

Voltage [mV] : 3232.9 [3100..3500]

Bias Current [mA] : 7.240,7.326,7.346,7.258 [5.492..8.5]

Rx Power Current [dBm] : -2,-2,-2,-2 [-12..6] <== check if RX level is proper or not.

Tx Power Current [dBm] : -2,-2,-2,0 [-14..6]

* How to troubleshoot

Symbol errors indicate a physical layer issue.  These are typically able to be cleared by reseating or replacing the cable.

#(ibportstate <lid number> <port number> reset)

Ex)

#Ibportstate 5 79 query

#ibportstate 5 79 disable

#ibportstate 5 79 enable

> After Cable replacement or port change, Please check the status.

    Mlxlink -d lid-<lid number> <portnumber>

Ex)

# mlxlink -d lid-3 -pc 36 <=== port counter clear

# mlxlink -d lid-3 -p 36 --show\_device  --show\_serdes\_tx     --show\_fec   --show\_ber\_monitor

# mlxlink -d lid-3 -m -c -e -p 36  ç== FEC uncorrectable/physical error counter

## **4.6 PortRcvErrors**

* How to troubleshoot

These are typically able to be cleared by reseating or replacing the cable.

# **5. PM tips for switches “IB/Ethernet/SX6036” from sysdump**

## **5.1 From : Ethernet/IB switch**

* From sysdump ‘Sysinfo.txt’

1. Please find sysinfo.txt file from sysdump

* **Ethernet switch**

Eth1/1:

Admin state : Enabled

Operational state : Up

Last change in operational status : 2w 1d and 2:07:50 ago (1 oper change)

Boot delay time : 0 sec

Description : N\A

Mac address : 24:8a:07:81:04:ec

MTU : 9000 bytes (Maximum packet size 9022 bytes)

Fec : auto

Flow-control : receive off send off

Actual speed : 40 Gbps

Auto-negotiation : Enabled

Width reduction mode : Unknown

Switchport mode : trunk

MAC learning mode : Enabled

Last clearing of "show interface" counters: 00:20:48

60 seconds ingress rate : 506180208 bits/sec, 63272526 bytes/sec, 41173 packets/sec 🡸= See ingress/ egress difference

60 seconds egress rate : 2349264 bits/sec, 293658 bytes/sec, 4457 packets/sec

Rx:

53228689 packets

53095932 unicast packets

132309 multicast packets

448 broadcast packets

80359739790 bytes

809713 discard packets 🡸==== check if it’s increases or not.

0 error packets 🡸====

0 fcs errors

0 undersize packets

0 oversize packets

0 pause packets

0 unknown control opcode

0 symbol errors 🡸====

Tx:

4493032 packets

4317677 unicast packets

175355 multicast packets

0 broadcast packets

1056276187 bytes

0 discard packets 🡸====

0 error packets

0 hoq discard packets 🡸====

* **IB Switch**

IB1/13 state:

Logical port state : Active

Physical port state : LinkUp

Current line rate : 56.0 Gbps

Supported speeds : sdr, ddr, qdr, fdr10, fdr

Speed : fdr

Supported widths : 1X, 4X

Width : 4X

Max supported MTUs : 4096

MTU : 4096

VL capabilities : VL0 - VL7

Operational VLs : VL0 - VL7

Description : ibs1#1

IB Subnet : infiniband-default

Phy-profile : high-speed-ber

Width reduction mode : Disabled

RX bytes : 631437720

RX packets : 7998549

RX errors : 0 🡸====

Symbol errors : 0 🡸====

VL15 dropped packets : 0

TX bytes : 7724808536

TX packets : 5211828

TX wait : 260003439 🡸===

TX discarded packets : 0

## **5.1 From : SX6036**

* From sdkdump

..................................................

Port Performance Counters

..................................................

Port 0x12300 - IEEE 802.3 Counters Group

==================================================

frames\_transmitted\_ok 0

frames\_received\_ok 0

frame\_check\_sequence\_errors 0

alignment\_errors 0

octets\_transmitted\_ok 0

octets\_received\_ok 0

multicast\_frames\_xmitted\_ok 0

broadcast\_frames\_xmitted\_ok 0

multicast\_frames\_received\_ok 0

broadcast\_frames\_received\_ok 0

in\_range\_length\_errors 0

out\_of\_range\_length\_field 0

frame\_too\_long\_errors 0

symbol\_error\_during\_carrier 0

mac\_control\_frames\_transmitted 0

mac\_control\_frames\_received 0

unsupported\_opcodes\_received 0

pause\_mac\_ctrl\_frames\_received 0

pause\_mac\_ctrl\_frames\_transmitted 0

Port 0x12300 - RFC 2863 Counters Group

==================================================

if\_in\_octets 0

if\_in\_ucast\_pkts 0

if\_in\_discards 0

if\_in\_errors 0

if\_in\_unknown\_protos 0

if\_out\_octets 0

if\_out\_ucast\_pkts 0

if\_out\_discards 0

if\_out\_errors 0

if\_in\_multicast\_pkts 0

if\_in\_broadcast\_pkts 0

if\_out\_multicast\_pkts 0

if\_out\_broadcast\_pkts 0

Port 0x12300 - RFC 2819 Counters Group

==================================================

ether\_stats\_drop\_events 0

ether\_stats\_octets 0

ether\_stats\_pkts 0

ether\_stats\_broadcast\_pkts 0

ether\_stats\_multicast\_pkts 0

ether\_stats\_crc\_align\_errors 0

ether\_stats\_undersize\_pkts 0

ether\_stats\_oversize\_pkts 0

ether\_stats\_fragments 0

ether\_stats\_jabbers 0

ether\_stats\_collisions 0

ether\_stats\_pkts64octets 0

ether\_stats\_pkts65to127octets 0

ether\_stats\_pkts128to255octets 0

ether\_stats\_pkts256to511octets 0

ether\_stats\_pkts512to1023octets 0

ether\_stats\_pkts1024to1518octets 0

ether\_stats\_pkts1519to2047octets 0

ether\_stats\_pkts2048to4095octets 0

ether\_stats\_pkts4096to8191octets 0

ether\_stats\_pkts8192to10239octets 0

Port 0x12300 - RFC 3635 Counters Group

==================================================

dot3stats\_alignment\_errors 0

dot3stats\_fcs\_errors 0

dot3stats\_single\_collision\_frames 0

dot3stats\_multiple\_collision\_frames 0

dot3stats\_sqe\_test\_errors 0

dot3stats\_deferred\_transmissions 0

dot3stats\_late\_collisions 0

dot3stats\_excessive\_collisions 0

dot3stats\_internal\_mac\_transmit\_errors 0

dot3stats\_carrier\_sense\_errors 0

dot3stats\_frame\_too\_longs 0

dot3stats\_internal\_mac\_receive\_errors 0

dot3stats\_symbol\_errors 0

dot3control\_in\_unknown\_opcodes 0

dot3in\_pause\_frames 0

dot3out\_pause\_frames 0

Port 0x12300 - CLI Counters Group

==================================================

port\_rx\_octets 0

port\_rx\_frames 0

port\_rx\_jumbo 0

port\_rx\_unicast 0

port\_rx\_multicast 0

port\_rx\_broadcast 0

port\_rx\_no\_buffer 0

port\_rx\_fcs\_errors 0

port\_rx\_runt 0

port\_rx\_other\_errors 0

port\_tx\_octets 0

port\_tx\_frames 0

port\_tx\_jumbo 0

port\_tx\_unicast 0

port\_tx\_multicast 0

port\_tx\_broadcast 0

port\_tx\_errors 0

Port 0x12300 - EXTENDED Counters Group

==================================================

tx\_wait 0

ecn\_marked 0

no\_buffer\_discard\_mc 0

rx\_ebp 0

tx\_ebp 0

rx\_buffer\_almost\_full 0

rx\_buffer\_full 0

tx\_stats\_pkts64octets 0

tx\_stats\_pkts65to127octets 0

tx\_stats\_pkts128to255octets 0

tx\_stats\_pkts256to511octets 0

tx\_stats\_pkts512to1023octets 0

tx\_stats\_pkts1024to1518octets 0

tx\_stats\_pkts1519to2047octets 0

tx\_stats\_pkts2048to4095octets 0

tx\_stats\_pkts4096to8191octets 0

tx\_stats\_pkts8192to10239octets 0

Port 0x12300 - DISCARD Counters Group

==================================================

ingress\_general 0

ingress\_policy\_engine 0

ingress\_vlan\_membership 0

ingress\_tag\_frame\_type 0

egress\_vlan\_membership 0

loopback\_filter 0

egress\_general 0

egress\_link\_down 0

egress\_hoq 0

port\_isolation 0

egress\_policy\_engine 0

ingress\_tx\_link\_down 0

egress\_stp\_filter 0

egress\_hoq\_stall 0

egress\_sll 0 🡸== “ingress\_discard\_all ” Please find below

Port 0x12300 - PER PRIO Counters Group

==================================================

Prio 0

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 1

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 2

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 3

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 4

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 5

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 6

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 7

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Port 0x12300 - PER TC Counters Group

==================================================

TC 0

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 1

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 2

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 3

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 4

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 5

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 6

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 7

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

prsunny@SYD26-0101-0708-15T1:~$ show int counters 🡸= check ingress\_discard\_all   
IFACE STATE RX\_OK RX\_BPS RX\_UTIL RX\_ERR RX\_DRP RX\_OVR TX\_OK TX\_BPS TX\_UTIL TX\_ERR TX\_DRP TX\_OVR  
----------- ------- --------- -------- --------- -------- -------- -------- --------- -------- --  
Ethernet0 U 77401644 N/A N/A 0 0 0 612546543 N/A N/A 0 0 0  
Ethernet4 U 82013052 N/A N/A 0 0 0 548287581 N/A N/A 0 0 0  
Ethernet8 U 75445567 N/A N/A 0 0 0 640693267 N/A N/A 0 0 0  
Ethernet12 U 75069284 N/A N/A 0 0 0 590180253 N/A N/A 0 0 0  
Ethernet16 U 81110674 N/A N/A 0 0 0 512705461 N/A N/A 0 0 0  
Ethernet20 U 74303484 N/A N/A 0 0 0 560971127 N/A N/A 0 0 0  
Ethernet24 U 83767004 N/A N/A 0 0 0 567398414 N/A N/A 0 0 0  
Ethernet28 U 89491970 N/A N/A 0 0 0 527621910 N/A N/A 0 0 0  
Ethernet32 U 77686352 N/A N/A 0 0 0 529482801 N/A N/A 0 0 0  
Ethernet36 U 77047095 N/A N/A 0 0 0 446032780 N/A N/A 0 0 0  
Ethernet40 U 81180399 N/A N/A 0 0 0 522866000 N/A N/A 0 0 0  
Ethernet44 U 83906899 N/A N/A 0 0 0 521436128 N/A N/A 0 0 0  
Ethernet48 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet52 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet56 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet60 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet64 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet68 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet72 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet76 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet80 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet84 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet88 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet92 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet96 U 869161966 N/A N/A 0 0 0 100500628 N/A N/A 0 0 0  
Ethernet100 U 860994567 N/A N/A 0 0 0 104002217 N/A N/A 0 0 0  
Ethernet104 U 901131138 N/A N/A 0 0 0 97657185 N/A N/A 0 0 0  
Ethernet108 U 815075863 N/A N/A 0 0 0 114106192 N/A N/A 0 0 0  
Ethernet112 U 823447607 N/A N/A 0 0 0 92292313 N/A N/A 0 0 0  
Ethernet116 U 857305589 N/A N/A 0 0 0 99577287 N/A N/A 0 0 0  
Ethernet120 U 828557210 N/A N/A 0 0 0 103516688 N/A N/A 0 0 0  
Ethernet124 U 728265885 N/A N/A 0 0 0 103008124 N/A N/A 0 0 0  
  
 Port 0x10300 - DISCARD Counters Group   
==================================================  
ingress\_general                          0  
ingress\_policy\_engine                    0  
ingress\_vlan\_membership                  0  
ingress\_tag\_frame\_type                   97164  
egress\_vlan\_membership                   0  
loopback\_filter                          0  
egress\_general                           0  
egress\_link\_down                         0  
egress\_hoq                               0  
port\_isolation                           0  
egress\_policy\_engine                     4158013  
ingress\_tx\_link\_down                     0  
egress\_stp\_filter                        0  
egress\_hoq\_stall                         0  
egress\_sll                               0  
ingress\_discard\_all                      346428 🡸=== SN2700s with the reason of egress\_policy\_engine. If possible, please modify the default ACL rule to permit all.

# **6. Causing Fabric degradation**

## **6.1 Link speed downgrade**

* From ibdiagnet2.log

With ‘ibdiagnet2.log’, sometimes we can see ‘link speed downgrade’. This can cause congestion.

**Speed / Width checks**

**-I- Link Speed Check (Compare to supported link speed)**

**-E- Links Speed Check finished with errors**

**-E- Link: S7cfe9003009e90a0/N7cfe9003009e90a0/P21<-->skyway-hpci:ib-gw/U1/P1 - Unexpected actual link speed 10 (enable\_speed1="2.5 or 5 or 10 or 14 or 25 or FDR10", enable\_speed2="2.5 or 5 or 10 or 14 or 25 or 50" therefore final speed should be 25)**