



*LET'S
BUILD
TOMORROW
TODAY*

Troubleshooting Nexus 5600/6000 series Switches

Prashanth Krishnappa CCIE#18057(DC, R&S)

BRKDCT-3100

Session Goals

- Learn about the Nexus 5600/6000 and NX-OS troubleshooting approach
- Learn about common Nexus 5600/6000 issues and how to troubleshoot them
- Learn about tools available in NX-OS to troubleshoot common issues



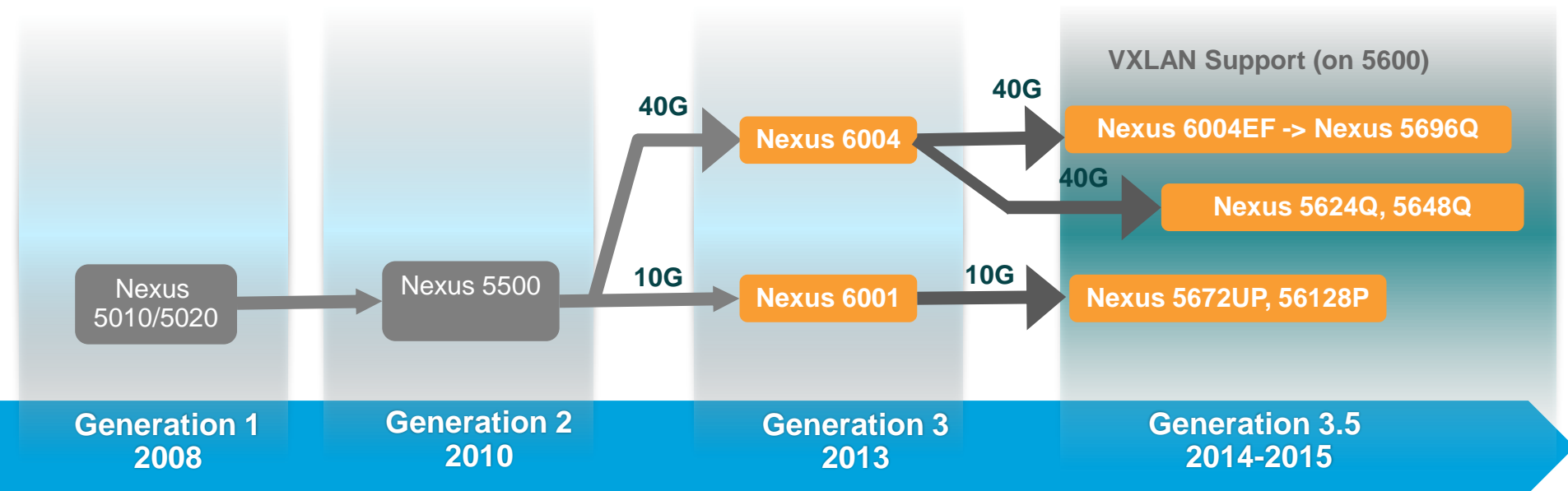
Related sessions

- **BRKARC-3452** - Cisco Nexus 5600/6000 Switch Architecture(6/11 10am)
- **BRKARC-3454** - In-depth and personal with the Cisco Nexus 2000 Fabric Extender Architectures, Features, and Topologies(6/9 8am, 6/10 1pm)
- **BRKDCT-2458** - Nexus 9000/7000/6000/5000 Operations and Maintenance Best Practices(6/9 8am)
- **BRKDCT-3346** - End-to-End QoS Implementation and Operation with Cisco Nexus Switches(6/9 1pm)
- **BRKDCT-1890** - Network visibility using advanced Analytics in Nexus switches(6/9 3:30pm)
- **BRKDCT-2378** - VPC Best Practices and Design on NX OS(6/8 10am)
- **BRKDCT-3313** - FabricPath Operation and Troubleshooting(6/10 8am)

Agenda

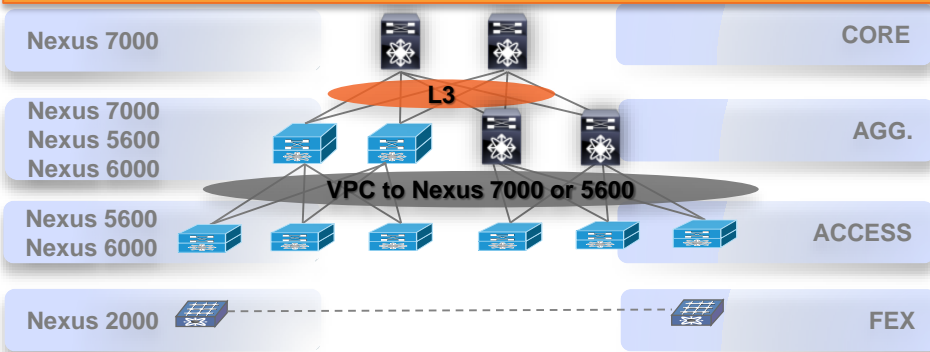
- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queueing
 - ELAM

Nexus 5000/6000 Evolution

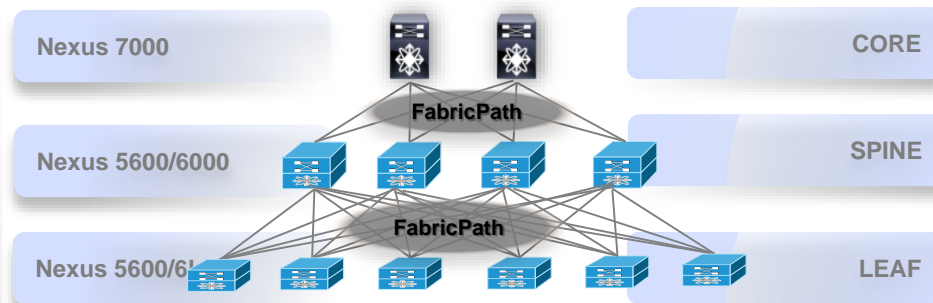


Nexus 5600/6000 Use-cases

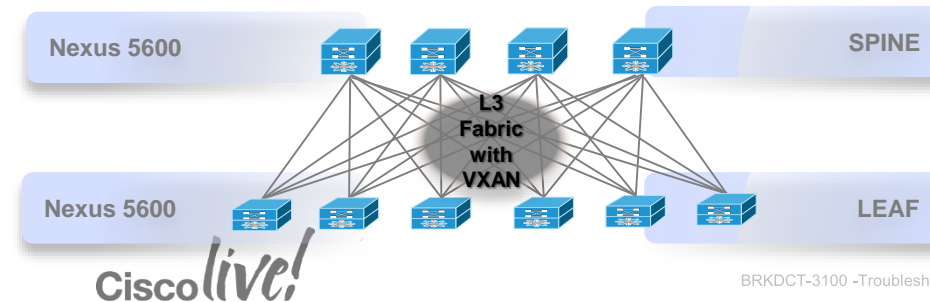
Classic 3-Tier with FEX



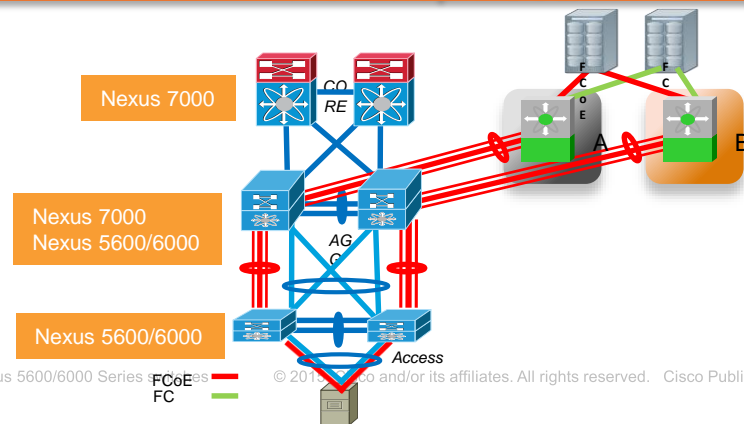
Fabricpath



VXLAN Fabric (5600 only)



Multi hop FCoE



NX-OS operation tips

- CLI list and grep

```
esc-5672-left# show cli list | grep "debug spanning"
no debug spanning-tree all
no debug spanning-tree bpdu_rx interface <if> tree <int:1-4095>
no debug spanning-tree bpdu_rx interface <if>
no debug spanning-tree bpdu_rx tree <int:1-4095>
no debug spanning-tree bpdu_rx
<snip>
```

```
esc-5672-left# show queuing interface | egrep -i ethernet1|discard|qos-group | exclude sched
Ethernet1/1 queuing information:
  qos-group 0
    Pkts discarded on ingress          : 2379491
Ethernet1/2 queuing information:
  qos-group 0
    Pkts discarded on ingress          : 0
Ethernet1/3 queuing information:
  qos-group 0
    Pkts discarded on ingress          : 367
<snip>
```


FSM

- NX-OS records the **finite state machine** for many important processes
- Using this event-history of FSM states and triggers, debugging can be done after a problem has occurred. Sooner the better
- Important to compare timestamps and watch for inter-process communication.
- Some common processes:
 - Ethpc – ethernet port client: responsible for talking to the mac and phy
 - Ethpm – ethernet port manager: responsible for translating between configuration and ethpc. ethpc would inform ethpm that link is up, and then ethpm will proceed to give instructions on what the configuration is for the port
 - Port-channel – port-channeling process responsible for aggregating physical links
 - lacp – 802.3ad standard for aggregating links

FSM

Example: An LACP Po12 flapped and we are tasked to find out why

```
2015 Apr 18 08:06:03 esc-5672-left %ETH_PORT_CHANNEL-5-FOP_CHANGED: port-channel12: first
operational port changed from Ethernet1/12 to none
2015 Apr 18 08:06:03 esc-5672-left %ETHPORT-5-IF_DOWN_PORT_CHANNEL_MEMBERS_DOWN: Interface
port-channel12 is down (No operational members)
<snip>
2015 Apr 18 08:06:18 esc-5672-left %ETH_PORT_CHANNEL-5-PORT_UP: port-channel12:
Ethernet1/12 is up
2015 Apr 18 08:06:18 esc-5672-left %ETH_PORT_CHANNEL-5-FOP_CHANGED: port-channel12: first
operational port changed from none to Ethernet1/12
```

FSM

EthPM event-history

```
esc-5672-left# show system internal ethpm event-history interface ethernet 1/12
369) FSM:<Ethernet1/12> Transition at 463177 usecs after Sat Apr 18 08:06:03 2015
    Previous state: [ETH_PORT_FSM_ST_BUNDLE_MEMBER_UP]
    Triggered event: [ETH_PORT_FSM_EV_EXTERNAL_REINIT_NO_FLAP_REQ]
    Next state: [FSM_ST_NO_CHANGE]
```

LACP event-history

```
esc-5672-left# show lacp internal event-history interface ethernet 1/12
64) FSM:<Ethernet1/12> Transition at 462569 usecs after Sat Apr 18 08:06:03 2015
    Previous state: [LACP_ST_PORT_MEMBER_COLLECTING_AND_DISTRIBUTING_ENABLED]
    Triggered event: [LACP_EV_RECEIVE_PARTNER_PDU_TIMED_OUT]
    Next state: [LACP_ST_PORT_IS_DOWN_OR_LACP_IS_DISABLED]
```

Po12 flapped due to lack of LACP PDU timeout

NX-OS operation tips

- debug logfile
- Print debugs to a file rather than terminal

```
esc-5672-left# debug logfile CiscoLive_debugs
esc-5672-left# debug spanning-tree bpdu_rx tree 10

esc-5672-left# dir log:
    2184   Apr 16 14:08:51 2015   CiscoLive_debugs
        31   Apr 14 14:38:42 2015   dmesg
         0   Apr 14 14:39:10 2015   libfipf.3842

esc-5672-left# undebug all
esc-5672-left# show debug logfile CiscoLive_debugs
2015 Apr 16 14:08:45.234274 stp: BPDU RX: vb 1 vlan 10, ifi 0x1600023b (port-channel572)
2015 Apr 16 14:08:45.234359 stp: BPDU Rx: Received BPDU on vb 1 vlan 10 port port-channel572 pkt_len 64
bpdu_len 42 netstack flags 0x80000ed enc_type sstp
2015 Apr 16 14:08:45.234468 stp: RSTP(10): msg on port-channel572

esc-5672-left# copy log:CiscoLive_debugs tftp:
```

Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

MTS

- NX-OS uses Message and Transaction Service(MTS) to communicate between processes.
- Useful to check when troubleshooting
 - high CPU
 - unresponsive CLI / timeout
 - control-plane disruption
- When troubleshooting a process, we may look for specific MTS messages queued.
- MTS messages may be coming in too fast, or there could be a message stuck at the top of the queue

MTS

- Observed impact.. Interface configuration timing out
- FEXes offline events

```
esc-6001-Leaf-B(config)# int ethernet 101/1/1
esc-6001-Leaf-B(config-if)# shutdown
```

Please check if command was successful using appropriate show commands

```
esc-6001-Leaf-B# show system internal mts buffers
MTS buffers in use = 89
esc-6001-Leaf-B# show system internal mts buffers summary
```

node	sapno	recv_q	pers_q	npers_q	log_q
sup	175	0	81	0	0
sup	619	0	0	0	2
sup	284	0	4	0	0
sup	179	0	2	0	0
sup	392	0	2	0	0

```
esc-6001-Leaf-B#
```

MTS

- persistent queue is generally seen growing old

```
esc-6001-Leaf-B# show system internal mts buffers details
```

Node/Sap/queue	Age (ms)	SrcNode	SrcSAP	DstNode	DstSAP	OPC	MsgId	MsgSize	RRToken	Offset	
sup/175/pers	2535070	0x3B02	181	0x101	0	8182	0x3fda8	74	0	0x2beee04	
sup/175/pers	2534659	0x3D02	181	0x101	0	8182	0x35227	74	0	0x2beec04	
sup/175/pers	2509199	0x101	450	0x101	175	61466	0x6e2dc5		1970	0x6e2dc5	0x2aad004
sup/175/pers	2509197	0x101	450	0x101	175	61466	0x6e2dc7		1970	0x6e2dc7	0x2bef004

<snip>

```
esc-6001-Leaf-B# show system internal mts sup sap 175 description
```

Ethpm SAP

```
esc-6001-Leaf-B# show system internal mts sup sap 181 description
```

Ethpc SAP

```
esc-6001-Leaf-B# show system internal mts sup sap 450 description
```

Mcecm SAP

```
esc-6001-Leaf-B# show system internal mts opcodes | grep 8182
```

8182 MTS_OPC_LINK_EVENT_DOWN: SYNC SEQNO

```
esc-6001-Leaf-B# show system internal mts opcodes | grep 61466
```

61466 MTS_OPC_ETHPM_API_PORT_REINIT: SYNC

- Ethpm is deadlocked on EthPC during a link down event
- Root cause was a bug getting triggered during A/A FEX link flap during vPC delay restore (CSCuh63423)

Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

Crashes

- Nexus 5600/6000 is a single-supervisor platform; critical processes require a system restart upon a crash.
- Some processes in Nexus 5600/6000 are able to be restarted in a stateful manner.
- NX-OS attempts to create a core file with information helpful to aid in finding and fixing the problem
- A syslog message is sent prior to crash(Saved in NVRAM log)

```
2015 Apr 16 10:38:48 esc-6004EF %$ VDC-1 %$ %SYSMGR-2-SERVICE_CRASHED: Service
"Century" (PID 3751) hasn't caught signal 6 (core will be saved).
```

- show version, show system reset-reason

```
esc-6004EF# show system reset-reason
----- reset reason for Supervisor-module 1 (from Supervisor in slot 1) ---
1) At 883318 usecs after Thu Apr 16 10:38:48 2015
   Reason: Reset triggered due to HA policy of Reset
   Service: Century hap reset
   Version: 7.1(1)N1(1)
```

Crashes

- Extracting core file

```
esc-6004EF# show cores
```

VDC	Module	Instance	Process-name	PID	Date (Year-Month-Day Time)
---	-----	-----	-----	-----	-----
1	18	1	Century	3751	2015-04-16 10:46:09

- Core in volatile memory. Copy off core file for analysis before further reload/reboots

```
esc-6004EF# copy core:?
```

```
core: Enter URL "core://<module-number>/<process-id>[/instance-num]"
```

```
esc-6004EF# copy core://18/3751/1 ?
```

```
bootflash: Select destination filesystem
```

```
ftp: Select destination filesystem
```

```
scp: Select destination filesystem
```

```
sftp: Select destination filesystem
```

```
tftp: Select destination filesystem
```

Crashes

```
esc-6004EF# show processes log
Process          PID      Normal-exit  Stack  Core  Log-create-time
-----
Century          3751             N      N      N  Thu Apr 16 10:38:48 2015
...
```

```
esc-6004EF# show processes log pid 3751
Service: Century
Description: Century USD
Executable: /isan/bin/century

Started at Thu Apr 16 10:20:02 2015 (960293 us)
Stopped at Thu Apr 16 10:38:48 2015 (87531 us)
Uptime: 18 minutes 46 seconds

Start type: SRV_OPTION_RESTART_STATELESS (23)
Death reason: SYSMGR_DEATH_REASON_FAILURE_SIGNAL (2)
...
```

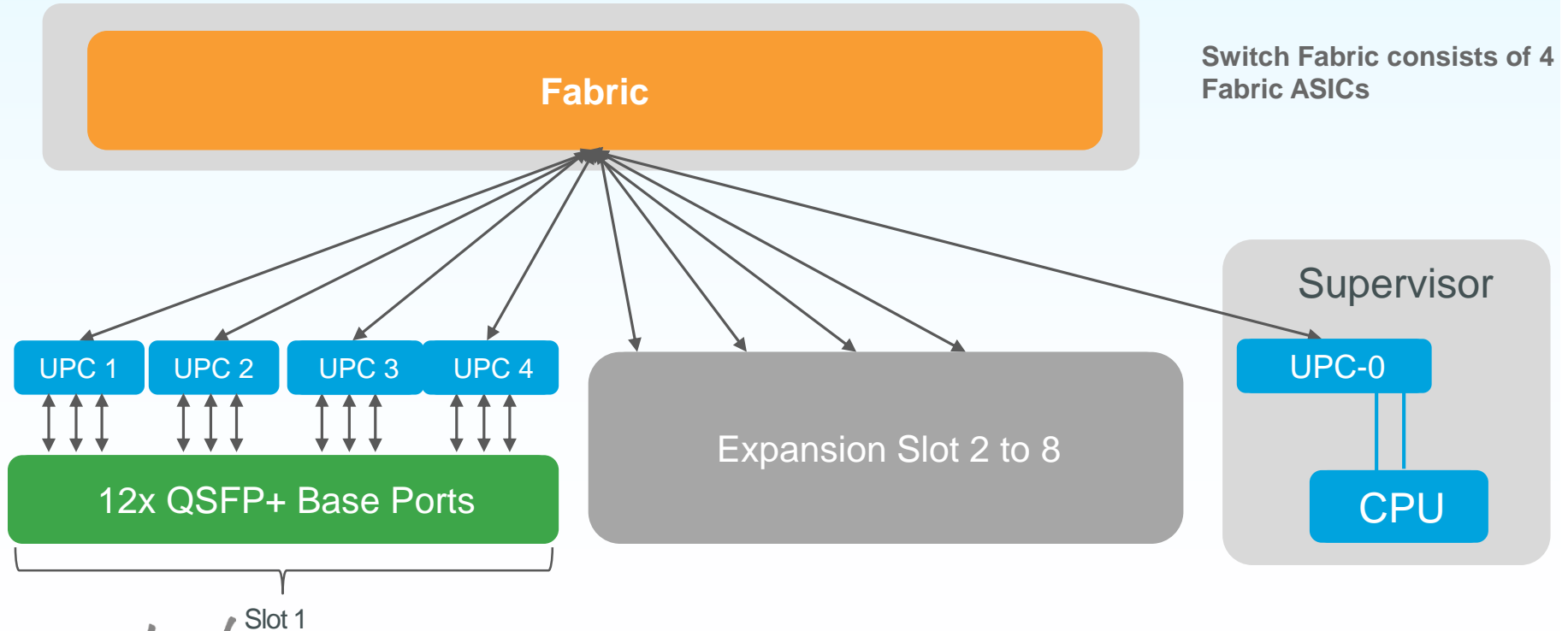
Crashes

- In addition to the core file, these details are essential:
 - Was there a configuration change?
 - Was there a physical topology change?
 - Can this be reproduced?
 - Was there a recent upgrade?
 - Are you using an uncommon configuration?
- The more details pointing to a root cause, the more feasible it is to find the problem, provide a workaround, and a fix.
- Root cause for Century process restart was due to OIR of 100G LEM into 6004 with wrong BIOS(documented in release notes)

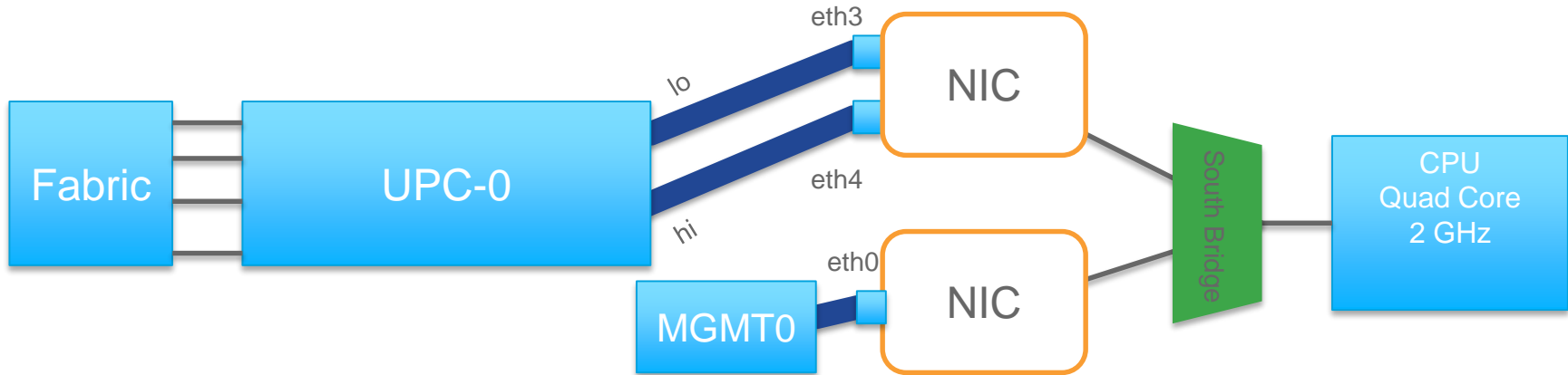
Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

Cisco Nexus 5696Q Internal Architecture



CPU Architecture



```
esc-5672-left# show hardware internal bigsur all-ports | grep sup
sup1      |0   |0   |0   | 0 - |48   |b3   |en |dn |15010000|pass| 0.00
sup0      |0   |0   |0   | 1 - |49   |b3   |en |dn |15020000|pass| 0.00
esc-5672-left#
```


Control Plane Policing(CoPP)

- CoPP protects CPU and protocols from inadvertent and/or malicious traffic attacks
- CoPP is on by default for eth3 and eth4 interfaces

```
esc-5672-left# show policy-map interface control-plane
```

```
Control Plane
```

```
  service-policy input: copp-system-policy-default
```

```
    class-map copp-system-class-igmp (match-any)
```

```
      match protocol igmp
```

```
      police cir 1024 kbps , bc 65535 bytes
```

```
        conformed 264 bytes; action: transmit
```

```
        violated 0 bytes;
```

```
    class-map copp-system-class-pim-hello (match-any)
```

```
      match protocol pim
```

```
      police cir 1024 kbps , bc 4800000 bytes
```

```
        conformed 858 bytes; action: transmit
```

```
        violated 0 bytes;
```

```
<snip>
```

Control Plane Policing(CoPP)

- CoPP policy in hardware

```
esc-5672-left# show platform afm info copp-tbls
```

```
COPP table
```

```
-----
```

```
* mgmt/ipv6-mgmt - http,snmp,ntp,telnet,ssh,ftp
```

```
** excp/ipv6-excp - mtu failure, martian address, punt, 13 header errors, icmp errors
```

Policer-Num	Name	CIR	Burst	Passed-Bytes	Dropped-Bytes
-------------	------	-----	-------	--------------	---------------

```
-----
```

0	default	64000	6250	116126770176	84197084277
---	---------	-------	------	--------------	-------------

1	stp	2500000	4687	275591604	0
---	-----	---------	------	-----------	---

2	lACP	128000	4687	279347880	0
---	------	--------	------	-----------	---

```
<snip>
```

21	mgmt/ipv6-mgmt*	1500000	4687	1122729398	0
----	-----------------	---------	------	------------	---

23	arp/ipv6-nd	8000	3515	4992509684	879588302
----	-------------	------	------	------------	-----------

```
<snip>
```

27	hsrp vrrp/ipv6-hsrp	128000	250	3634691228	173000820
----	---------------------	--------	-----	------------	-----------

Control Plane Policing(CoPP)

- Pre-Defined CoPP templates
 - Default CoPP Policy (copp-system-policy-default)
 - Scaled Layer 2 CoPP Policy (copp-system-policy-scaled-l2)
 - Scaled Layer 3 CoPP Policy (copp-system-policy-scaled-l3)
 - Customized CoPP Policy (copp-system-policy-customized)
- Individual/custom class-maps cannot be attached to customized CoPP policy
- CoPP template/policy can only be changed but not completely removed

```
esc-5672-left(config)# control-plane
esc-5672-left(config-cp)# service-policy input copp-system-policy-customized
esc-5672-left(config-cp)# show copp status
Last Config Operation: service-policy input copp-system-policy-customized
Last Config Operation Timestamp: 05:31:55 EDT Jun  4 2015
Last Config Operation Status: Success
Policy-map attached to the control-plane: copp-system-policy-customize
```

High CPU

- Hopefully you have a baseline to compare the current CPU trends with a known nominal state
- Always gather 3 commands repeating frequently
 - `show processes cpu sort | exclude 0.0`
 - `show system resources`
 - `show processes cpu history`

High CPU

- Note the difference between *, maximum CPU and #, average CPU
- Focus on extended high average CPU periods

```
esc-5672-left# show processes cpu history
```

```
      1  1      1      1      1  1      11
789509607796857706878950694778698849688895079850886958858500
753105000482598603786430941227125016911055026100692801248500
100   ** * *      *      *      *      * * * * *      **
 90   ** ** * * * * * * * * * * * * * * * * * * * * * *
 80   *** ** * * * * * * * * * * * * * * * * * * * * *
 70   *** ** ***** * * * * * * * * * * * * * * * * *
 60   *** ***** * * * * * * * * * * * * * * * * *
 50   ***** * * * * * * * * * * * * * * * * *
 40   ***** * * * * * * * * * * * * * * * * *
 30   ***** * * * * * * * * * * * * * * * * *
 20   *#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
 10   #####*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
 0...5...1...1...2...2...3...3...4...4...5...5...
      0      5      0      5      0      5      0      5
      CPU% per minute (last 60 minutes)
      * = maximum CPU%   # = average CPU%
```

High CPU

```
esc-5672-left# show system resources
```

```
Load average:   1 minute: 0.95   5 minutes: 1.54   15 minutes: 1.46
```

```
Processes      :   468 total, 1 running
```

```
CPU states    :   26.7% user,   26.7% kernel,   46.5% idle
```

```
<snip>
```

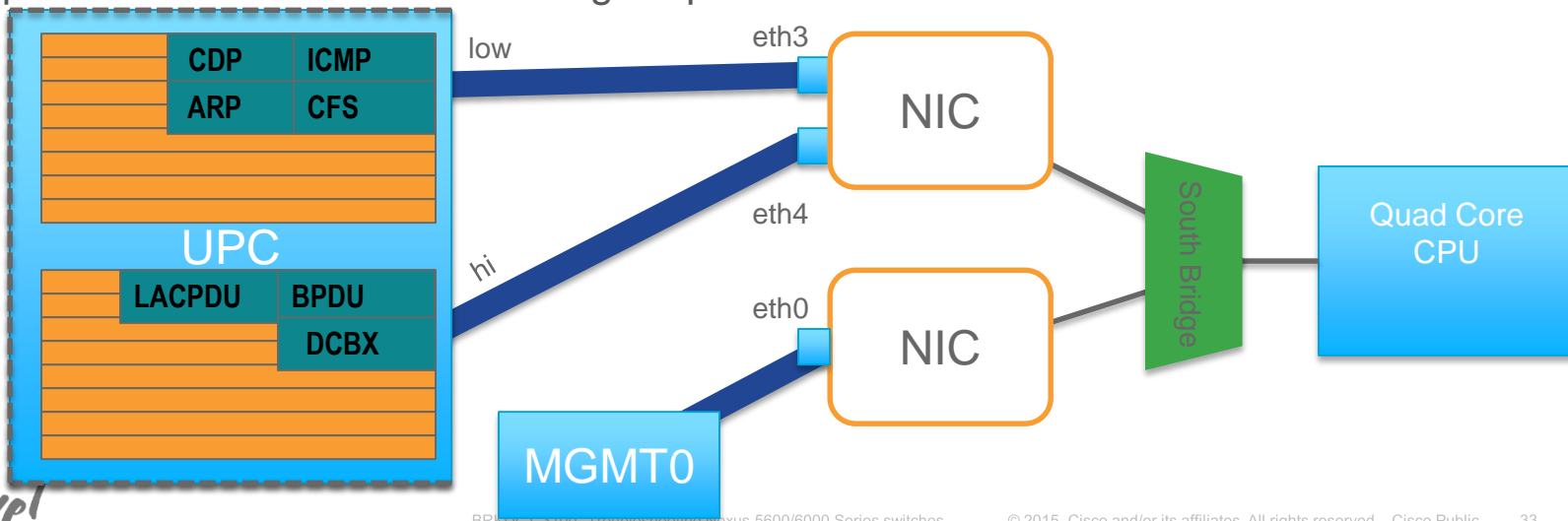
```
Memory usage:  8243352K total,   2962280K used,   5281072K free
```

```
esc-5672-left# show processes cpu sort | exclude 0.0
```

PID	Runtime(ms)	Invoked	uSecs	1Sec	Process
3744	20141	96368	209	22.0%	snmpd
4249	3272768	1540296	2124	1.9%	fcpc
4682	1894770	2869436	660	1.2%	fcpc_thread_sm/0
3700	1378506	2232990	617	0.2%	bigsurusd

Ethanalalyzer

- Capturing and displaying control-plane frames with built-in Ethanalalyzer utility
 - based on wireshark project, NX-OS command frontend
 - Can display like tshark, or capture to .pcap file to analyze elsewhere
 - Can be used on mgmt0 as well as eth3 or eth4, the low and high priority CPU queues and to redirect forwarding drops



Ethalyzer options

```
esc-5672-left# ethalyzer local interface ?
```

inbound-hi	Inbound(high priority) interface
inbound-low	Inbound(low priority) interface
mgmt	Management interface

```
esc-5672-left# ethalyzer local interface inbound-hi ?
```

```
<snip>
```

autostop	Capture autostop condition
capture-filter	Filter on ethalyzer capture
capture-ring-buffer	Capture ring buffer option
decode-internal	Include internal system header decoding
detail	Display detailed protocol information
display-filter	Display filter on frames captured
limit-captured-frames	Maximum number of frames to be captured (default is 10)
limit-frame-size	Capture only a subset of a frame
raw	Hex/Ascii dump the packet with possibly one line summary
write	Filename to save capture to

Ethalyzer Example

- capture mgmt0 traffic and save to as a PCAP file on bootflash
- File can be read on the switch or exported

```
esc-5672-left# ethalyzer local interface mgmt write bootflash:mgmt.pcap
```

```
Capturing on mgmt0
```

```
10
```

```
esc-5672-left# dir bootflash: | grep mgmt.
```

```
1051 Apr 28 21:54:02 2015 mgmt.pcap
```

```
esc-5672-left# ethalyzer local read bootflash:mgmt.pcap
```

```
2015-04-28 21:54:01.585437 10.116.200.148 -> 172.18.118.38 SNMP get-next-request 1.3.6.1.2.1.2.2.1.22.16957440
```

```
2015-04-28 21:54:01.586573 172.18.118.38 -> 10.116.200.148 SNMP get-response 1.3.6.1.2.1.2.2.1.22.16961536
```

```
2015-04-28 21:54:01.673769 10.116.200.148 -> 172.18.118.38 SNMP get-next-request 1.3.6.1.2.1.2.2.1.22.16961536
```

```
2015-04-28 21:54:01.674887 172.18.118.38 -> 10.116.200.148 SNMP get-response 1.3.6.1.2.1.2.2.1.22.16965632
```

```
2015-04-28 21:54:01.965024 172.18.118.39 -> 172.18.118.38 UDP Source port: 3200 Destination port: 3200
```

Ethalyzer Example

- Detailed view of an ARP request

```
esc-5672-left# ethalyzer local interface inbound-low display-filter arp limit-captured-frames 1 detail
Capturing on inband
Frame 1 (66 bytes on wire, 66 bytes captured)
  Arrival Time: Apr 28, 2015 22:05:19.157809000
  <snip>
  [Protocols in frame: eth:vlan:arp]
  Ethernet II, Src: 00:10:94:10:10:01 (00:10:94:10:10:01), Dst: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)
    Destination: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)
      Address: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)
        .... 1 .... = IG bit: Group address (multicast/broadcast)
        .... 1. .... = LG bit: Locally administered address (this is NOT the factory default)
    Source: 00:10:94:10:10:01 (00:10:94:10:10:01)
      Address: 00:10:94:10:10:01 (00:10:94:10:10:01)
        .... 0 .... = IG bit: Individual address (unicast)
        .... 0. .... = LG bit: Globally unique address (factory default)
    Type: 802.1Q Virtual LAN (0x8100)
    802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 94
      000. .... = Priority: 0
      ...0 .... = CFI: 0
      .... 0000 0101 1110 = ID: 94
    Type: ARP (0x0806)
```

Internal
VLAN

Ethalyzer

- N5K/6K uses internal VLANs for communication within the switch
- VLAN ID seen in ethalyzer will be internal VLAN ID.
- External VLAN to internal VLAN mapping can be obtained

```
esc-5672-left# show platform afm info global | begin Vlan
```

```
Vlan mapping table
```

```
-----
```

```
Ext-vlan: 1      -  Int-vlan: 106
```

```
Ext-vlan: 2      -  Int-vlan: 102
```

```
Ext-vlan: 3      -  Int-vlan: 101
```

```
Ext-vlan: 4      -  Int-vlan: 100
```

```
Ext-vlan: 5      -  Int-vlan: 99
```

```
Ext-vlan: 6      -  Int-vlan: 98
```

```
Ext-vlan: 7      -  Int-vlan: 97
```

```
Ext-vlan: 8      -  Int-vlan: 96
```

```
Ext-vlan: 9      -  Int-vlan: 95
```

```
Ext-vlan: 10     -  Int-vlan: 94
```

Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - **CRC Errors**
 - Forwarding
 - Buffering/Queuing
 - ELAM

Cut-through Mode and CRC Errors

- Cut-through switching changes how we troubleshoot problems in the switch.
 - Ethernet CRC is at the end of the frame, so even a CRC error cannot cause a drop on a cut-through port.
 - We are already forwarding the frame by the time the ingress mac can read the CRC value.



Example of received corrupted frame.

- Traffic flow is from Eth1/6 to Eth1/32
- Frames already coming in corrupted into switch

```
esc-5672-right# show interface e1/6
```

```
...
```

```
RX
```

```
2 unicast packets 202 multicast packets 0 broadcast packets
```

```
200204 input packets 100869552 bytes
```

```
0 jumbo packets 0 storm suppression bytes
```

```
0 runs 0 giants 200000 CRC 0 no buffer
```

```
200000 input error 0 short frame 0 overrun 0 underrun 0 ignored
```

```
esc-5672-right# show hardware internal bigsur port ethernet 1/6 counters rx | inc CRC
```

```
RX_PKT_CRC_NOT_STOMPED | 0 | 0 | 0
```

```
RX_PKT_CRC_STOMPED | 200000 | 200000 | 0
```

```
esc-5672-right#
```

STOMPED

- In older code, input error/CRC were seen as output error on egress in show interface
- 6.0(2)N2(1) and later, counted as TX Frame errors

```
esc-5672-right# show hardware internal bigsur port ethernet 1/32 counters rx | inc FRAME
```

```
TX_PKT_FRAME_ERROR | 200000 | 200000 | 0
```

```
esc-5672-right#
```

Example of switch stomping due to MTU violation.

- Traffic flow is from Eth1/6 to Eth1/32
- 9000 byte frames already coming in on Eth1/6 but switch MTU is 1500 bytes

MTU violation

```
esc-5672-right# show queuing interface ethernet 1/6 | inc MTU
q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
esc-5672-right# show hardware internal bigsur port ethernet 1/6 counters rx | grep
RX_PKT_SIZE_IS_819
RX_PKT_SIZE_IS_8192_TO_9216 | 200000 | 200000 | 0
esc-5672-right# show hardware internal bigsur asic 1 counters interrupt | grep -i mtu
big_bmin_cll_INT_pl_err_ig_mtu_vio | 3 | 0 | 3 | 0
```

CRC Stomp

- Packets are truncated to 1500 bytes and switched but stomped with CRC

```
esc-5672-right# show hardware internal bigsur port ethernet 1/32 counters rx | grep
TX_PKT_SIZE_IS_1519
TX_PKT_SIZE_IS_1519_TO_2047 | 200000 | 200000 | 0
esc-5672-right# show hardware internal bigsur asic 3 counters interrupt | grep crc
big_bmen_glb_INT_pcl_p2_norm_crc_stomp | 3 | 0 | 3 | 0
big_fwe_psrl_P2_INT_pkt_err_eth_crc_stomp | 3 | 0 | 3 | 0
esc-5672-right# show hardware internal bigsur port ethernet 1/32 counters rx | inc FRAME
TX_PKT_FRAME_ERROR | 200000 | 200000 | 0
```

Finding the Source of CRC Errors

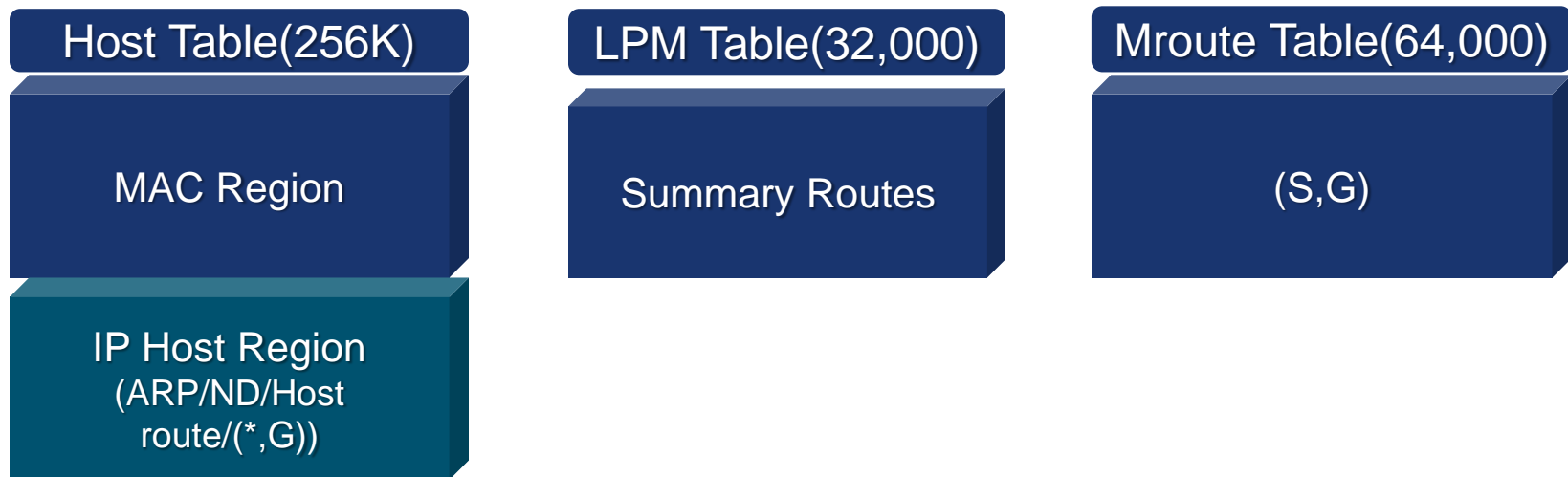
- CRC errors are introduced in 3 ways:
 - Bad physical connection
 - copper, fiber, transceiver, phy
 - “stomping” due to intentionally **originated** errors
 - Received bad CRC “stomped” from neighboring cut-through switch.
- Start by finding any RX CRC counters.
 - If none, then this switch is responsible for originating
 - Use **interrupt counters** to find the reason and port, if intentional
- Log in to next switch upstream of CRC counters, check for RX CRC there.
 - Use the above logic to determine if this switch is originating any errors.
 - Finally, inspect optics/pluggables, fiber/cables and troubleshoot as a Layer 1 issue. Change cable and port to find where the problem follows.
- Store & Forwarding mode can be configured

```
esc-5672-right(config)# hardware ethernet store-and-fwd-switching  
Enabling store-and-forward switching. Please copy the configuration and reload the switch  
esc-5672-right(config)#
```

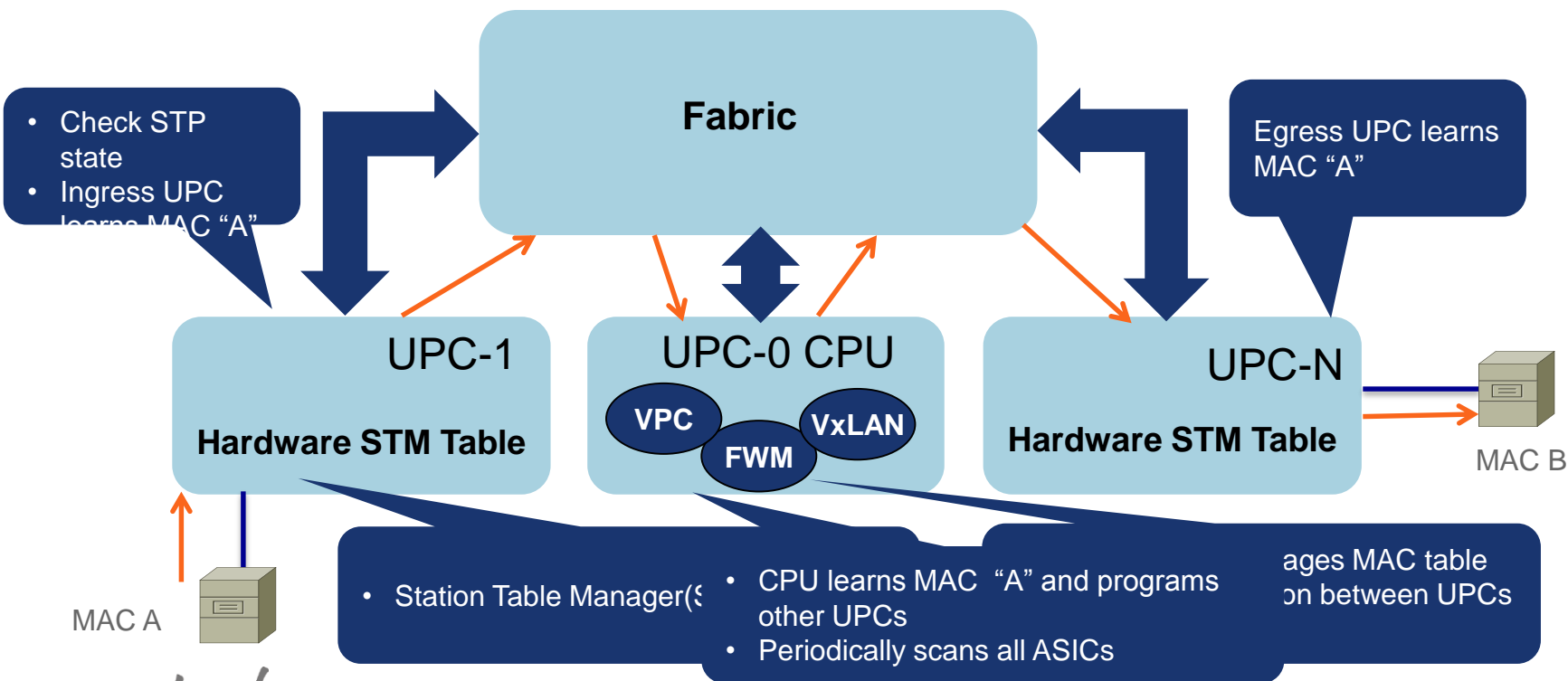

Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

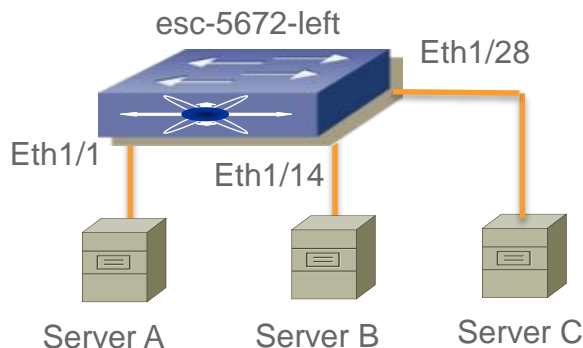
Key Forwarding tables



Nexus 5600/6000 L2 Unicast Forwarding



Nexus 5600/6000 L2 Unicast Forwarding



Scenario:

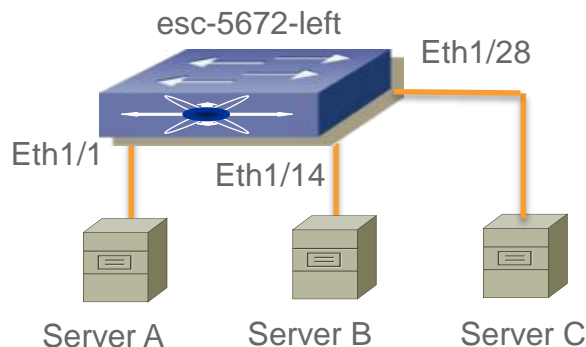
- Server A,B and C are servers in VLAN 10

Given:

- Correct configuration in place and servers sending traffic to each other
- All servers have had resolved ARP entries resolved.
- We will verify programming state for L2 Forwarding

Nexus 5600/6000 L2 Unicast Forwarding

- Get front panel to internal port ASIC mapping



```
esc-5672-left# show hardware internal bigsur all-ports | egrep name|1/1|1/14|1/28
```

name	idx	slot	asic	eport	logi	flag	adm	opr	if_index	diag	ucVer
1gb1/1	1	0	1	0 p	0	b3	en	up	1a000000	pass	0.00
1gb1/14	2	0	2	1 p	13	b3	en	up	1a00d000	pass	0.00
1gb1/28	3	0	3	3 p	27	b3	en	up	1a01b000	pass	0.00

Nexus 5600/6000 L2 Unicast Forwarding

- Check for STP and MAC address table

```
esc-5672-left# show spanning-tree vlan 10
```

```
VLAN0010
```

```
Spanning tree enabled protocol rstp
```

```
Root ID      Priority    4106
```

```
Address      001b.54c2.44c2
```

```
Cost         1
```

```
Port         4767 (port-channel672)
```

```
Hello Time   2 sec Max Age 20 sec Forward Delay 15 sec
```

```
Bridge ID    Priority    32778 (priority 32768 sys-id-ext 10)
```

```
Address      002a.6af9.737c
```

```
Hello Time   2 sec Max Age 20 sec Forward Delay 15 sec
```

```
Interface    Role Sts Cost      Prio.Nbr Type
```

```
-----  
Po572        Desg FWD 1          128.4667 (vPC peer-link) Network P2p
```

```
Po672        Root FWD 1          128.4767 (vPC) P2p
```

```
Eth1/1       Desg FWD 4          128.129 Edge P2p
```

```
Eth1/14      Desg FWD 4          128.142 Edge P2p
```

```
Eth1/28      Desg FWD 4          128.156 Edge P2p
```

```
esc-5672-left# sh mac address-table dynamic vlan 10 | inc 0010
```

```
* 10          0010.9410.0011    dynamic    70          F    F    Eth1/1
```

```
* 10          0010.9410.0114    dynamic    10          F    F    Eth1/14
```

```
* 10          0010.9410.0128    dynamic    50          F    F    Eth1/28
```

```
esc-5672-left#
```

Nexus 5600/6000 L2 Unicast Forwarding

- Check for STP and MAC address table in hardware

```
esc-5672-left# show platform fwm info vlanif 10 ethernet 1/1
vlanif vlan 1.10 if 1a000000 stp state: forwarding
esc-5672-left# show platform fwm info vlanif 10 ethernet 1/14
vlanif vlan 1.10 if 1a00d000 stp state: forwarding
esc-5672-left# show platform fwm info vlanif 10 ethernet 1/28
vlanif vlan 1.10 if 1a01b000 stp state: forwarding
```

```
esc-5672-left# show platform fwm info hw-stm asic 0 | grep 0010.9410
1.10    0010.9410.0011    Eth1/1        1:5469:0  1:0:1    2.a.bc.0.0.3 (e:0)
1.10    0010.9410.0114    Eth1/14       1:8546:0  1:0:1    2.a.bc.0.0.4 (e:0)
1.10    0010.9410.0128    Eth1/28       1:9954:0  1:0:1    2.a.bc.0.0.5 (e:0)
esc-5672-left# show platform fwm info hw-stm asic 1 | grep 0010.9410
1.10    0010.9410.0011    Eth1/1        1:5469:0  1:0:1    2.a.bc.0.0.3 (e:0)
1.10    0010.9410.0114    Eth1/14       1:8546:0  1:0:1    2.a.bc.0.0.4 (e:0)
1.10    0010.9410.0128    Eth1/28       1:9954:0  1:0:1    2.a.bc.0.0.5 (e:0)
esc-5672-left# show platform fwm info hw-stm asic 2 | grep 0010.9410
1.10    0010.9410.0011    Eth1/1        1:5469:0  1:0:1    2.a.bc.0.0.3 (e:0)
1.10    0010.9410.0114    Eth1/14       1:8546:0  1:0:1    2.a.bc.0.0.4 (e:0)
1.10    0010.9410.0128    Eth1/28       1:9954:0  1:0:1    2.a.bc.0.0.5 (e:0)
esc-5672-left#
```

Nexus 5600/6000 L2 Unicast Forwarding

- MAC address event history

```
esc-5672-left# show platform fwm info mac 0010.9410.0011 10
mac vlan 1.10 mac 0010.9410.0011: vlan 1.10
mac vlan 1.10 mac 0010.9410.0011: learned-on Eth1/1 age 110 ref_map = 'vlan if'
mac vlan 1.10 mac 0010.9410.0011: nohit_count 0 hw_programmed 1 mac_clone 0
mac vlan 1.10 mac 0010.9410.0011: old_if_index 'null'
mac vlan 1.10 mac 0010.9410.0011: pss_flags 0
mac vlan 1.10 mac 0010.9410.0011 cfg attrs - not-cli-cfg not-static movable no-drop no-regmac non-netstack-learnt
not-secure not-src-drop
mac vlan 1.10 mac 0010.9410.0011: mcec_flags 0x1, mac_info_flags 0, rem_if 0, sync_count 1 rcv_count 0
mac vlan 1.10 mac 0010.9410.0011: CDCE Address 3:0:0:bc:a:2
Mac history (Last 35 operations):
Total operations: 4:
  Operation: Mac create (9)
    (flags: Loc (0x1) mac_info_flags (0x0) if: 0x1a000000 hint: 0)
    at Sat May 2 04:23:58 2015
  Operation: Mac learned from hw (40)
    (flags: Loc (0x1) mac_info_flags (0x0) if: 0x1a000000 hint: 0)
    at Sat May 2 04:23:58 2015
  Operation: Mac sent to peer on local learn (15)
    (flags: Loc (0x1) mac_info_flags (0x0) if: 0x1a000000 hint: 0)
    at Sat May 2 04:23:58 2015
  Operation: Mac sent to peer on local learn (15)
    (flags: Loc (0x1) mac_info_flags (0x0) if: 0x1a000000 hint: 0)
    at Sat May 2 04:27:46 2015
```


Nexus 5600/6000 L2 Unicast Forwarding

- Check FWM for any drops on interface

```
esc-5672-left# show platform fwm info pif ethernet 1/1 | inc asic
Eth1/1 pd: slot 0 logical port num 0 slot_asic_num 1 global_asic_num 1 fw_inst 0 phy_fw_inst 0 fc 0
esc-5672-left# show platform fwm info pif ethernet 1/1 | inc drop
Eth1/1 pd: tx stats: bytes 171600001 frames 700102 discard 0 drop 0
Eth1/1 pd: rx stats: bytes 2072557027 frames 1613373 discard 0 drop 137487
esc-5672-left# show platform fwm info pif ethernet 1/1 | inc drop
Eth1/1 pd: tx stats: bytes 171600123 frames 700103 discard 0 drop 0
Eth1/1 pd: rx stats: bytes 2072557027 frames 1613373 discard 0 drop 138964
```

FWM Drops

- Drops can be seen due to configuration but further investigation needed

```
esc-5672-left# show platform fwm info asic-errors 1
<snip>
Printing non zero Carmel error registers - 32 bits:
BIG_DROP_IDS_CODE_0_1: res = 146207 [0]
esc-5672-left# show platform fwm info asic-errors 1
<snip>
Printing non zero Carmel error registers - 32 bits:
BIG_DROP_IDS_CODE_0_1: res = 150432 [0]
```

FWM ASIC
error

Nexus 5600/6000 L2 Unicast Forwarding

- Drops are due to FWM IDS check failure
- FWM dropped packet can be redirected to Sup for further inspection

```
esc-5672-left# debug platform fwm pkt-drop-redirect drop-condition IDS_CODE_0_1 asic-id 1
esc-5672-left# ethanalyzer local interface inbound-low display-filter ip.addr==192.168.10.38 detail
<snip>
Header checksum: 0x9301 [incorrect, should be 0x385c]
    [Good: False]
    [Bad : True]
        [Expert Info (Error/Checksum): Bad checksum]
            [Message: Bad checksum]
            [Severity level: Error]
            [Group: Checksum]
Source: 192.168.10.38 (192.168.10.38)
Destination: 192.168.10.128 (192.168.10.128)
User Datagram Protocol, Src Port: 1024 (1024), Dst Port: 1024 (1024)
<snip>
esc-5672-left# no debug platform fwm pkt-drop-redirect asic-id 1
```

- Drops in this case were due to host sending frames with incorrect IPv4 checksum

Nexus 5600/6000 Host table exhaustion

- Host table exhaustion

```
2015 May 4 12:09:17 esc-5672-left %FWM-2-STM LIMIT REACHED: Unicast station table dynamic capacity reached (limit 123305) -
creating mac 0020.0000.bcal on port Eth1/14 and vlan 10 disabling dynamic learn notifications for 180 seconds or till capacity
reaches 1500 entries
```

```
esc-5672-left# show mac address-table count
```

```
MAC Entries for all vlans:
```

```
Dynamic Address Count: 123301
```

```
Static Address (User-defined) Count: 0
```

```
Multicast MAC Address Count: 0
```

```
Total MAC Addresses in Use: 123301
```

```
Total PVLAN Clone MAC Address Count: 0
```

```
esc-5672-left# show platform fwm info stm-stats
```

```
Global level learning: disabled
```

```
Vlan level learning: enabled
```

```
MAC Stats: (learning_disable ucast 1 mcast 0 learn_on_exceptions 0)
```

```
STM Threshold - total ucast entries : 123305
```

```
STM Threshold - total mcast entries : 0
```

```
STM Threshold - dynamic ucast entries : 123305
```

```
STM Threshold - dynamic mcast entries(excl. cloned) : 0
```

```
STM Threshold - dynamic cloned mcast entries : 0
```

```
STM Threshold - dynamic mcast entries(combined) : 0
```

```
STM Threshold - ucast cloned entries : 0
```

```
STM Threshold - mcast cloned entries : 0
```

```
STM Threshold - ucast cloned adds : 0
```

```
STM Threshold - ucast cloned destroys : 0
```

```
STM Threshold - total limit : 131072
```

```
STM Threshold - dynamic ucast limit : 123305
```

MAC/ARP Resource Carving CLI

- Specify the resource template to use

```
esc-5672-left(config)# hardware profile route resource service-template ?
hrt-128-stm-128  Hrt: 128k, Stm: 128k (default size)
hrt-224-stm-32   Hrt: 224k, Stm: 32k
hrt-32-stm-224   Hrt: 32k, Stm: 224k
hrt-64-stm-192   Hrt: 64k, Stm: 192k
hrt-96-stm-160   Hrt: 96k, Stm: 160k

esc-5672-left(config)#
```

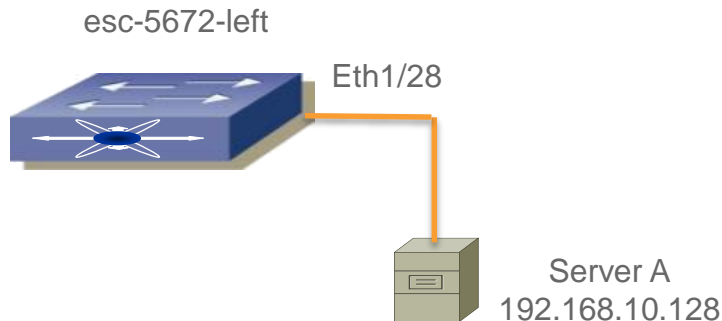
- Need to save the configuration and reload the switch to be applied.
- Show commands:
 - show hardware profile route resource template
 - show hardware profile route resource template default

Nexus 5600/6000 L2 Multicast Forwarding

- IP based multicast forwarding
- No concern of overlapping Multicast addresses
- By default 4000 IGMP snooping groups are supported
- Can be increased to 16000

```
esc-5672-left(config)# hardware multicast snooping group-limit ?  
  <100-16000> Specify a value between 100-16000  
esc-5672-left(config)# hardware multicast snooping group-limit 16000
```

Nexus 5600/6000 L2 Multicast Forwarding



Scenario:

- Server A in VLAN 10 is receiver sending IGMP membership reports to 238.1.1.1

Given:

- IGMP querier is present and on Po672

Will verify state and programming for the group

Nexus 5600/6000 L2 Multicast Forwarding

- IGMP Snooping verification

```
esc-5672-left# show ip igmp snooping vlan 10
IGMP Snooping information for vlan 10
  IGMP snooping enabled
  Lookup mode: IP
  Optimised Multicast Flood (OMF) disabled
  IGMP querier present, address: 192.168.10.157, version: 3, i/f Po672
  Switch-querier disabled
  IGMPv3 Explicit tracking enabled
  IGMPv2 Fast leave disabled
  IGMPv1/v2 Report suppression enabled
  IGMPv3 Report suppression disabled
  Link Local Groups suppression enabled
  Router port detection using PIM Hellos, IGMP Queries
  Number of router-ports: 2
  Number of groups: 2
  VLAN vPC function enabled
  Active ports:
    Po572   Po672   Eth1/1  Eth1/14  Eth1/28
```

```
esc-5672-left# show ip igmp snooping mrouter vlan 10
Type: S - Static, D - Dynamic, V - vPC Peer Link
      I - Internal, F - Fabricpath core port
      C - Co-learned, U - User Configured
      P - learnt by Peer

Vlan  Router-port  Type      Uptime      Expires
10   Po572         SV        2w1d        never
10   Po672         D         2w1d        00:04:46
```

Nexus 5600/6000 L2 Multicast Forwarding

- IGMP control plane troubleshooting

```
esc-5672-left# show ip igmp snooping groups vlan 10
Type: S - Static, D - Dynamic, R - Router port, F - Fabricpath core port
Vlan Group Address      Ver  Type  Port list
10   */*                -    R     Po572 Po672
```

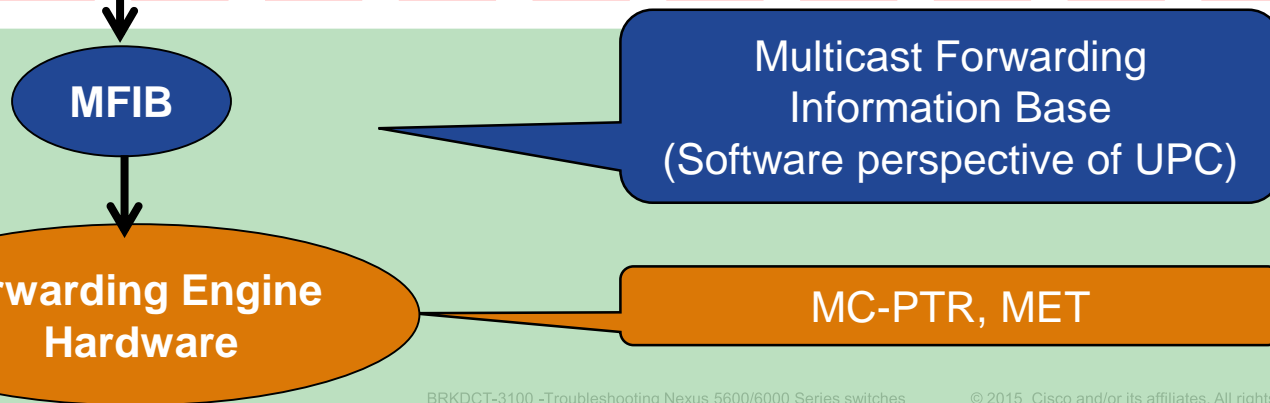
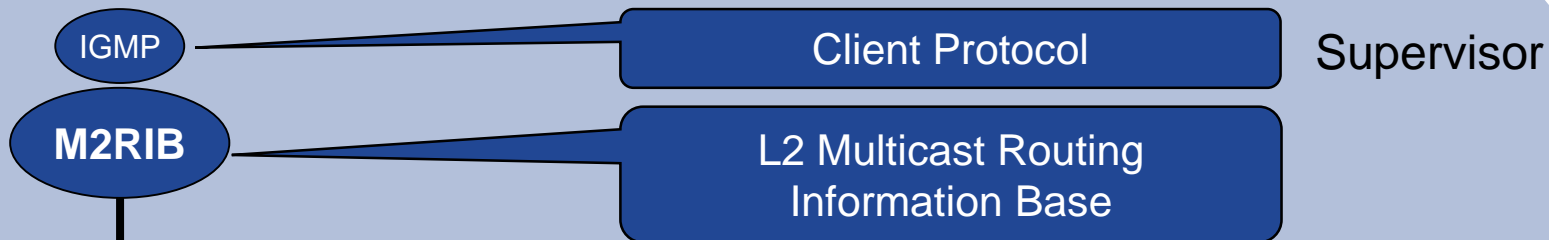
```
esc-5672-left# ethanalyzer local interface inbound-low display-filter igmp limit-captured-frames 0
Capturing on inband
2015-05-19 22:35:24.823030 192.168.10.128 -> 238.1.1.1    IGMP V2 Membership Report / Join group 238.1.1.1
1 packet captured
```

```
esc-5672-left# show ip igmp snooping internal event-history vlan | grep 238.1.1.1

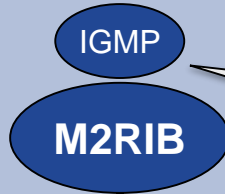
2015 May 19 22:35:24.822701 igmp [3977]: [4839]: SN: <10> Forwarding report for (*, 238.1.1.1) came on Eth1/28
2015 May 19 22:35:24.822688 igmp [3977]: [4839]: SN: <10> Updated oif Eth1/28 for (*, 238.1.1.1) entry
2015 May 19 22:35:24.822337 igmp [3977]: [4839]: SN: <10> Created IGMPv2 oif Eth1/28 for (*, 238.1.1.1)
2015 May 19 22:35:24.822269 igmp [3977]: [4839]: SN: <10> Received v2 report: group 238.1.1.1 from 192.168.10.128 on
Eth1/28
2015 May 19 22:35:24.822103 igmp [3977]: [4839]: SN: <10> Received v2 report with group(238.1.1.1) and
destination(238.1.1.1) address, from_cfs = 0
```

```
esc-5672-left# show ip igmp snooping groups vlan 10
Type: S - Static, D - Dynamic, R - Router port, F - Fabricpath core port
Vlan Group Address      Ver  Type  Port list
10   */*                -    R     Po572 Po672
10   238.1.1.1          v2    D     Eth1/28
```


Layer 2 Data Plane Troubleshooting: L2 Multicast



Layer 2 Data Plane Troubleshooting: L2 Multicast



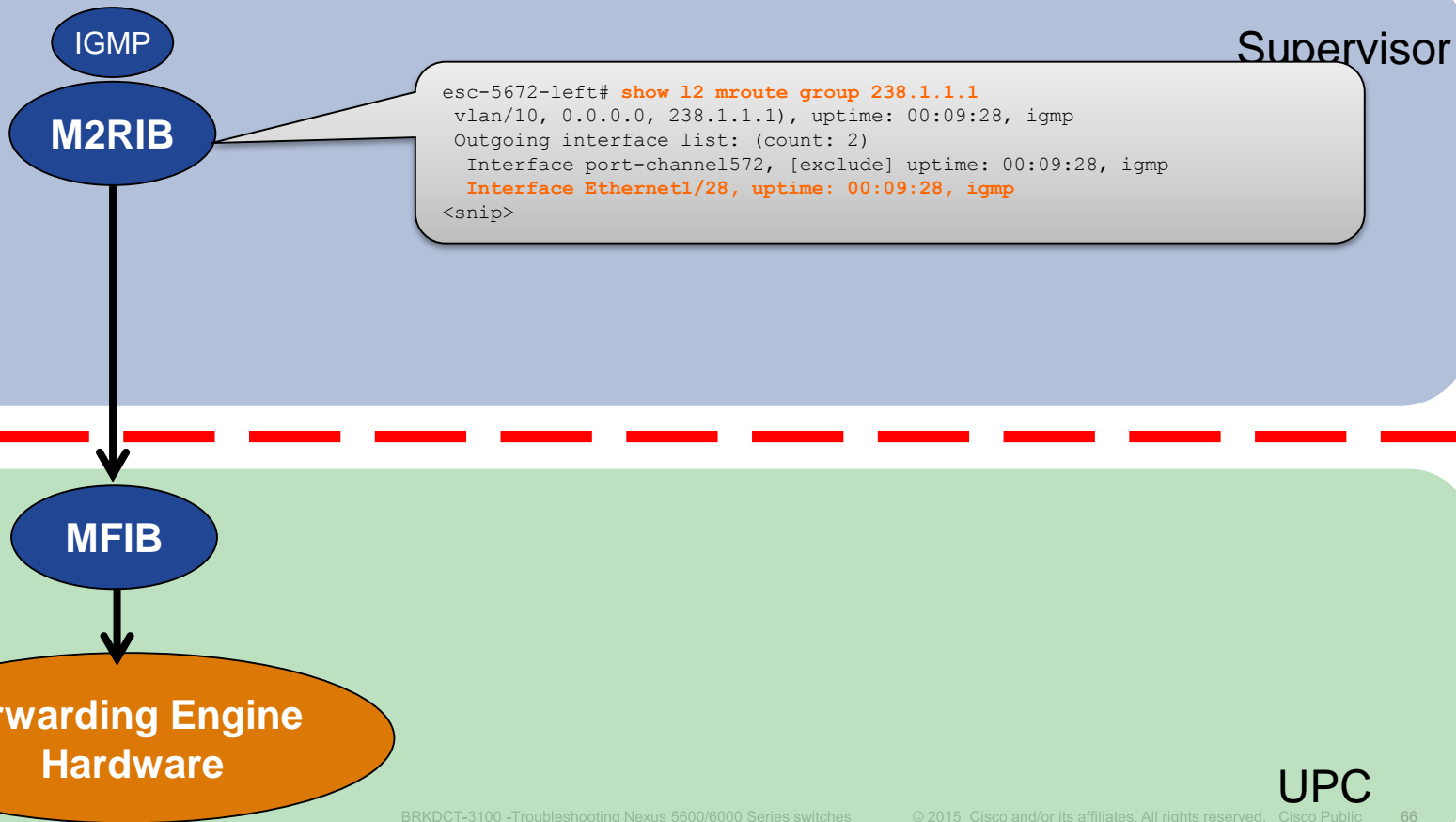
Supervisor

```
esc-5672-left# show ip igmp snooping groups vlan 10
Type: S - Static, D - Dynamic, R - Router port, F - Fabricpath core
port
Vlan  Group Address      Ver  Type  Port list
10    */*                  -    R     Po572 Po672
10    238.1.1.1          v2   D     Eth1/28
```

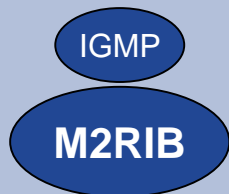
MFIB

Forwarding Engine
Hardware

Layer 2 Data Plane Troubleshooting: L2 Multicast



Layer 2 Data Plane Troubleshooting: L2 Multicast



Supervisor



Forwarding Engine
Hardware

```
esc-5672-left# show system internal forwarding ipfib pd_tree ip-addr 238.1.1.1
ip-address      mc_tag      hit_addr_ind  ref_count  starg_valid
-----
238.1.1.1      94             0x215a7       1           1
esc-5672-left#
```

```
esc-5672-left# show system internal forwarding ipfib sg_tree mc_tag 94
mc_tag source-ip      group-ip      mc_idx  src_hit_idx  grp_hit_idx  sg_idx  sg_only
-----
94      0.0.0.0          238.1.1.1    219     0x0          0x215a7     136615 0
esc-5672-left#
```

Layer 2 Data Plane Troubleshooting: L2 Multicast

Supervisor

IGMP

M2RIB

MFIB

Forwarding Engine
Hardware

```
esc-5672-left# show platform fwm info sg 0.0.0.0 238.1.1.1 10
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: Num interfaces:3.
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: vlan 1.10 pss_flags 2
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: sg ifindex list rcvd from PI (not
pss'ed) - 0x1600023b,0x1600029f,0x1a01b000
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: sg down/deleted ifindex list -
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: sg iod list - 13-14,43
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: sg iod oifl list - 13-14,43
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: gpinif 0
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: fwm new sg oifl 0
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: fwm sg oifl 219
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: fwm new mcec_sg oifl 0
sg vlan 1.10 sgp sip 0.0.0.0 gip 238.1.1.1: fwm mcec sg oifl 0
```

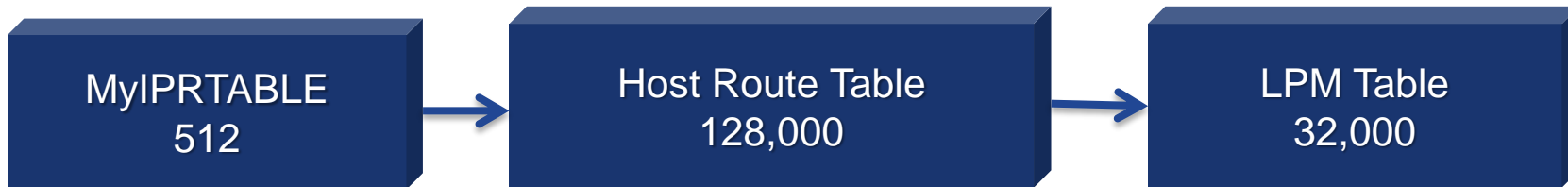
```
esc-5672-left# show platform fwm info oifl 219
oifl 219 vdc 1 oifl 219: vdc 1 gpinif 0, mcast idx 219(alt:0), oifl_type 4
oifl 219 vdc 1 oifl 219: oifl iods 13-14,43
oifl 219 vdc 1 oifl 219: max_iod 8192, ref count 1 num_oifs 3, seq_num 312
oifl 219 vdc 1 oifl 219: hw pgmd: 1 msg present: 0 born @: 1342992134 msecs
oifl 219 vdc 1 oifl 219: oifl_type 4, l2_bum_ref_cnt 0, l3_macg_ref_cnt 1,
l2_ref_cnt 1
oifl 219 vdc 1 oifl 219: if_indexs - Po572 Po672 Eth1/28
```

Nexus 5600/6000 L3 Unicast Forwarding

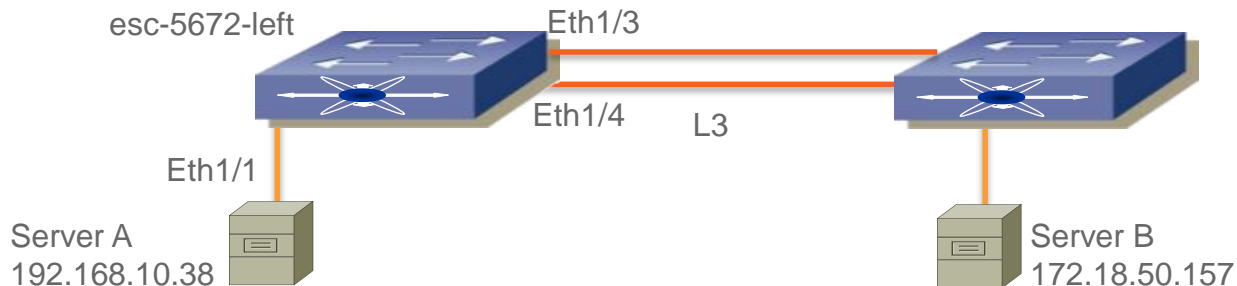
- L3 look up capability is built into the hardware forwarding pipeline
- L3 License required to activate L3 features
 - LAN Base Service License
 - LAN Enterprise Services License
- ND ISSU not possible with L3 License installed
- MyIPRTTable contains the list of MAC addresses the switch can route for
- If the look up is a hit against MyIPRTTable, packets are routed, if not they are bridged

N5600/6000 L3 table look up order

- MyIPRTTable is looked to see if packet needs to be routed.
- If it is a hit, Host Route Table(HRT) is looked up next
- If no hit in HRT table, Longest Prefix Match(LPM) table is looked up next



Nexus 5600/6000 L3 Unicast Forwarding



Goal:

- To check for L3 unicast routing information in software and hardware

Given:

- Server A(192.168.10.38) is in VLAN 10
- esc-5672-left has SVI in VLAN 10 with HSRP and OSPF configured
- Server B(172.18.50.157) is learnt via ECMP OSPF route

Nexus 5600/6000 L3 Unicast Forwarding

- Check for local interface/HSRP state, MyIPRTTable

```
esc-5672-left# show hsrp brief
*:IPv6 group    #:group belongs to a bundle
                P indicates configured to preempt.
                |
Interface      Grp  Prio P State    Active addr    Standby addr    Group addr
Vlan10         10  100 Active local     192.168.10.39   192.168.10.1    (conf)
esc-5672-left# show int vlan 10
Vlan10 is up, line protocol is up
  Hardware is EtherSVI, address is 002a.6af9.737c
  Internet Address is 192.168.10.37/24
  MTU 1500 bytes, BW 1000000 Kbit, DLY 10 usec
esc-5672-left# show hsrp interface vlan 10 | inc Virtual
  Virtual IP address is 192.168.10.1 (Cfged)
  Virtual mac address is 0000.0c9f.f00a (Default MAC)
esc-5672-left# show platform fwm info l3lif vlan 10 | inc int-vlan|mac
Vlan10: iftype SVI: int-vlan 94 l3-vdc-vlan 10 fhrp_enable 1 num_fhrp_grps 1
Vlan10: mac-address: 002a.6af9.737c
Vlan10: fhrp-mac:0000.0c9f.f00a, l2-fm-state:L2FM_MAC_STATE_ACTIVE, remote l2-fm-state:L2FM_MAC_STATE_STANDBY
```

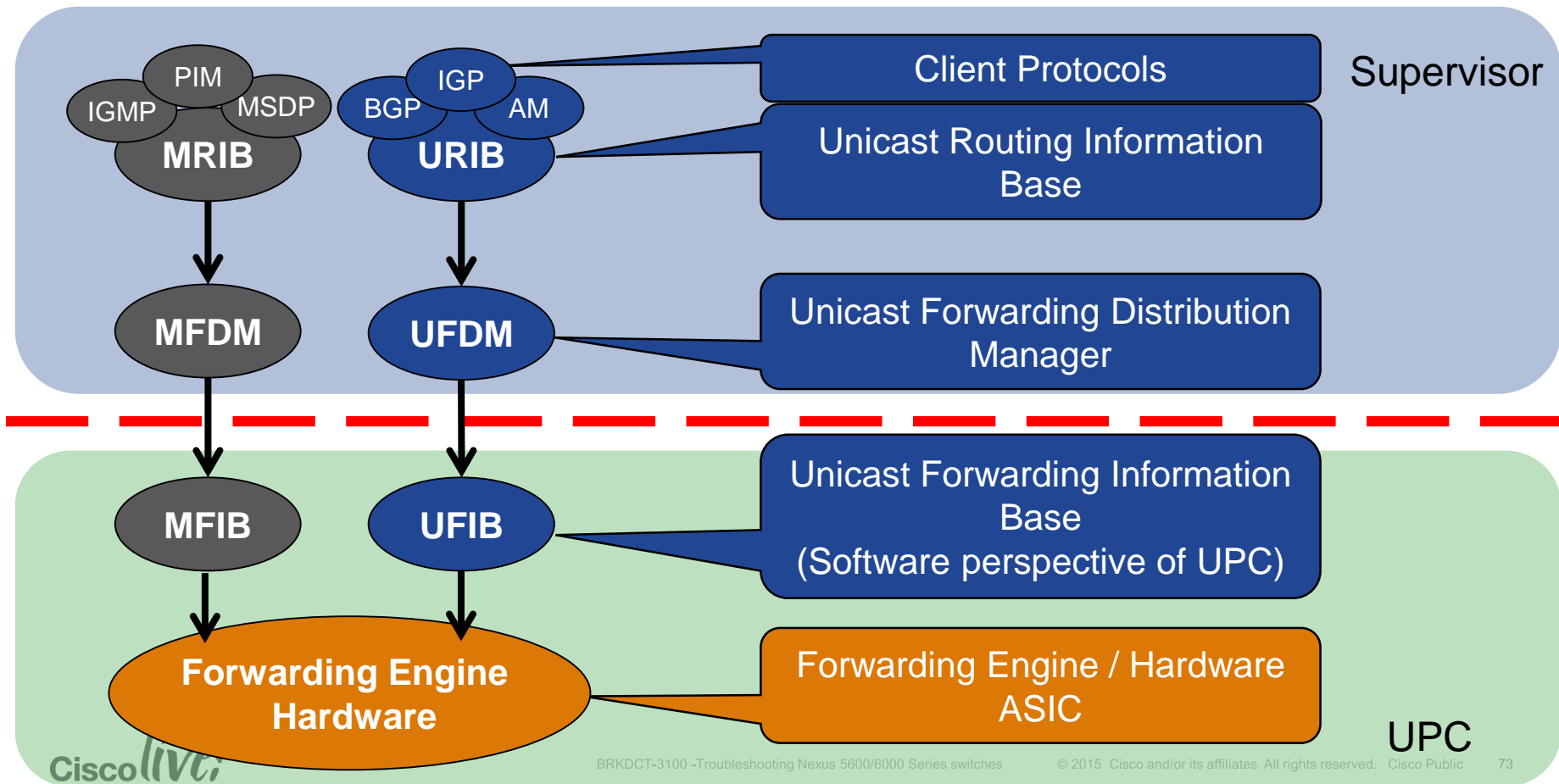
```
esc-5672-left# show system internal forwarding myiprttable
```

Index	BD No	Mac Addr	Ref Count
0	50	002a.6af9.737c	1
1	0	002a.6af9.737c	1
<snip>			
4	3	0000.0c9f.f00a	1

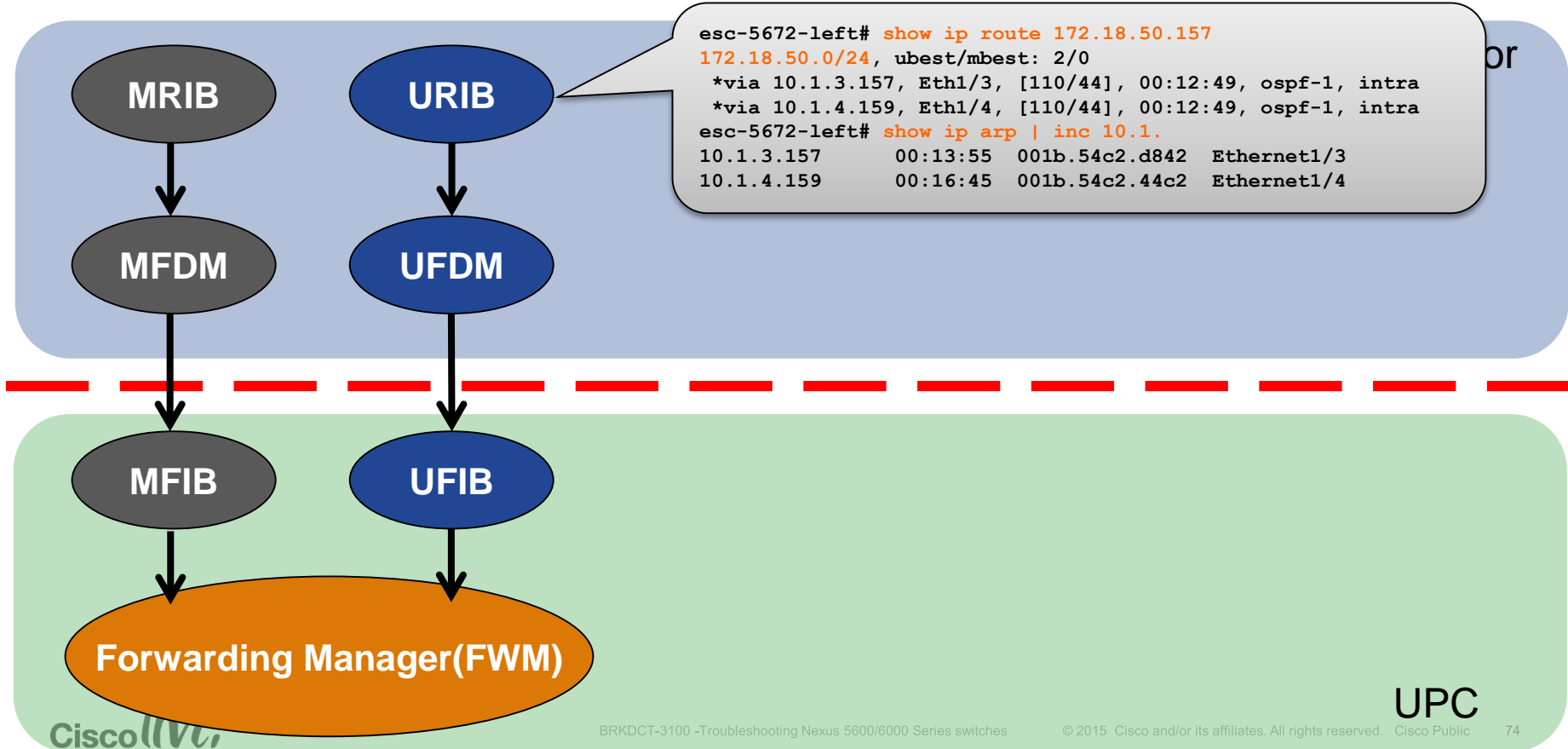
SVI MAC

HSRP
VMAC

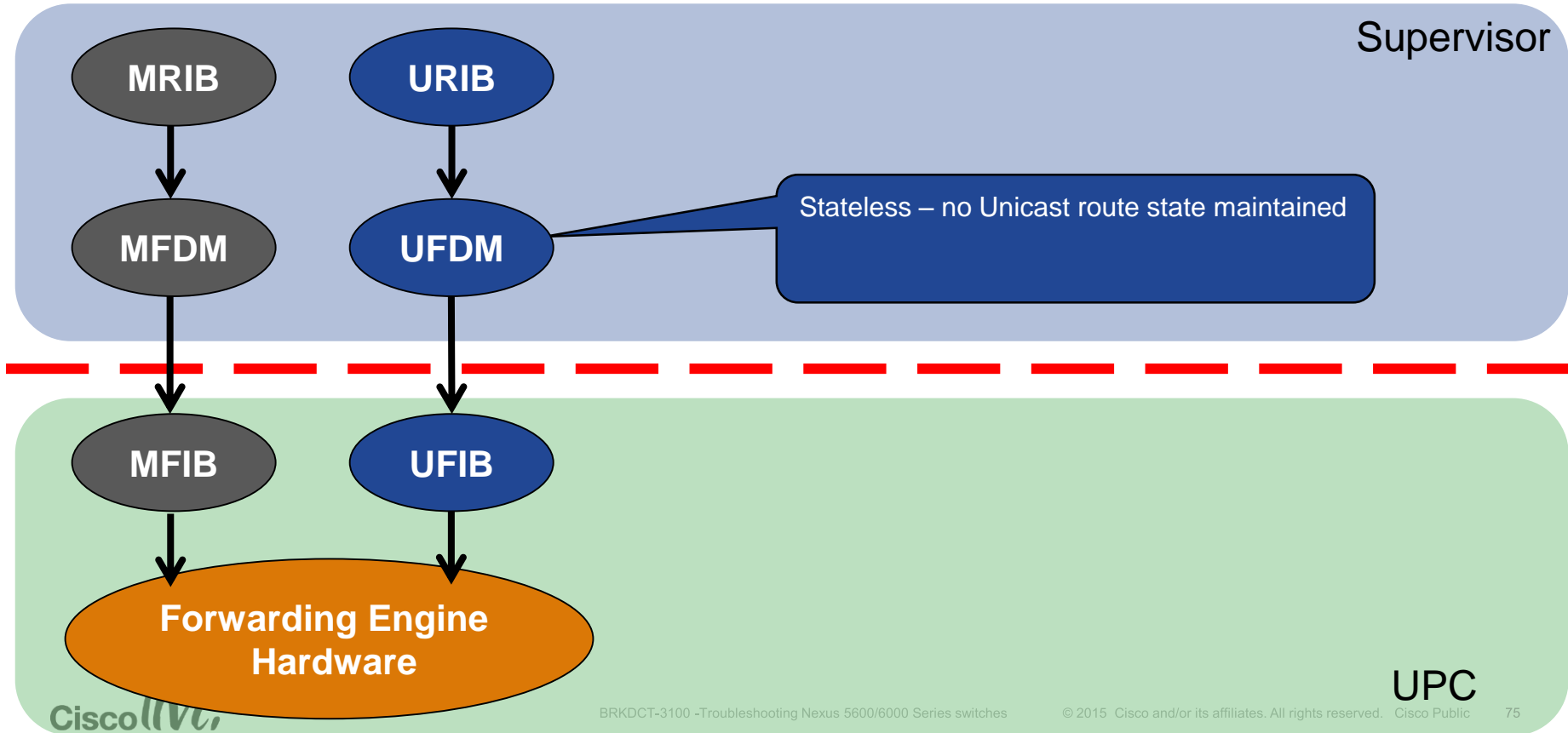
Layer 3 Data Plane Troubleshooting: Components



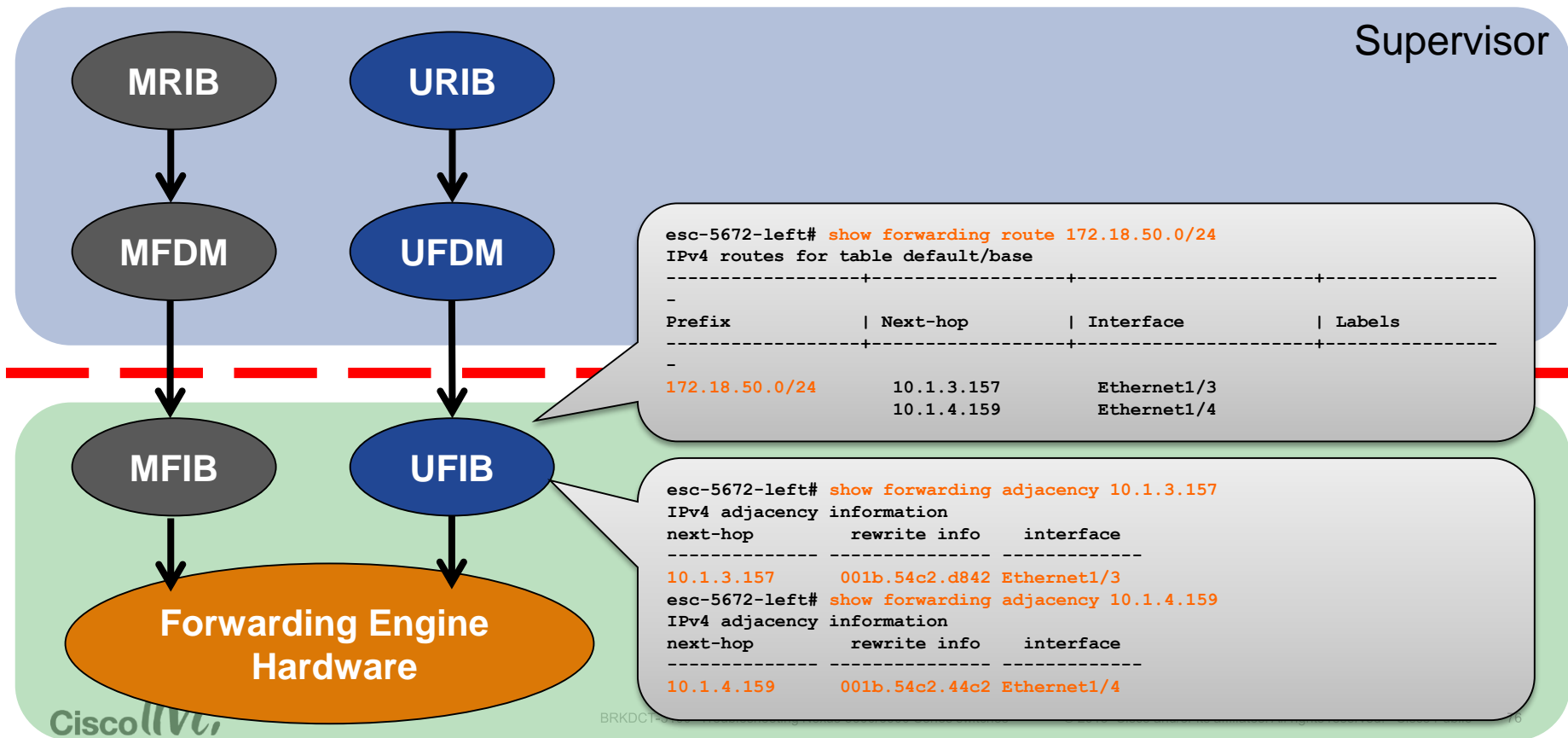
Layer 3 Data Plane Troubleshooting: Unicast



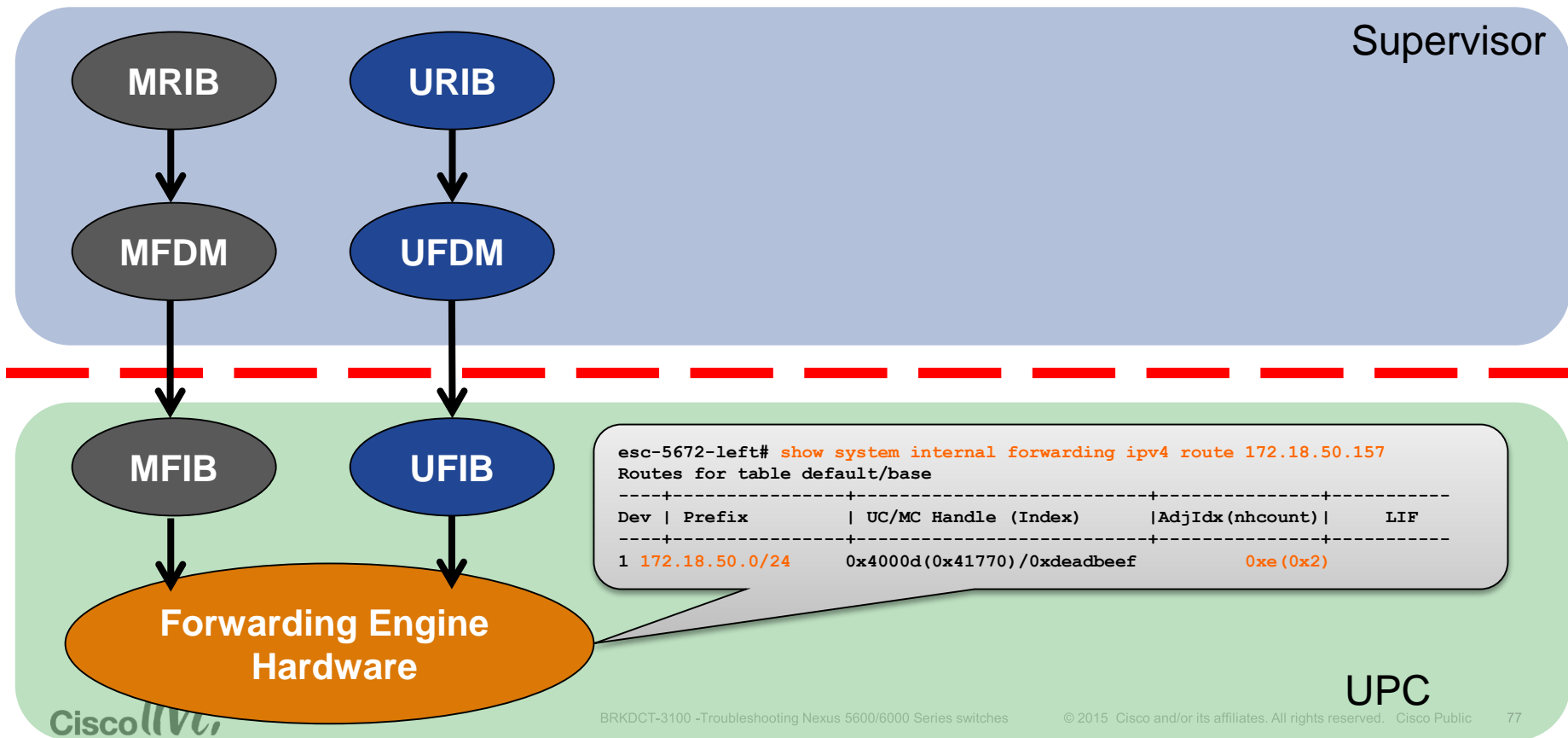
Layer 3 Data Plane Troubleshooting: Unicast



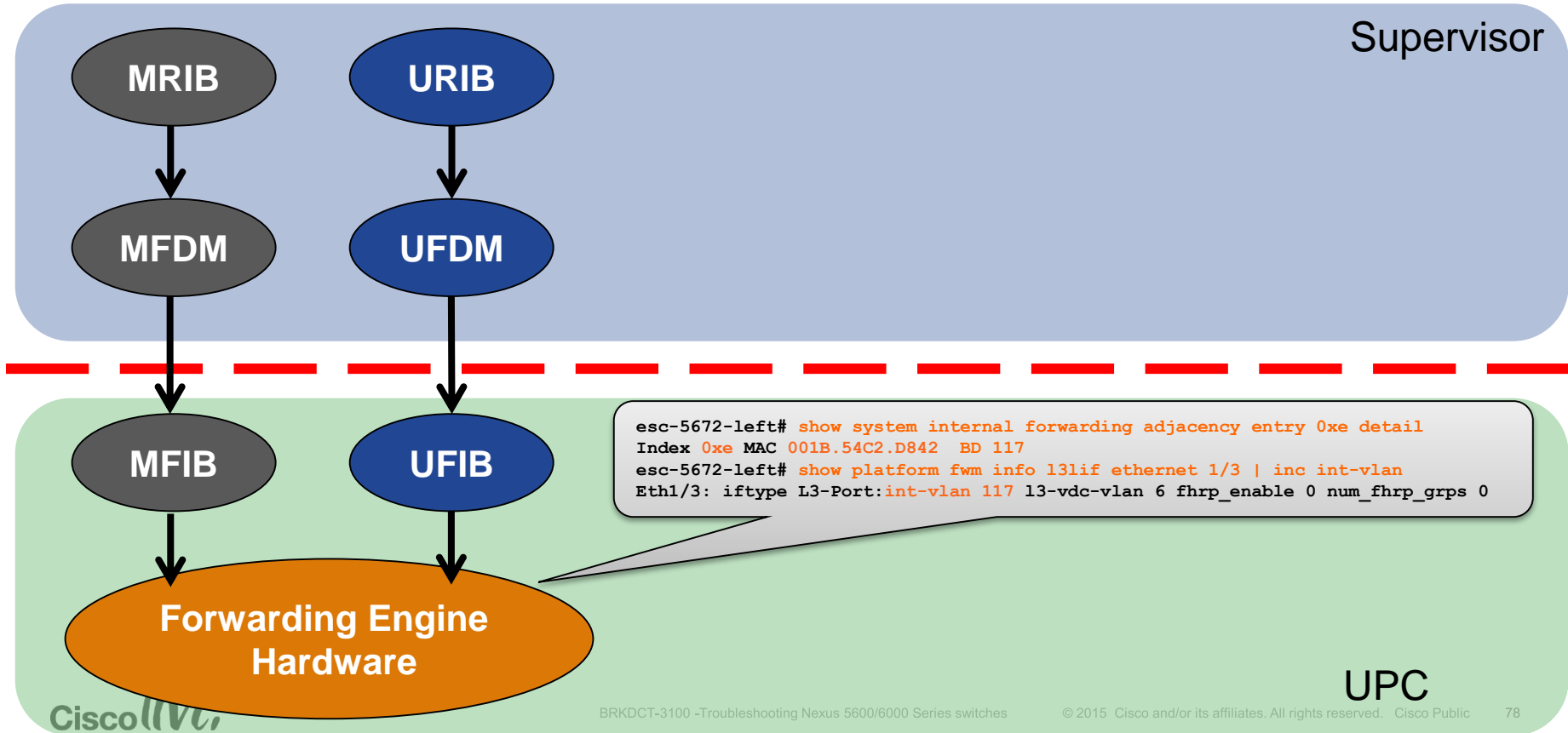
Layer 3 Data Plane Troubleshooting: Unicast



Layer 3 Data Plane Troubleshooting: Unicast



Layer 3 Data Plane Troubleshooting: Unicast



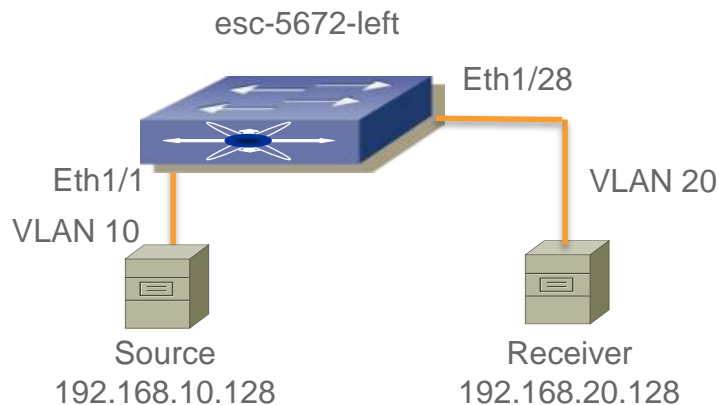
Nexus 5600/6000 L3 Unicast Forwarding

- LPM TCAM exhaustion

```
esc-5672-left#
2015 May  5 20:39:21 esc-5672-left %FWM-4-FIB_TCAM_RESOURCE_WARNING: FIB TCAM usage is at 90 percent
2015 May  5 20:39:29 esc-5672-left %FWM-2-FIB_TCAM_RESOURCE_EXHAUSTION: FIB TCAM exhausted, 1.122.48.0
prefix insert failed
esc-5672-left# show ip route summary
IP Route Table for VRF "default"
Total number of routes: 33031
Total number of paths: 33044
Best paths per protocol:      Backup paths per protocol:
  am           : 3             None
  local        : 4
  direct       : 4
  broadcast    : 9
  ospf-1       : 33024
Number of routes per mask-length:
 /8 : 2      /24: 33009  /32: 20

esc-5672-left# show hardware profile status
Max Mcast Routes = 8192.
Used Mcast Routes = 5.
Total LPM Entries = 32384.
Used Unicast IPv4 LPM Entries = 31291.
Used Unicast IPv6 LPM Entries = 2.
<snip>
```


Nexus 5600/6000 L3 Multicast Forwarding



Scenario:

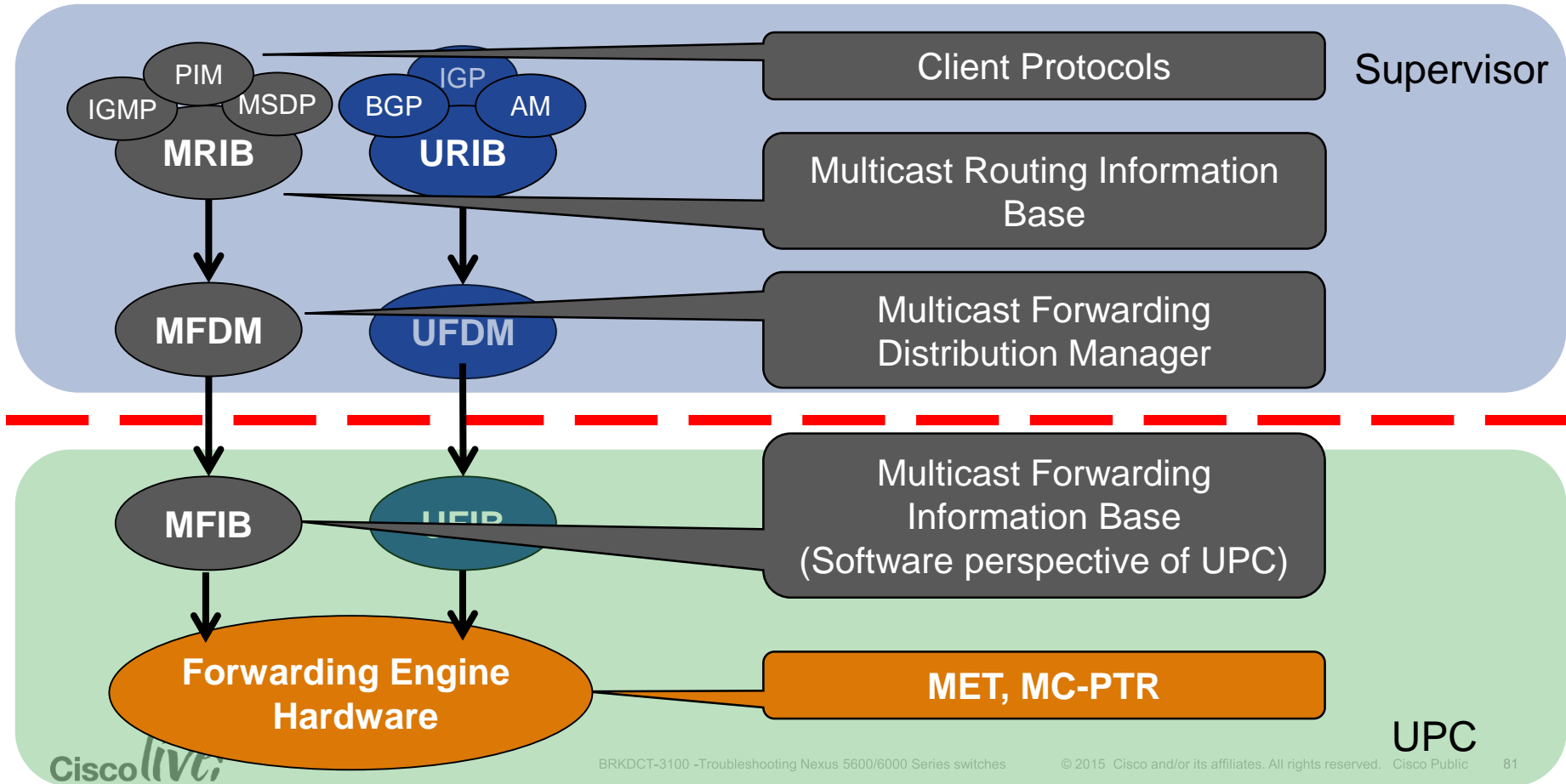
- Source server in VLAN 10 is sending multicast traffic to group 238.1.1.1
- Receiver in VLAN 20 is sending IGMP membership reports to group 238.1.1.1

Given:

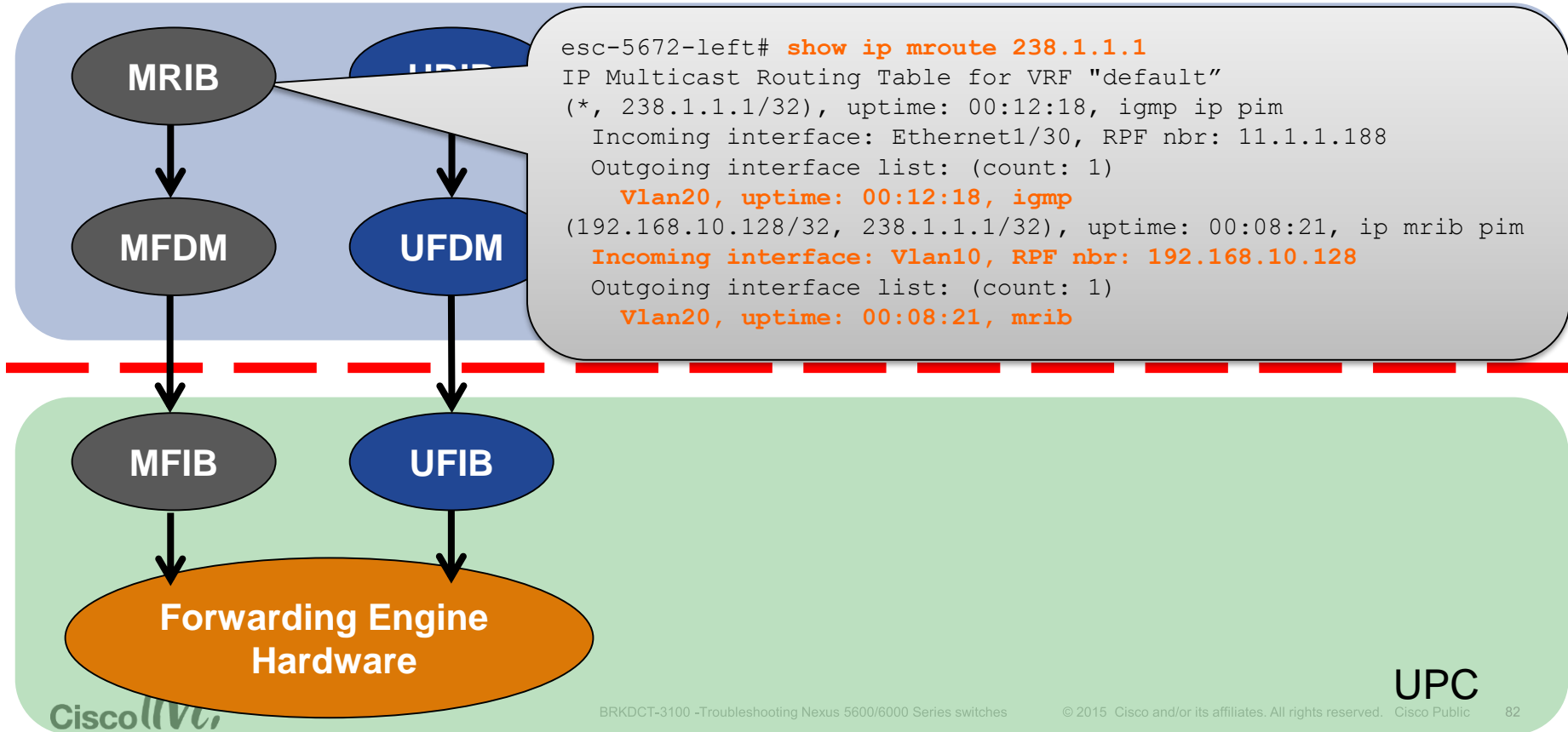
- SVI for VLAN 10, 20 and PIM is configured under it. RP exists in the network

Will verify state and programming for the group

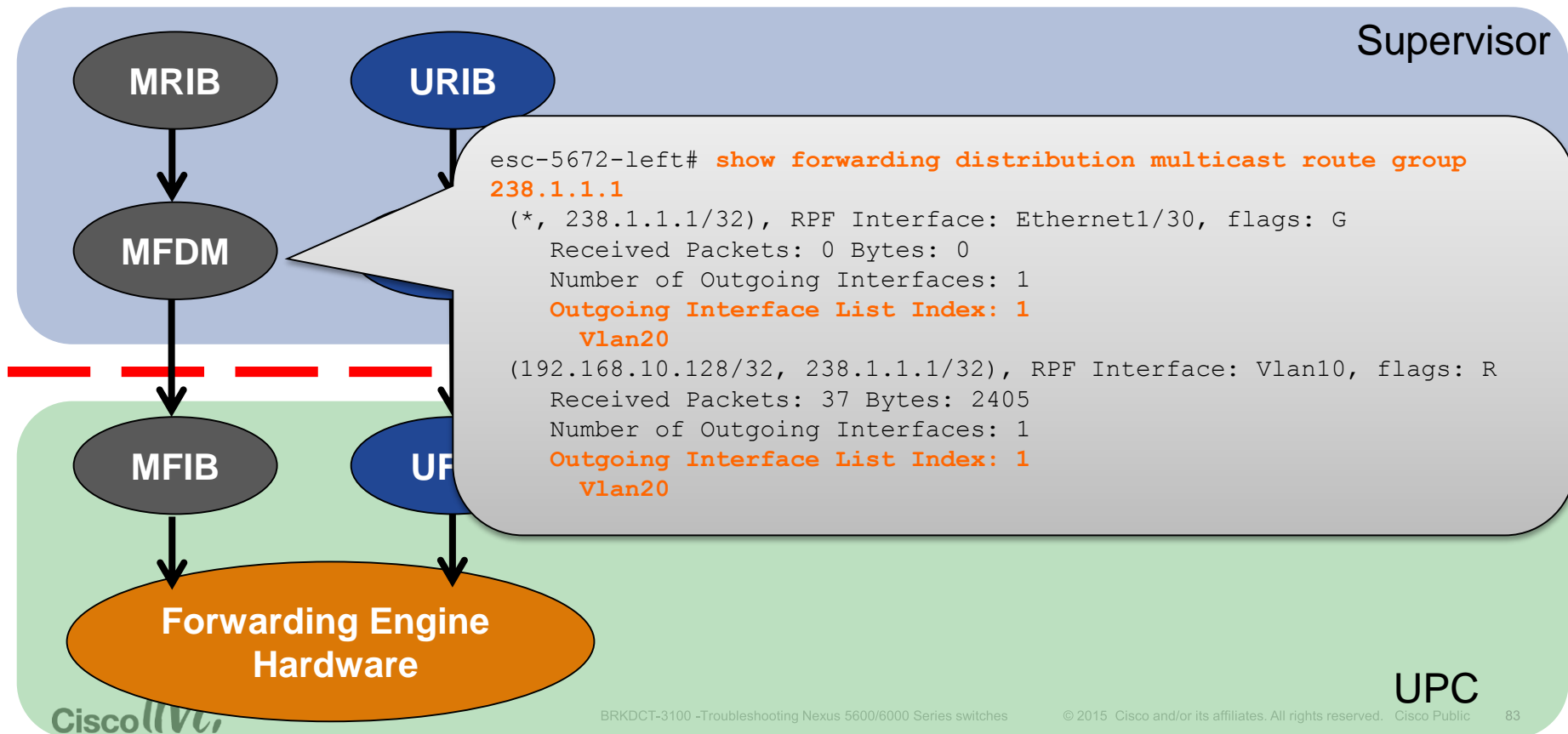
Layer 3 Data plane Troubleshooting: Multicast



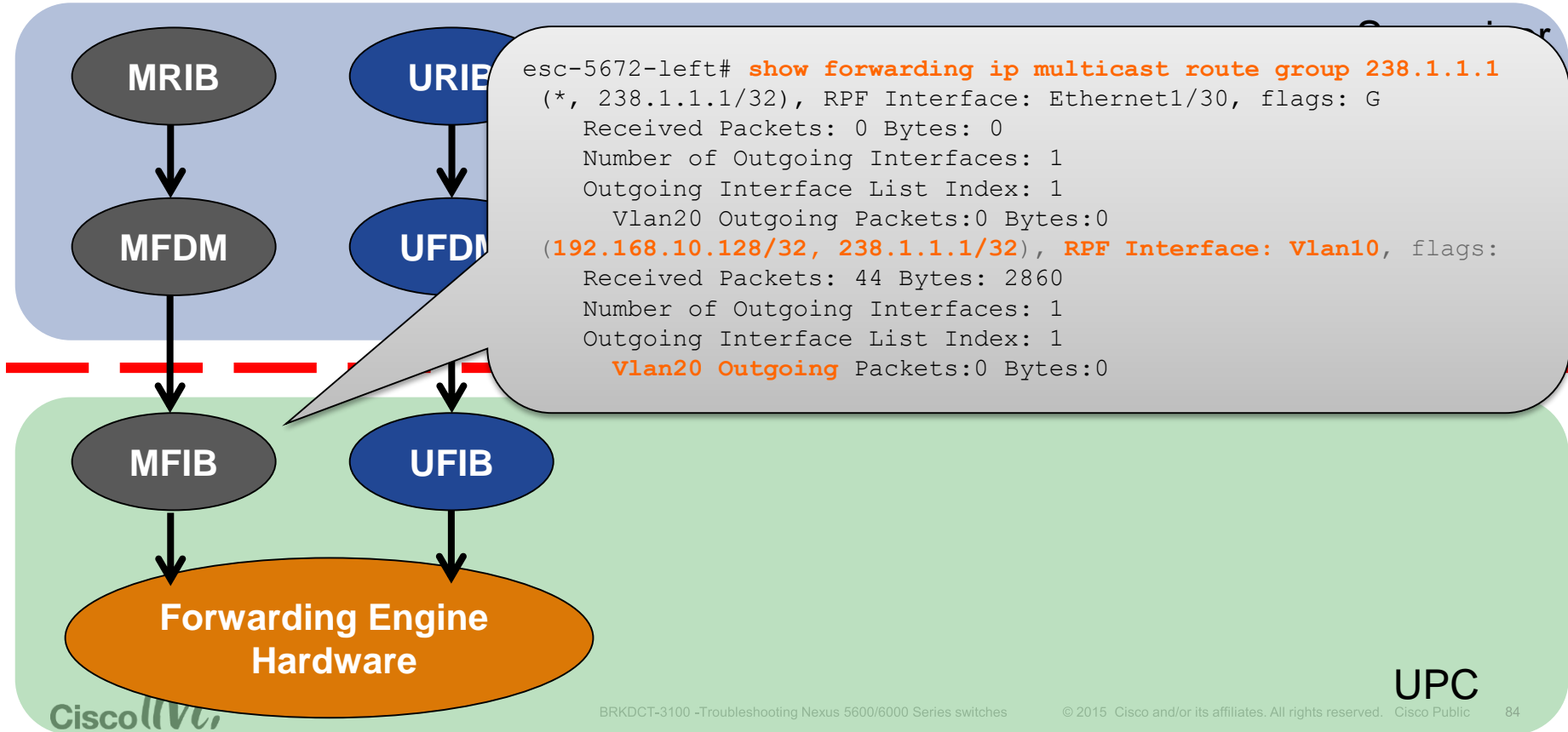
Layer 3 Data Plane Troubleshooting: Multicast



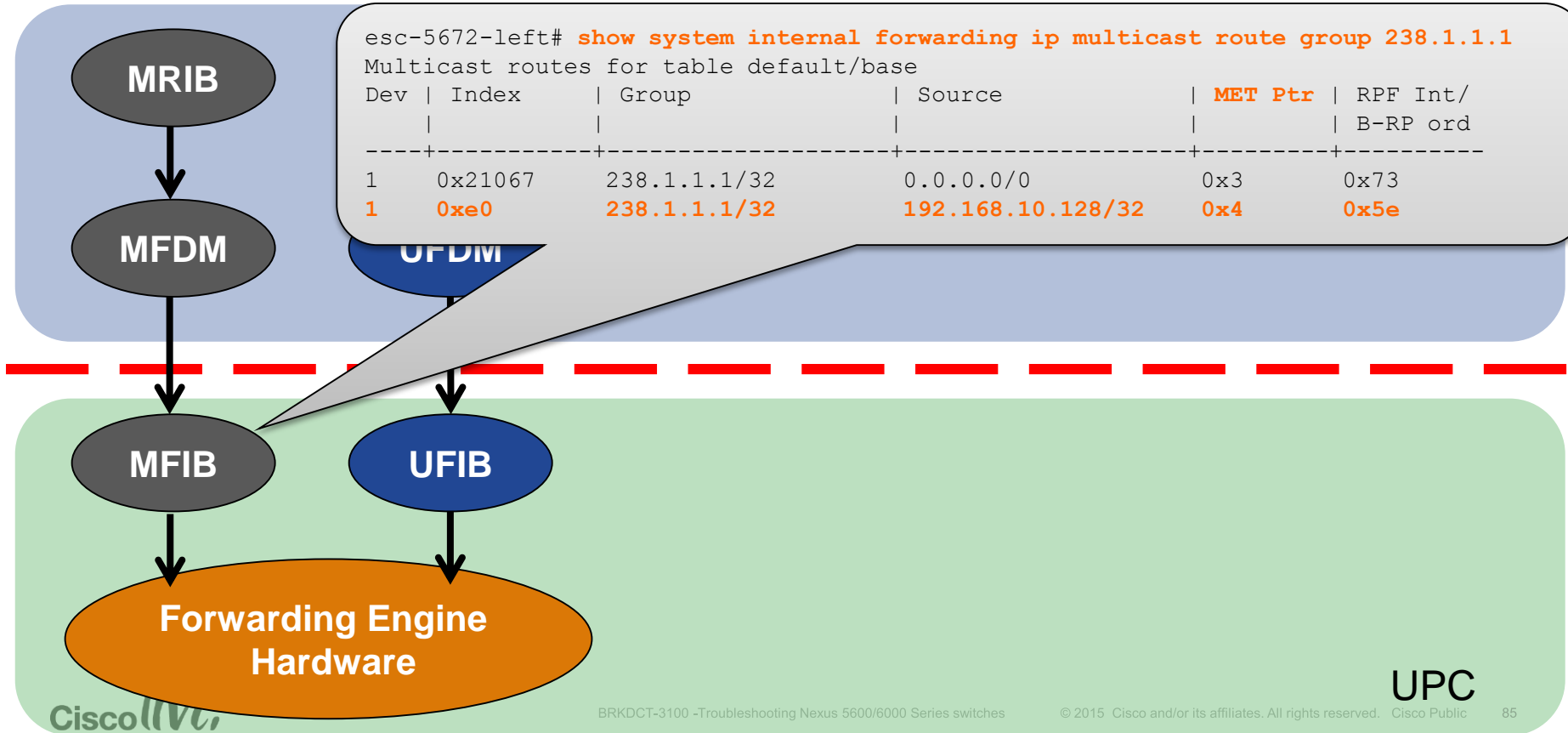
Layer 3 Data Plane Troubleshooting: Multicast



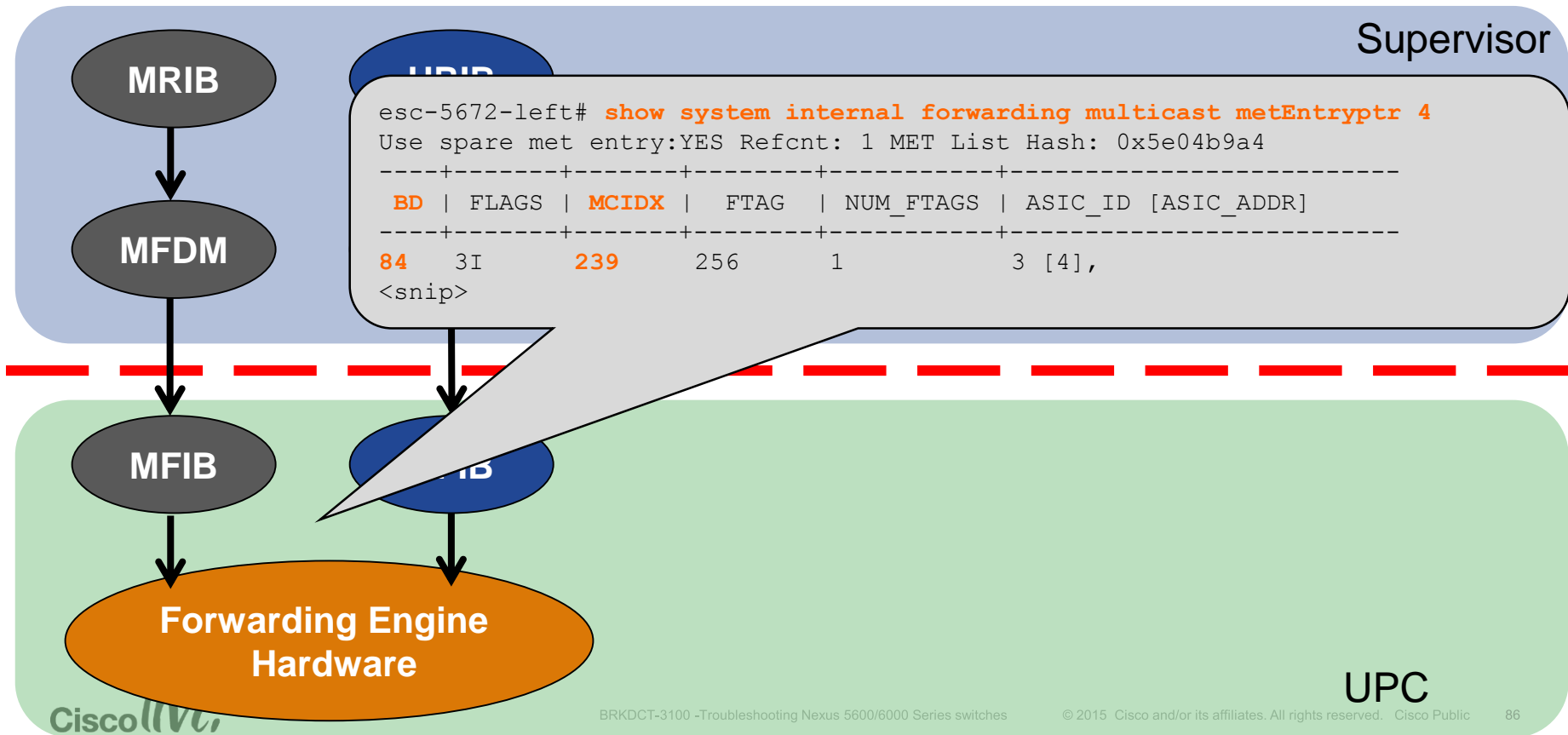
Layer 3 Data Plane Troubleshooting: Multicast



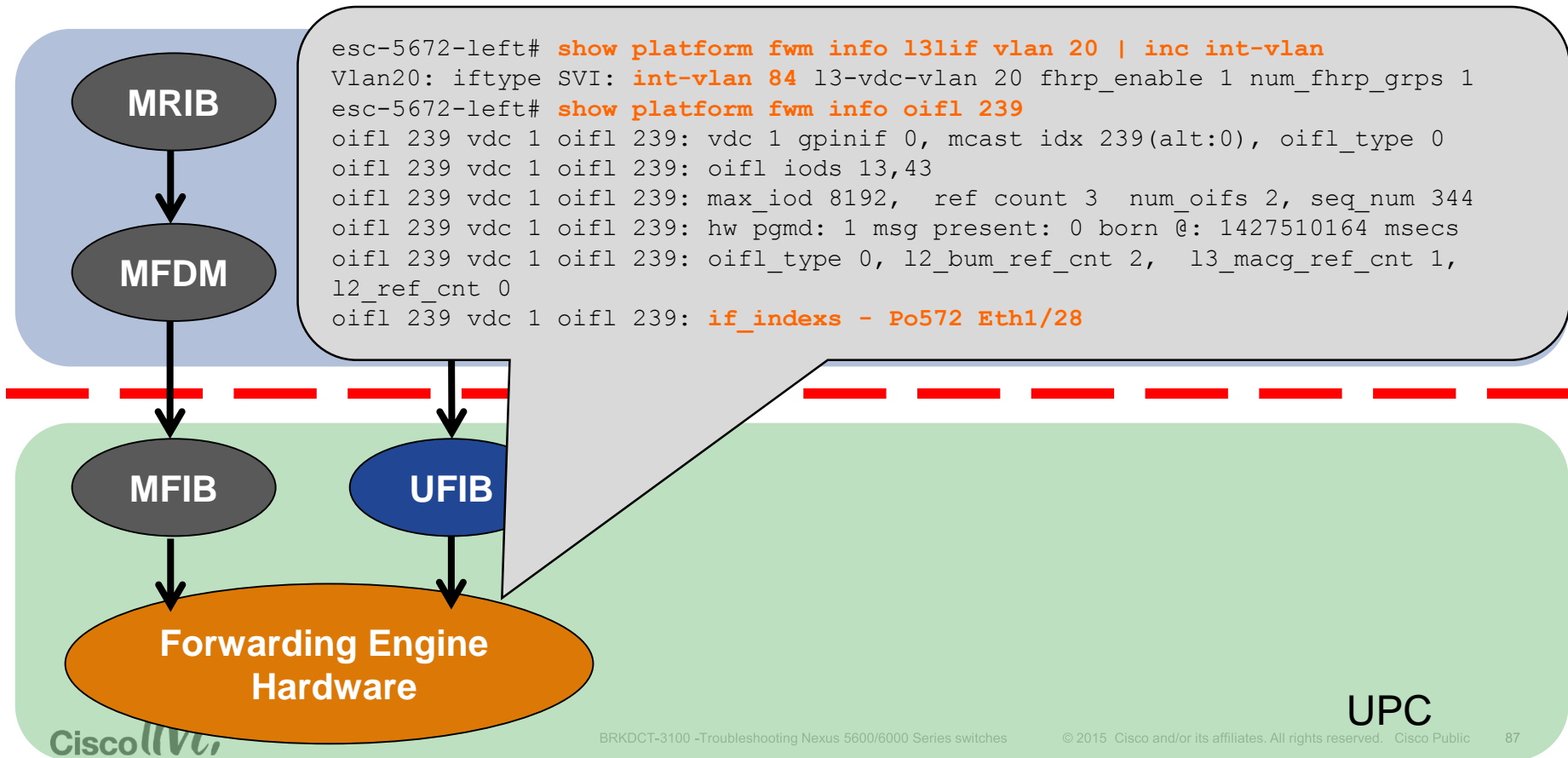
Layer 3 Data Plane Troubleshooting: Multicast



Layer 3 Data Plane Troubleshooting: Multicast



Layer 3 Data Plane Troubleshooting: Multicast

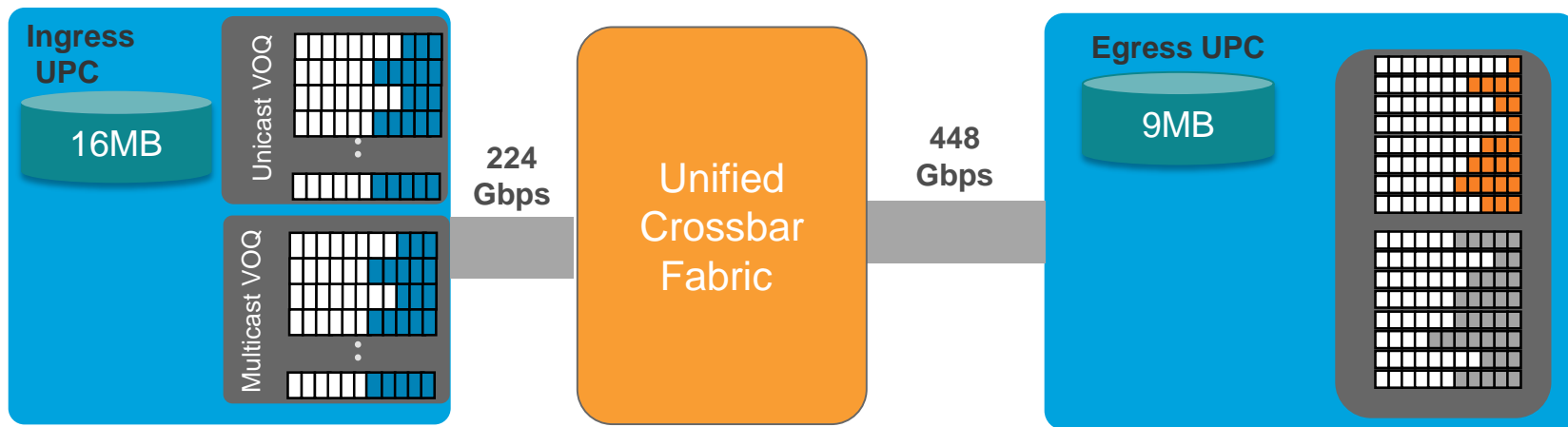


Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

Packet Buffering

- 25MB packet buffer is shared by every three 40 GE ports or twelve 10 GE ports.
- Buffer is 16MB at ingress and 9MB at egress.
- Unicast packet can be buffered at both ingress and egress.
- Multicast Buffered at egress only



Flexible Buffer Management

Default Ingress Buffer Allocation with default QoS

Buffer Pool	10 GE Port	40 GE Port
Control traffic (per port)	64 KB	67.2 KB
SPAN (per port)	38.4 KB	153.6 KB
Class default (per port)	100 KB	100 KB
Shared buffer	13.2 MB	14.7 MB

```
esc-5672-left(config)# policy-map type network-qos custom-queue-limit
esc-5672-left(config-pmap-nq)# class type network-qos class-default
esc-5672-left(config-pmap-nq-c)# queue-limit ?
<20480-1005440> Queue size in bytes(max-limit for user-defined class-905280)
esc-5672-left(config-pmap-nq-c)# queue-limit 1005440 bytes
esc-5672-left(config-pmap-nq-c)# system qos
esc-5672-left(config-sys-qos)# service-policy type network-qos custom-queue-limit
esc-5672-left(config-sys-qos)#
```

Flexible Buffer Management

Default Egress Buffer Allocation with default QoS

Buffer Pool	10 GE Port	40 GE Port
Unicast (per port)	363 KB	650KB with 10G fabric mode 635KB with 40G fabric mode
Multicast (per ASIC)	4.3 MB	6.6 MB

Multicast buffer tuning coming in NX-OS 7.2

```
esc-5672-left(config)# hardware multicast-buffer-tune ?  
<CR>  
esc-5672-left(config)# hardware multicast-buffer-tune
```

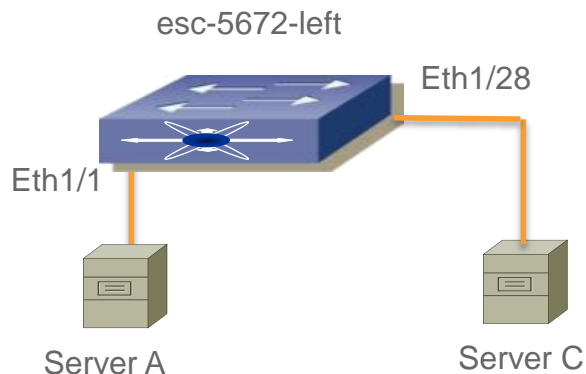
Nexus 5600/6000 Queuing

Queuing implication on troubleshooting:

For unicast traffic, congestion drops occur at **INGRESS!**

For multicast(flooded traffic), congestion drops occur at **EGRESS!**

Nexus 5600/6000 Unicast Queuing



Problem:

- Server A traffic is getting dropped/Poor performance

Given:

- Server A is sending traffic toward Server C. Possibly other servers too
- All servers have had resolved ARP entries resolved.
- All servers are configured to be in the same VLAN

Nexus 5600/6000 Unicast Queuing

- Server A sending line rate bursts of 100000 packets to Server C

```
esc-5672-left# show queuing interface ethernet 1/1 ; sh queuing interface ethernet 1/28
```

```
Ethernet1/1 queuing information:
```

```
<snip>
```

```
RX Queuing
```

```
qos-group 0
```

```
q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
```

```
drop-type: drop, xon: 0, xoff: 0
```

```
Statistics:
```

```
Pkts received over the port : 100000
```

```
Ucast pkts sent to the cross-bar : 100000
```

```
Mcast pkts sent to the cross-bar : 0
```

```
Ucast pkts received from the cross-bar : 0
```

```
Pkts sent to the port : 30
```

```
Pkts discarded on ingress : 0
```

```
Per-priority-pause status : Rx (Inactive), Tx (Inactive)
```

```
Ethernet1/28 queuing information:
```

```
<snip>
```

```
RX Queuing
```

```
qos-group 0
```

```
q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
```

```
drop-type: drop, xon: 0, xoff: 0
```

```
Statistics:
```

```
Pkts received over the port : 0
```

```
Ucast pkts sent to the cross-bar : 0
```

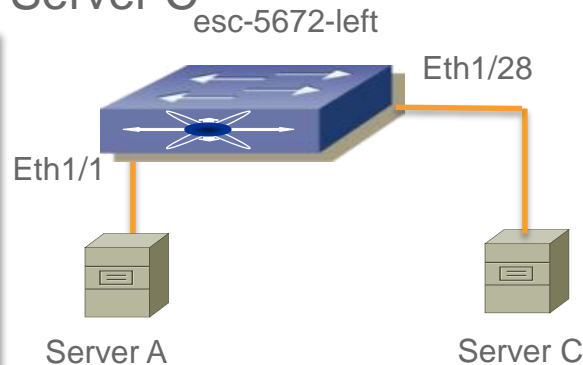
```
Mcast pkts sent to the cross-bar : 0
```

```
Ucast pkts received from the cross-bar : 100000
```

```
Pkts sent to the port : 100028
```

```
Pkts discarded on ingress : 0
```

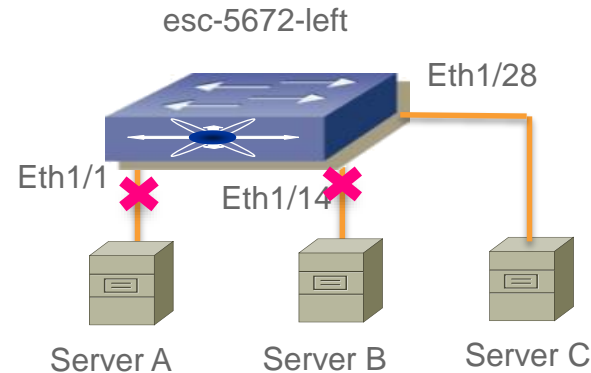
```
Per-priority-pause status : Rx (Inactive), Tx (Inactive)
```



Nexus 5600/6000 Unicast Queuing

- Server B comes online and getting complaints about performance problems
- You are seeing packet discards on ingress

```
esc-5672-left# show queuing interface ethernet 1/1
Ethernet1/1 queuing information:
  TX Queuing
    qos-group    sched-type  oper-bandwidth
    0            WRR         100
  RX Queuing
    qos-group 0
    q-size: 100160, q-size-40g: 100160, HW MTU: 1500 (1500 configured)
    drop-type: drop, xon: 0, xoff: 0
  Statistics:
    Pkts received over the port          : 100000
    Ucast pkts sent to the cross-bar      : 59007
    Mcast pkts sent to the cross-bar      : 0
    Ucast pkts received from the cross-bar : 0
    Pkts sent to the port                 : 14
    Pkts discarded on ingress             : 40993
    Per-priority-pause status             : Rx (Inactive), Tx (Inactive)
esc-5672-left# sh int ethernet 1/1 | inc discard|input
100000 input packets 51200000 bytes
0 input with dribble 40993 input discard
esc-5672-left# sh int ethernet 1/14 | inc discard|input
100000 input packets 51200000 bytes
0 input with dribble 40928 input discard
```



Nexus 5600/6000 Unicast Queuing

- Ingress discards typically are due to an egress congestion
- Common causes
 - Multiple interfaces sending bursts to one interface
 - Speed mismatch(Ex: 40G interface sending traffic to 10G host)
 - Ether-channel hash collision
 - Microbursts filling up ingress buffers
- Nexus 5600/6000 has a rich suite of Data Analytics
 - Microburst monitoring
 - SPAN/ERSPAN on drop
 - Latency monitoring
 - Buffer usage monitoring
 - Capability to identify congested egress

Nexus 5600/6000 Unicast Queuing

- Eth1/28(ASIC#3) congested due to line rate bursts from Eth1/1(ASIC#1) and Eth1/14(ASIC#2)

```
esc-5672-left# show platform software qd info counters voq ASIC-num 1
```

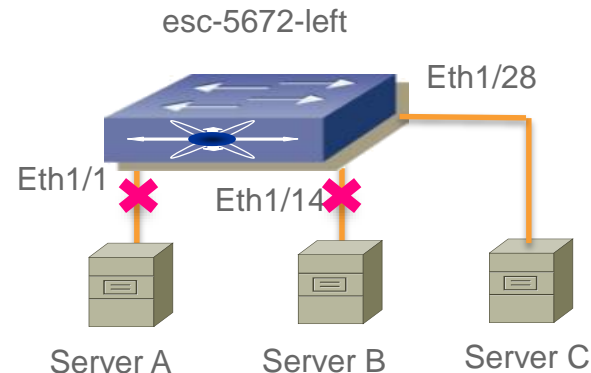
port	TRANSMIT	TAIL DROP	HEAD DROP
Eth1/28			
QUEUE-3	59007	40993	0

```
esc-5672-left# show platform software qd info counters voq ASIC-num 2
```

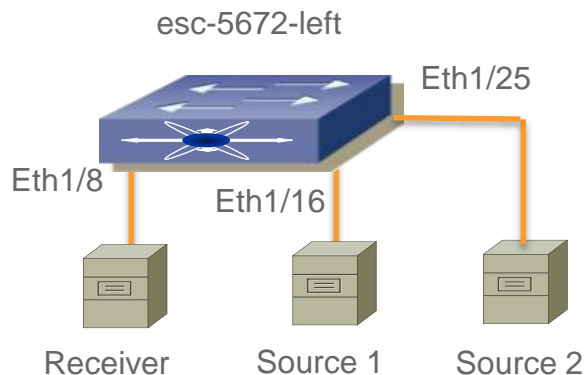
port	TRANSMIT	TAIL DROP	HEAD DROP
Eth1/28			
QUEUE-3	59072	40928	0

```
esc-5672-left# show platform software qd info counters voq interface ethernet 1/28
```

slot	asic	TRANSMIT	TAIL DROP	HEAD DROP
0	1			
	QUEUE-3	59007	40993	0
0	2			
	QUEUE-3	59072	40928	0



Nexus 5600/6000 Multicast Queuing



Problem:

- Multicast Application is seeing gaps/poor performance

Given:

- Receiver has sent IGMP reports for and configuration is correct
- Source 1 is sending line rate burst of 500K packets to 238.1.1.1
- Source 2 is sending line rate bursts of 500K packets to 239.1.1.1

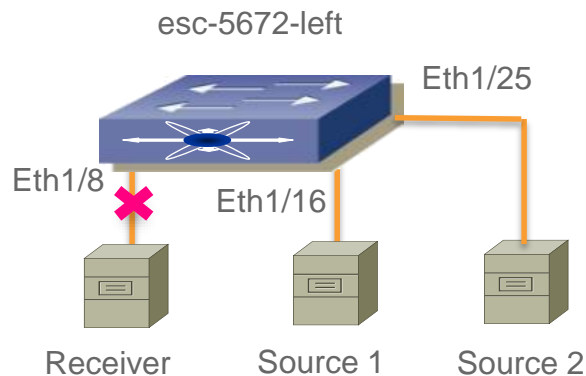
Nexus 5600/6000 Multicast Queuing

- Multicast queuing happens on egress.
- Drops seen as discards on egress

```
esc-5672-left# show interface ethernet 1/8 | inc discard|multicast|X
RX
 0 unicast packets  3 multicast packets  0 broadcast packets
 0 input with dribble  0 input discard
TX
 0 unicast packets  636576 multicast packets  0 broadcast packets
 0 lost carrier  0 no carrier  0 babble  363484 output discard
esc-5672-left#
```

```
esc-5672-left# show interface ethernet 1/16 | inc discard|multicast|X
RX
 0 unicast packets  500000 multicast packets  0 broadcast packets
 0 input with dribble  0 input discard
```

```
esc-5672-left# show interface ethernet 1/25 | inc discard|multicast|X
RX
 0 unicast packets  500000 multicast packets  0 broadcast packets
 0 input with dribble  0 input discard
```



Multicast buffer tuning parameters for better burst absorption

```
esc-5672-left(config)# hardware multicast-buffer-tune
esc-5672-left(config)# hardware multicast-port-prune-threshold ?
<720-8000> Threshold in Kbytes
```

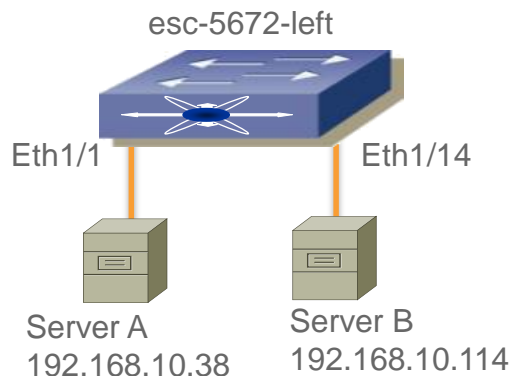
Agenda

- Introduction
- Platform Overview and Troubleshooting
 - MTS
 - Crashes
 - CPU/Etheralyzer
 - CRC Errors
 - Forwarding
 - Buffering/Queuing
 - ELAM

Nexus 5600/6000 ELAM

- Embedded Logic Analyzer(ELAM) is supported starting 7.x
- Captures the first packet which matches trigger
- Implemented in hardware with a parallel snoop process of actual packet decision process
- Cannot be used to troubleshoot packet loss/performance problems.
- There is no impact/penalty to switch or traffic due to ELAM
- Avoids need for SPAN packet captures
- Meant to be used for troubleshooting by TAC/development

Nexus 5600/6000 ELAM Scenario



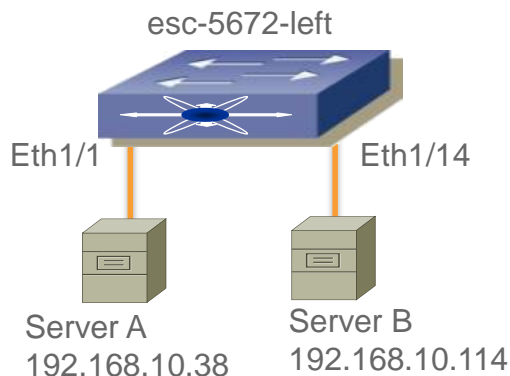
Goal:

- Trace pings from server A to server B

Given:

- Server A is sending ICMP traffic toward Server B.
- Servers have resolved ARP, no other apparent problems seen in switch

Nexus 5600/6000 ELAM Scenario



```
esc-5672-left# show hardware internal bigsur all-ports | egrep name|1/1|1/14
```

name	idx	slot	asic	eport	logi	flag	adm	opr	if_index	diag	ucVer
1gb1/1	1	0	1	0 p	0	b3	en	up	1a000000	pass	0.00
1gb1/14	2	0	2	1 p	13	b3	en	up	1a00d000	pass	0.00

Nexus 5600/6000 ELAM

- Multiple options available for ELAM
- Be as specific as possible but ELAM can be triggered for all slots/ASIC instances, combination of IP/MAC addresses, VLAN, L3-4 protocol types etc
- Here ingress ELAM is being set to trigger on a specific source/destination IP address

```
esc-5672-left# elam slot 1 asic bigsur instance 1
esc-5672-left(bigsur-elam) # trigger lu ingress ?
    arp    ARP Frame Format
    ce     CE Frame Format
    fc     FC Frame Format
    ipv4   IPv4 Frame Format
esc-5672-left(bigsur-elam) # trigger lu ingress ipv4 if source-ipv4-address_ipv4
192.168.10.38 destination-ipv4-address_ipv4 192.168.10.114
esc-5672-left(bigsur-elam) # start capture
esc-5672-left(bigsur-elam) # show capture lu
ELAM: Nothing captured
esc-5672-left(bigsur-elam) #
```

Nexus 5600/6000 ELAM

- ELAM gets triggered when traffic hits the ASIC it is configured on

```
esc-5672-left(bigsur-elam)# show capture lu
Ingress Interface: Ethernet1/1 IS NOT A PC
+-----+
|                Lookup Vector                |
+-----+
| Field          | Raw Value          |
+-----+
| SID            | 0                  |
| PKT_ID         | 15                 |
<snip>
| CE_DA          | 0x001094100114    |
| CE_SA          | 0x001094100011    |
<snip>
| L3_IPV6        | 0                  |
| L3_SA          | 192.168.10.38     |
| L3_DA          | 192.168.10.114    |
| L3_TOS         | 0                  |
| L3_FRAG        | 0                  |
| L3_MF          | 0                  |
| L3_TTL         | 64                 |
| <snip>         |                    |
| L3_ESP         | 0                  |
| L3_PROT        | 1                  |
| L3_LENGTH      | 84                 |
<snip>
esc-5672-left(bigsur-elam)#
```

Nexus 5600/6000 ELAM

- Switching decision(result vector) and packet can be displayed

```
esc-5672-left(bigsur-elam)# show capture rs
Egress Interface: Ethernet1/14 IS NOT A PC
+-----+
|               Result Vector               |
+-----+
| Field          | Raw Value      |
+-----+
| NSH_WORD2      | 0x5e0040       |
| CE_DA          | 0x001094100114 |
| CE_DA_RW       | 0              |
| CE_SA          | 0x001094100011 |
| CE_SA_RW       | 0              |
<snip>
| L3_DA          | 192.168.10.114 |
| L3_DA_RW       | 0              |
| L3_SA          | 192.168.10.38  |
| L3_SA_RW       | 0              |
| L3_TTL         | 64             |
<snip>
| EXT_VLAN       | 10             |
<snip>
|
+-----+
esc-5672-left(bigsur-elam)#
```

Nexus 5600/6000 ELAM

- ELAM on Egress ASIC(Eth1/28 ASIC 2)

```
esc-5672-left(bigsur-elam)# elam slot 1 asic bigsur instance 2
esc-5672-left(bigsur-elam)# trigger lu egress ipv4 if source-ipv4-address_ipv4 192.168.10.38 destination-ipv4-
address_ipv4 192.168.10.114
esc-5672-left(bigsur-elam)# start capture
esc-5672-left(bigsur-elam)# show capture lu
ELAM: Nothing captured
esc-5672-left(bigsur-elam)# show capture lu
Egress Interface: Ethernet1/14 IS NOT A PC
```

Egress

```
+-----+
|                Lookup Vector                |
+-----+
|      Field      |      Raw Value      |
+-----+
| SID              | 3                    |
| PKT_ID           | 10                   |
<snip>
| NSH_WORD2        | 0x5e0040             |
| CE_DA            | 0x001094100114       |
| CE_SA            | 0x001094100011       |
<snip>
| L3_SA            | 192.168.10.38        |
| L3_DA            | 192.168.10.114       |
| L3_TOS           | 0                     |
```

Session Goals

- Learn about the Nexus 5600/6000 and NX-OS troubleshooting approach ✓
- Learn about common Nexus 5600/6000 issues and how to troubleshoot them ✓
- Learn about tools available in NX-OS to troubleshoot common issues ✓



Participate in the “My Favorite Speaker” Contest

Promote Your Favorite Speaker and You Could Be a Winner

- Promote your favorite speaker through Twitter and you could win \$200 of Cisco Press products (@CiscoPress)
- Send a tweet and include
 - Your favorite speaker’s Twitter handle @prkrishn
 - Two hashtags: #CLUS #MyFavoriteSpeaker
- You can submit an entry for more than one of your “favorite” speakers
- Don’t forget to follow @CiscoLive and @CiscoPress
- View the official rules at <http://bit.ly/CLUSwin>

Complete Your Online Session Evaluation

- Give us your feedback to be entered into a Daily Survey Drawing. A daily winner will receive a \$750 Amazon gift card.
- Complete your session surveys though the Cisco Live mobile app or your computer on Cisco Live Connect.



Don't forget: Cisco Live sessions will be available for viewing on-demand after the event at [CiscoLive.com/Online](https://cislive.com/online)

Continue Your Education

- Demos in the Cisco campus
- Walk-in Self-Paced Labs
- Table Topics
- Meet the Engineer 1:1 meetings
- Related sessions

Thank you



TOMORROW starts here.