
TOMORROW starts here.
cisco
Cisco live!

UCS Performance Troubleshooting

BRKCOM-3002

Greg Scarlett

Technical Service Engineer

CCIE Data Centre #42291

Cisco live!

Agenda

- Troubleshooting Methodology and Processes
- Path Tracing
- LAN Performance
- SAN Performance
- Compute Performance
- Testing Tools

"The accomplishment of a given task **measured** against preset **known** standards of **accuracy**, completeness, cost, and **speed**."

"Our Mail is slow to open" Anonymous Users

TAC Case
March 2013

Survey

- Networking Problem?
- Storage Problem?
- Compute Problem? (BIOS, Memory?)
- Operating System?
- External?

What Affects Performance?

Rx Queues

UIF

Firmware

Driver

B2B Credits

Power

VIF

LSO

Heat

Round Robin

CoS

LIF

Hashing

Buffers

RSS

TCP Offload

Congestion

QoS

Pinning

MRU

Multipathing

Fixed Path

Arbitration

Flow Control

Troubleshooting Methodology

Before You Start.

- Troubleshooting is an Art
- Establish Baselines pre/post production
- Use all available resources – Free or Paid
- Document Changes
  - Network/Topology
  - Configuration

Troubleshooting Process

Build The Picture

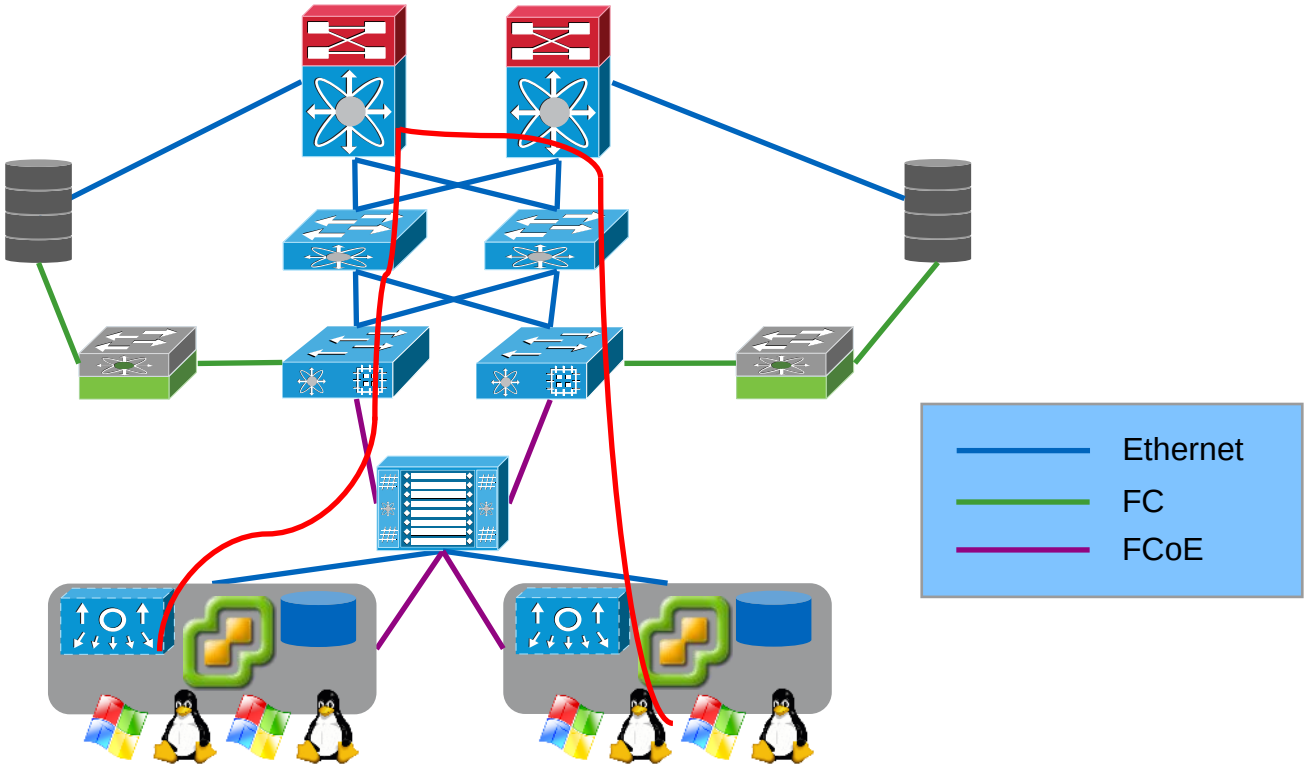
- Define the Problem – What Is vs What Is Not
  - Document end to end. FW, Drivers, OS
  - Identify and Isolate traffic path
  - Create a Diagram.
  - Reference diagrams
- One change at a time
  - No Shotgun troubleshooting
  - Consistency in testing

"Replication between Exchange Mailboxes is performing slowly" Exchange Administrator

TAC Case
March 2013

Troubleshooting Process

Build The Picture



Divide & Conquer

UCS Performance Areas can be categorised into the following areas:

Infrastructure

- Fabric Interconnects
- IOMs
- Adapters
- SPFs/Cables

Platform

- BIOS
- Chipset
- Adapter Settings

OS Specific

- Windows vs. Linux
- TCP vs. UDP vs. Multicast
- RSS
- CPU Affinity
- Interrupts

We'll focus on these areas

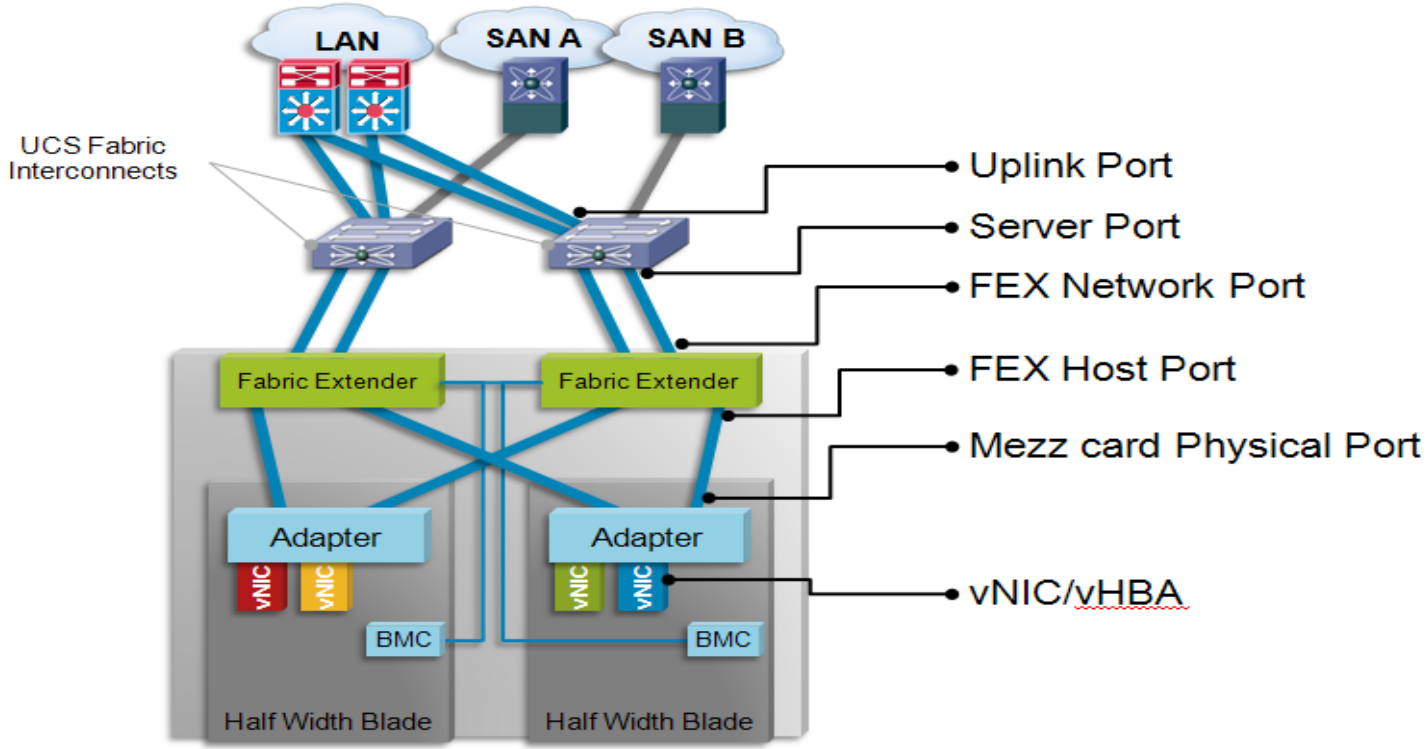
Cisco live!

“Traffic between the VM’s is slow” Server Administrator

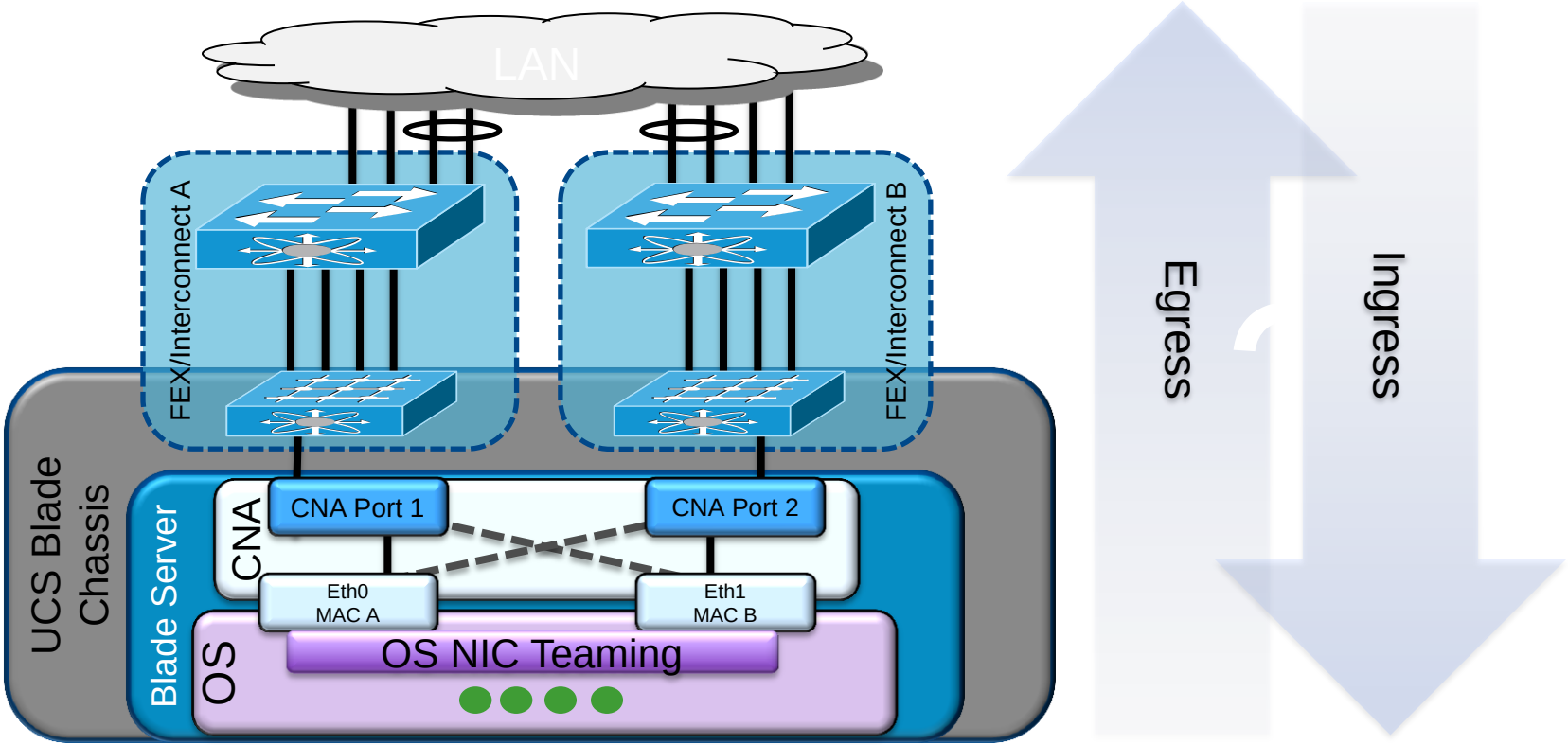
TAC Case
March 2013

Infrastructure Path Tracing

System Components – Hop By Hop

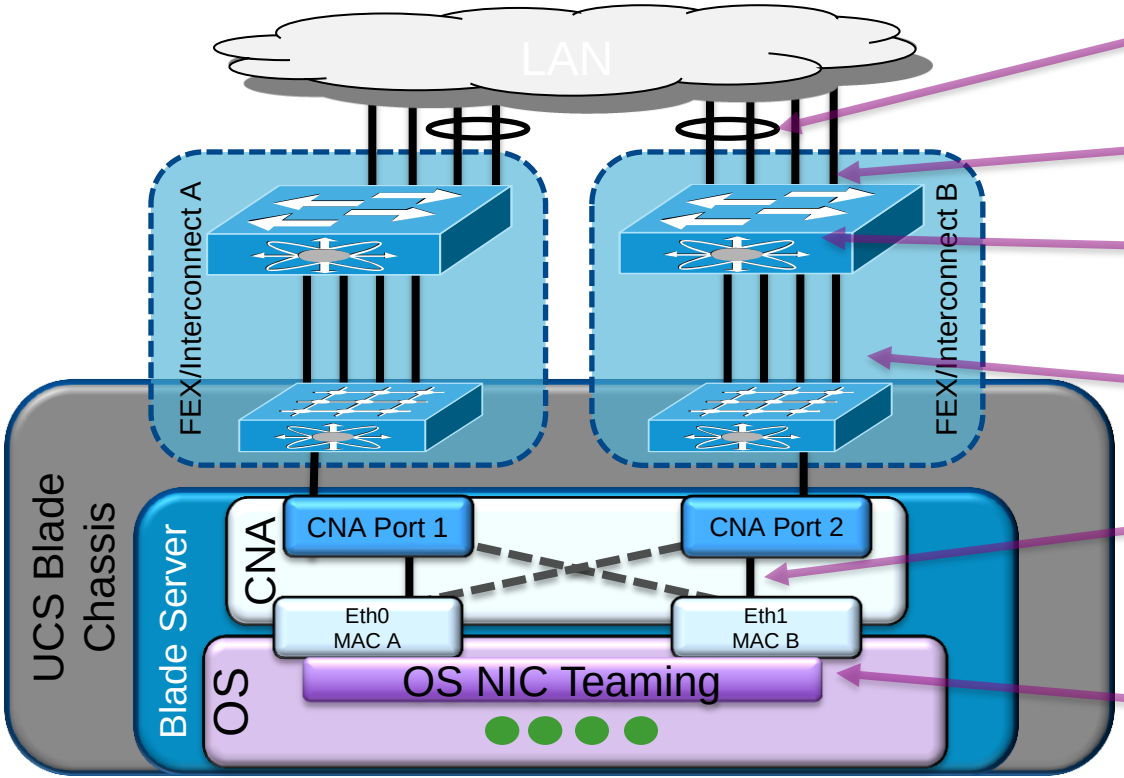


Which Path Will UCS Choose?



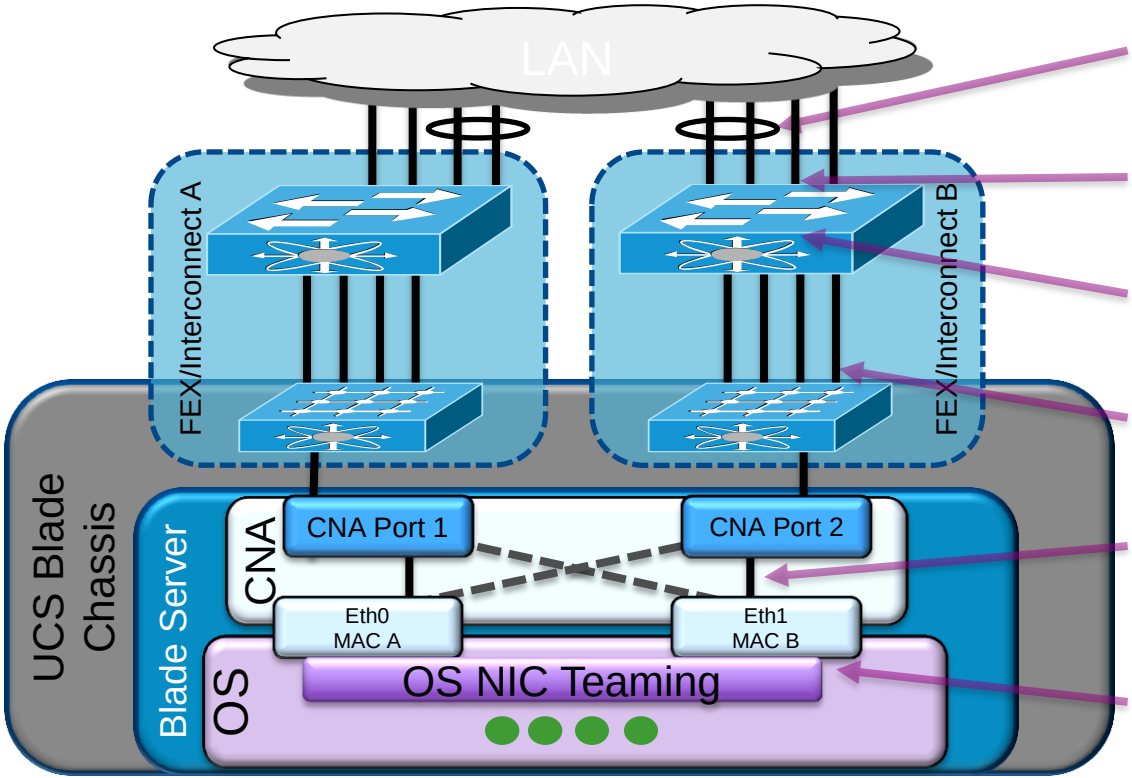
UCS Frame Flow Decisions

Egress



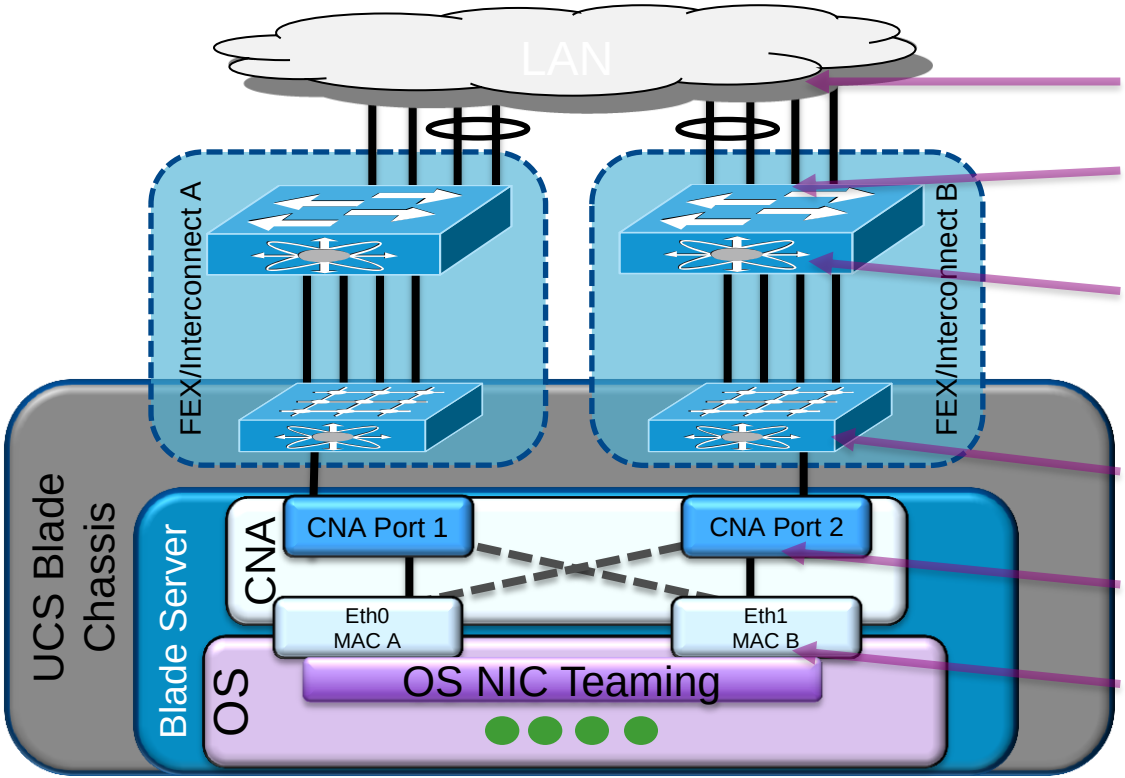
- Which port in the port channel?
  (2-8 choices)
- Which uplink/border port or port channel?
  (many choices)
- Local or remote destination?
  (2 choices)
- Which Fabric Port?
  (4 choices)
- Which CNA Port?
  (2 choices)
- Which PCIe Ethernet Interface?
  (1-58 choices depending on CNA)

UCS Frame Flow Decisions
Egress



- Port Channelling Algorithm
- Border Port Pinning
- L2 Switching in FIs
- Fabric Port Pinning
- UCS Fabric Failover
- OS Routing Table or OS NIC Teaming

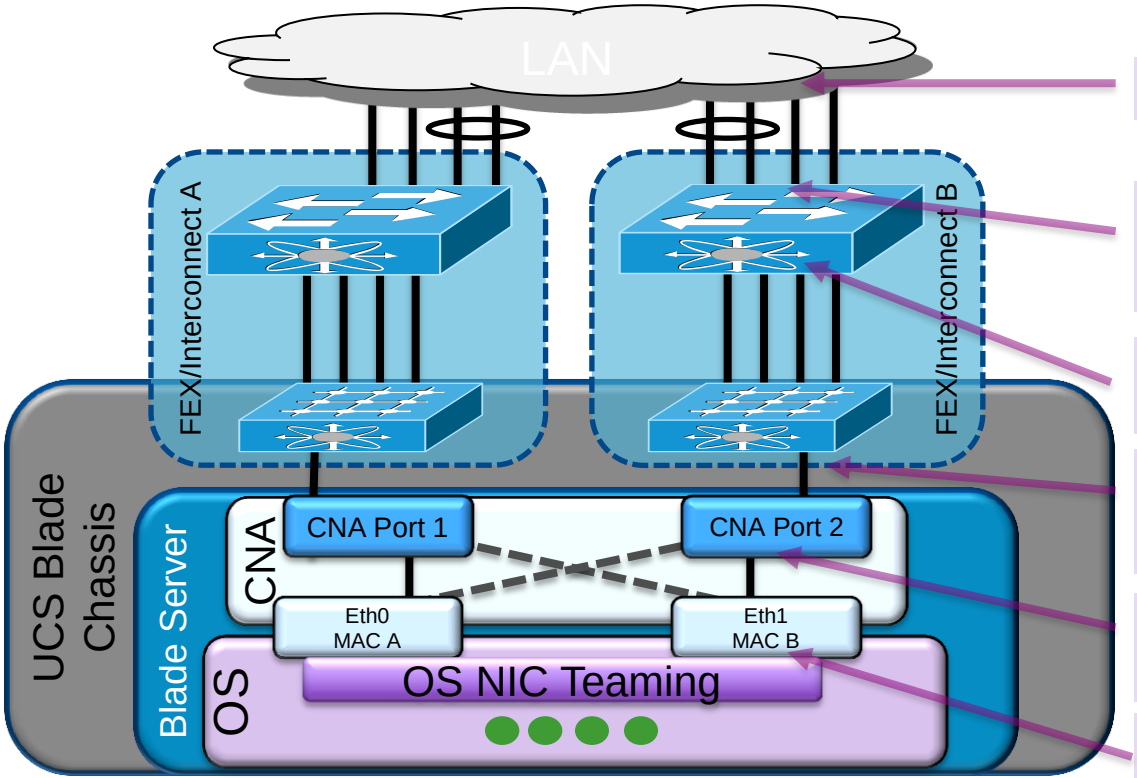
UCS Frame Flow Decisions
Ingress



- Which downlink or port channel?
- Allow the frame inbound?
  (decision depends on ‘switch mode’ vs. ‘end host mode’)
- Which Fabric Extender Port?
- Which Server Bay Port?
  (8 choices)
- Which PCIe Device (vNIC)?
  (1-58 choices depending on CNA)
- Pass frame to OS?

UCS Frame Flow Decisions

Ingress



(Upstream Switch Decides)

Déjà vu, RPF, border port pinning

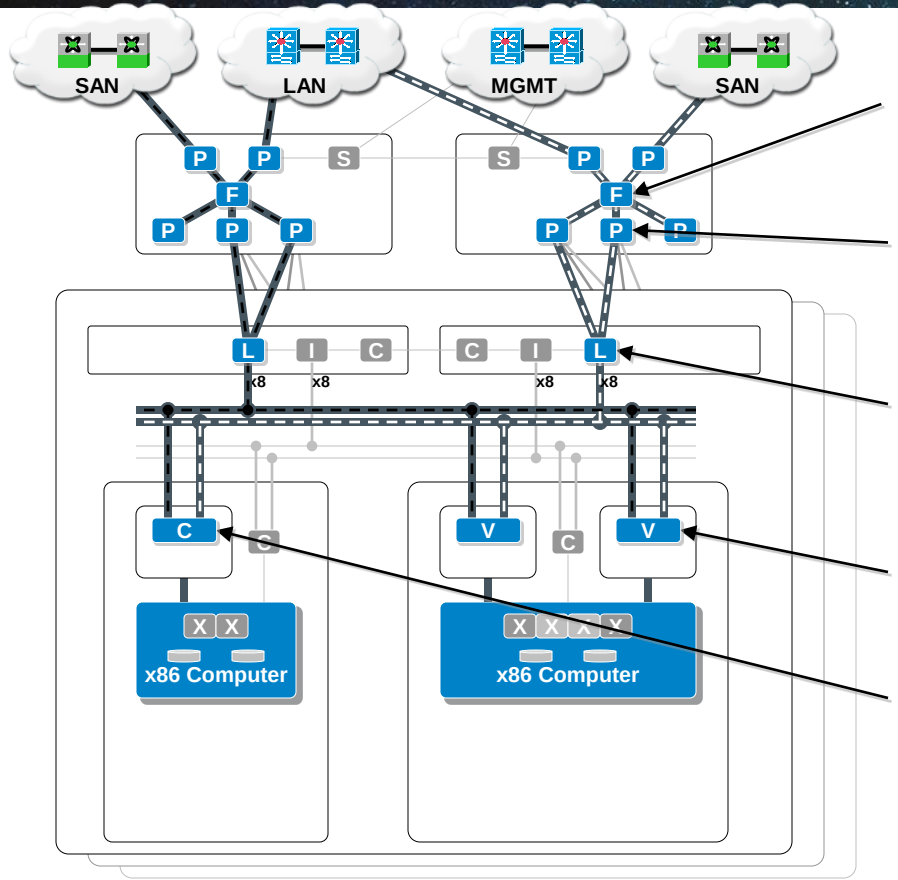
Fabric Port Pinning

VNTag + Offset
(MAC Learning on FIs)

VNTag Identifier

Dest. MAC and Ethertype binding

System Components – ASICs (Gen 1 vs. Gen 2)



- Fabric ASIC : Altos/Sunnyvale
- Port ASIC : Gatos/Carmel
- FEX ASIC : Redwood/Woodside
- VIC ASIC : Palo/Sereno
- Gen-1 CNA ASIC : Menlo

Why Do I Care About ASIC Names?

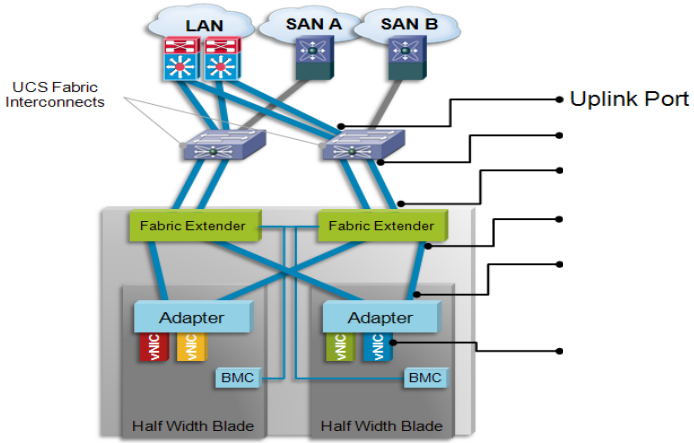
```
fex-1# show platform software woodside rate

fex-1# show platform software redwood sts

TSI-UCS-A(nxos)# show hardware internal carmel crc

TSI-UCS-A(nxos)# show hardware internal sunny event-history
errors
```

Narrowing Down The Problem

- Define the problem
  - From which point to what other point is the problem?
  - Do we see the problem in one direction or both?
- Eliminate variables
  - Is the problem seen between traffic traversing the same fabric?
  - Is the problem only happening on a specific path?
- List all the ports in the traffic path
  - VIFs, FEX, HIFs, NIFs, Fabric and Uplink ports

Defining The Ports

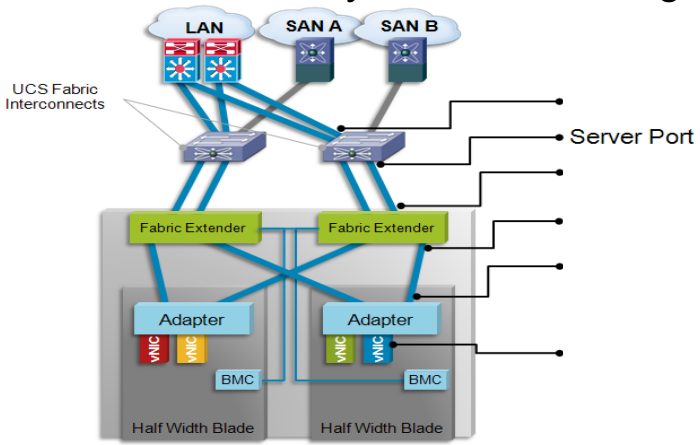
- FI Uplink/Trunk Port
  - The Fabric Interconnect defines Uplink ports as those ports connecting to the LAN
  - Always in trunk mode (no such thing as **mode access** configuration)
  - VLAN 1 is default (native) & can be changed
  - Port-channel configuration allowed (LACP only)
  - There is currently no vPC or Fabric Path feature in the FI



Defining The Ports

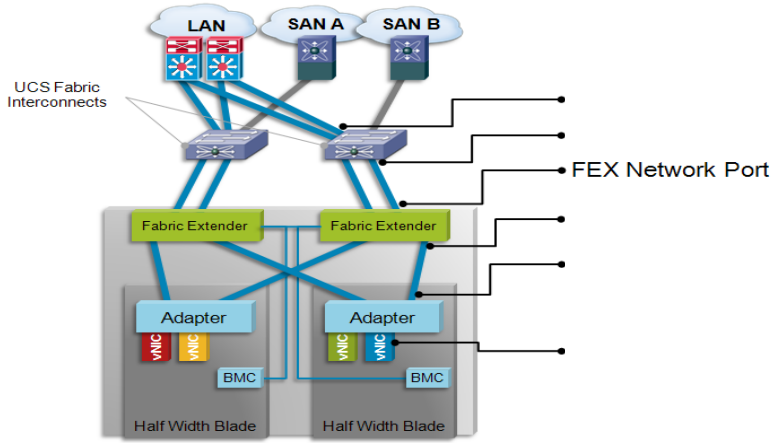
- Fabric Interconnect FEX-Fabric aka Server Interfaces (SIF)
  - The Fabric Interconnect (FI) defines fex-fabric ports as those ports connecting to the IOMs in the chassis
  - IOM Host Interfaces (HIFs) ports are statically pinned to FEX-fabric ports (SIF)
  - Same concept Nexus FEXs use with Satellite ports.

Note: The term “FEX” and “IOM” are commonly used interchangeably



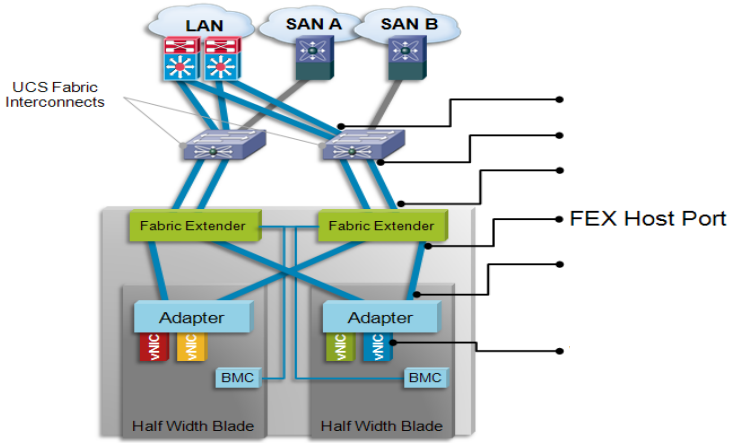
Defining The Ports

- IOM Network Interfaces (NIF)
  - The IOM defines these ports which are external connecting the IOM to the FI.
  - NIF port are either configured as individual or channeled to the FI's as server ports (SIF) – depends on model of IOM.
  - Same concept Nexus FEXs use with Satellite ports.



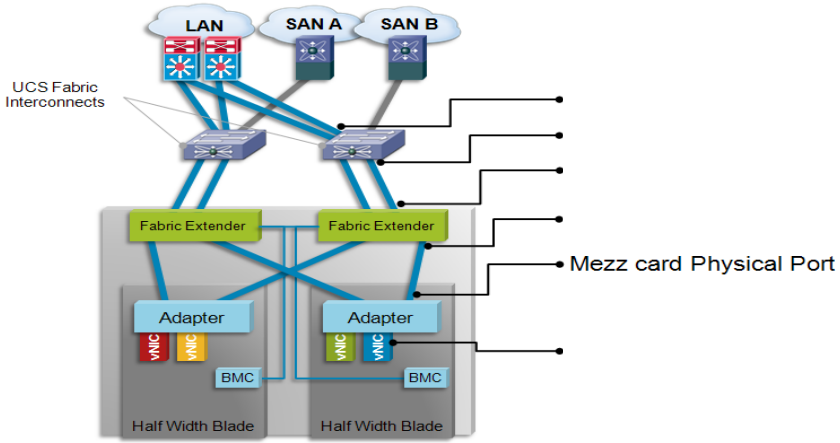
Defining The Ports

- IOM Host Interfaces (HIFs)
  - Each IOM provides a number of internal ports per blade
  - IOM model 2104XP provides 8x internal ports (one for each blade)
  - IOM model 2204XP provides 16x internal ports (two for each blade)
  - IOM model 2208XP provides 32x internal ports (four for each blade)
  - Each HIF is defined by three different values, EthX/Y/Z. Chassis/Adapter/Slot



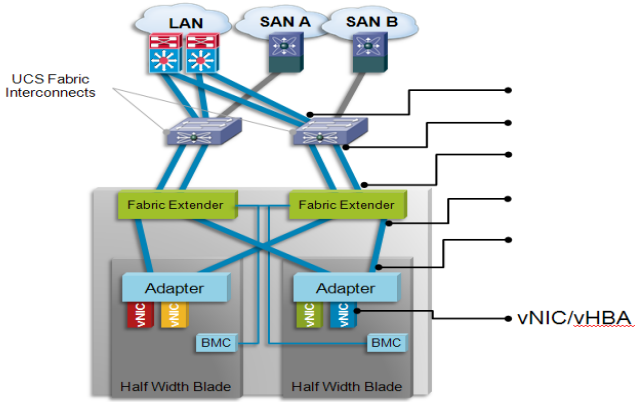
Defining The Ports

- Adapter Uplink Interface (UIFs)
  - Each Adapter has 2 physical uplinks, one to each uplink
  - References as 0 and 1
  - These are also known as the Data Centre Ethernet (DCE) Interfaces



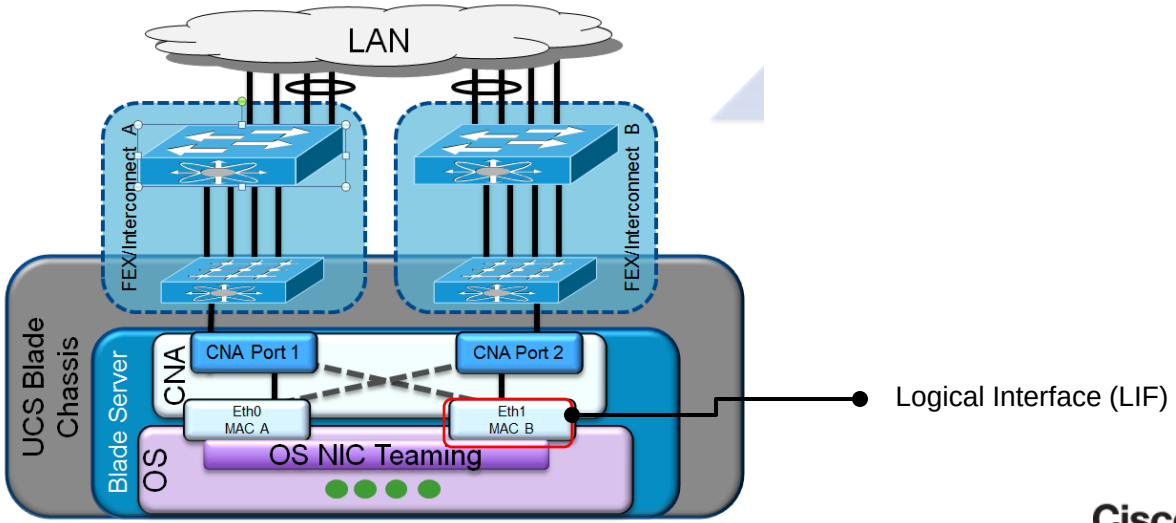
Defining The Ports

- Virtual Interface (VIF)
  - Defined as Ethernet (veth) or Fibre Channel (vfc)
  - A vNIC with Fabric Failover enabled will have two VIFs assigned (Primary & Backup)
  - Represent the vNIC or vHBA on the compute blade towards OS
  - Pinned automatically or manually (pin groups) to border port or FC uplink ports
  - veth and vfc numbers are dynamically assigned
  - System automatically allocates a certain number of VIFs per service-profile for its own management/control traffic



Defining The Ports

- Logical Interfaces (LIF)
  - Represent the logical interface of a VIF pair (those with Fabric Failover enabled)
  - LIF indexes are managed at the adapter level
  - Not visible within UCSM



"VM's are hosted on NFS storage and use iSCSI volumes on the VM"

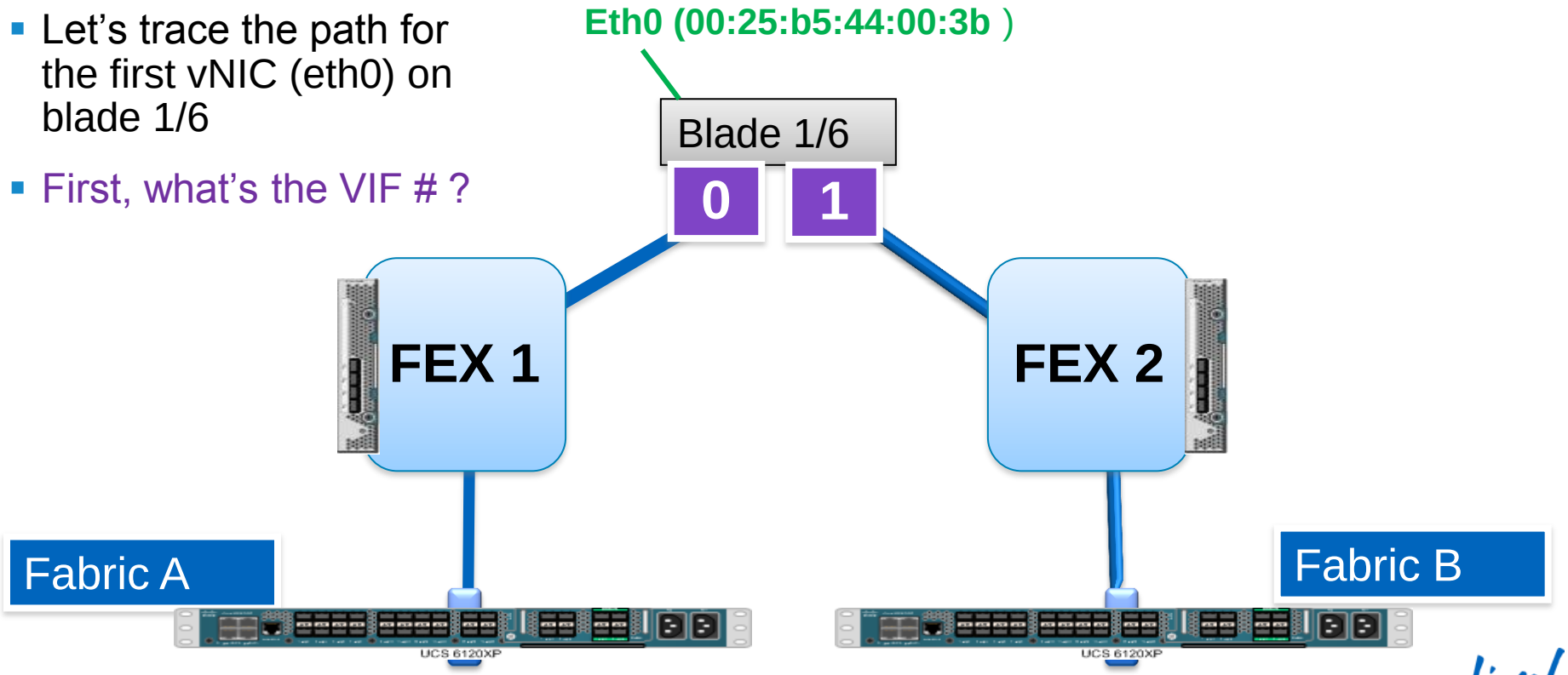
Server Administrator

TAC Case

March 2013

Trace Example

- Let's trace the path for the first vNIC (eth0) on blade 1/6
- First, what's the VIF # ?

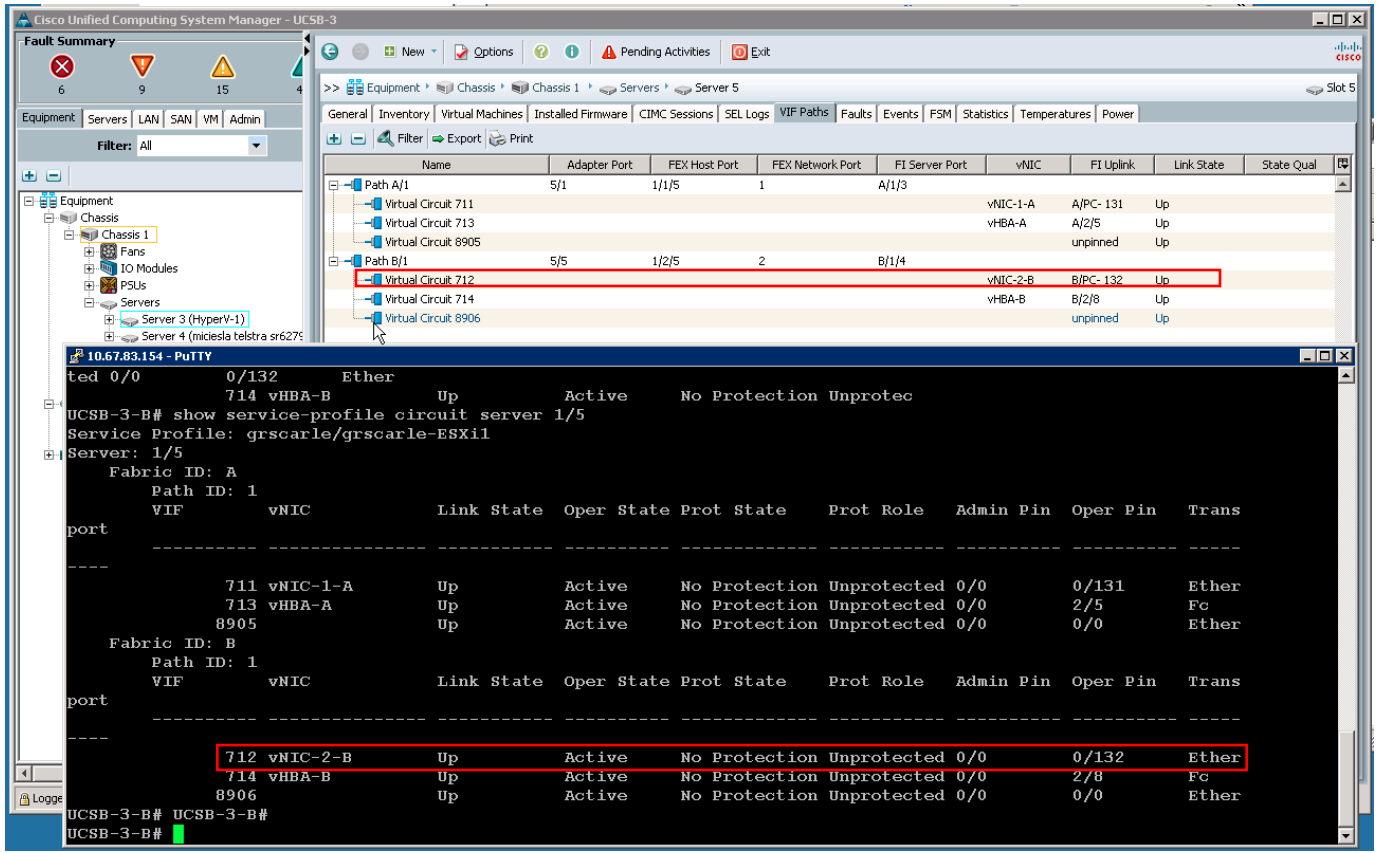


VIF Pinning – Service Profile View

- UCSM top level : **show service-profile circuit server <chassis#>/<slot#>**

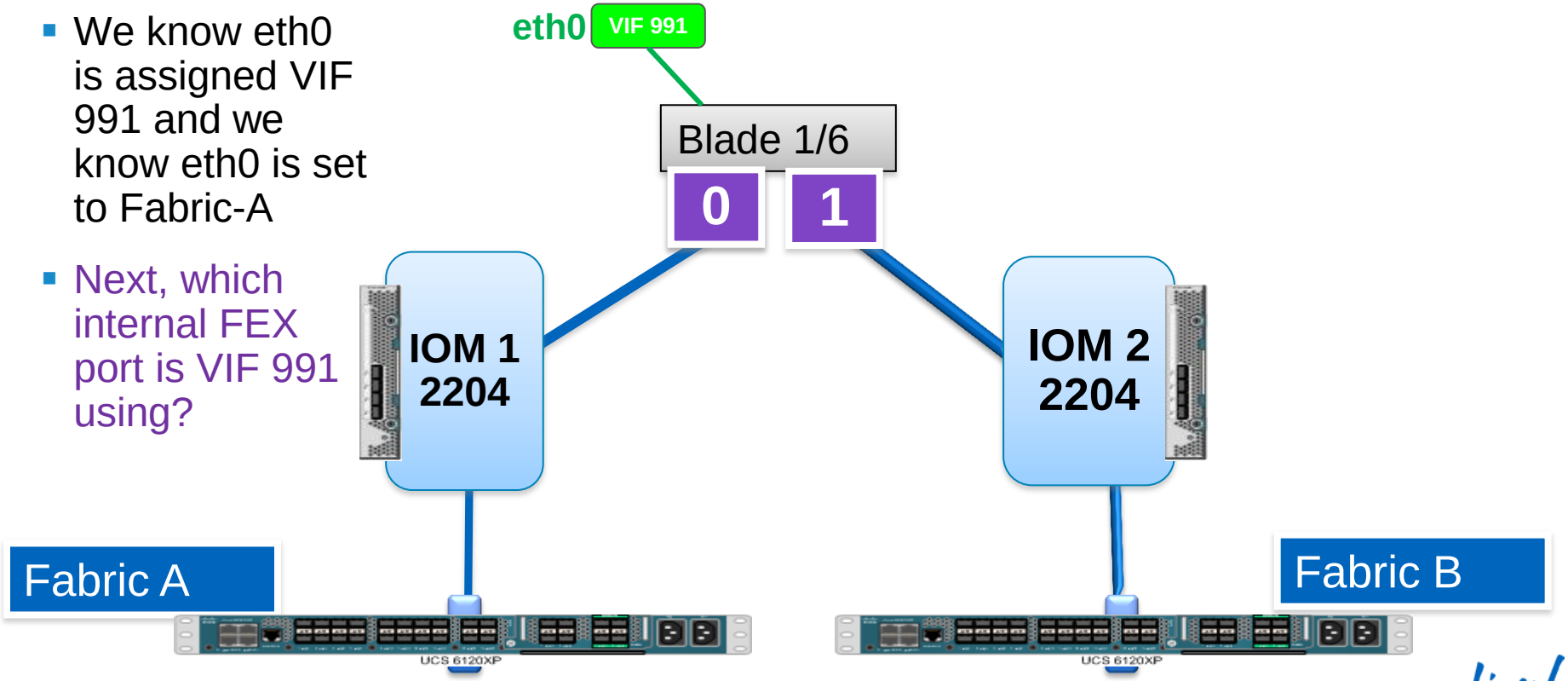
```
UCS-A# show service-profile circuit server 1/6
Service Profile: grscarle/Perf-Test-3
Server: 1/6
    Fabric ID: A
        VIF        vNIC            Link State  Oper State Prot State    Prot Role   Admin Pin  Oper Pin   Transport
        ---------- --------------- ----------- ---------- ------------ ----------- ---------- ---------- ---------
             9178                  Up          Active     No Protection Unprotected 0/0        0/0        Ether
              986 fc0              Up          Active     No Protection Unprotected 0/0        0/0        Fc
              988 eth1             Up          Active     Passive       Backup      0/0        1/7        Ether
              990 eth3             Up          Active     Passive       Backup      0/0        1/7        Ether
              991 eth0             Up          Active     Active        Primary     0/0        1/7        Ether
              993 eth2             Up          Active     Active        Primary     0/0        1/7        Ether
    Fabric ID: B
        <snip>
```

VIF Pinning – GUI vs CLI



Trace Example

- We know eth0 is assigned VIF 991 and we know eth0 is set to Fabric-A
- Next, which internal FEX port is VIF 991 using?



IOM Internal Port Information – 2100XP

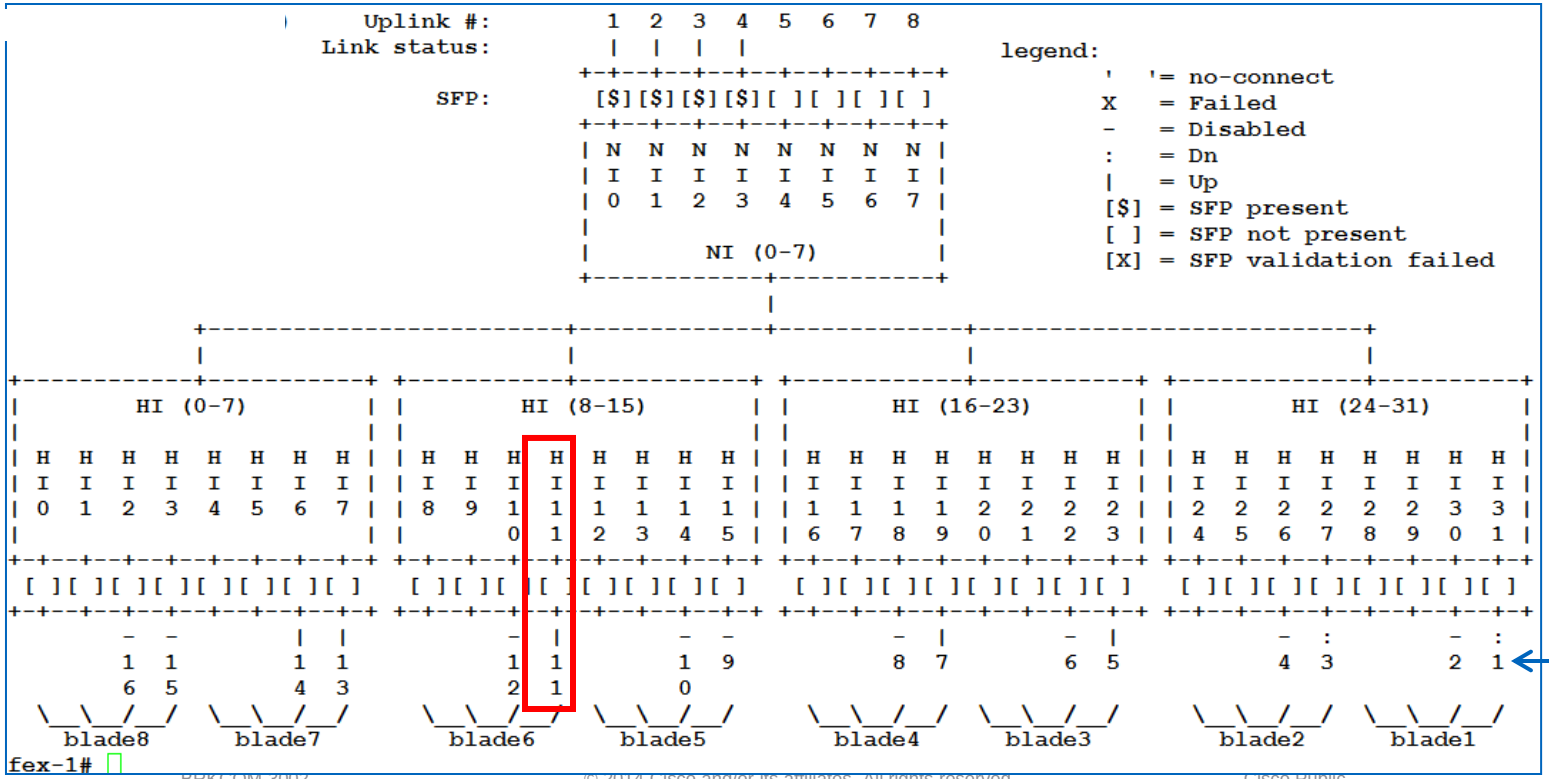
- connect iom <chassis #>
- show platform software redwood sts

```
Attaching to FEX 2 ...
To exit type 'exit', to abort type '$.'
Bad terminal type: "xterm". Will assume vt100.
fex-1# show platform software redwood sts
Board Status Overview:
 legend:
         ' '= no-connect
         X  = Failed
         -  = Disabled
         :  = Dn
         |  = Up
         ^  = SFP+ present
         v  = Blade Present
--------------------------------

         +---+----+----+----+
         |   |    |    | [$]|
         +---+----+----+----+
           :    :    :    |
         +-+----+----+----+-+
         | 0    1    2    3 |
         | I    I    I    I |
         | N    N    N    N |
         |                  |
         |      ASIC 0      |
         |                  |
         | H H H H H H H H  |
         | I I I I I I I I  |
         | 0 1 2 3 4 5 6 7  |
         +-+-+-+-+-+-+-+-+--+
           - : - - - : | :
          +-+-+-+-+-+-+-+-+
          |v|v|-|-|v|v|v|v|
          +-+-+-+-+-+-+-+-+
Blade:      8 7 6 5 4 3 2 1
fex-1#
```

```
legend:
            = no-connect
        X   = Failed
        -   = Disabled
        :   = Dn
        |   = Up
        [$] = SFP present
        [ ] = SFP not present
        [X] = SFP validation failed
```

IOM Internal Port Information – 2200XP

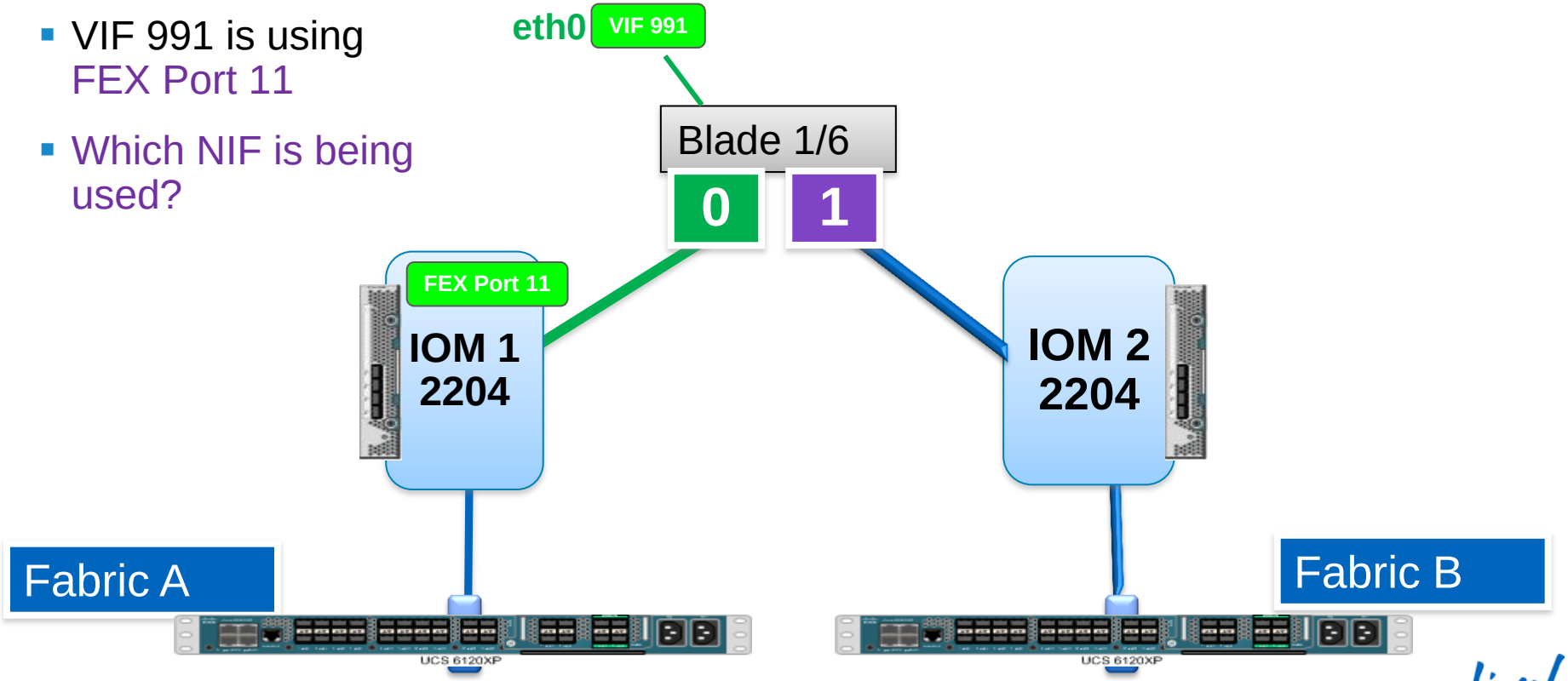
- `show platform software woodside sts`



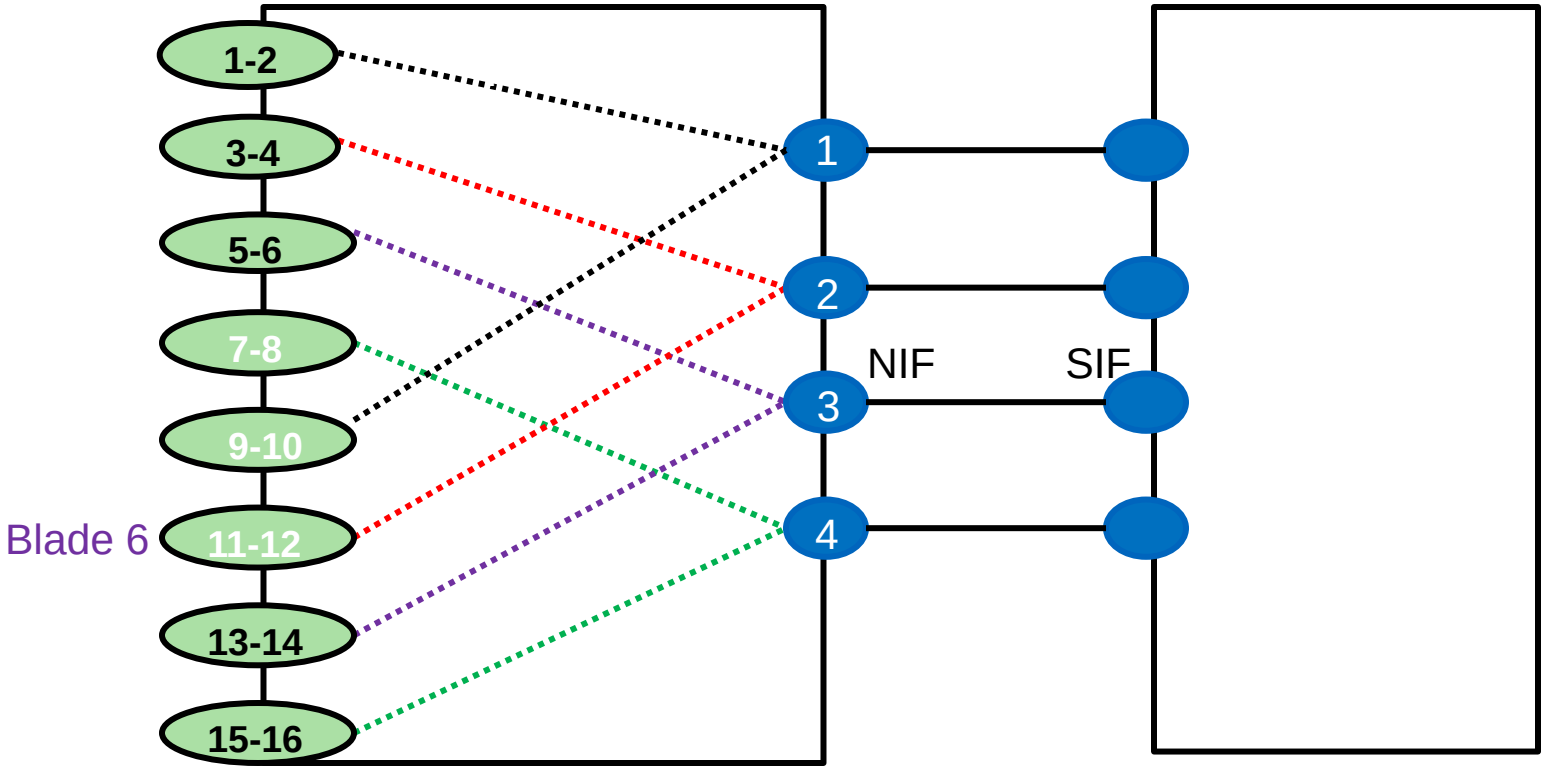
FEX Ports

Trace Example

- VIF 991 is using FEX Port 11
- Which NIF is being used?

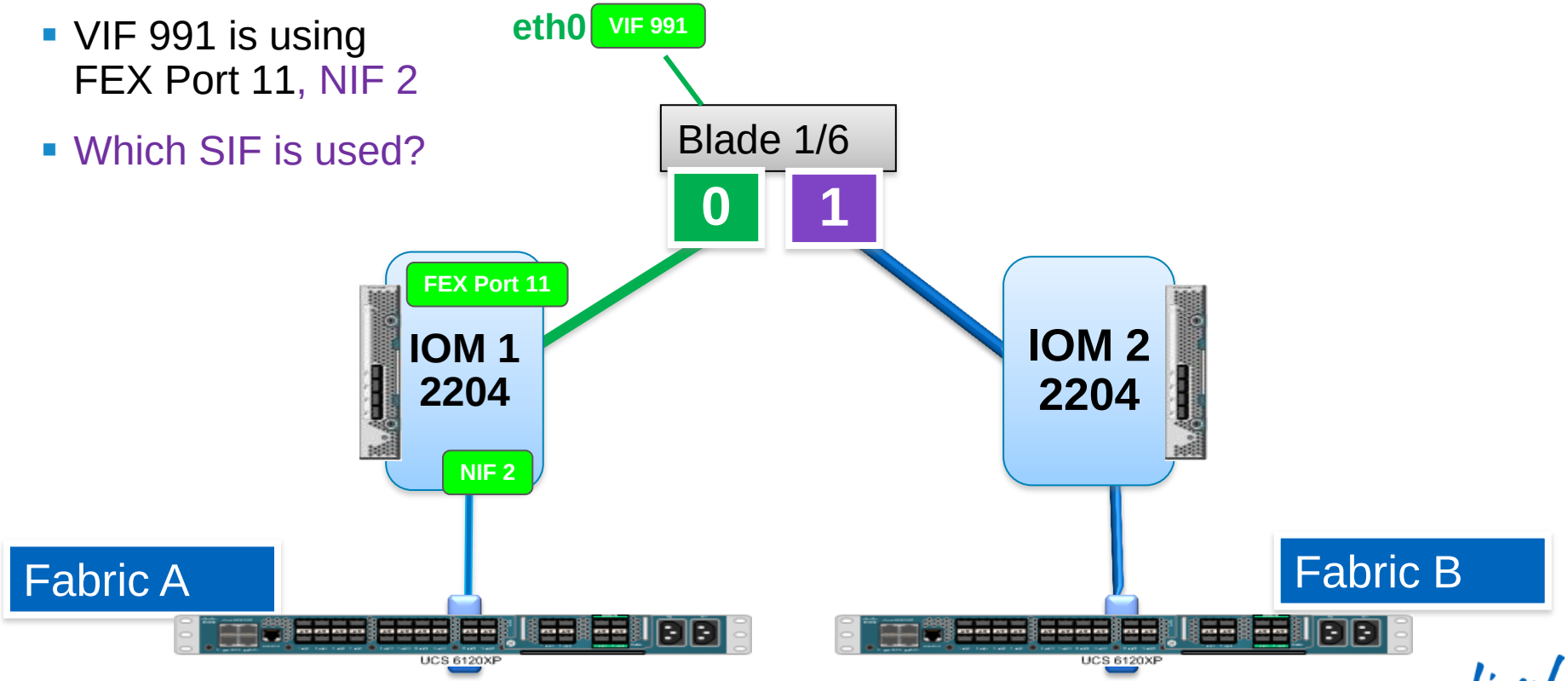


FEX To Fabric Port Pinning (2204XP)



Trace Example

- VIF 991 is using FEX Port 11, NIF 2
- Which SIF is used?



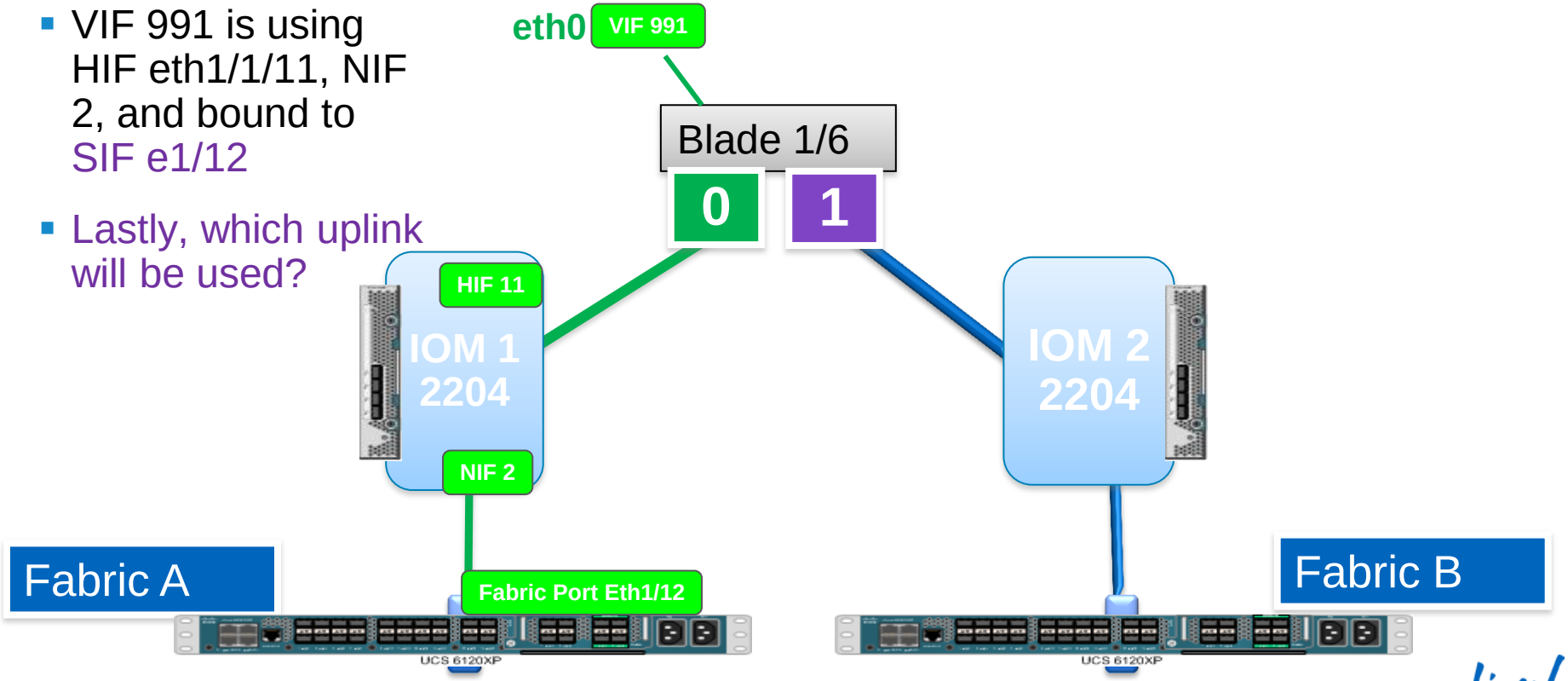
IOM Port Information

- Connect nxos : **show fex <chassis#> detail**

```
FEX: 1 Description: FEX0001   state: Online
  FEX version: 5.0(3)N2(2.11d) [Switch version: 5.0(3)N2(2.11d)]
  FEX Interim version: 5.0(3)N2(2.11d)
  Switch Interim version: 5.0(3)N2(2.11d)
  Chassis Model: N20-C6508,  Chassis Serial: FOX1326G5KH
  Extender Model: UCS-IOM-2204XP,  Extender Serial: FCH154176G0
  Part No: 73-14488-01
  Card Id: 184, Mac Addr: cc:ef:48:1f:dc:2a, Num Macs: 38
  Module Sw Gen: 21  [Switch Sw Gen: 21]
  post level: complete
 pinning-mode: static    Max-links: 1
  Fabric port for control traffic: Eth1/13
  Fabric interface state:
    Eth1/11 - Interface Up. State: Active
    Eth1/12 - Interface Up. State: Active
    Eth1/13 - Interface Up. State: Active
    Eth1/14 - Interface Up. State: Active
  Fex Port        State  Fabric Port
        Eth1/1/1  Down     Eth1/11
        Eth1/1/2  Down        None
        Eth1/1/3  Down     Eth1/12
        Eth1/1/4  Down        None
        Eth1/1/5    Up     Eth1/13
        Eth1/1/6  Down        None
        Eth1/1/7    Up     Eth1/14
        Eth1/1/8  Down        None
        Eth1/1/9  Down        None
       Eth1/1/10  Down        None
       Eth1/1/11    Up     Eth1/12
       Eth1/1/12  Down        None
       Eth1/1/13    Up     Eth1/13
       Eth1/1/14    Up     Eth1/13
       Eth1/1/15  Down        None
       Eth1/1/16  Down        None
       Eth1/1/17    Up     Eth1/14
```

Trace Example

- VIF 991 is using HIF eth1/1/11, NIF 2, and bound to SIF e1/12
- Lastly, which uplink will be used?



VIF Pinning – Fabric Interconnect View

- Connect nxos : **show pinning border-interface active**

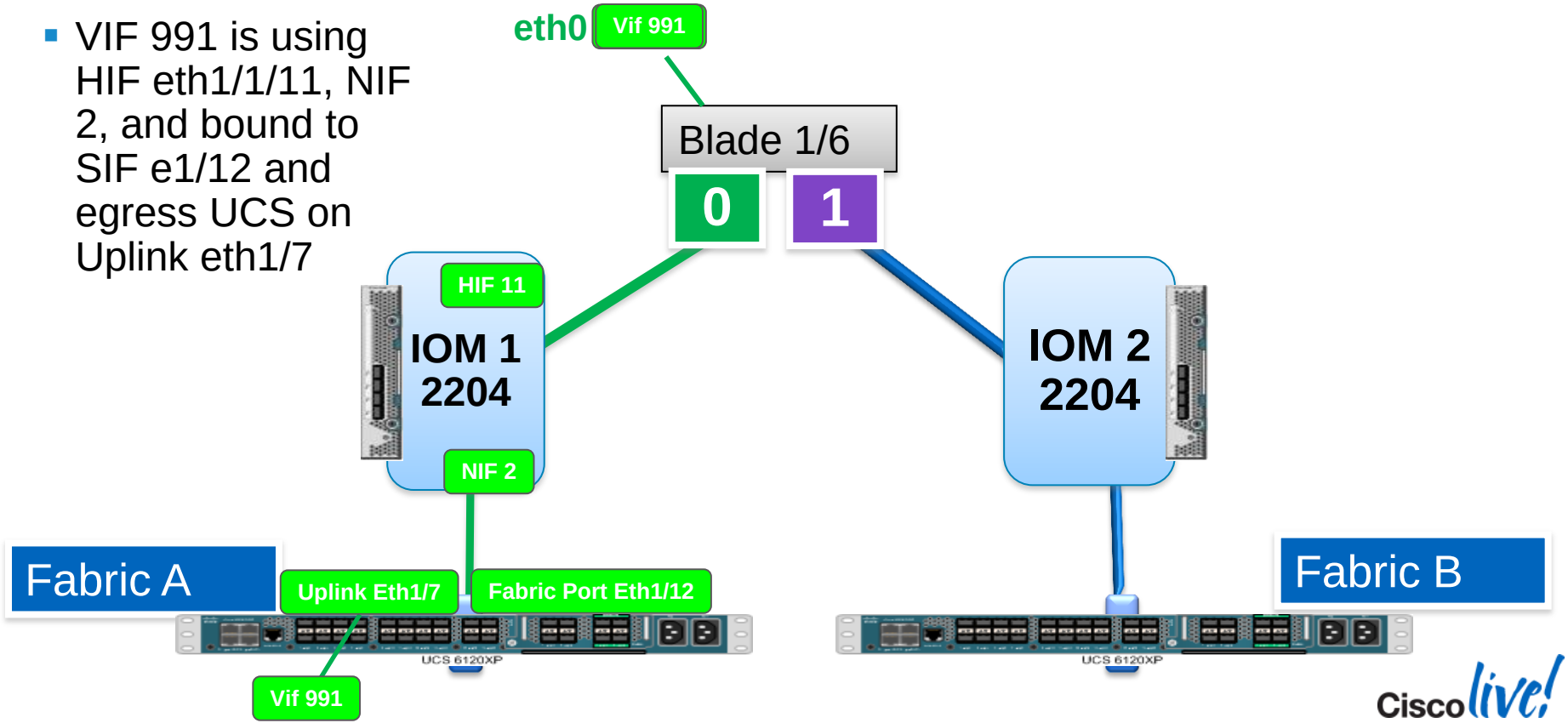
```
UCS-A(nxos)# show pinning border-interfaces active
--------------------+---------+--------------------------------------
Border Interface     Status    SIFs
--------------------+---------+--------------------------------------
Eth1/7               Active    Veth988         Veth991 Veth993
Eth1/8               Active    Veth963 Veth974 Eth1/1/3 Eth2/1/7
Total Interfaces : 2
```

- Connect nxos : **show pinning server-interfaces**

```
UCS-A(nxos)# show pinning server-interfaces | i Veth
Veth956          No                -                          -
Veth963          No                Eth1/8                     2:27:23
Veth974          No                Eth1/8                     2:27:23
Veth988          No                Eth1/7                     2:27:23
                                                              2:27:23
Veth991          No                Eth1/7                     2:27:23
Veth993          No                Eth1/7                     2:27:23
```

Trace Example

- VIF 991 is using HIF eth1/1/11, NIF 2, and bound to SIF e1/12 and egress UCS on Uplink eth1/7



Narrowing Down The Problem

- Define the problem
  - From which point to what other point is the problem?
  - Do we see the problem in one direction or both?
- Eliminate variables
  - Is the problem seen between traffic traversing the same fabric?
  - Is the problem only happening on a specific fabric path?
- List all the ports in the traffic path
  - VIFs, FEX, HIFs, NIFs, Fabric and Uplink ports

Blade 1/6
vNIC: eth0
VIF: 991
DCE: 0
FEX: 1/1/11

HIF: 11
NIF: 2
SIF: Eth 1/12
Uplink: Eth 1/7

LAN Performance

Performance 101

Throughput

- In data transmission, throughput is the amount of data transferred successfully over a link from one end to another in a given period of time. It is usually expressed in a magnitude of bits per second *(Gbps/Mbps)*.
- Refers to how fast a device is actually sending data over the communication channel
- Also known as "Consumed Bandwidth"

Bandwidth

- Refers to how fast a device can send data over a single communication channel
- Also known as "Maximum Throughput"

Performance Analogy

Using an example of cars on a highway, the highway would represent available **Bandwidth** allowing a max # of cars to travel across it at a max speed limit. The cars would represent packets or **Throughput.** Throughput on a highway can be limited by various factors such as accidents or construction. In networking this could be due to congestion or bad frames (pot holes!).

Throughput <= Bandwidth

Performance Tools – Free vs. Paid

No Charge/Free Tools

Iperf	Ttcp
Jperf	Netcps
Netperf	Qcheck
Ntttcp	Ostinato
Nettcp	etc

Paid Tools

IxChariot
Spirient
Agileload
etc.

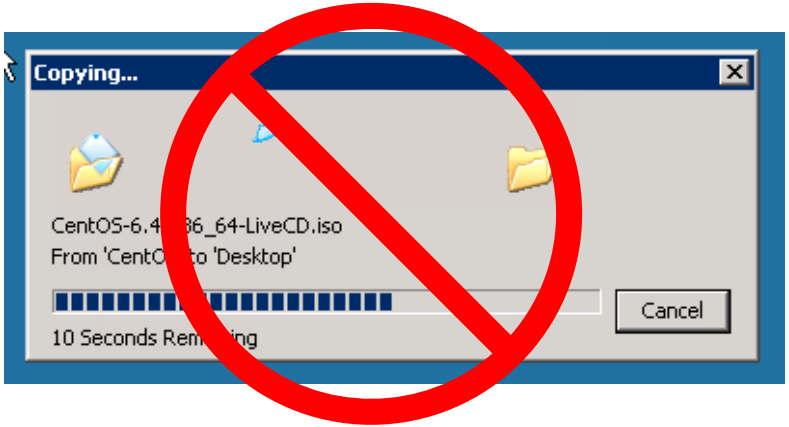
Note: All variations of ttcp/iperf report **payload** or user data rates, i.e. no overhead bytes from headers (TCP, UDP, IP, etc.) are included in the reported data rates. When comparing to "line" rates or "peak" rates, it is important to consider all of this overhead.

Tools Compared

Tool	Type	Platform	Protocols
Iperf/Jperf	Client/Server	Cross	TCP/UDP
NetPerf	Client/Server	Cross	TCP/UDP
Ntttcp	Client/Server	Windows	TCP/UDP

Performance Tools – Bad/Problem Tools

- SCP/SFTP
  - Encrypted overhead
- Windows Shares
  - ‘Chatty’ protocol.
  - Masks underlying file systems



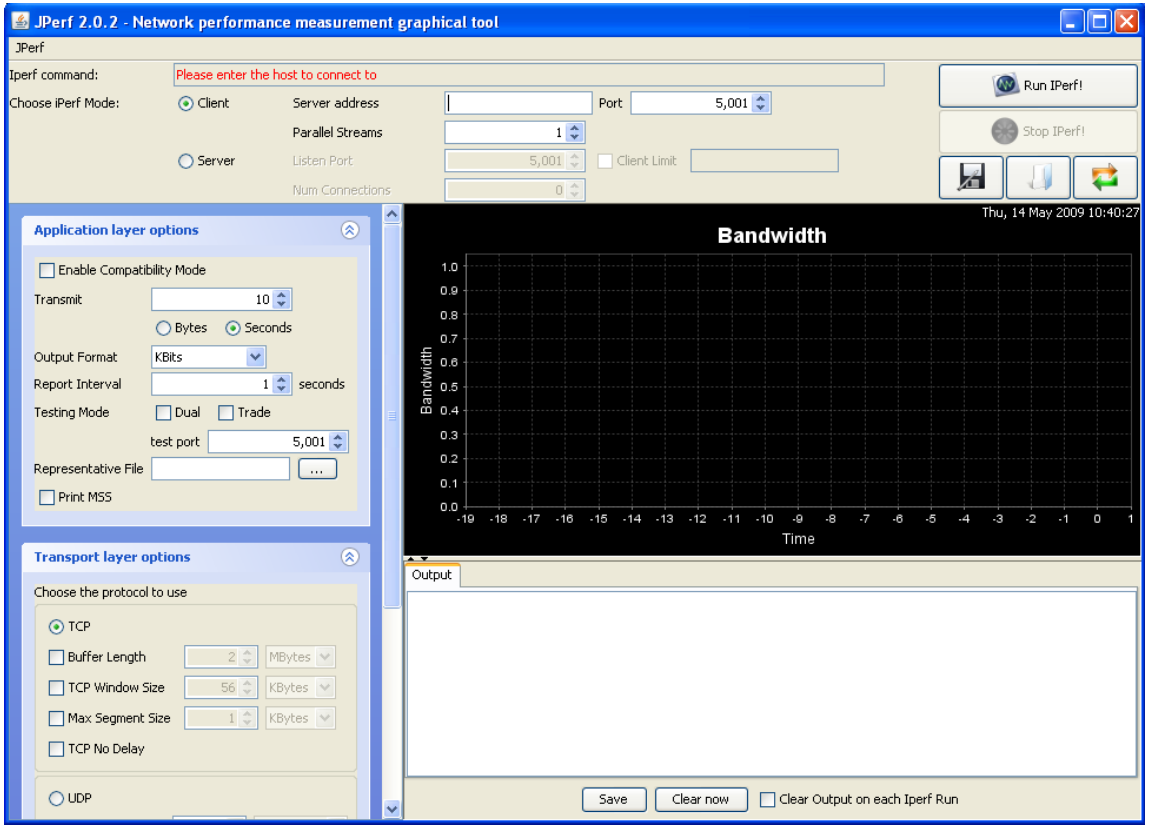
Simple Test

- Running iperf on two blades, different Chassis
- Server: iperf –s -B 192.168.10.1 -m
- Client: iperf -c 192.168.10.1 -t 300 -i 10 -m
- This will test max TCP throughput between the two nodes
- Reporting Interval every 10s for 300s duration
- Uses the default windows size
- Uses the default port of 5001
- Prints the max MTU (less headers)

IPERF Test Results

Test	Source	Receiver	MTU	Protocol	Streams	Test Parameters	Adapter Policy	BIOS	Results - Gbps
1	perf-test-1	perf-test-2	1500/1448	TCP	1	iperf -c 192.168.10.2 -m -t 120 -i 10	Linux Default	Defaults	**8.85**
2	perf-test-1	perf-test-2	1500/1448	TCP	1	iperf -c 192.168.10.2 -m -t 120 -i 10	Linux Default	Defaults	**8.87**
3	perf-test-1	perf-test-2	1500/1448	TCP	1	iperf -c 192.168.10.2 -m -t 120 -i 10	Linux Default	Defaults	**8.80**
4	perf-test-1	perf-test-2	1500/1448	TCP	2	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 2	Linux Default	Defaults	**9.35**
5	perf-test-1	perf-test-2	1500/1448	TCP	2	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 2	Linux Default	Defaults	**9.35**
6	perf-test-1	perf-test-2	1500/1448	TCP	2	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 2	Linux Default	Defaults	**9.35**
7	perf-test-1	perf-test-2	1500/1448	TCP	5	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 5	Linux Default	Defaults	**9.35**
8	perf-test-1	perf-test-2	1500/1448	TCP	5	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 5	Linux Default	Defaults	**9.35**
9	perf-test-1	perf-test-2	1500/1448	TCP	5	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 5	Linux Default	Defaults	**9.35**
10	perf-test-1	perf-test-2	1500/1448	TCP	10	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 10	Linux Default	Defaults	**9.35**
11	perf-test-1	perf-test-2	1500/1448	TCP	10	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 10	Linux Default	Defaults	**9.35**
12	perf-test-1	perf-test-2	1500/1448	TCP	10	iperf -c 192.168.10.2 -m -t 120 -i 10 -P 10	Linux Default	Defaults	**9.35**

JPerf



Baseline Testing

- Controlled environment
- Repeat tests at min. 3 times
- Test both directions Sender ⇔ Receiver
- Try different size MTU ie. Jumbo frames if using iSCSI / IP Storage.
- Ensure test duration is >3mins. Allows for TCP windowing adjustments

Monitoring Performance

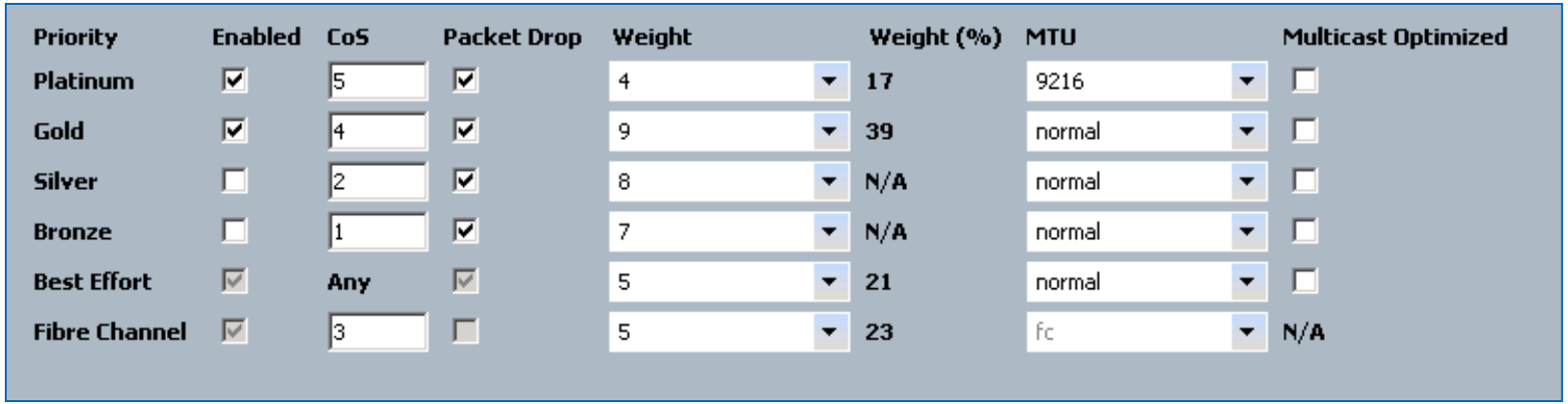
Looking For Congestion

```
UCS-A(nxos)# show interface ethernet 1/1/11 priority-flow-control
================================================================
Port                Mode Oper(VL bmap)  RxPPP       TxPPP
================================================================

Ethernet1/1/11       Auto Off           0           0
UCS-A(nxos)# show interface ethernet 1/12 priority-flow-control
================================================================
Port                Mode Oper(VL bmap)  RxPPP       TxPPP
================================================================

Ethernet1/12        Auto Off           0           0
UCS-A(nxos)#
```

Any pause frames on the FEX or Fabric Interfaces?

QoS Considerations

- CoS/QoS within UCS is simple to configure
- Needs to be configured End-to-End
- Can do more harm than good if configured incorrectly

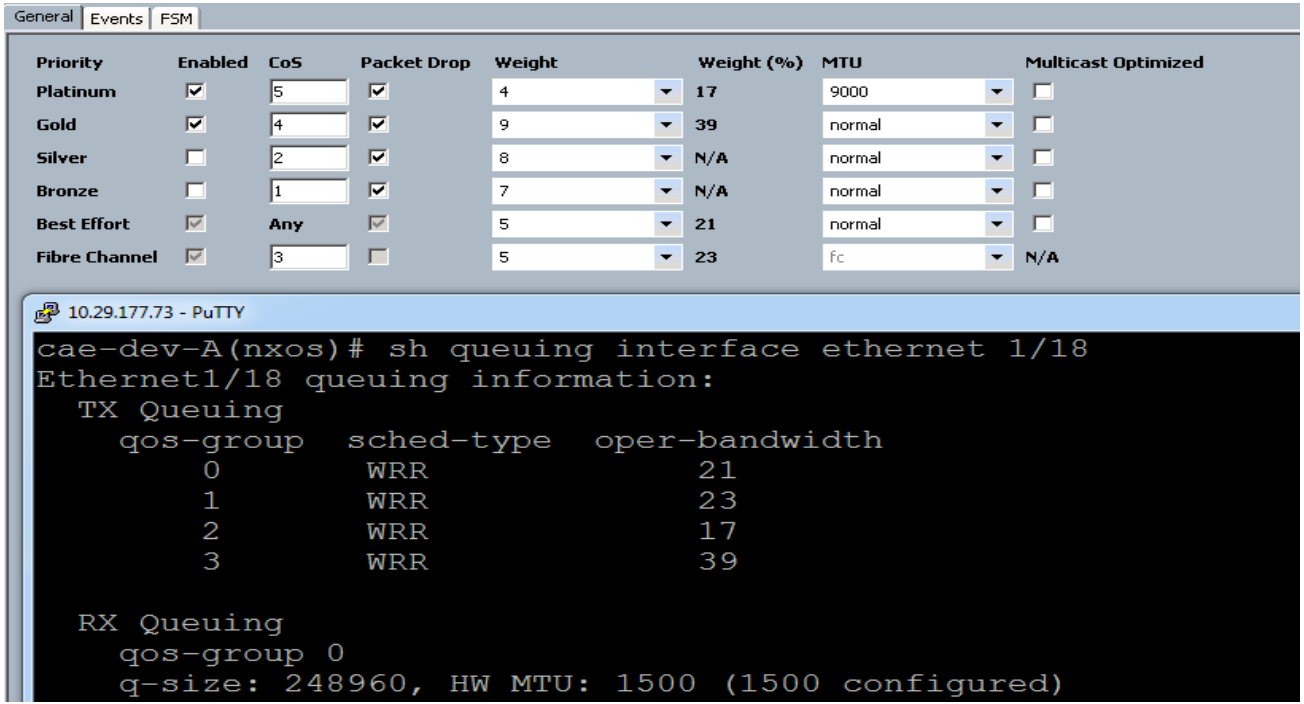


Priority	Enabled	CoS	Packet Drop	Weight	Weight (%)	MTU	Multicast Optimized
Platinum	☑	5	☑	4	17	9216	☐
Gold	☑	4	☑	9	39	normal	☐
Silver	☐	2	☑	8	N/A	normal	☐
Bronze	☐	1	☑	7	N/A	normal	☐
Best Effort	☑	Any	☑	5	21	normal	☐
Fibre Channel	☑	3	☐	5	23	fc	N/A

QoS Queing GUI vs. CLI

- Connect nxos

show queuing interface eth x/y



QoS – Misconfiguration

```
show queuing interface ethernet 1/5
Ethernet1/5 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             50
        1       WRR             50
   RX Queuing
    qos-group 0
    q-size: 360960, HW MTU: 9216 (9216 configured)
    drop-type: drop, xon: 0, xoff: 360960
    Statistics:
        Pkts received over the port             : 0
        Ucast pkts sent to the cross-bar        : 0
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 0
        Pkts sent to the port                   : 0
        Pkts discarded on ingress               : 0
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

QoS – Misconfigured

```
show queuing interface ethernet 1/5 – cont'd
qos-group 1
    q-size: 79360, HW MTU: 2158 (2158 configured)
    drop-type: no-drop, xon: 20480, xoff: 40320
    Statistics:
        Pkts received over the port             : 809739
        Ucast pkts sent to the cross-bar        : 743529
        Mcast pkts sent to the cross-bar        : 0
        Ucast pkts received from the cross-bar  : 67599
        Pkts sent to the port                   : 67599
        Pkts discarded on ingress               : 66210
        Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
```

- If QoS/CoS values aren't correctly set on both sides of a link, this could result in unnecessarily dropped frames.

Adapter Commands (VIC)

Adapter Specific Commands

- Based on the Adapter used, there are various commands we can leverage.
- Cisco VIC allows to attach to the Master Control Program (MCP) to view verbose enic stats & counters, or Fabric Layer Services (FLS) to view fnic (FC) stats & counters. We will focus on the VIC command sets.
- For Non-Cisco adapters (M71, M72, M73, M61 etc) We have a different subset of commands

VIF Details

- Connect adapter x/y/z (Chassis, Blade, Adapter)

```
UCS-A# connect adapter 1/6/1
adapter 1/6/1 # connect
adapter 1/6/1 (top):1# attach-mcp
adapter 1/6/1 (mcp):1# vnic
<snip>
---------------------------------------- -------- -------------------------
                vnic                      lif              vif
id  name            type    bb:dd.f state lif state uif  ucsm   idx vlan state
--- --------------- ------- ------- ----- --- ----- --- ----- ----- ---- -----
 13 vnic_1          enet    06:00.0 UP      2 UP    =>0   991    91    1 UP
                                                    - 1   992    84    1 UP
 14 vnic_2          enet    07:00.0 UP      3 UP    - 0   987    92    1 UP
                                                    =>1   988    85    1 UP
 15 vnic_3          enet    08:00.0 UP      4 UP    =>0   993    93    1 UP
                                                    - 1   994    86    1 UP
 16 vnic_4          fc      0a:00.0 UP      5 UP    =>1   985    87  200 UP
 17 vnic_5          fc      0b:00.0 UP      6 UP    =>0   986    94  100 UP
```

Indicates which Fabric Failover enabled interface is active

VIF Details

- Connect adapter x/y/z (Chassis, Blade, Adapter)

```
UCS-A# connect adapter 1/6/1
adapter 1/6/1 # connect
adapter 1/6/1 (top):1# attach-mcp
adapter 1/6/1 (mcp):1# vif
------- ----- ----- ---- --------------- -----
        vif
lif.uif index pri   hash state           flags
------- ----- ----- ---- --------------- -----
  2.0      91     0   91 UP              NIV, CREATED, VIFHASH, VUP, VIFINFO, DCXUP
  2.1      84     0   84 UP              NIV, CREATED, VIFHASH, VUP, STANDBY, VIFINFO, DCXUP
  3.0      92     0   92 UP              NIV, CREATED, VIFHASH, VUP, STANDBY, VIFINFO, DCXUP
  3.1      85     0   85 UP              NIV, CREATED, VIFHASH, VUP, VIFINFO, DCXUP
  4.0      93     0   93 UP              NIV, CREATED, VIFHASH, VUP, VIFINFO, DCXUP
  4.1      86     0   86 UP              NIV, CREATED, VIFHASH, VUP, STANDBY, VIFINFO, DCXUP
  5.0      94     0   94 UP              NIV, CREATED, VIFHASH, VUP, STANDBY, VIFINFO, DCXUP
  5.1      87     0   87 UP              NIV, CREATED, VIFHASH, VUP, VIFINFO, DCXUP
  6.1      88     0   88 UP              NIV, CREATED, VIFHASH, VUP, VIFINFO
  7.0      95     0   95 UP              NIV, CREATED
```

DCE (UIF) Stats

```
adapter 1/6/1 (mcp):1# dcem-macstats [UIF#]
```

```
   1061 Tx frames len == 64
    168 Tx frames 64 < len <= 127
   5647 Tx frames 128 <= len <= 255
      6 Tx frames 256 <= len <= 511
     16 Tx frames 512 <= len <= 1023
      8 Tx frames 1024 <= len <= 1518
   6906 Tx total packets
1143159 Tx bytes
   6906 Tx good packets
   1445 Tx unicast frames
   5423 Tx multicast frames
     38 Tx broadcast frames
```

```
  42954 Rx Frames 64 < len <= 127
   2644 Rx Frames 128 <= len <= 255
  85018 Rx Frames 256 <= len <= 511
     16 Rx Frames 512 <= len <= 1023
      1 Rx Frames 1024 <= len <= 1518
      1 Rx Frames 1519 <= len <= 2047
 130634 Rx total received packets
32292176 Rx bytes
 130634 Rx good packets
   1485 Rx unicast frames
  27672 Rx multicast frames
 101477 Rx broadcast frames
1143159 Rx bytes for good packets
114.638bps  Tx Rate
 3.238kbps  Rx Rate
```

IO Module Commands

IOM Commands

- Two different methods to pull IOM counters.
- Option 1:

```
UCS-A# connect iom 1
Attaching to FEX 1 ...
To exit type 'exit', to abort type '$.'
fex-1# show platform software [redwood][woodside] rate
```

- Option 2:

```
UCS-A# connect iom 1
Attaching to FEX 1 ...
To exit type 'exit', to abort type '$.'
fex-1# dbgexec woo
woo> rate
woo> help
```

Type “Ctrl+C” to exit

Produces same output

Monitoring IOM Interface Rates

- While running a load scenario between blades
- connect iom <chassis#>
- show platform software [redwood][woodside] rate

```
fex-1# show platform software woodside 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|
+--------++------------+-----------+------------++------------+-----------+------------+-------+-------+---+
| 0-BI   ||         47 |         9 |   7.94Kbps ||         42 |         8 |   8.59Kbps |    85 |   107 |   |
| 0-CI   ||          8 |         1 |   8.49Kbps ||          6 |         1 |   7.88Kbps |   644 |   801 |   |
| 0-NI3  ||    3806308 |    761261 |   9.41Gbps ||      73159 |     14631 |  11.70Mbps |  1525 |    80 |   |
| 0-NI2  ||          1 |         0 |   1.74Kbps ||          2 |         0 |   2.13Kbps |  1072 |   648 |   |
| 0-NI1  ||          1 |         0 |   1.74Kbps ||          9 |         1 |   5.74Kbps |  1072 |   378 |   |
| 0-NI0  ||          1 |         0 |   1.74Kbps ||          2 |         0 |   2.13Kbps |  1072 |   648 |   |
| 0-HI19 ||      73113 |     14622 |  11.69Mbps ||    3806252 |    761250 |   9.41Gbps |    79 |  1525 |   |
| 0-HI11 ||          8 |         1 |   4.04Kbps ||          0 |         0 |   0.00 bps |   296 |     0 |   |
| 0-HI7  ||          1 |         0 | 440.00 bps ||          0 |         0 |   0.00 bps |   259 |     0 |   |
+--------++------------+-----------+------------++------------+-----------+------------+-------+-------+---+
```

Monitoring IOM Interface Stats

- connect iom <chassis#>
- show platform software [redwood][woodside] rmon 0 <HIF# | NIF#>

```
fex-1# show platform software woodside rmon 0 ni3
```

```
fex-1# show plat sof woodside rmon 0 ni3
+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+
| TX                    |        Current        |         Diff          | RX                    |        Current        |         Diff          |
+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+
| TX_PKT_LT64           |                     0 |                     0 | RX_PKT_LT64           |                     0 |                     0 |
| TX_PKT_64             |                 15371 |                     1 | RX_PKT_64             |                    14 |                     0 |
| TX_PKT_65             |              17275405 |                     0 | RX_PKT_65             |              93398689 |                     2 |
| TX_PKT_128            |                903036 |                     1 | RX_PKT_128            |                481998 |                     0 |
| TX_PKT_256            |               2391483 |                     0 | RX_PKT_256            |                106504 |                     0 |
| TX_PKT_512            |               2550287 |                     0 | RX_PKT_512            |                530444 |                    27 |
| TX_PKT_1024           |               3931780 |                    25 | RX_PKT_1024           |                 32774 |                     0 |
| TX_PKT_1519           |            4163102089 |                     0 | RX_PKT_1519           |             438772852 |                     0 |
| TX_PKT_2048           |                     0 |                     0 | RX_PKT_2048           |                     0 |                     0 |
| TX_PKT_4096           |                     0 |                     0 | RX_PKT_4096           |                     0 |                     0 |
| TX_PKT_8192           |                     0 |                     0 | RX_PKT_8192           |                     0 |                     0 |
| TX_PKT_GT9216         |                     0 |                     0 | RX_PKT_GT9216         |                     0 |                     0 |
| TX_PKTTOTAL           |            4190169451 |                    27 | RX_PKTTOTAL           |             533323275 |                    29 |
| TX_OCTETS             |         6370843967079 |                 27006 | RX_OCTETS             |          678490434653 |                 17636 |
| TX_PKTOK              |            4190169451 |                    27 | RX_PKTOK              |             533323275 |                    29 |
| TX_UCAST              |            4189675980 |                     2 | RX_UCAST              |             531847545 |                     2 |
| TX_MCAST              |                493344 |                    25 | RX_MCAST              |               1474949 |                    27 |
| TX_BCAST              |                   127 |                     0 | RX_BCAST              |                   781 |                     0 |
| TX_VLAN               |                     0 |                     0 | RX_VLAN               |                     0 |                     0 |
| TX_PAUSE              |                     0 |                     0 | RX_PAUSE              |                     0 |                     0 |
| TX_USER_PAUSE         |                     0 |                     0 | RX_USER_PAUSE         |                     0 |                     0 |
| TX_FRM_ERROR          |                     0 |                     0 |                       |                       |                       |
|                       |                       |                       | RX_OVERSIZE           |                     0 |                     0 |
|                       |                       |                       | RX_TOOLONG            |                     0 |                     0 |
|                       |                       |                       | RX_DISCARD            |                     0 |                     0 |
|                       |                       |                       | RX_UNDERSIZE          |                     0 |                     0 |
|                       |                       |                       | RX_FRAGMENT           |                     0 |                     0 |
|                       |                       |                       | RX_CRC_NOT_STOMPED    |                     0 |                     0 |
|                       |                       |                       | RX_CRC_STOMPED        |                     0 |                     0 |
|                       |                       |                       | RX_INRANGEERR         |                     0 |                     0 |
|                       |                       |                       | RX_JABBER             |                     0 |                     0 |
| TX_OCTETSOK           |         6370843967079 |                 27006 | RX_OCTETSOK           |          678490434653 |                 17636 |
+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+
fex-1#
```

- Note these commands return a “snapshot” of the system. Repeat a few times and monitor the “Diff” columns to view incremental changes

Monitoring IOM Interface Drops

- connect iom <chassis#>
- show platform software [redwood][woodside] drops 0 <HIF# | NIF#>

```
fex-1# show platform software woodside drops 0 ni3
fex-1# show plat soft woodside drops 0 HI3
WOO_BI_CNT_RX_FWD_DROP                    [40204]: 93
WOO_HI_CT_CNT_MUX_TX_FLUSHED              [f1648]: 1 HI7
WOO_HI_CT_CNT_MUX_TX_FLUSHED              [271648]: 2 HI31
fex-1# show plat soft woodside drops 0 NI1
WOO_BI_CNT_RX_FWD_DROP                    [40204]: 0
WOO_HI_CT_CNT_MUX_TX_FLUSHED              [f1648]: 1 HI7
WOO_HI_CT_CNT_MUX_TX_FLUSHED              [271648]: 2 HI31
```

Monitoring IOM Interface Logs

- connect iom <chassis#>
- show platform software [redwood][woodside] elog

```
fex-1# show platform software woodside elog
06/27/2013 18:59:55.483836 - 0-NI0 : SFP+ Inserted
06/27/2013 18:59:55.519156 - 0-NI1 : SFP+ Inserted
06/27/2013 18:59:55.552643 - 0-NI2 : SFP+ Inserted
06/27/2013 18:59:55.586038 - 0-NI3 : SFP+ Inserted
06/27/2013 18:59:55.619470 - 0-NI4 : SFP+ Inserted
06/27/2013 18:59:55.652929 - 0-NI5 : SFP+ Inserted
06/27/2013 18:59:55.686370 - 0-NI6 : SFP+ Inserted
06/27/2013 18:59:55.719795 - 0-NI7 : SFP+ Inserted
06/27/2013 18:59:58.243035 - 0-NI0 : Admin state changed to Enbl
06/27/2013 18:59:58.265628 - 0-NI1 : Admin state changed to Enbl
06/27/2013 18:59:58.290202 - 0-NI2 : Admin state changed to Enbl
<snip>
```

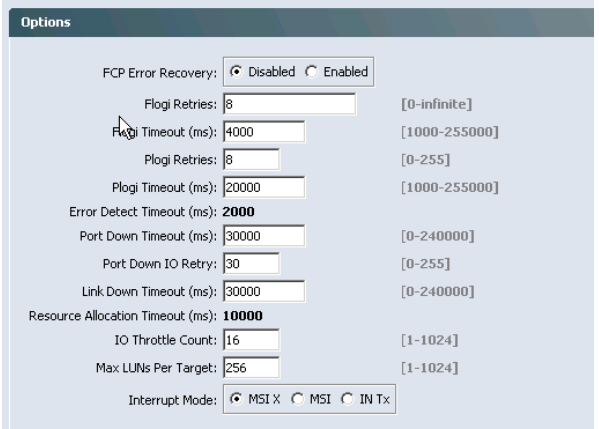
"iPerf testing between the VM's looks good. It looks like a storage problem.." Network Administrator

TAC Case
March 2013

SAN Performance

SAN Performance

- Most SAN related issues are due to Array limitations more often than host side.
  - Engage SAN Vendor
- Default Queues are set according to OS vendor recommendations
- Rx/Tx Queues can be adjusted but not recommended unless application or storage array vendor recommended



What To Look For

- Are seeing the issue with only certain hosts?
- If so, are there any commonalities between these hosts?
  - Adapter model
  - Driver & Firmware Versions
  - Chassis ID
  - FC uplink Pinning

What To Look For

- B2B Credit depletion/exhaustion

```
UCS-A(nxos)# show int fc1/33 bbcredit
fc1/33 is trunking
    Transmit B2B Credit is 250
    Receive B2B Credit is 16
    Receive B2B Credit performance buffers is 0
      16 receive B2B credit remaining
      250 transmit B2B credit remaining
      0 low priority transmit B2B credit remaining

UCS-A(nxos)# show int fc1/33 counters | i transitions
     0 BB credit transitions from zero
```

What To Look For

- Counters: Drop, Discards, Errors (CRC)

```
UCS-A(nxos)# show int fc1/33 counters
fc1/33
    1 minute input rate 88 bits/sec, 11 bytes/sec, 0 frames/sec
    1 minute output rate 88 bits/sec, 11 bytes/sec, 0 frames/sec
    401580 frames input, 22505468 bytes
      0 discards, 0 errors, 0 CRC
      0 unknown class, 0 too long, 0 too short
    401611 frames output, 22513040 bytes
      0 discards, 0 errors
    0 input OLS, 1 LRR, 0 NOS, 0 loop inits
    1 output OLS, 1 LRR, 0 NOS, 0 loop inits
    0 link failures, 0 sync losses, 0 signal losses
     0 BB credit transitions from zero
      16 receive B2B credit remaining
      250 transmit B2B credit remaining
      0 low priority transmit B2B credit remaining
```

What To Look For

- Transceiver Info

```
UCS-A(nxos)# show int fc1/33 transceiver detail

fc1/33 sfp is present
    name is CISCO-FINISAR
    part number is FTLF8524P2BNL-C2
    revision is B
    serial number is FNS104618KP
    FC Transmitter type is short wave laser w/o OFC (SN)
    FC Transmitter supports intermediate distance link length
    Transmission medium is multimode laser with 62.5 um aperture (M6)
    Supported speeds are - Min speed: 1000 Mb/s, Max speed: 4000 Mb/s
    Nominal bit rate is 4300 MBits/sec
    Link length supported for 50/125mm fiber is 150 m(s)
    Link length supported for 62.5/125mm fiber is 70 m(s)
    cisco extended id is unknown (0x0)

    No tx fault, no rx loss, in sync state, diagnostic monitoring type is 0x68
    SFP Diagnostics Information:
-------------------------------------------------------------------------------
                                    Alarms                  Warnings
                                  High        Low         High          Low
-------------------------------------------------------------------------------
  Temperature  40.92 C          89.00 C     -9.00 C     85.00 C       -5.00 C
  Voltage       3.29 V           3.60 V      3.00 V      3.50 V        3.10 V
  Current       7.67 mA         17.00 mA     1.00 mA    14.00 mA       2.00 mA
  Tx Power     -4.37 dBm         1.00 dBm  -13.57 dBm   -3.00 dBm     -9.51 dBm
  Rx Power     -4.93 dBm         4.00 dBm  -21.55 dBm    0.00 dBm    -16.99 dBm
  Transmit Fault Count = 0
-------------------------------------------------------------------------------
  Note: ++  high-alarm; +  high-warning; --  low-alarm; -  low-warning
```

SAN Performance Tools – Free vs. Paid

No Charge/Free Tools

- **dd**
- **iometer**
- **SQLio**
- copy/cp

Paid Tools

Solarwinds

Spirient

SAN Vendor tools

etc.

Simple Test – dd On Linux

- ‘dd’
  - Widely available
  - Highly customisable

Example:

‘Input File’ ‘Output File’ ‘Block Size’ ‘Sync Data before exit’

```
[root@localhost ~]# dd if=/dev/zero of=/root/file.big bs=1M count=1000 conv=fdatasync
1000+0 records in
1000+0 records out
1048576000 bytes (1.0 GB) copied, 0.830429 s, 1.3 GB/s
```

Other Usage:

`if=/dev/urandom` **Random Data**

Cisco live!

"Disk/LUN performance is fast and we don't see any problems on the Array side" Storage Administrator

TAC Case

March 2013

BIOS Settings & Performance Impact

BIOS Settings

- Each generation of processor will add new chipset features
- BIOS tokens are added to manage BIOS settings from UCSM (BIOS Policy)
- Adjustments to these settings should only be made by the recommendation of the OS or platform vendor
- Many times it's a decision between performance and power efficiencies. Many settings are default for balanced power saving.

Intel SpeedStep / SpeedBoost

- SpeedStep allows the CPU's clock frequency to be adjusted in real time.
- During period of light load, the CPU frequency is lowered thus lowering the power usage.
- SpeedBoost goes to the opposite extreme and allows the system to overclock itself assuming there is available power
- Useful for latency sensitive workloads on high utilisation system.
- Dependent on SpeedStep being enabled.

Processor C3 and C6 States

- These are two states or levels of halt & sleep the processor can enter into when not busy.
- Used to improve power efficiency
- Drawback is there is added overhead when processors “Wake up” and exit these states.
- C states range from 0 – 6.
  - 0 is a fully powered CPU
  - 1 is the halt state. The CPU is not currently executing instructions.
  - 3 is deep sleep. All internal clocks are stopped
  - 6 is deep power down. Reduces internal voltage
- C states are transitional.
- For max performance, these states can be disabled.

Hyperthreading

- Enables additional parallelisation of processing by allowing two processes to leverage the same resource
- Useful to applications that can take advantage of multi-threaded instructions
- Requires Operating System (OS) support.
- If your OS has not been optimised for Hyperthreading, it should be disabled.
- Recommendation to run baseline test against your applications with HT enabled & disabled to gauge impact.

Memory Performance

- All UCS memory sold is dual voltage memory.
- Memory can run at 1.35V or 1.5V
- Voltage affects the speed at which DIMMs operate, 800Mhz – 1600Mhz+
- Requires CPU to support the max DIMM speed
- BIOS setting for Power Saving or Performance set via BIOS policy

Non Uniform Memory Access (NUMA)

- Addresses the latest server chipset designs
- Each processor has access to dedicated banks of memory
- Allows the system to access memory belonging to the other CPUs but adds a “cost” to doing do, minimising this action when necessary.
- Confirm with OS vendor support
- Most hypervisors recommend enabling

"Network, Disk and Compute are all clear. We only see issues performing the Mailbox Replication." Admin Team

TAC Case

March 2013




Recap

What Have We Learned

- Understanding of the various hops & interfaces within the UCS
- The affect various BIOS settings can have on performance
- How to trace the exact path for VIF through FI uplink egress
- Where to look for congestion & throughput on various components
- Importance of baseline testing & Network documentation




Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2014 Polo Shirt!

Complete your Overall Event Survey and 5 Session Evaluations.

- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

Polo Shirts can be collected in the World of Solutions on Friday 21 March 12:00pm - 2:00pm

Learn online with Cisco Live!

Visit us online after the conference for full access to session videos and presentations.
www.CiscoLiveAPAC.com


CISCO
TM