

What You Make Possible



Troubleshooting IOS-XR

BRKSPG-3612

What's in this session?

- Abstract

This session focuses on the IOS-XR Software Architecture as well as the Hardware and Forwarding Architectures for the platforms that run IOS-XR. Initially the System Architectures will be explained, followed by troubleshooting scenarios for both Software and Operating System problems as well as some of the main features of IOS-XR such as Distributed Hardware Forwarding, MPLS Virtual Private Networks (MPLS-VPN) and Multicast.

- This is not an introductory session for IOS XR
- The session intended for an audience operating ASR9K & CRS
- Questions as we go are ok but please try keep it on topic, there will be an opportunity at the end to ask questions

Agenda Overview

- OS & Software Architecture
 - OS & Processes
 - Tools
- Platform & Forwarding Architecture
 - Platform / Hardware
 - Forwarding



Agenda

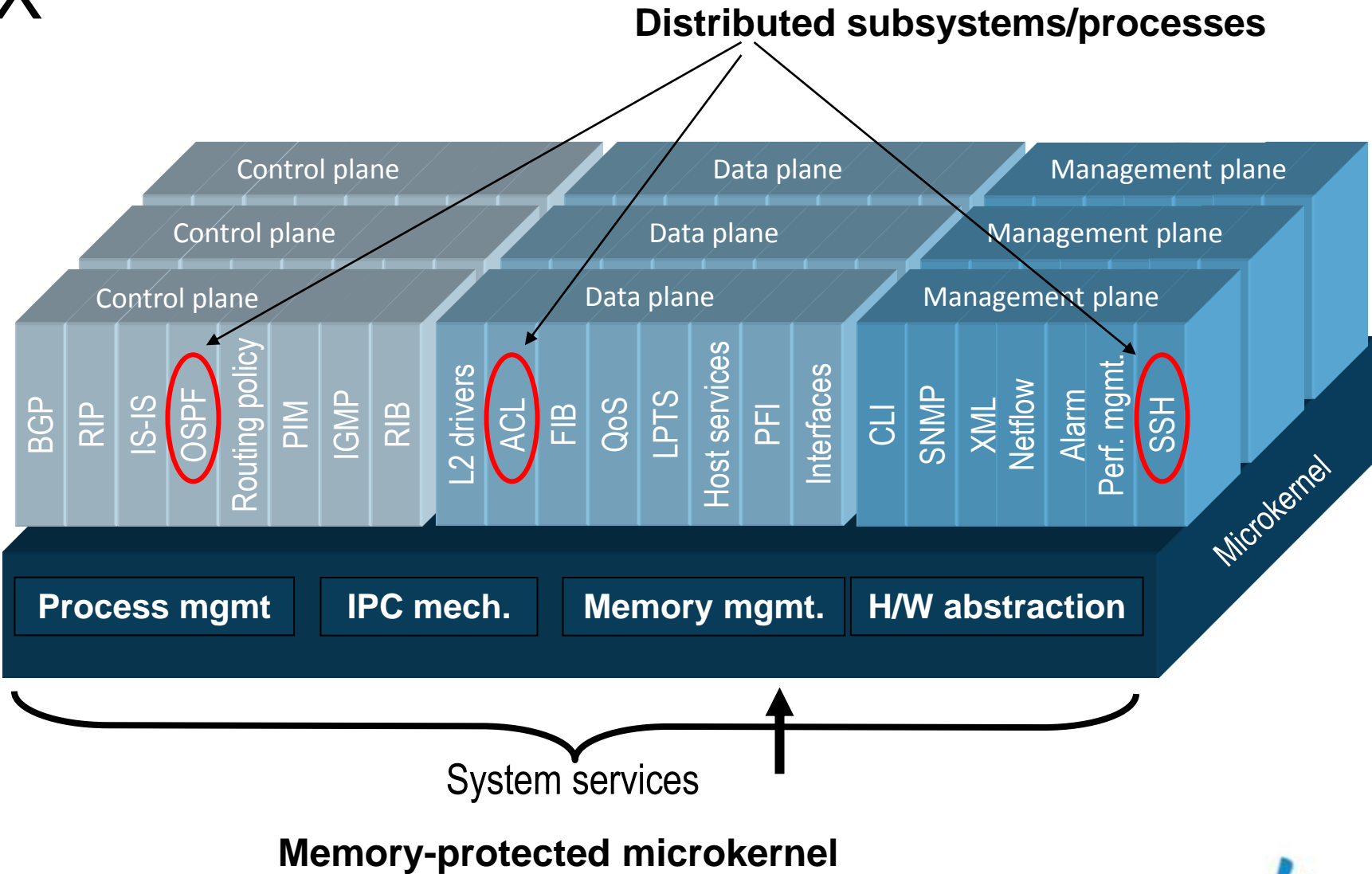
- **OS & Software Architecture**
 - OS & Processes
 - Tools
- **Platform & Forwarding Architecture**
 - Platform / Hardware
 - Forwarding

PROCC

IOS-XR

- IOS-XR runs on top of a UNIX based “micro kernel” called QNX

- What is does a “micro kernel” do?



Quick Tip – Diffing output

- Because XR runs on a UNIX based kernel there are lots of ways to get creative – bash, perl, diff

```
ASR2#show run | utility wc -l
Building configuration...
      366
```

```
HQ1#show int mgmtEth 0/RP0/CPU0/0 | file harddisk:/siMgmt0_1
HQ1#show int mgmtEth 0/RP0/CPU0/0 | file harddisk:/siMgmt0_2
HQ1#run diff /harddisk:/siMgmt0_1 /harddisk:/siMgmt0_2
<      975828 packets input, 65081860 bytes, 0 total input drops
---
>      975883 packets input, 65085172 bytes, 0 total input drops
<SNIP>
```

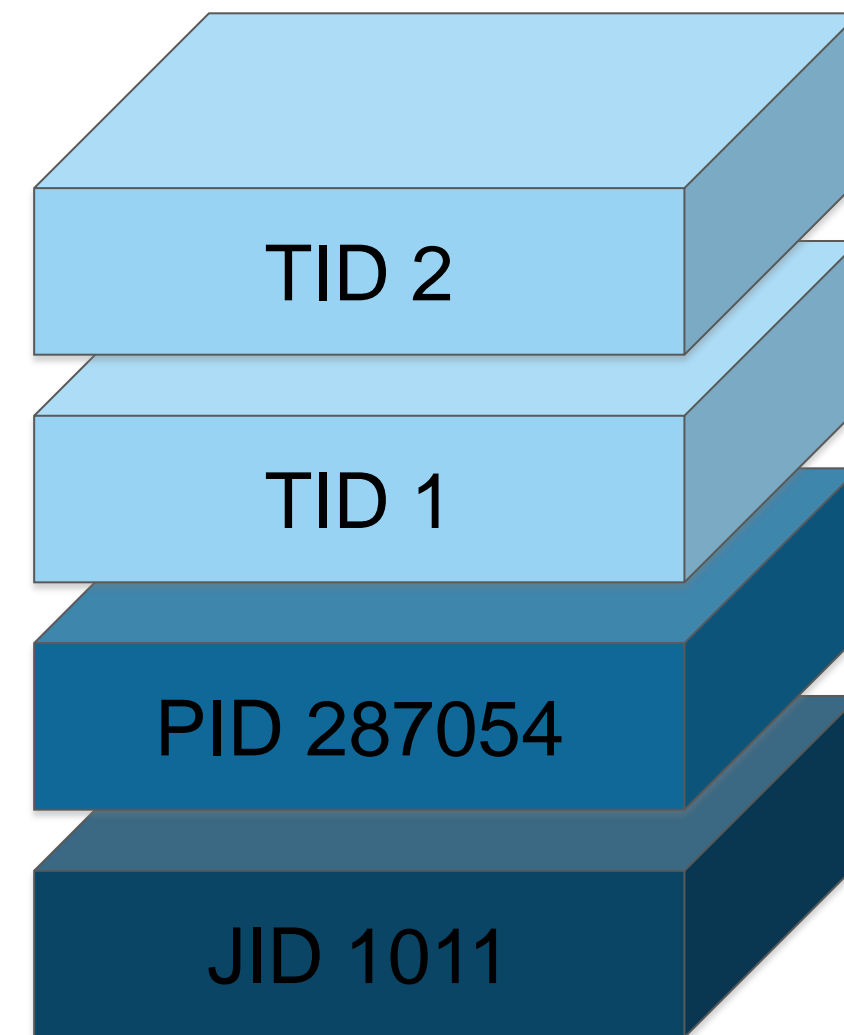


Processes – definitions

Terminology

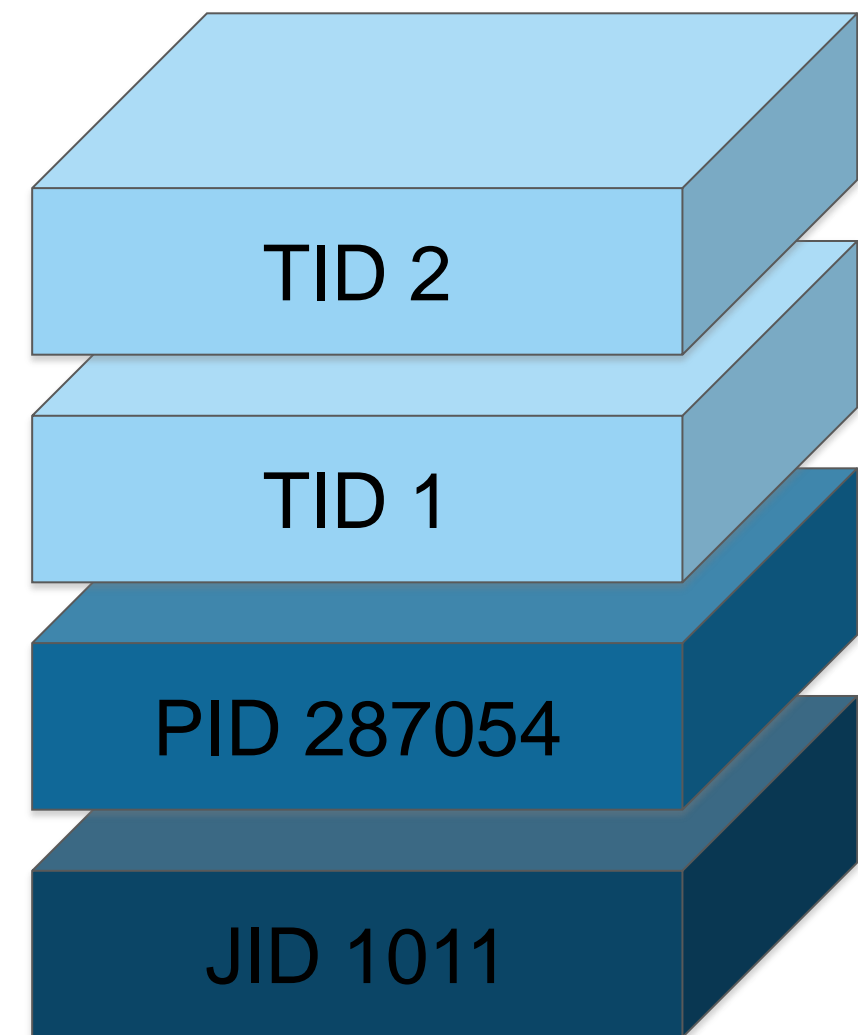
- Threads
- Processes

- Process ID (PID)
- Thread ID (TID)
- Job ID (JID)



Process Management Command

- Basic command
 - `show process`
- Monitor commands:
 - `monitor processes`
 - `monitor threads`
- Troubleshooting commands:
 - `show process blocked`
 - `show context`



Processes – show process

```
ASR1#show processes ospf
```

```
    Job Id: 1011
```

```
    PID: 287054
```

```
    Executable path: /disk0/iosxr-routing-4.2.0/bin/ospf
```

```
    Instance #: 1
```

```
    Version ID: 00.00.0000
```

```
    Respawn: ON
```

```
    Respawn count: 1
```

```
    Max. spawns per minute: 12
```

```
    Last started: Wed Jan 25 18:53:10 2012
```

```
    Process state: Run
```

```
    Package state: Normal
```

JID	TID	CPU	Stack	pri	state	TimeInState	HR:MM:SS:MSEC	NAME
1011	1	0	140K	10	Receive	0:00:00:0437	0:00:03:0149	ospf
1011	2	0	140K	10	Receive	354:32:47:0801	0:00:00:0000	ospf
1011	3	0	140K	10	Receive	354:32:47:0786	0:00:00:0000	ospf
1011	4	1	140K	10	Receive	0:00:30:0594	0:00:00:0064	ospf

```
<SNIP>
```

Processes – show process threadname

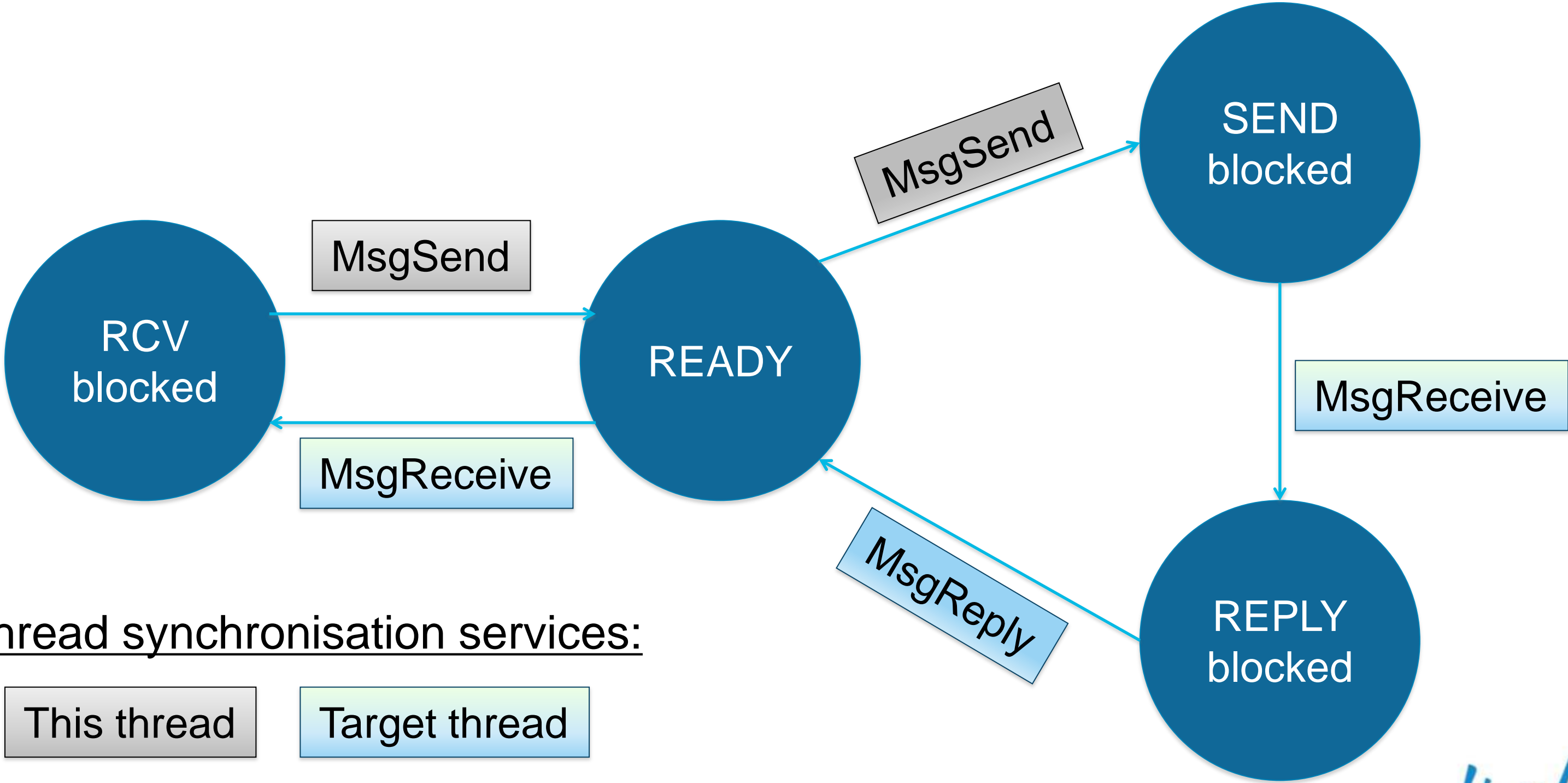
```
ASR1#show processes threadname 1011
```

JID	TID	ThreadName	pri	state	TimeInState	NAME
1011	1	Router-Thread	10	Receive	0:00:00:0963	ospf
1011	2	ITAL Server Thr	10	Receive	0:00:14:0211	ospf
1011	3	chkpt_evm	10	Receive	0:00:06:0247	ospf
1011	4	async	10	Receive	95:53:48:0681	ospf
1011	5		10	Receive	0:00:00:0056	ospf
1011	6		10	Sigwaitinfo	335:21:44:0125	ospf
1011	7	Hello-Thread	10	Receive	0:00:01:0874	ospf
1011	8	EDM-Thread	10	Receive	0:00:04:0756	ospf
1011	9	RIB-Thread	10	Condvar	93:53:34:0077	ospf
1011	10	TE-Thread	10	Receive	335:21:43:0759	ospf
1011	11	NSR-Thread	10	Receive	335:19:43:0758	ospf
1011	12	Protct-Thread	10	Receive	24:52:31:0499	ospf
1011	13		10	Receive	335:21:43:0703	ospf

Processes – Blocking and Thread Sync

- Processes can either block **by design** or as a result of thread synchronisation services breaking down.
- The kernel provides thread synchronisation services:
 - **MUTEX** and **CONDVAR** on the same CPU (non-network synchronisation).
 - **SEND, RECEIVE, REPLY** synchronisation services between threads on a network (different nodes)

Processes – Blocking and Thread Sync



Processes – Blocked

REPLY blocked can be normal

SEND blocked is unusual

HQ2#**show processes blocked**

Jid	Pid	Tid	Name	State	TimeInState	Blocked-on
65546	12298	1	ksh	Reply	26:20:52:0675	12296 devc-cona
53	69664	4	attachd	Reply	26:22:17:035	16397 mqueue
52	69675	2	attach_server	Reply	26:22:18:000	16397 mqueue
347	229468	1	tftp_server	Reply	26:20:15:0776	16397 mqueue
252	327887	2	lpts_fm	Reply	2:48:00:0848	282754 lpts_pa
65754	577754	1	exec	Reply	0:00:00:0108	1 kernel
65831	741671	1	more	Reply	0:00:00:0107	16395 pipe
65832	741672	1	show_processes	Reply	0:00:00:0000	1 kernel
65808	560419088	1	ksh	Send	0:19:02:0011	204906 locald_DSC
65809	558256401	1	exec	Send	5:24:53:0897	204906 locald_DSC
65809	558256401	2	exec	Send	4:58:00:0618	204906 locald_DSC

Example - Can't log in?!

- Telnet/ssh/console to router shows login prompt
- After login banner is shown but no CLI prompt appears, even when using the AUX port only the following syslogs appear



```
RP/0/4/CPU0:May 29 11:05:53.023 AEST: tty_exec_launcher[65839]: %MGBL-TTY-3-SET_PROCESS_GROUP : Failed to set '/dev/aux0' as the controlling terminal for the process: 'Inappropriate I/O control operation'
```

```
RP/0/4/CPU0:May 29 11:05:58.137 AEST: tty_exec_launcher[65839]: %MGBL-TTY-3-SET_PROCESS_GROUP : Failed to set '/dev/aux0' as the controlling terminal for the process: 'Inappropriate I/O control operation'
```

```
RP/0/4/CPU0:May 29 11:06:03.251 AEST: tty_exec_launcher[65839]: %MGBL-TTY-3-SET_PROCESS_GROUP : Failed to set '/dev/aux0' as the controlling terminal for the process: 'Inappropriate I/O control operation'
```

Quick Tip – Translate CLI to KST

- Describe shows what the kernel invokes when the CLI command is run
- Requires taskgroup “cisco-support”

```
XR12K#describe show processes blocked location 0/4/CPU0
```

```
<SNIP>
```

```
Spawn the process:
```

```
show_processes -b -n 64
```



Example - Can't log in?!

- Break into the kernel shell with the sequence

ESC, k, s, h

- Have a look for blocked processes

```
# show_processes -b -n 64
```

Jid	Pid	Tid	Name	State	TimeInState	Blocked-on
65548	16396	1	ksh	Send	0:07:08:0603	118884 locald_DSC
92	28686	1	timesync_client	Reply	1350:15:21:0729	12295 mbus-prp-b
66	20497	2	mbus_text	Reply	0:00:00:0310	12295 mbus-prp-b
63	45097	3	mbi-hello	Reply	1350:15:14:0290	12295 mbus-prp-b
<SNIP>						
291	118884	5	locald_DSC	Reply	427:52:09:0247	168231 tacacsd
291	118884	6	locald_DSC	Reply	243:01:32:0487	168231 tacacsd
291	118884	7	locald_DSC	Reply	210:13:19:0418	168231 tacacsd
291	118884	8	locald_DSC	Reply	334:42:57:0779	168231 tacacsd
291	118884	9	locald_DSC	Reply	1092:08:52:0019	168231 tacacsd
<SNIP>						
65755	815210715	1	ksh	Send	0:07:09:0654	118884 locald_DSC
65841	599527729	1	exec	Send	159:29:00:0687	118884 locald_DSC

Example - Can't log in?!

- Process “tacacsd” wasn't blocked, what was it doing?

```
# sysmgr_show -o -p tacacsd -n 64
```

```
Job Id: 1113
```

```
PID: 168231
```

```
<SNIP>
```

JID	TID	Stack	pri	state	TimeInState	HR:MM:SS:MSEC	NAME
1113	1	92K	16	Sigwaitinfo	1350:10:42:0393	0:00:00:0100	tacacsd
1113	2	92K	10	Nanosleep	0:00:26:0354	0:00:00:0007	tacacsd
1113	3	92K	10	Receive	0:01:14:0939	0:01:13:0612	tacacsd
1113	4	92K	16	Receive	1350:10:42:0406	0:00:00:0000	tacacsd
1113	5	92K	10	Receive	0:01:14:0940	0:00:51:0899	tacacsd
1113	6	92K	10	Receive	210:15:20:0284	0:01:50:0978	tacacsd
1113	7	92K	10	Receive	300:05:07:0179	0:01:15:0541	tacacsd

- IOS-XR even allows the user to “follow” a process, listing a stack trace for each thread!

Example - Can't log in?!

- In this case the problem was restored by issuing a process restart from the kernel shell

```
# sysmgr_control -r tacacsd -n 64
```

- Residual data recovered from traces were used to provide a known defect as the root cause
- For the best chance of getting root cause don't recover a problem before engaging Cisco TAC

Dumping a process core

- A “process core” is a copy of the process memory, written to the local disk
- Processes can crash leaving other processes untouched
- Processes will automatically dump the core when crashing, but this can also be forced for debugging purposes.



CRASH – information collection

- Check logs

```
RP/0/RSP0/CPU0:Feb  9 11:50:32.046 : dumper[59]: %OS-DUMPER-7-  
DUMP_ATTRIBUTE : Dump request with attribute 7 for process  
pkg/bin/bgp
```

```
RP/0/RSP0/CPU0:Feb  9 11:50:32.271 : dumper[59]: %OS-DUMPER-5-  
CORE_FILE_NAME : Core for process pkg/bin/bgp at  
harddisk:/dumper/first.bgp_1047.by.bgp.node0_RSP0_CPU0.ppc.Z on  
local_node
```

```
RP/0/RSP0/CPU0:Feb  9 11:50:36.437 : dumper[59]: %OS-DUMPER-5-  
DUMP_SUCCESS : Core dump success
```

CRASH – information collection

- Check context

```
ASR1#show context
```

```
node:          node0_RSP0_CPU0
```

```
-----  
Core for pid = 19009866 (pkg/bin/bgp)
```

```
Core dump time: Sat Feb 11, 2012: 22:00:11
```

```
Core for process at
```

```
harddisk: /dumper/bgp_1047.by.dumper_gen.node0_RSP0_CP  
U0.ppc.Z
```

Stack Trace

```
#0 0x4c2deebc
```

```
#1 0x4c259304
```

CRASH – information collection

- Obtain the files indicated in the logs

```
ASR1#dir harddisk:/dumper | i bgp
```

```
202490      -rw-   9508581      Thu Feb  9 11:50:34 2012
```

```
first.bgp_1047.by.bgp.node0_RSP0_CPU0.ppc.Z
```

```
202662      -rw-   16153      Thu Feb  9 11:50:34 2012
```

```
first.bgp_1047.by.bgp.node0_RSP0_CPU0.ppc.txt
```

```
202634      -rw-   133284      Thu Feb  9 11:50:35 2012
```

```
first.bgp_1047.by.bgp.node0_RSP0_CPU0.ppc.cpu_info.Z
```

CRASH – information collection

- Capture some more information to assist in crash analysis

```
ASR1#show install which file bgp location
```

```
Node 0/RSP0/CPU0 has file bgp for boot package /disk0/asr9k-os-mbi-4.2.0/0x100000/mbiasr9k-rp.vm from iosxr-routing
```

```
Package:
```

```
  iosxr-routing
```

```
    iosxr-routing V4.2.0[00]  IOS-XR Routing Package Definition
```

```
    Vendor   : Cisco Systems
```

```
    Desc     : IOS-XR Routing Package Definition
```

```
    Build    : Built on Mon Dec 19 18:22:13 AEST 2011
```

```
    Source   : By iox-bld2 in
```

```
/auto/srcarchive6/production/4.2.0/all/workspace for pie
```


Agenda

- **OS & Software Architecture**

- OS & Processes
- **Tools**

- **Platform & Forwarding Architecture**

- Platform / Hardware
- Forwarding



Traces

- Always on debug, no impact to the router
- Residual traces will remain on the router until wrapped

```
RP/0/RSP0/CPU0:ASR2#show cef trace errors unique
```

```
4 unique entries (2048 possible, 66 filtered)
```

```
Dec 14 04:43:57.707 fib/common/evt_err 0/RSP0/CPU0 8# t3 Registration failed.  
retry type 14. Cerr='FIB' detected the 'warning' condition 'Requested data exists  
Duplicates not allowed'
```

```
Dec 14 04:44:38.745 fib/ipv4/evt_err 0/RSP0/CPU0 3# t3 BCDL batched ADD for 1  
tables failed. Cerr='FIB' detected the 'try again' condition 'Temporary failure  
Try again later'
```

```
Dec 18 02:39:08.723 fib/ipv4/evt_err 0/RSP0/CPU0 3# t3 NHIdbExtLkup failed.  
Cerr='FIB' detected the 'warning' condition 'no protocol extension for ifhandle'
```

```
Dec 19 06:20:07.272 fib/ipv4/evt_err 0/RSP0/CPU0 3# t3  
Path={NhAddr:123.123.123.123/32,Nhifh:0} resolution failed. Cerr=No error
```

Example – OSPF & traces

- Logging shows OSPF neighbor is stuck in EXSTART

```
RP/0/RSP0/CPU0:ASR2#show logging process ospf
```

```
Syslog logging: enabled (0 messages dropped, 0 flushes, 0 overruns)
```

```
Console logging: level debugging, 5926 messages logged
```

```
Monitor logging: level debugging, 5476 messages logged
```

```
Trap logging: level informational, 0 messages logged
```

```
Buffer logging: level debugging, 5989 messages logged
```

```
RP/0/RSP0/CPU0:Dec 19 06:32:35.250 : ospf[1011]: %ROUTING-OSPF-5-ADJCHG :  
Process 1, Nbr 123.123.123.123 on GigabitEthernet0/2/0/0 in area 0 from DOWN  
to DOWN, Neighbor Down: dead timer expired,vrf default vrfid 0x60000000
```

```
RP/0/RSP0/CPU0:Dec 19 06:33:43.693 : ospf[1011]: %ROUTING-OSPF-5-ADJCHG :  
Process 1, Nbr 123.123.123.123 on GigabitEthernet0/2/0/0 in area 0 from  
EXSTART to DOWN, Neighbor Down: too many DBD retransmissions,vrf  
default vrfid 0x60000000
```

Example – OSPF & traces

- OSPF has many types of traces, we're interested in errors

```
RP/0/RSP0/CPU0:ASR2#show ospf trace
```

```
OSPF Trace Summary (1, RP/0/RSP0/CPU0:ASR2, 0M)
```

Trace Name	Size	Count	Description
1. adj	8192	775	adjacency
2. adj_cycle	8192	1527	dbd/flood events/pkts
3. config	2048	20	config events
4. errors	8192	122	errors
5. events	4096	119	mda/rtrid/bfd/vrf
6. ha	8192	400	startup/HA/NSF

Example – OSPF & traces

- Errors show us what we'd normally find in a debug
- This could have happened long before we came along

```
RP/0/RSP0/CPU0:ASR2#show ospf trace errors
```

```
Traces for OSPF 1 (Wed Dec 19 06:37:15)
```

```
Traces returned/requested/available: 110/8192/110
```

```
Trace buffer: errors
```

```
109 Dec 19 06:35:40.492 ospf_rcv_dbd: WARN nbr 123.123.123.123 larger  
MTU dbd_if_mtu 1500 oi_ip_mtu 1486
```

```
110 Dec 19 06:35:45.276 ospf_rcv_dbd: WARN nbr 123.123.123.123 larger MTU  
dbd_if_mtu 1500 oi_ip_mtu 1486
```

Quick Tip – XR MTU

- IOS-XR MTU commands are a little different to IOS

```
interface GigabitEthernet0/2/0/0
```

```
  mtu 1500
```

```
interface GigabitEthernet0/2/0/0.100
```

```
  mtu 1500
```

```
RP/0/RSP0/CPU0:ASR2#show ipv4 int gi 0/2/0/0 | i MTU
```

```
MTU is 1500 (1486 is available to IP)
```

```
RP/0/RSP0/CPU0:ASR2#show ipv4 int gi 0/2/0/0.100 | i MTU
```

```
MTU is 1500 (1482 is available to IP)
```



CRS Punt Traffic Capture

- Firstly, must configure for capture

```
HQ2 (config)#int TenGigE 0/2/0/0  
HQ2 (config-if)#capture software packets
```

- Then you can view

```
HQ2#show captured packets ingress interface  
ten0/2/0/0 loc 0/2/cpu0
```

CRS Punt Traffic Capture

```
HQ2#show captured packets ingress int ten0/2/0/0 location 0/2/cpu0
```

```
[20] Feb  4 12:51:26.388, len: 209, hits: 1, i/p i/f: TenGigE0/2/0/2
```

```
[punt reason: CDP] [PPE used: cluster=2 ppe=11]
```

```
[ether dst: 0100.0ccc.cccc src: 0024.98ea.cd51 type/len: 0xc3]
```

```
aaaa0300 000c2000 02b47df9 00010008 41535232 00030012 54656e47 69674530  
2f302f30 2f310002 00110000 00010101 cc00042e 2e2e0600 04000800 00000100  
05005b43 6973636f 20494f53 20585220 536f6674 77617265 2c205665 7273696f
```


ASR9K Traffic Mirroring (SPAN)

- First configure a monitor-session with a destination

```
ASR2 (config) #monitor-session CiscoLIVE destination ?  
interface      Specify a destination interface  
pseudowire    Specify a pseudowire
```

- Then apply the monitor-session under the interesting interface

```
interface GigabitEthernet0/0/0/0  
monitor-session CiscoLIVE  
acl  
mirror first 64  
ipv4 access-list CiscoLIVE  
10 permit pim any any capture  
20 permit ipv4 any any
```

Agenda

- OS & Software Architecture

- OS & Processes
- Tools

- Platform & Forwarding Architecture

- Platform / Hardware
- Forwarding



Platform

IOS-XR Hardware



XR12000
Multi-Service Edge

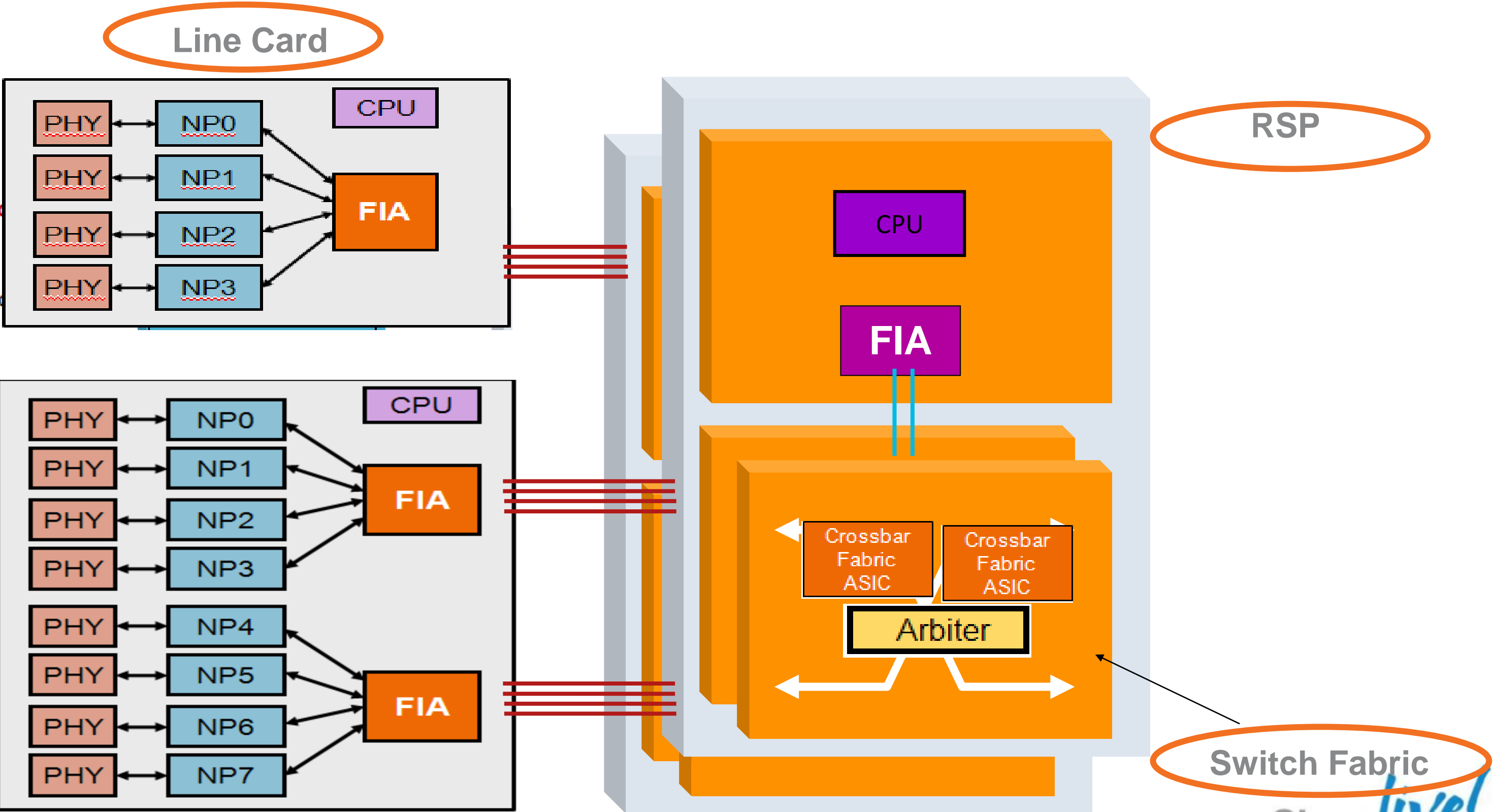


CRS
Core

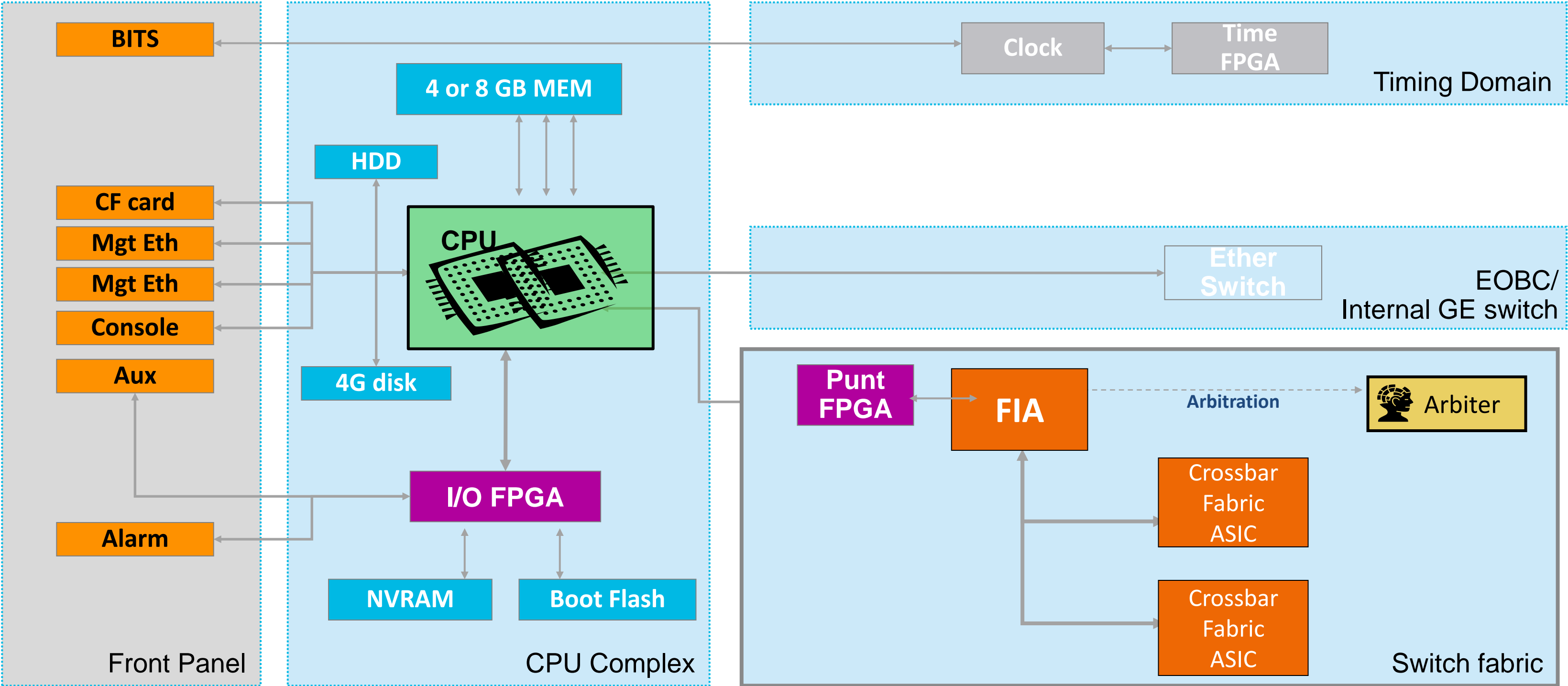
ASR 9000
Aggregation Services



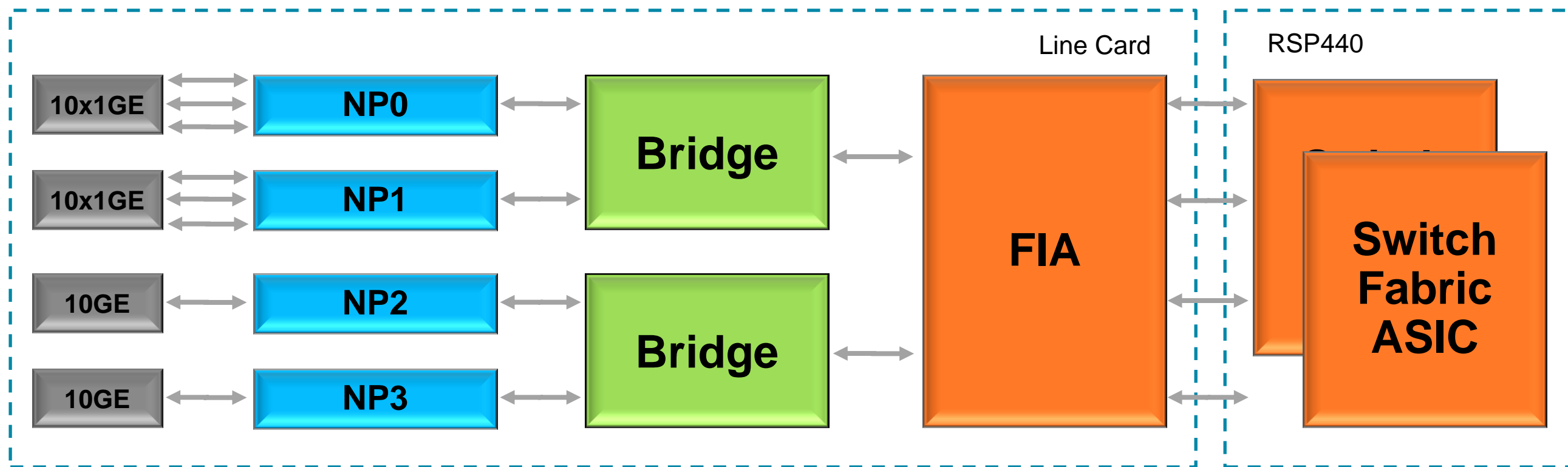
ASR9K - System Architecture "At-a-Glance"



ASR9K - RSP Engine Architecture



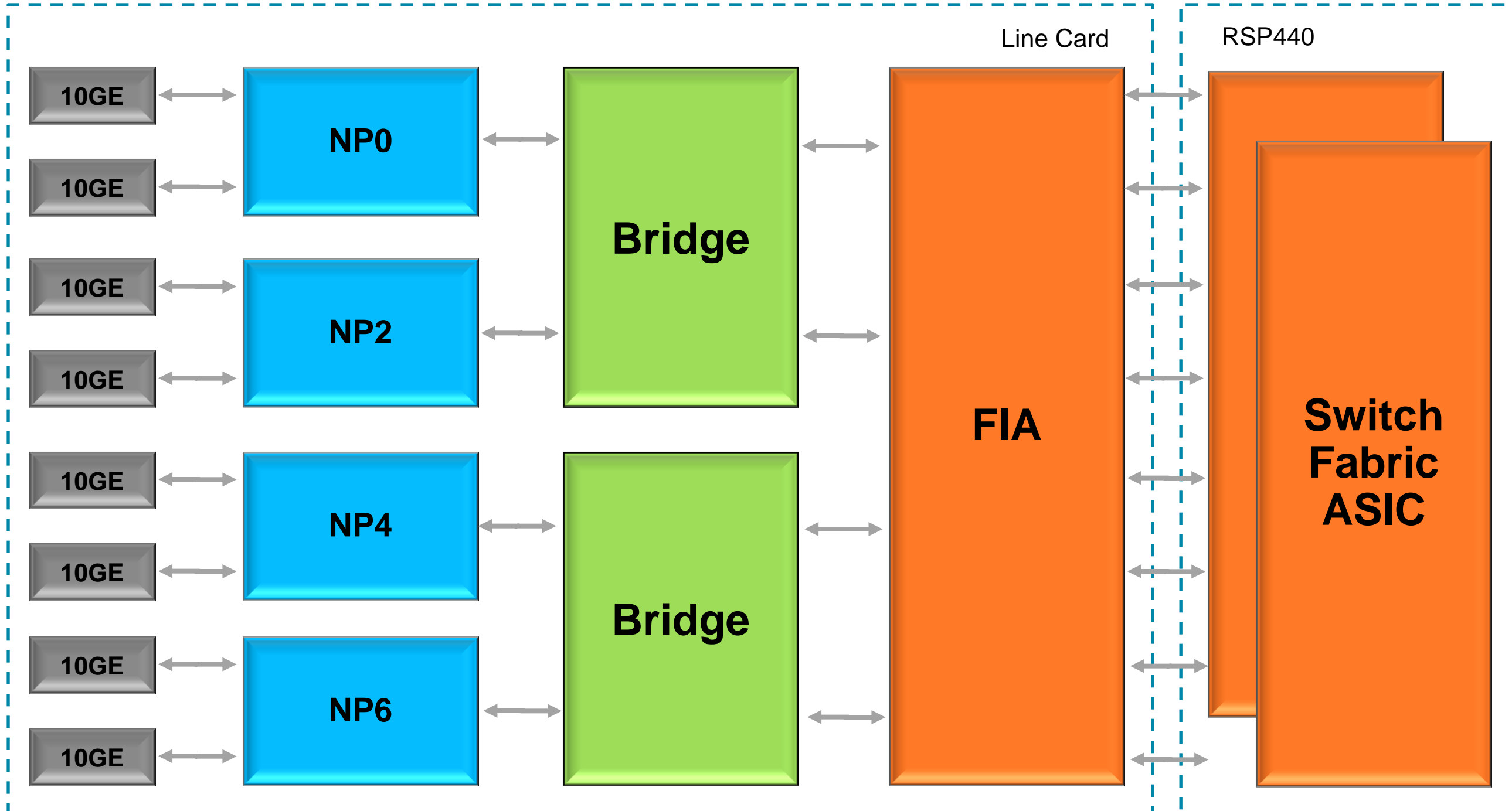
ASR9K – Trident Line Card Architecture



Example: A9K-2T20G-E

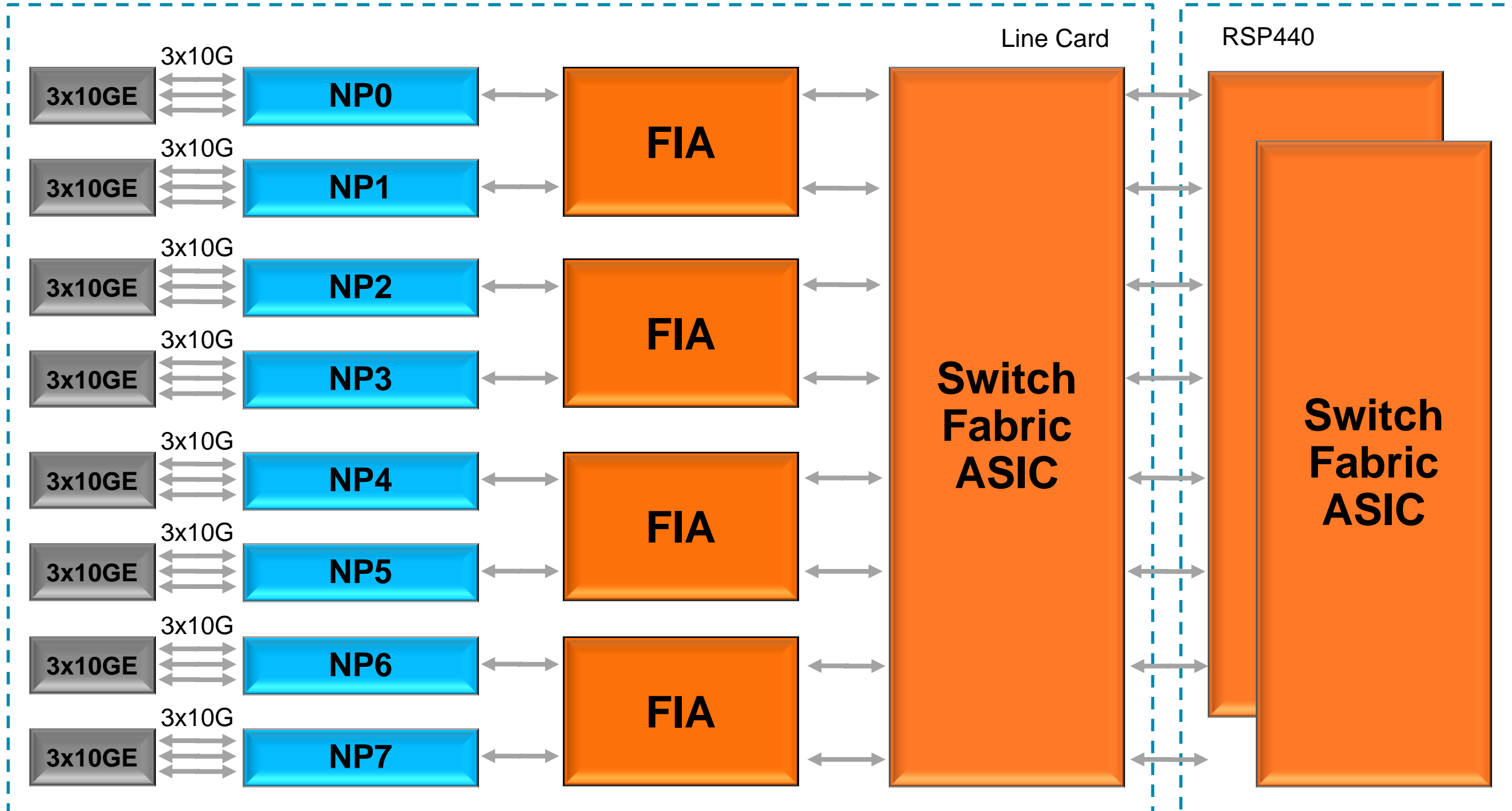
- Every line card has a CPU which runs distributed control plane protocols like BFD & ARP, even the FIB is calculated independently
- Multicast replication can occur at every stage in the forwarding

ASR9K – Trident Line Card Architecture



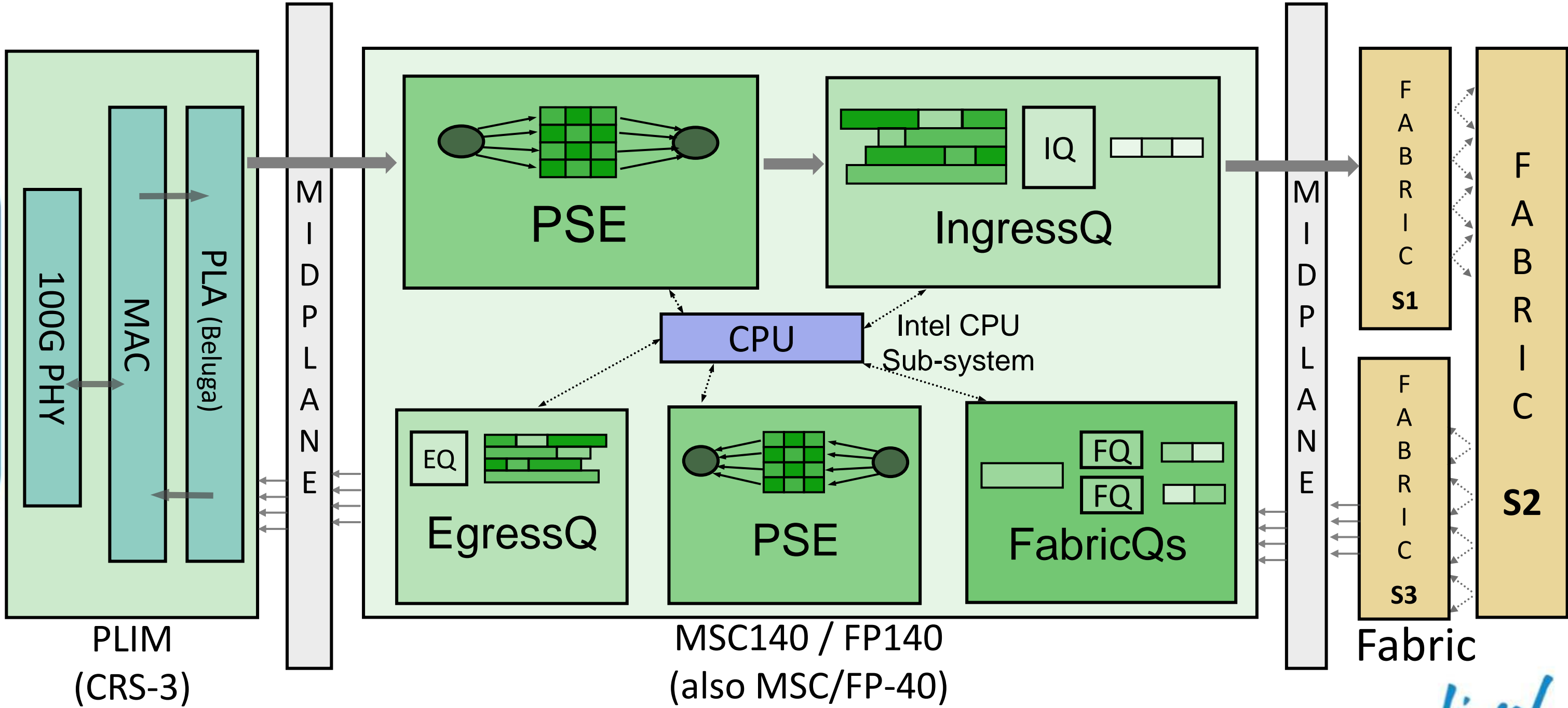
Example: A9K-8T/4-E

ASR9K – Typhoon Line Card Architecture



Example: A9K-24x10G

CRS Line Card Architecture



Example - Line card is crashing

- Sometimes in life it can all feel a little overwhelming

```
LC/0/1/CPU0:Dec 30 13:54:05.145 GMT-7: wdsysmon[306]: Process asic_scan_server pid 36909 prio 22 using 88 percent is the top user of CPU
LC/0/1/CPU0:Dec 30 13:54:05.145 GMT-7: wdsysmon[306]: Top CPU user process asic_scan_server pid 36909 is a known hog.
LC/0/1/CPU0:Dec 30 13:54:05.744 GMT-7: wdsysmon[306]: Process asic_scan_server pid 36909 prio 22 using 87 percent is the top user of CPU
LC/0/1/CPU0:Dec 30 13:54:05.744 GMT-7: wdsysmon[306]: Top CPU user process asic_scan_server pid 36909 is a known hog.
LC/0/1/CPU0:Dec 30 13:54:06.751 GMT-7: asic_scan_server[102]: %L2-ASIC_SCAN_SERVER-2-NODE_BRINGDOWN : reason CHIP FREEZE status SUCCESS. Scanning complete. Node is being brought down : pkg/bin/asic_scan_server : (PID=36909) : -Traceback= 40100f6c 401016f4 4a1652f0 4a16311c 40101b1c 40107044
LC/0/1/CPU0:Dec 30 13:54:05.824 GMT-7: fabricq_mgr[136]: %FABRIC-FABRICQ-3-PCL_PKT : Minor error in PCL of fabricq asic 0. PCL UC Partial Packet: CAOPCI: 0x28 (0/8, UC, LO)
LC/0/1/CPU0:Dec 30 13:54:05.873 GMT-7: ingressq[163]: %FABRIC-INGRESSQ_ALARM-2-DEVICE_HALT : INGRESSQ: device halted, sub-block: SP_TOP_INTR, halted entity: FQM : pkg/bin/ingressq : (PID=41015) : -Traceback= 4a9ecc08 4a5a07a4 4a5983f8 4a1652f0 4a16311c 4a598800 4a1cffb0
LC/0/1/CPU0:Dec 30 13:54:05.875 GMT-7: ipv4_mfwd_partner[155]: %ROUTING-IPV4_MFWD-4-FROM_MRIB_UPDATE : MFIB couldn't process update from MRIB : Cannot process 0xe0000000:(10.0.0.228,239.1.1.190/32) - Table lookup returned Bad file descriptor
LC/0/1/CPU0:Dec 30 13:54:05.940 GMT-7: discovery[123]: %PLATFORM-DISCOVERY-4-PCI_LATCH : ASIC-ERR: PCI0 error #1: Name = MErrMsg #1: Description = a PCI master receiving message error (info latched and follows)
LC/0/1/CPU0:Dec 30 13:54:06.791 GMT-7: ingressq[163]: %FABRIC-INGRESSQ_DLL-4-LNS_LOP_DROP : Error: Cells dropped due to low availability of planes, aggr cell drop count: 404
RP/0/RP0/CPU0:Dec 30 13:54:08.961 GMT-7: shelfmgr[356]: %PLATFORM-SHELFMGR-3-NODE_RESET_ALARM : Node 0/1/SP in critical alarms, request to reset ...
LC/0/1/CPU0:Dec 30 13:54:06.794 GMT-7: discovery[123]: %PLATFORM-DISCOVERY-7-ASMP_PCI_DETAILS : ASIC-ERR: PCI0 error #1: Command: Split Completion, Error Address: 00000000 13000000, Error Attribute: 62000804
LC/0/1/CPU0:Dec 30 13:54:06.800 GMT-7: discovery[123]: %PLATFORM-DISCOVERY-7-ASMP_ERR_REGS : ASIC-ERR: Addr High 00000000 Addr Low 13000000 Err Attribute 04080062, Cause 70004000, Command 0000000C
LC/0/1/CPU0:Dec 30 13:54:06.806 GMT-7: sysmgr[77]: %OS-SYSMGR-2-MANAGED_REBOOT : reboot to be managed by process (hfr_pm_common) reason (ASIC asic_scan instance 0 in critical alarm)
```

Quick Tip – ASIC errors

- All sorts of ASIC errors get logged by the system SBE, MBE and many more!

```
HQ1#show asic-errors all detail location 0/5/CPU0
```

```
<SNIP>
```

```
Name           : Spa Port 1 TSI SPI4 Status OOF
```

```
Thresh/period(s) : 1/0      Alarm state: OFF
```

```
Error count     : 12
```

```
-----  
First N errors.
```

```
@Time, Error-Data
```

```
-----  
Dec 10 00:52:11.723: TSI SPI4 Status OOF
```

```
Dec 10 00:52:13.764: TSI SPI4 Status OOF
```

```
Dec 10 00:52:15.766: TSI SPI4 Status OOF
```



Example - Line card is crashing

- Zeroing in on the logs of interest we can see..

```
LC/0/1/CPU0:Dec 30 13:54:05.873 GMT-7: ingressq[163]:  
%FABRIC-INGRESSQ_ALARM-2-DEVICE_HALT : INGRESSQ: device  
halted, sub-block: SP_TOP_INTR, halted entity: FQM :  
pkg/bin/ingressq : (PID=41015) : -Traceback= 4a9ecc08  
4a5a07a4 4a5983f8 4a1652f0 4a16311c 4a598800  
4a1cffb0
```

```
LC/0/1/CPU0:Dec 30 13:54:05.940 GMT-7: discovery[123]: %PLATFORM-  
DISCOVERY-4-PCI_LATCH : ASIC-ERR: PCI0 error #1: Name = MErrMsg #1:  
Description = a PCI master receiving message error (info latched  
and follows)
```

```
RP/0/RP0/CPU0:Dec 30 13:54:08.961 GMT-7: shelfmgr[356]: %PLATFORM-  
SHELFMGR-3-NODE_RESET_ALARM : Node 0/1/SP in critical alarms, requ  
est to reset ...
```

Example - Line card is crashing

- Showing of asic-errors reveals exactly what happened

```
HQ2#show asic-errors ingressq 0 all location
```

```
*****  
*                               *  
*           Multiple Bit Errors           *  
*                               *  
*****  
Name                : FCRAM memory Double bit error  
Base address        : 0x0  
Node Key            : 0x1021421  
Thresh/period(s)   : 2/60    Alarm state: OFF  
Error count         : 1  
Last clearing       : Thu Dec 30 13:54:06 2010  
Last N errors       : 1  
-----
```



Agenda

- OS & Software Architecture

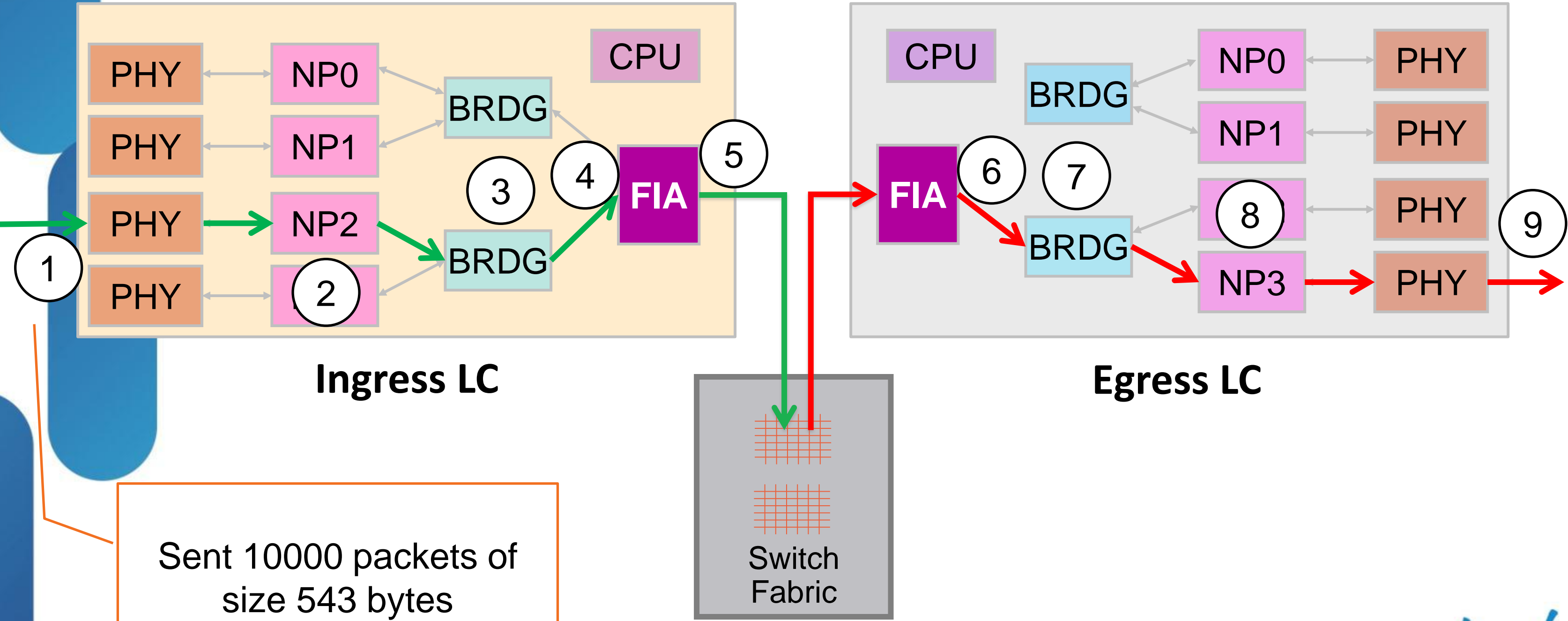
- OS & Processes
- Tools

- **Platform & Forwarding Architecture**

- Platform / Hardware
- **Forwarding**

PracK

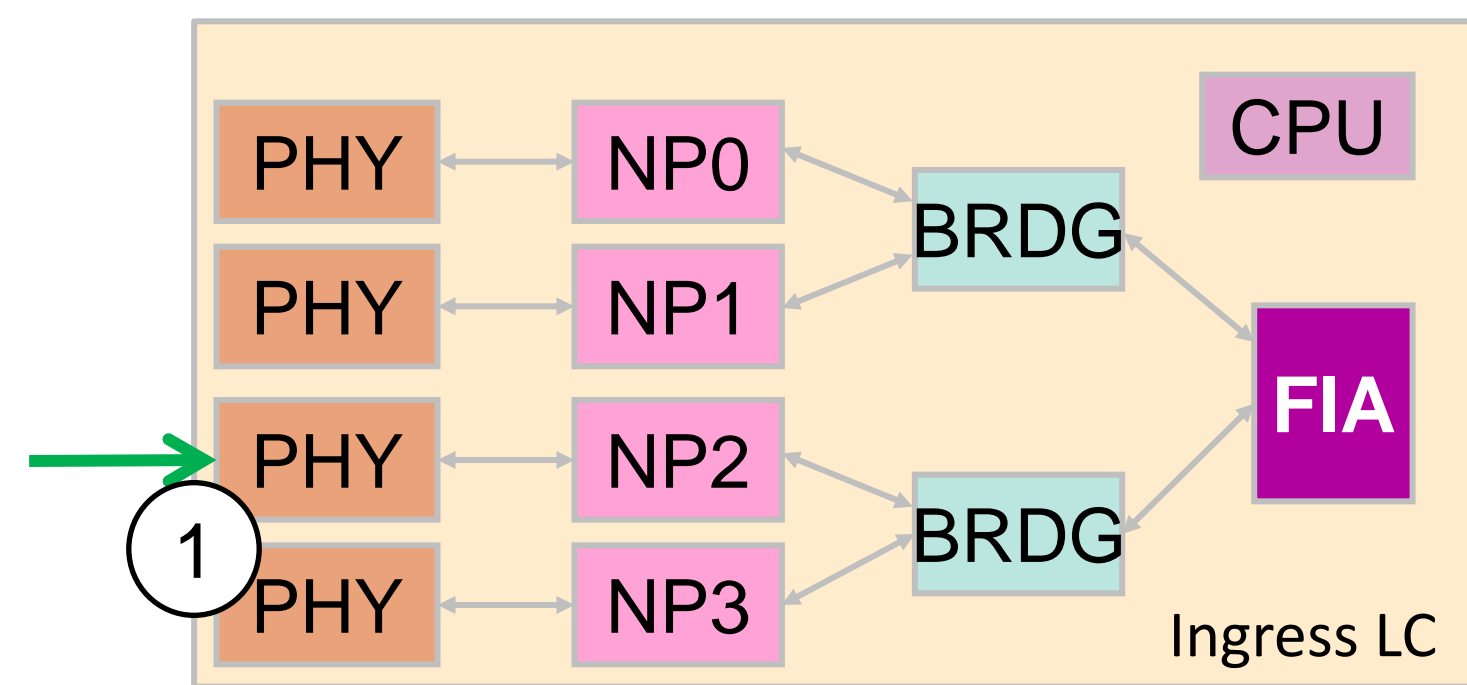
Unicast Forwarding



Sent 10000 packets of size 543 bytes

Unicast Forwarding

Ingress PHY



```
ASR2#show controllers tenGigE 0/1/0/0 stats
```

```
Ingress:
```

```
Input total bytes = 5628648
```

```
Input good bytes = 5628648
```

```
Input total packets = 10046
```

```
Input pkts 512-1023 bytes = 10000
```

```
Input drop overrun = 0
```

```
Input drop abort = 0
```

```
Input error CRC = 0
```

```
Input error collisions = 0
```


Quick Tip – Real time monitoring

```
ASR2#monitor interface TenGigE 0/0/0/0
```

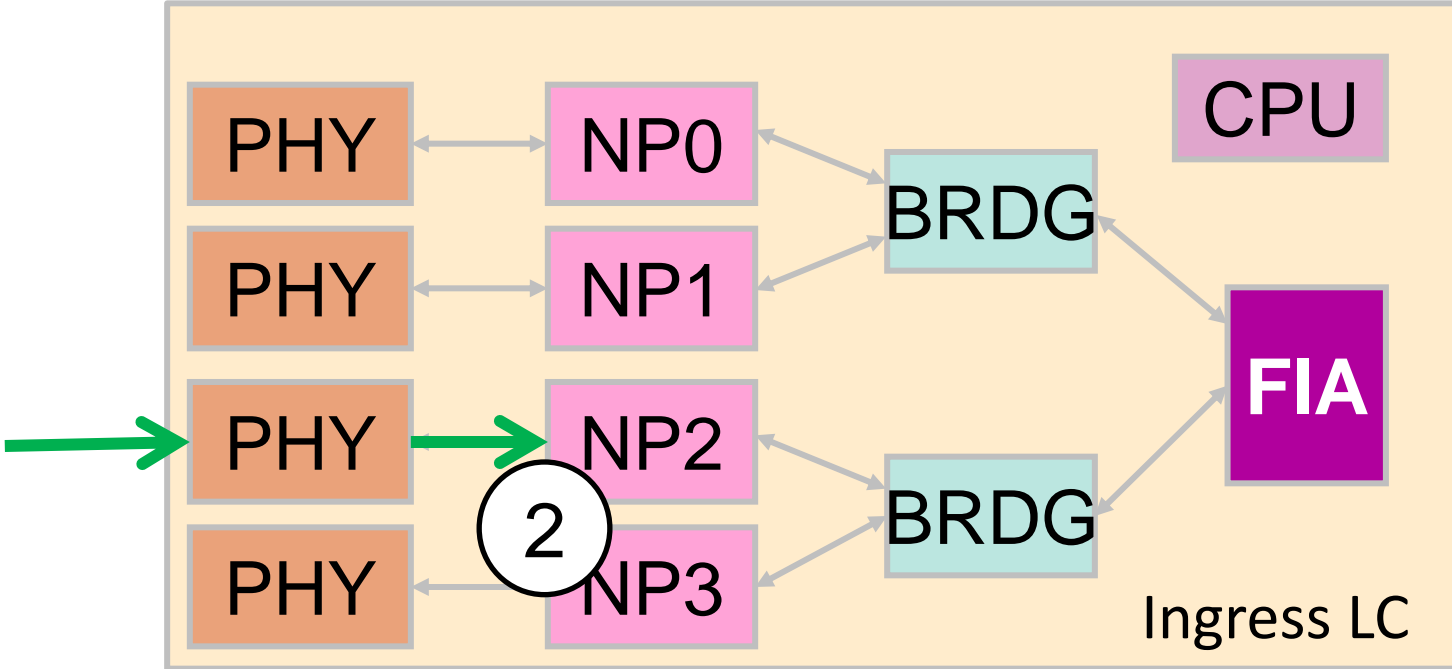
```
Traffic Stats: (2 second rates)                                Delta
Input  Packets:                                                69          4
Input  pps:                                                    2
Input  Bytes:                                                  58217       4670
Input  Kbps (rate):                                           13          ( 0%)
Output Packets:                                               7011       379
Output pps:                                                  24
Output Bytes:                                                  3937905    213117
Output Kbps (rate):                                           110        ( 0%)

Errors Stats:
Input  Total:                                                  0           0
Input  CRC:                                                    0           0
Input  Frame:                                                  0           0
Input  Overrun:                                                0           0
Output Total:                                                  0           0
Output Underrun:                                               0           0
```



Unicast Forwarding

Ingress NP



```
ASR2#show controllers np counters np2 location 0/0/cpu0
```

Read 22 non-zero NP counters:

Offset	Counter	FrameValue	Rate (pps)
22	PARSE_ENET_RECEIVE_CNT	10125	48
29	MODIFY_FABRIC_TRANSMIT_CNT	10000	48

Quick Tip – Which NP?



- Interface, NP, Bridge & FIA mapping

```
ASR2#sh controllers np ports all location 0/0/cpu0
```

```
Node: 0/0/CPU0:
```

NP	Bridge	Fia	Ports
0	0	0	TenGigE0/0/0/3, TenGigE0/0/0/7
1	0	0	TenGigE0/0/0/2, TenGigE0/0/0/6
2	1	0	TenGigE0/0/0/1, TenGigE0/0/0/5
3	1	0	TenGigE0/0/0/0, TenGigE0/0/0/4

Unicast Forwarding Ingress Bridge

```
ASR2#show controllers fabric fia
bridge stats location 0/0/cpu0
```

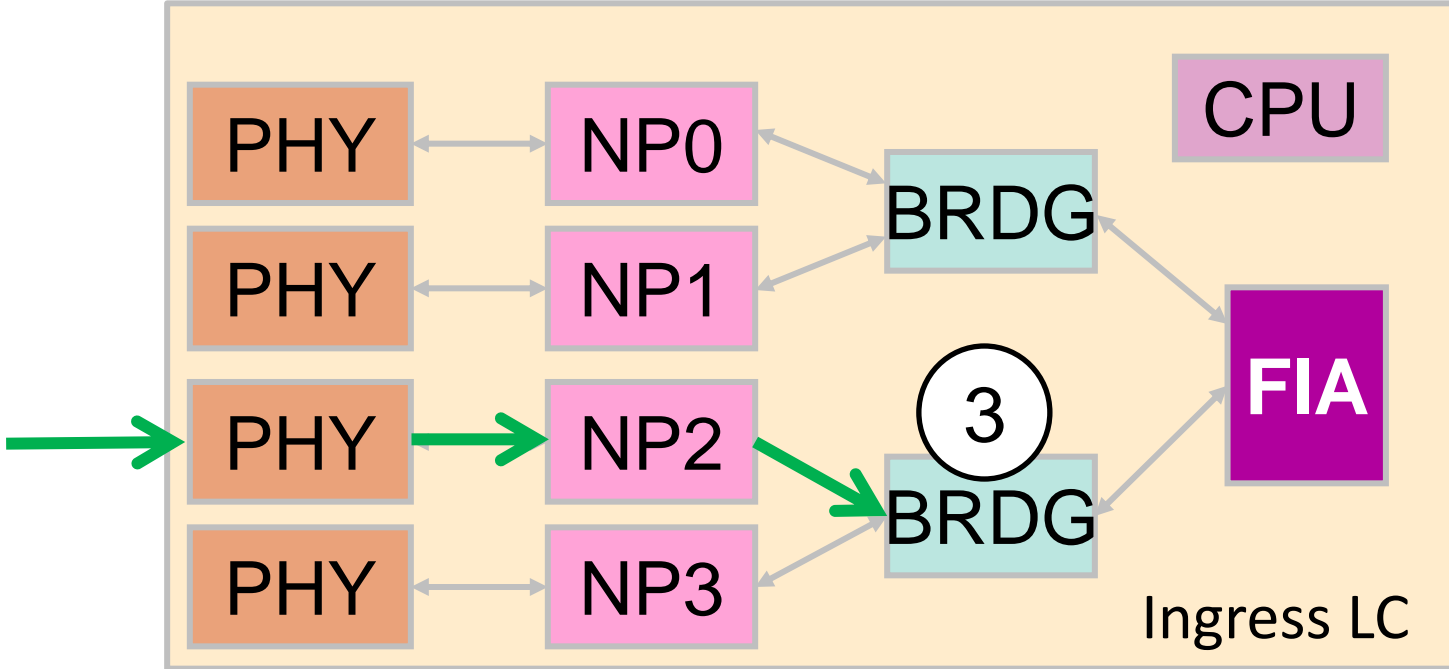
Cast/ Prio	Packet Direction	Packet Count	

Unicast Ingress Stats			

UC HP	NP-2 to Fabric	298	
UC LP	NP-2 to Fabric	10191	
Ingress Drop Stats (MC & UC combined)			

Priority	Packet Direction	Error Drops	Threshold Drops

LP	NP-2 to Fabric	0	0
HP	NP-2 to Fabric	0	0

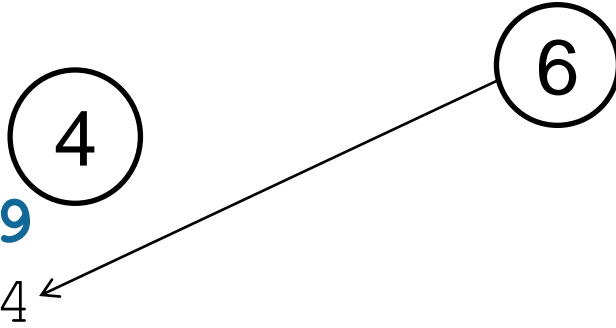


Unicast Forwarding FIA/Fabric

```
ASR2#show controllers fabric
fia stats location 0/0/cpu0
```

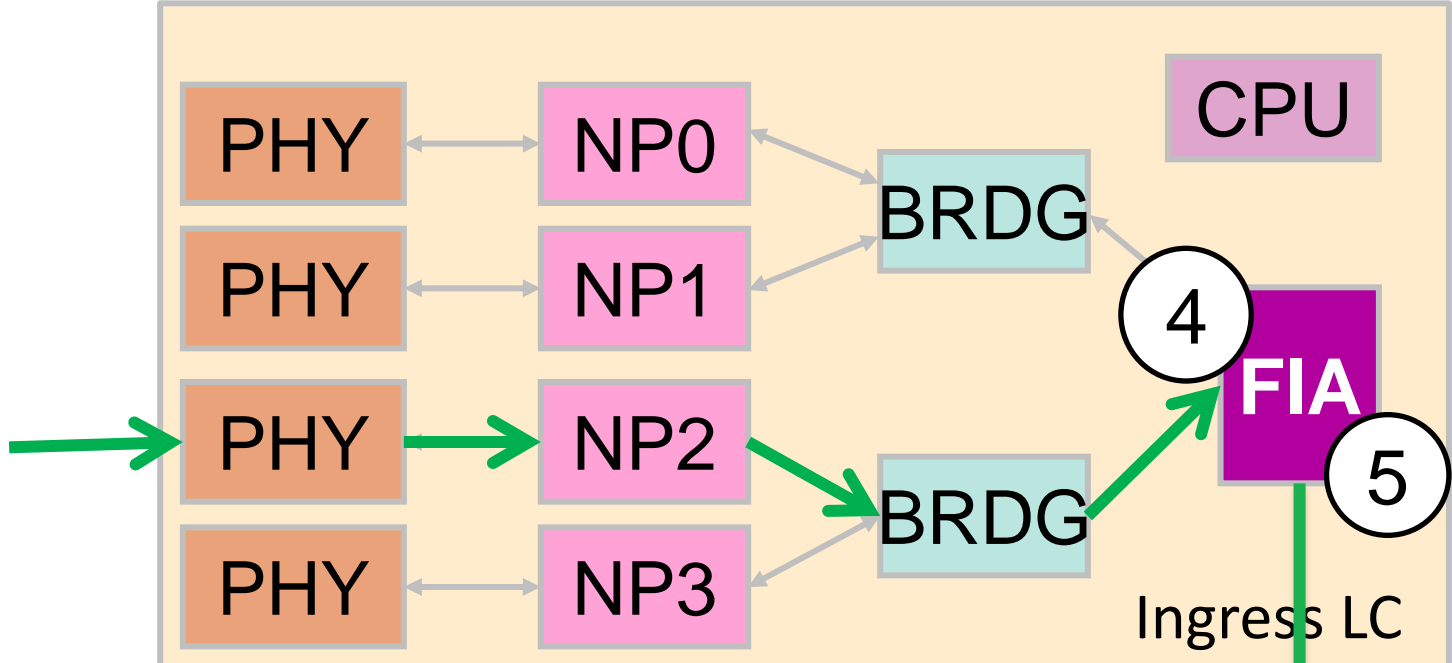
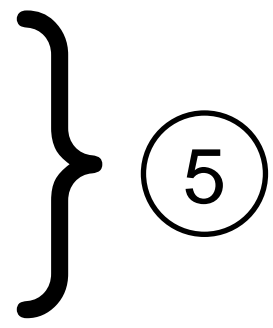
FIA:0 DDR Packet counters:

```
=====
From Bridge#[0] 545
To Bridge #[0] 30
From Bridge#[1] 12349
To Bridge #[1] 11274
```

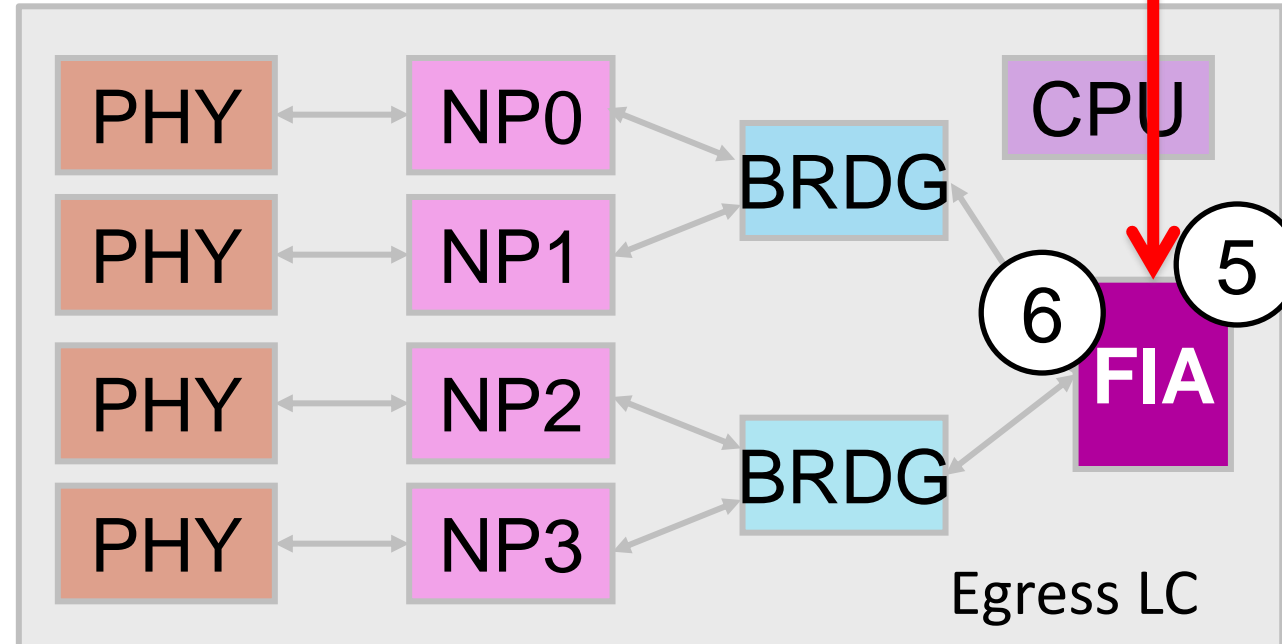
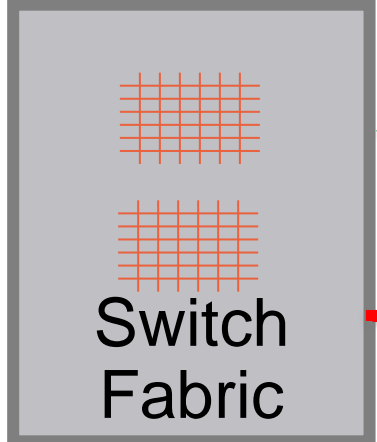


FIA:0 SuperFrame counters:

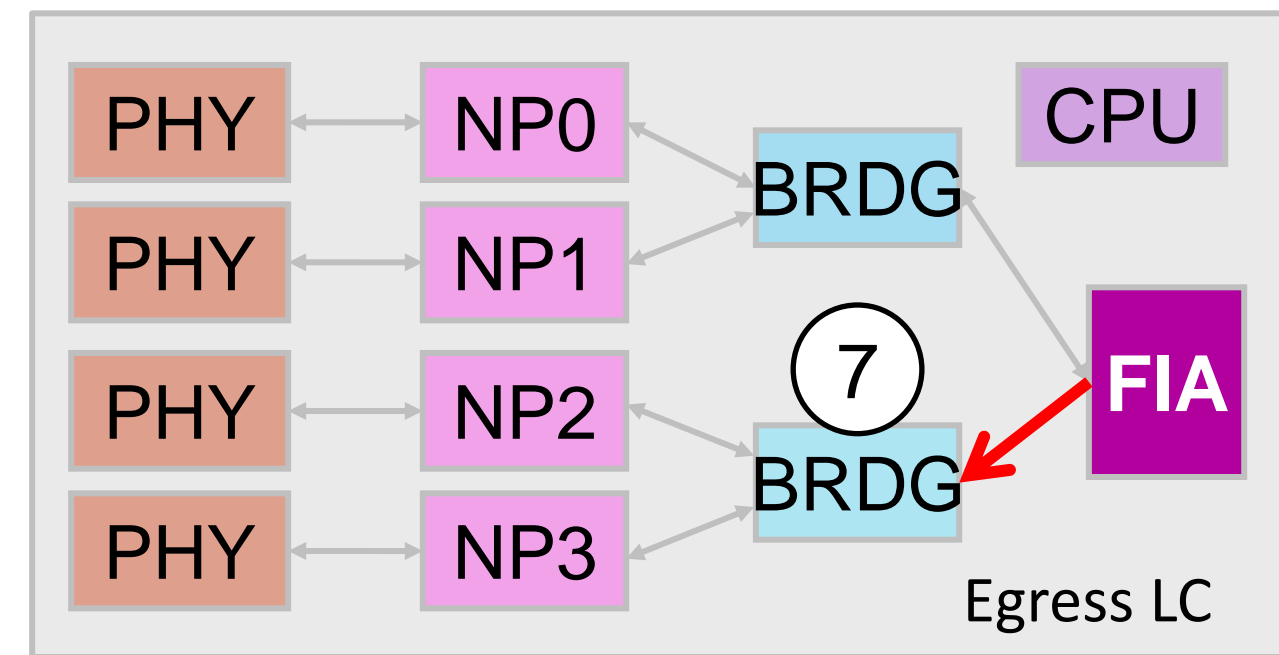
```
=====
To Unicast Xbar[0] 7981
To Unicast Xbar[1] 4912
From Unicast Xbar[0] 6392
From Unicast Xbar[1] 4912
```



Note location should be egress LC to show stats for 6.



Unicast Forwarding Egress Bridge



```
ASR2#ASR2#show controllers fabric fia bridge stats
location 0/0/cpu0
```

Cast/ Prio	Packet Direction	Packet Count	Error Drops	Threshold Drops

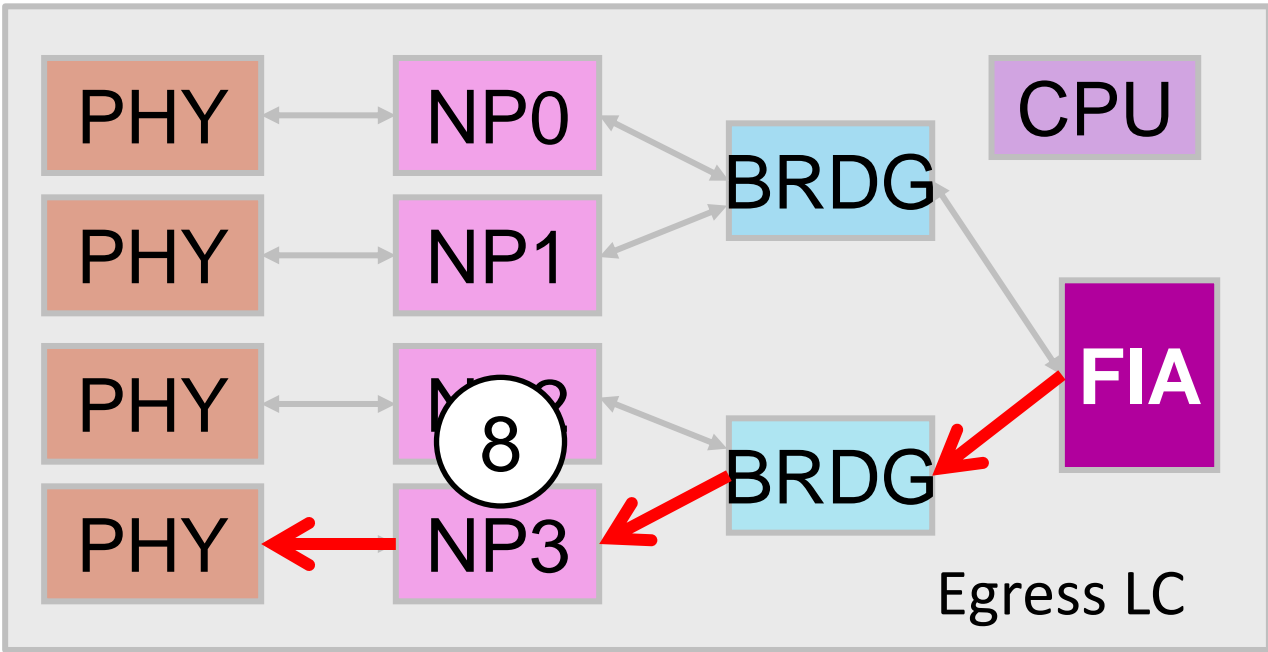
Unicast Egress Stats				

UC HP	Fabric to NP-3	503	0	0
UC LP	Fabric to NP-3	10056	0	0

UC	Total Egress	11179	0	0

Unicast Forwarding

Egress NP

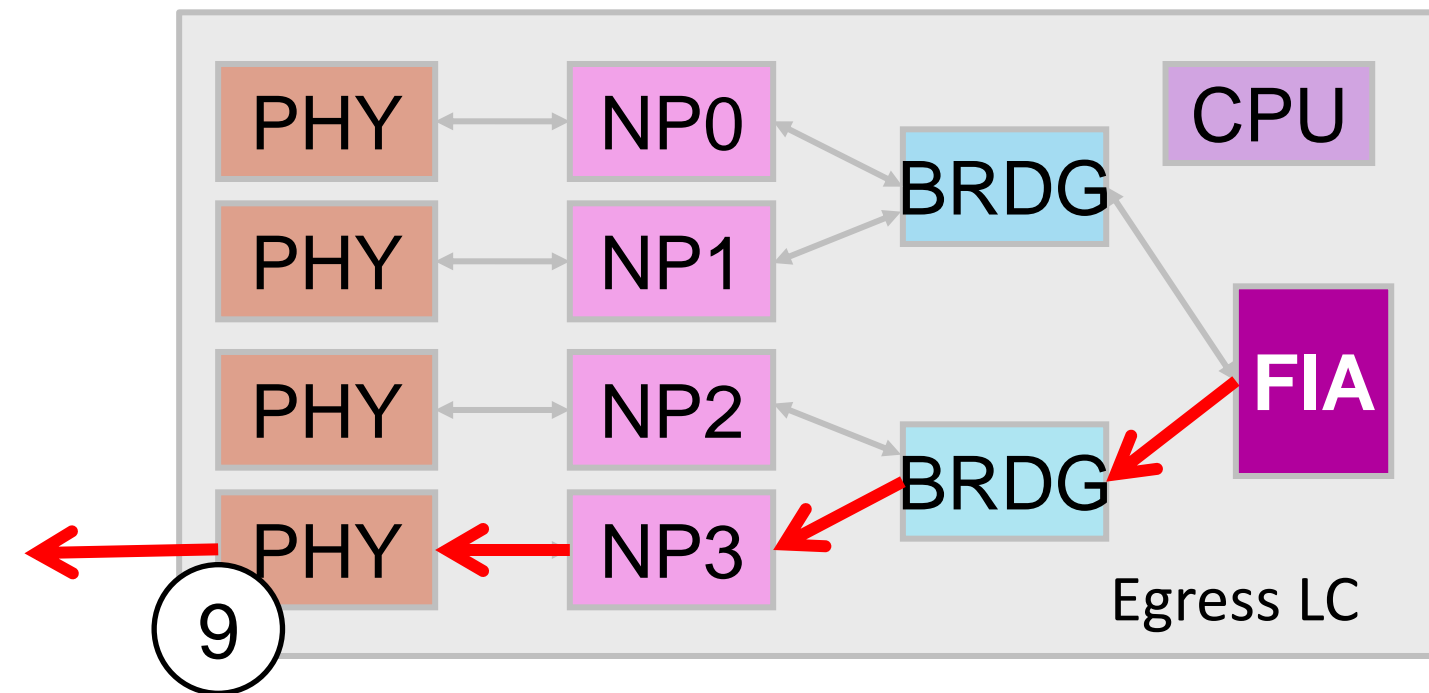


```
ASR2#show controllers np counters np3 location 0/0/cpu0
```

Read 28 non-zero NP counters:

Offset	Counter	FrameValue	Rate (pps)
23	PARSE_FABRIC_RECEIVE_CNT	10266	24
30	MODIFY_ENET_TRANSMIT_CNT	10266	24

Unicast Forwarding Egress PHY



```
ASR2#show controllers tenGigE 0/0/0/0 stats
```

Egress:

Output total bytes	=	5648557
Output good bytes	=	5648557
Output total packets	=	10045
Output pkts 512-1023 bytes	=	10000
Output drop underrun	=	0
Output drop abort	=	0
Output drop other	=	0
Output error other	=	0

Drop Counters – NP, Bridge, FIA

- Can show drops in NP, Bridge, and FIA with one command

```
ASR2#show drops location 0/0/CPU0
```

```
Node: 0/0/CPU0:
```

```
-----  
NP 3 Drops
```

```
-----  
PARSE_INGRESS_DROP_CNT          63  
IN_UIDB_DOWN                      63  
-----
```

```
No Bridge Drops
```

```
-----  
No FIA Drops  
-----
```

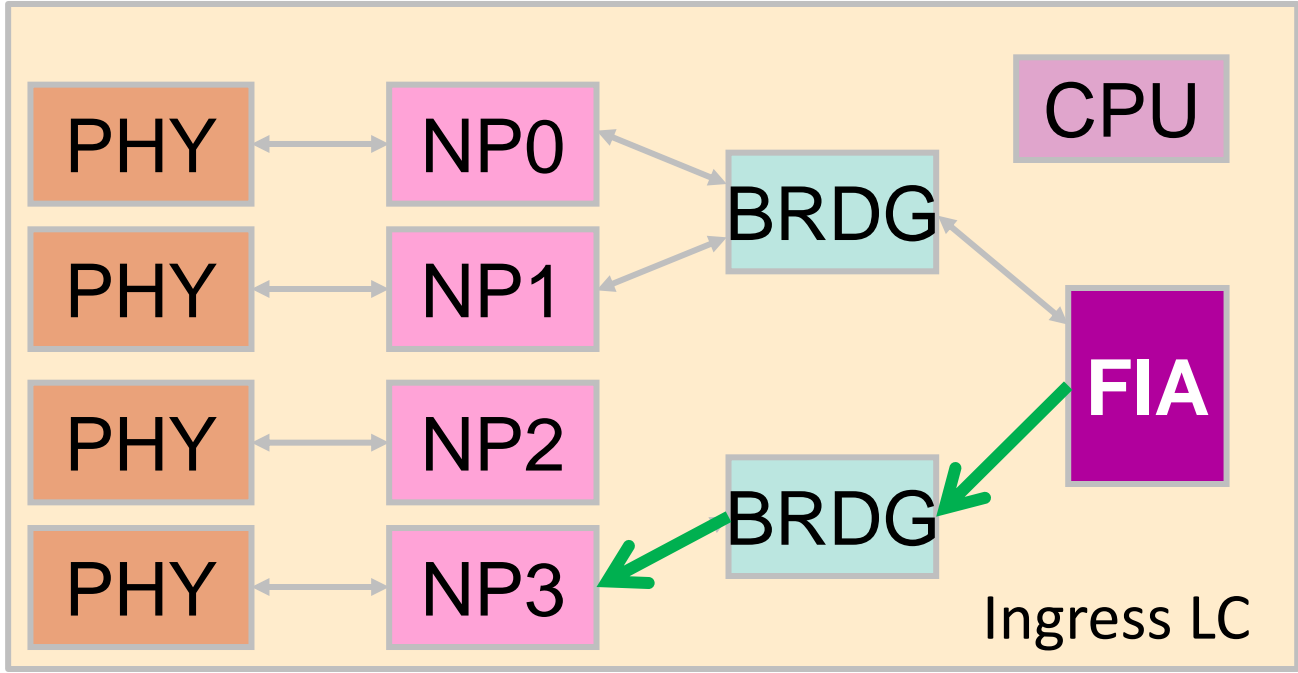
Drop Counters – NP, Bridge, FIA

- Bridge non-blocking architecture, drops are very rare.
- Threshold drops indicates backpressure downstream

```
ASR2#show controllers fabric fia bridge flow-control location 0/0/CPU0
```

```
Ingress BackPressure Counters
*****
FIA to bridge-0          0
FIA to bridge-1          0

All      Bridge to NP-0    0
All      Bridge to NP-1    0
```



Quick Tip – But I bought CRS??

- CRS architecture is a lot simpler because there is only one L2/L3 lookup engine per line card (per direction)

RP/0/RP1/CPU0:FF2#**show controllers** ?

egressq

Show Egress Queue Manager information

fabricq

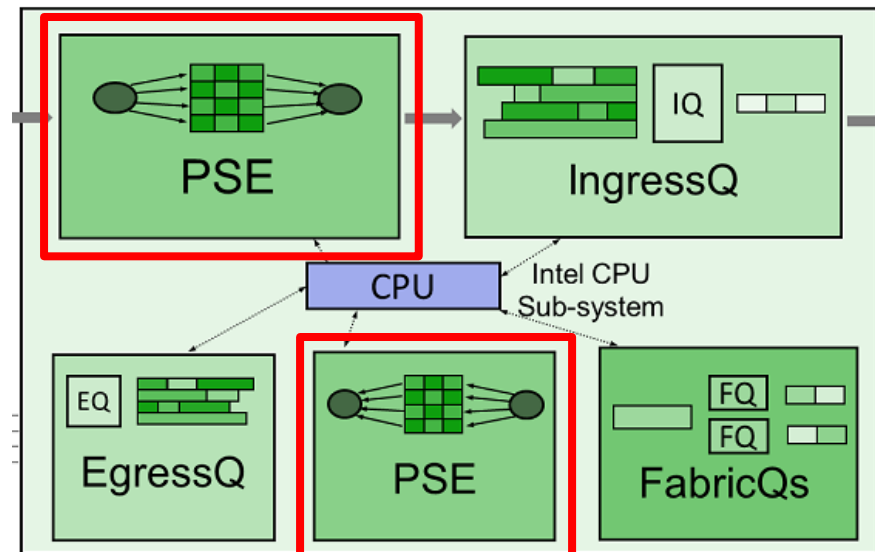
Fabric queue ASIC show screens.

ingressq

Show commands for the ingressq queueing ASIC.

pse

Packet Switching Engine information



Example – Unicast Forwarding (MPLSVPN)

- ASR9K is deployed as MPLS VPN PE using IRB(BVI)
- From remote pinging connected works
- Can't ping to routed addresses



Example – Unicast Forwarding (MPLSVPN)

- PE-CE routing looks ok

```
RP/0/RSP0/CPU0:ASR2#show route vrf CiscoLIVE
```

```
S    10.0.0.1/32 [1/0] via 192.168.123.2, 00:08:01
```

```
B    172.16.123.1/32 [200/0] via 3.3.3.3 (nexthop in vrf default), 00:10:37
```

```
C    192.168.123.0/30 is directly connected, 00:14:14, BVI123
```

```
L    192.168.123.1/32 is directly connected, 00:14:14, BVI123
```

- Remote ping across MPLS cloud to routed address is failing

```
RP/0/RP0/CPU0:HQ1#ping vrf CiscoLIVE 192.168.123.2
```

```
!!!!
```

```
RP/0/RP0/CPU0:HQ1#ping vrf CiscoLIVE 10.0.0.1
```

```
.....
```

Example – Unicast Forwarding (MPLSVPN)

- Let's clear drop counters on the ASR9K

```
RP/0/RSP0/CPU0:ASR2#clear controller np counters all
```

- Then send some ping traffic to help us spot anything abnormal in the counters

```
RP/0/RP0/CPU0:HQ1#ping vrf CiscoLIVE 10.0.0.1 count 1000 time 0
```

```
Type escape sequence to abort.
```

```
Sending 1000, 100-byte ICMP Echos to 10.0.0.1, timeout is 0 seconds:
```

```
.....
```

```
<SNIP>
```

Example – Unicast Forwarding (MPLSVPN)

- The drops shown below are from LC0/6 NP1, this just happens to be the NP servicing the core facing interfaces where the MPLS packet ingresses

```
RP/0/RSP0/CPU0:ASR2#show drops
```

```
<SNIP>
```

```
Node: 0/6/CPU0:
```

```
NP 1 Drops:
```

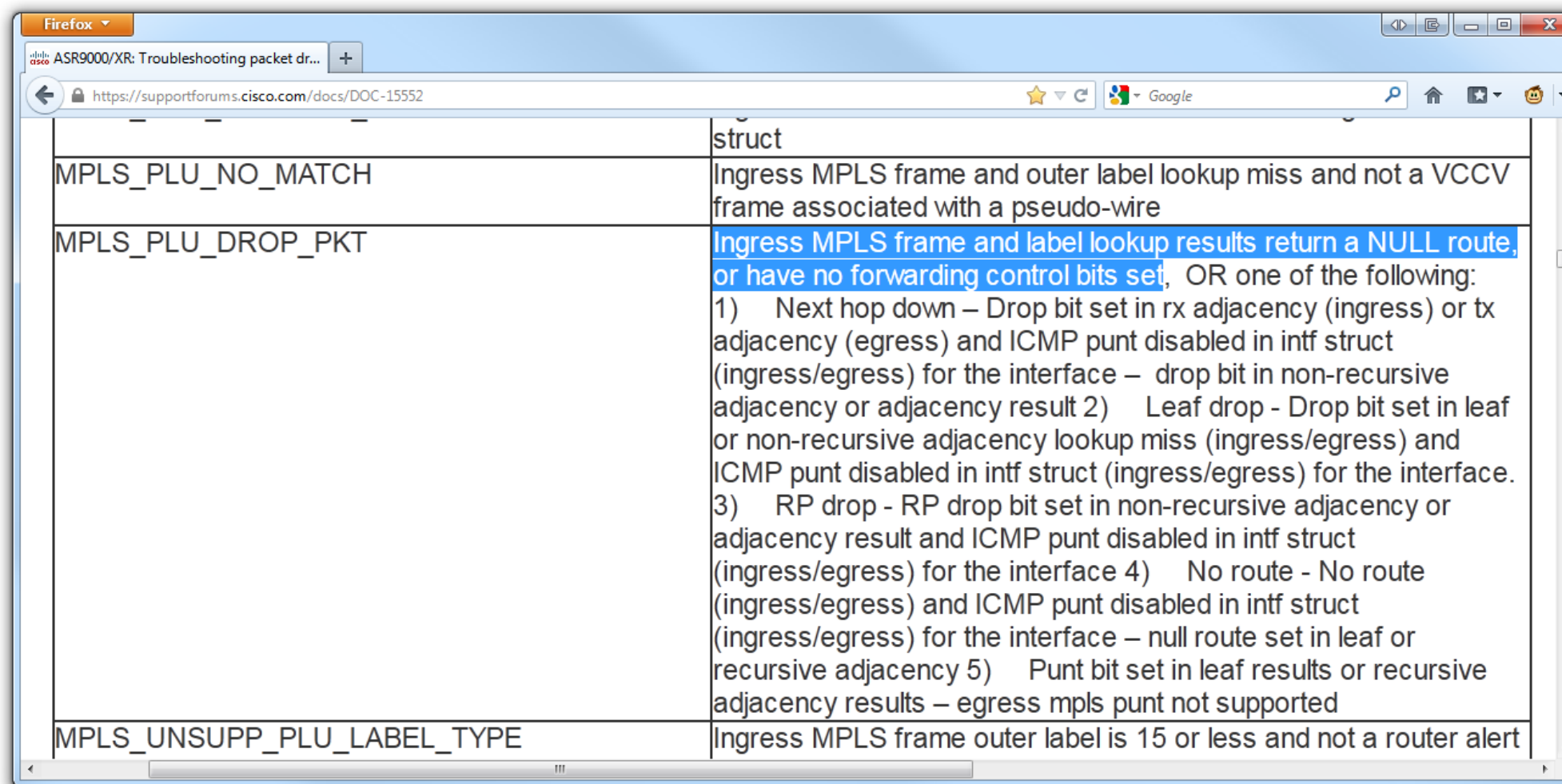
```
-----  
RESOLVE_INGRESS_DROP_CNT                               1000  
MPLS_PLU_DROP_PKT                                   1000  
-----
```

Quick Tip – What does it all mean?

- Cisco Support Forums reference article

ASR9000/XR: Troubleshooting packet drops and understanding NP drop counters

<https://supportforums.cisco.com/docs/DOC-15552>



	struct
MPLS_PLU_NO_MATCH	Ingress MPLS frame and outer label lookup miss and not a VCCV frame associated with a pseudo-wire
MPLS_PLU_DROP_PKT	Ingress MPLS frame and label lookup results return a NULL route, or have no forwarding control bits set, OR one of the following: 1) Next hop down – Drop bit set in rx adjacency (ingress) or tx adjacency (egress) and ICMP punt disabled in intf struct (ingress/egress) for the interface – drop bit in non-recursive adjacency or adjacency result 2) Leaf drop - Drop bit set in leaf or non-recursive adjacency lookup miss (ingress/egress) and ICMP punt disabled in intf struct (ingress/egress) for the interface. 3) RP drop - RP drop bit set in non-recursive adjacency or adjacency result and ICMP punt disabled in intf struct (ingress/egress) for the interface 4) No route - No route (ingress/egress) and ICMP punt disabled in intf struct (ingress/egress) for the interface – null route set in leaf or recursive adjacency 5) Punt bit set in leaf results or recursive adjacency results – egress mpls punt not supported
MPLS_UNSUPP_PLU_LABEL_TYPE	Ingress MPLS frame outer label is 15 or less and not a router alert

Example – Unicast Forwarding (MPLSVPN)

- Looking up documentation reveals a feature restriction, we must use per-VRF label allocation which isn't the default behaviour for BGP

Additional IPv4-Specific Environments Supported for IRB

- Configuration of up to a maximum of 2000 BVIs.
- Up to a maximum of 128k IPv4 adjacencies.
- Layer 3 IP multicast, with ability to take ingress IP multicast traffic and bridge it to multiple Layer 2 subinterfaces (Ethernet flow points) on a bridge domain that are part of multicast groups.
- **VRFs for IPv4 (Per-VPN label VRFs only—not per prefix).**

http://www.cisco.com/en/US/docs/routers/asr9000/software/asr9k_r4.1/interfaces/configuration/guide/hc41irb.html#wp1030591

Example – Unicast Forwarding (MPLSVPN)

- Configuration change resolves the issue!

```
RP/0/RSP0/CPU0:ASR2#configure
```

```
RP/0/RSP0/CPU0:ASR2(config)#router bgp 123
```

```
RP/0/RSP0/CPU0:ASR2(config-bgp)#vrf CiscoLIVE
```

```
RP/0/RSP0/CPU0:ASR2(config-bgp-vrf)#label-allocation-mode per-vrf
```

```
RP/0/RSP0/CPU0:ASR2(config-bgp-vrf)#commit
```

```
RP/0/RSP0/CPU0:ASR2(config-bgp-vrf)#end
```

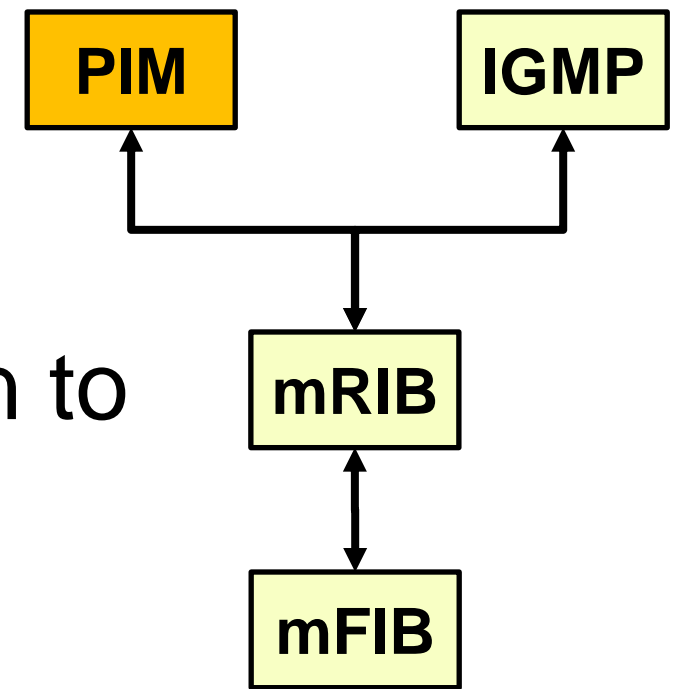
```
RP/0/RSP0/CPU0:ASR2#clear bgp vpnv4 unicast * soft
```

```
RP/0/RP0/CPU0:HQ1#ping vrf CiscoLIVE 10.0.0.1
```

!!!!

Multicast Infrastructure

- PIM and IGMP send their mroute, olist information to mRIB



```
RP/0/RSP0/CPU0:ASR2#show pim topology 10.1.1.1 224.0.1.39 detail
```

```
(10.1.1.1,224.0.1.39) SPT DM Up: 00:45:24
```

```
JP: Null(never) RPF: GigabitEthernet0/2/0/1,123.123.123.2 Flags: KAT(00:00:15)
```

```
Up: MT clr (00:00:00) MDT: JoinSend N, Cache N/N, Misc (0x0,0/0)
```

```
Cache: Add 00:00:00, Rem 00:00:00. MT Cnt: Set 0, Unset 0. Joins sent 0
```

```
MDT-ifh 0x0/0x0 MT Slot none/ none
```

RPF Table: IPv4-Unicast-default

TenGigE0/0/0/0	00:36:38	fwd
TenGigE0/0/0/3	00:36:38	fwd
TenGigE0/0/0/5	00:36:38	fwd
GigabitEthernet0/2/0/1	00:45:24	fwd

Quick Tip – Huh? xGID

- FGID is a bit mask programmed into fabric hardware. It indicates which slots should receive a multicast flow

FGID: 0x5
 / \
... 0101 = slot 2 and 0

- MGID performs the same function for components within the linecard (ASR9K only)

MGID: 0x4205
 = decimal 16901

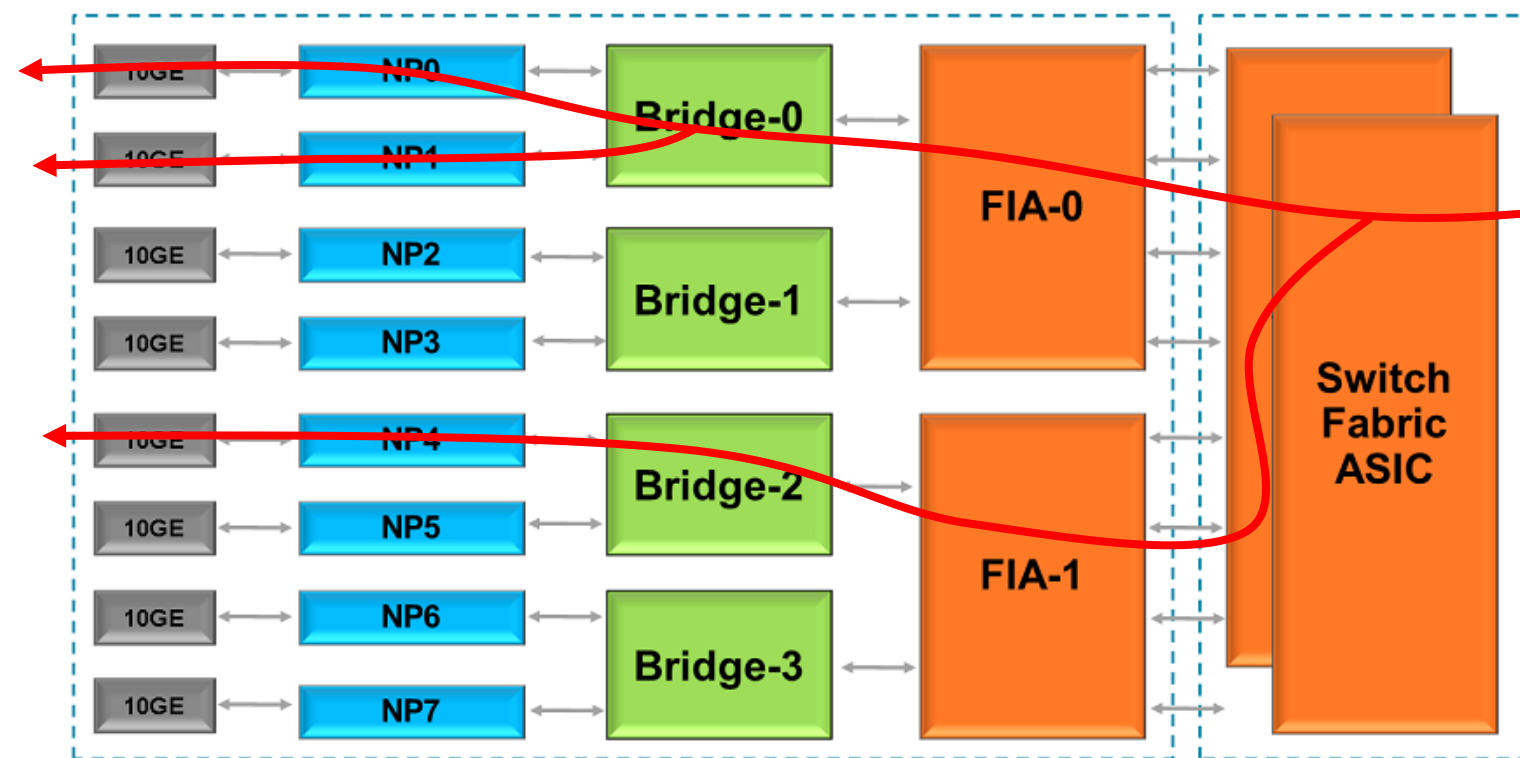


Quick Tip – Huh? xGID

- MGID mappings can be determined by querying the decimal value on the linecard CPU

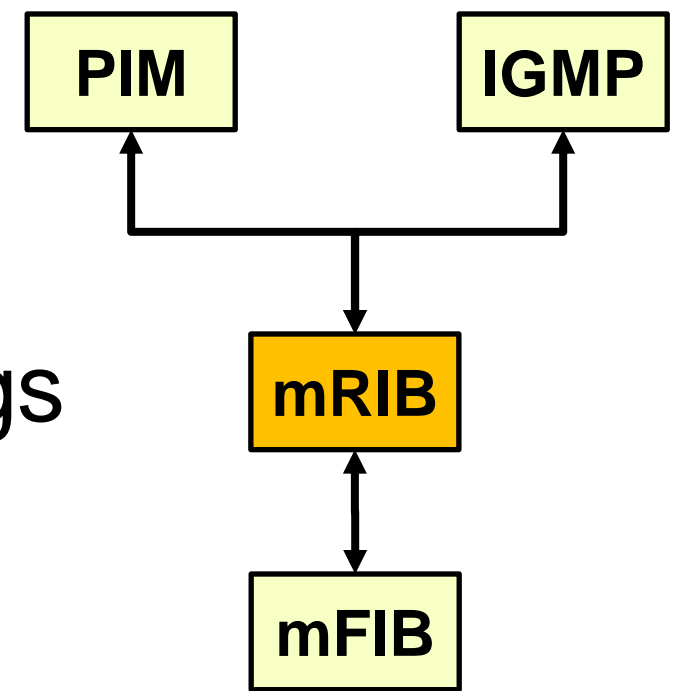
```
ASR2#show controllers mgidprgm mgidindex 16901 location 0/0/CPU0
```

Device	MGID-Bits
=====	=====
FIA-0	1
FIA-1	1
Bridge-0	11
Bridge-1	0
Bridge-2	1
Bridge-3	0



Multicast Infrastructure

- mRIB is the central repository for routes, olist, flags
- Internal MGID/FGID masks are built



```
RP/0/RSP0/CPU0:ASR2#show mrib route 10.1.1.1 224.0.1.39 detail
```

```
(10.1.1.1,224.0.1.39) Ver: 0x5e4b RPF nbr: 123.123.123.2 Flags:, FMA: 0x501c26  
FGID: 0x5 MGID: 0x4205
```

```
Up: 00:09:53
```

Incoming Interface List

```
GigabitEthernet0/2/0/1 Flags: F A, Up: 00:09:53
```

Outgoing Interface List

```
TenGigE0/0/0/0 Flags: F, Up: 00:01:07
```

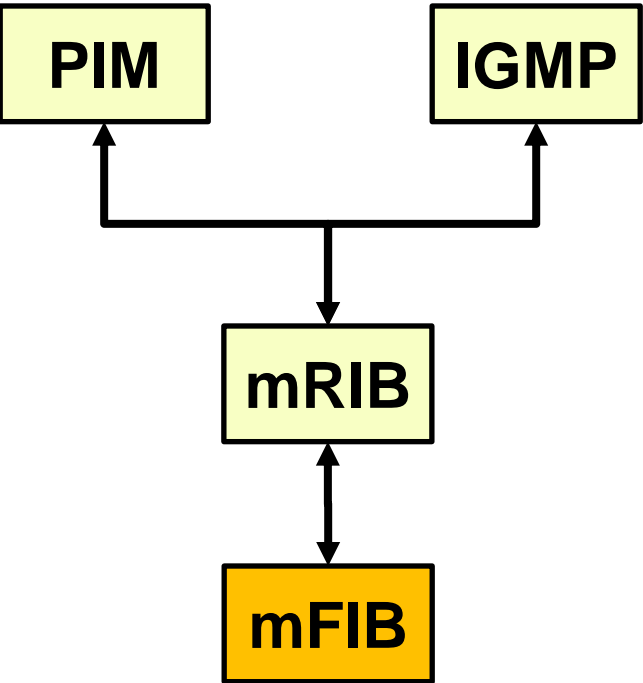
```
TenGigE0/0/0/3 Flags: F, Up: 00:01:07
```

```
TenGigE0/0/0/5 Flags: F, Up: 00:01:07
```

```
GigabitEthernet0/2/0/1 Flags: F A, Up: 00:09:53
```

Multicast Infrastructure

- mFIB interfaces with the mRIB but also performs hardware programming



```

ASR2#show mfib hardware route olist 10.1.1.1 224.0.1.39 location 0/0/CPU0
LC Type: Trident
Source: 10.1.1.1          Group: 224.0.1.39          RPF Int: Gi0/2/0/1
  
```

Route Information

C	IX	IC	IP	ID	IS	IU	IF	IR	IA	IG	ET	EO	ER	EC	BS
0	F	F	F	F	F	F	F	0x40000c0	0x5	0x4205	1	T	10	1	0x3687d
1	F	F	F	F	F	F	F	0x40000c0	0x5	0x4205	1	T	10	1	0x3687d

<SNIP>



Verifying multicast forwarding

- Statistics shown per NP below can be both received on ingress or from bridge/fabric, likewise with forwarded

```
ASR2#show mfib hardware route statistics 10.1.1.1 224.0.1.39 location 0/0/CPU0
```

```
S: 10.1.1.1 G: 224.0.1.39 Pr:64
```

```
-----  
C      R (packets:bytes) / F (packets:bytes) / P (packets) / ID (packets) / ED (packets)  
-----
```

```
0      351:16848 / 351:16848 / 0 / 351 / 0
```

```
1      352:16896 / 352:16896 / 0 / 352 / 0
```

```
2      0:0 / 0:0 / 0 / 0 / 0
```

```
3      0:0 / 0:0 / 0 / 0 / 0
```

```
4      354:16992 / 354:16992 / 0 / 354 / 0
```

```
5      0:0 / 0:0 / 0 / 0 / 0
```

```
6      0:0 / 0:0 / 0 / 0 / 0
```

```
7      0:0 / 0:0 / 0 / 0 / 0
```

```
<continued on next slide>
```

————— ID = Ingress Drops!!

Verifying multicast forwarding

- Statistics shown per interface only display traffic leaving the router

<continued from previous slide>

Interface Statistics:

C	Interface	F/P/D (packets:bytes)
0	Te0/0/0/5	351:16848 / 0:0 / 0:0
1	Te0/0/0/3	352:16896 / 0:0 / 0:0
4	Te0/0/0/0	354:16992 / 0:0 / 0:0

ASR2#show mfib hardware route statistics 10.1.1.1 224.0.1.39 location 0/2/CPU0

S: 10.1.1.1 G: 224.0.1.39 Pr:64

<SNIP>

3 4016:192768 / 4017:192816 / 19 / 0 / 0

<SNIP>

3 Gi0/2/0/1 0:0 / 0:0 / 0:0

Verifying multicast forwarding

- Checking the relevant NP counters shows packets are dropping

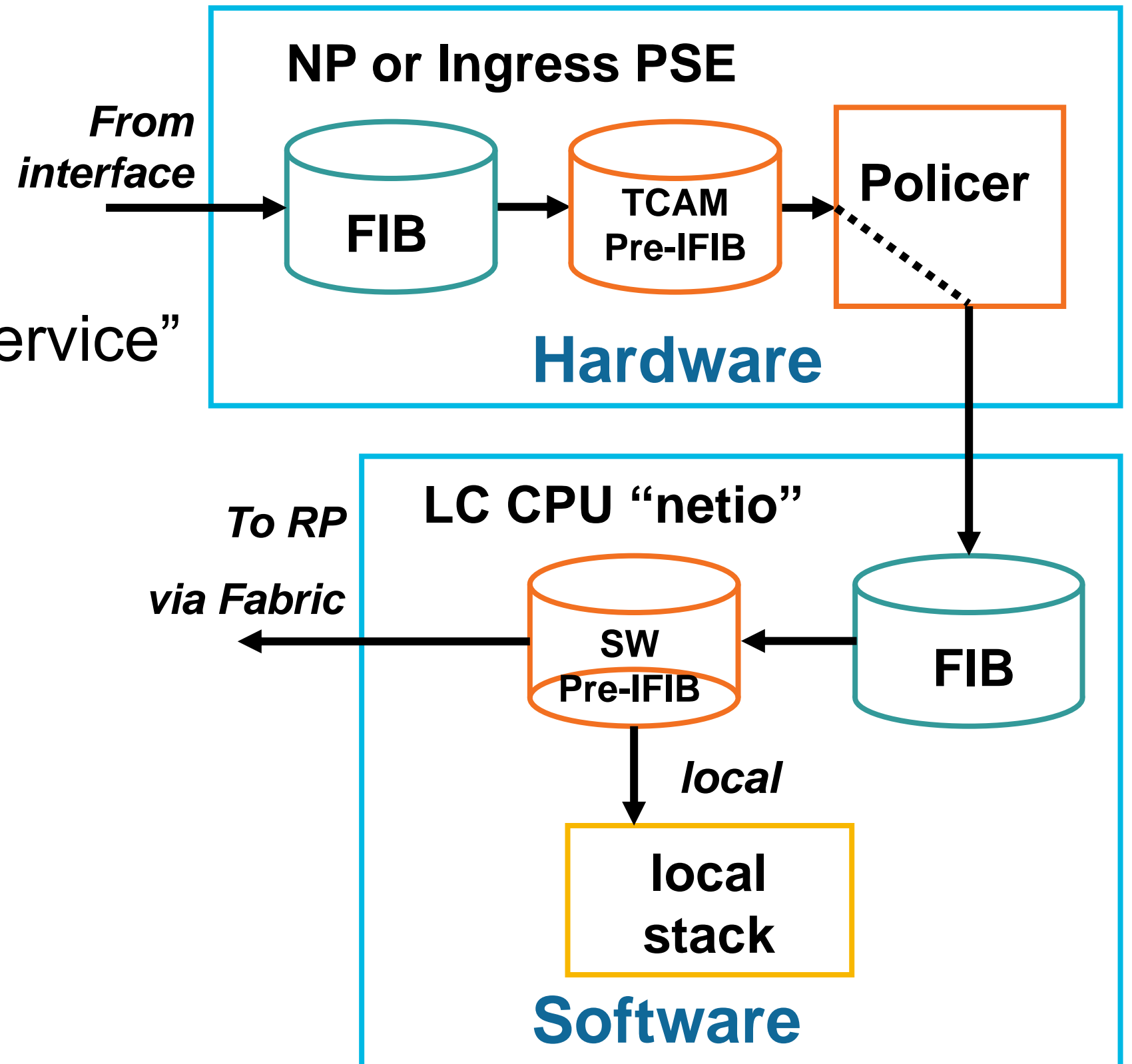
```
ASR2#show drops np np0 location 0/0/CPU0
 36  MODIFY_INGRESS_DROP_CNT          351          1
 47  MODIFY_RPF_FAIL_DROP_CNT       351         1
```

- These interfaces were looped, causing the egressing packet to forward back onto the NP and fail RPF (expected)

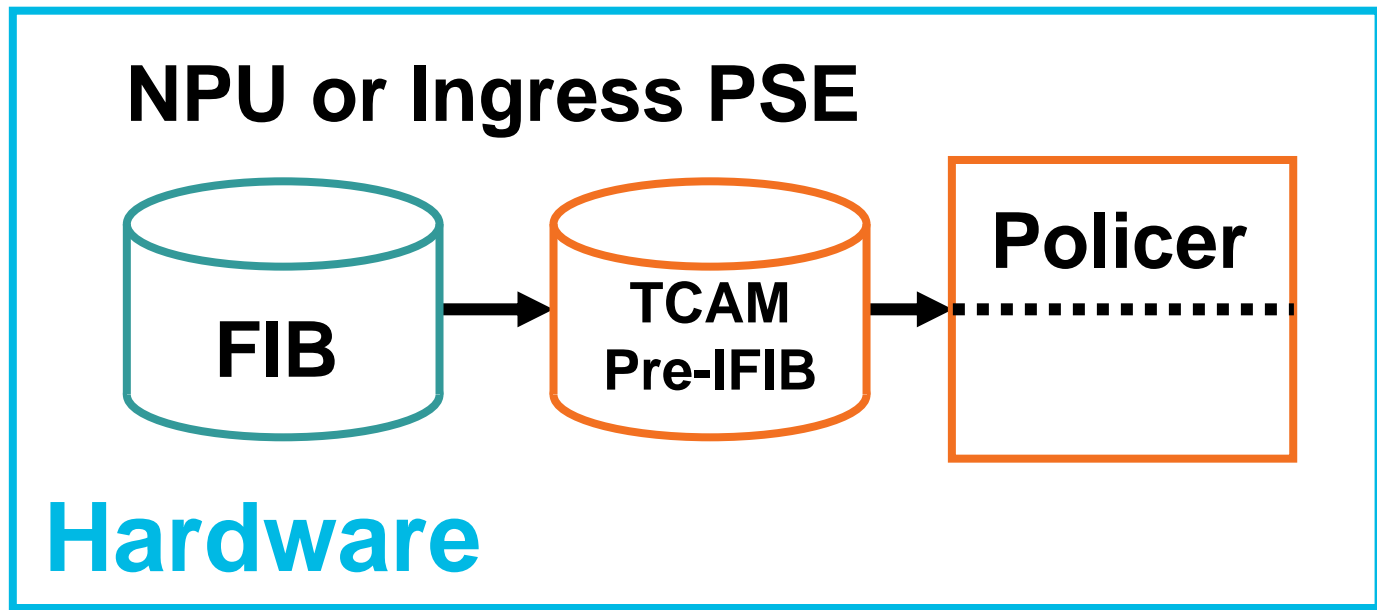
```
ASR2#show run interface TenGigE 0/0/0/5
interface TenGigE0/0/0/5
  ipv4 address 10.5.0.1 255.255.255.0
  loopback internal
```

LPTS

- “Local Packet Transport Service”
- Punt
- Policing & Discards
- Reassembly



LPTS – TCAM Pre-IFIB



```
ASR2#show lpts pifib hardware
entry statistics location 0/0/CPU0
```

```
Node: 0/0/CPU0:
```

```
-----
L3 - L3 Protocol;L4 - Layer4 Protocol; Intf - Interface;
Dest - Destination Node;
na - Not Applicable or Not Available
```

L3	VRD id	L4	Intf	Dest	Pkts/Drops	laddr,Port	raddr,Port
IPV4	*	OSPF	BE56	48	456/0	224.0.0.5,any	any,any
IPV4	*	ICMP	any	Local	0/0	any,any	any,UNREACH
IPV4	*	ICMP	any	Local	0/0	any,any	any,TIMXCEED
IPV4	*	ICMP	any	Local	100/0	any,any	any,ECHO

LPTS – Hardware Police

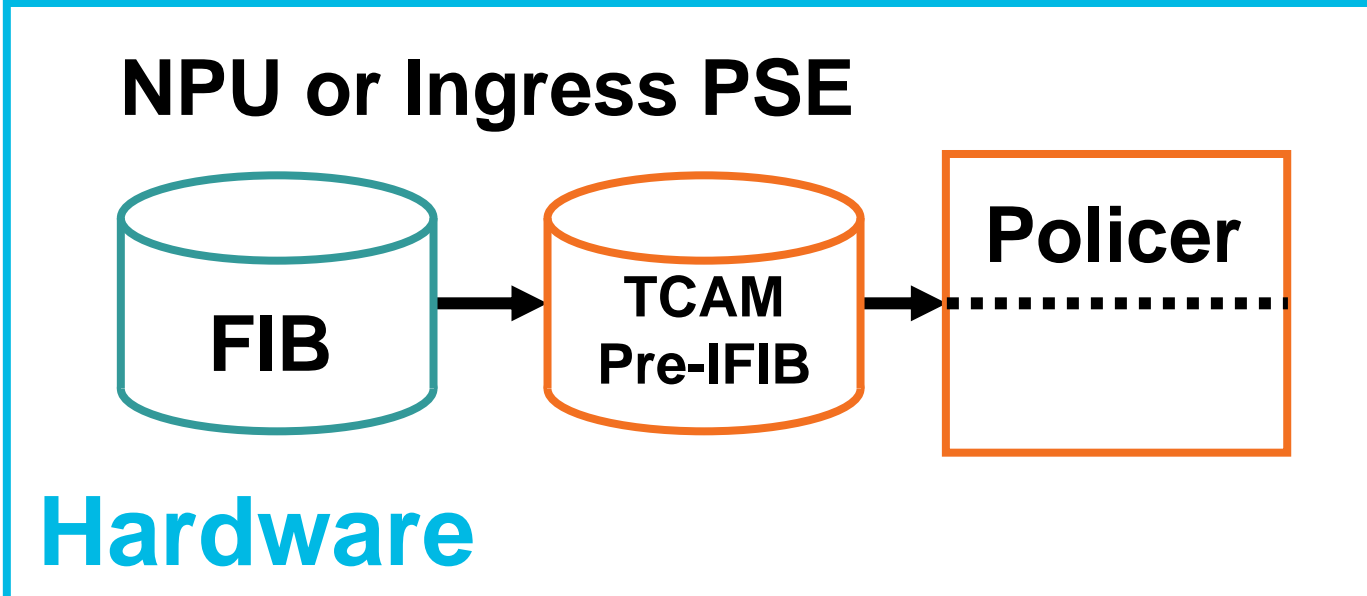
```
ASR2#show lpts pifib hardware
police location 0/0/CPU0
```

Wed Feb 8 22:59:41.284 AEST

Node 0/0/CPU0:

Burst = 100ms for all flow types

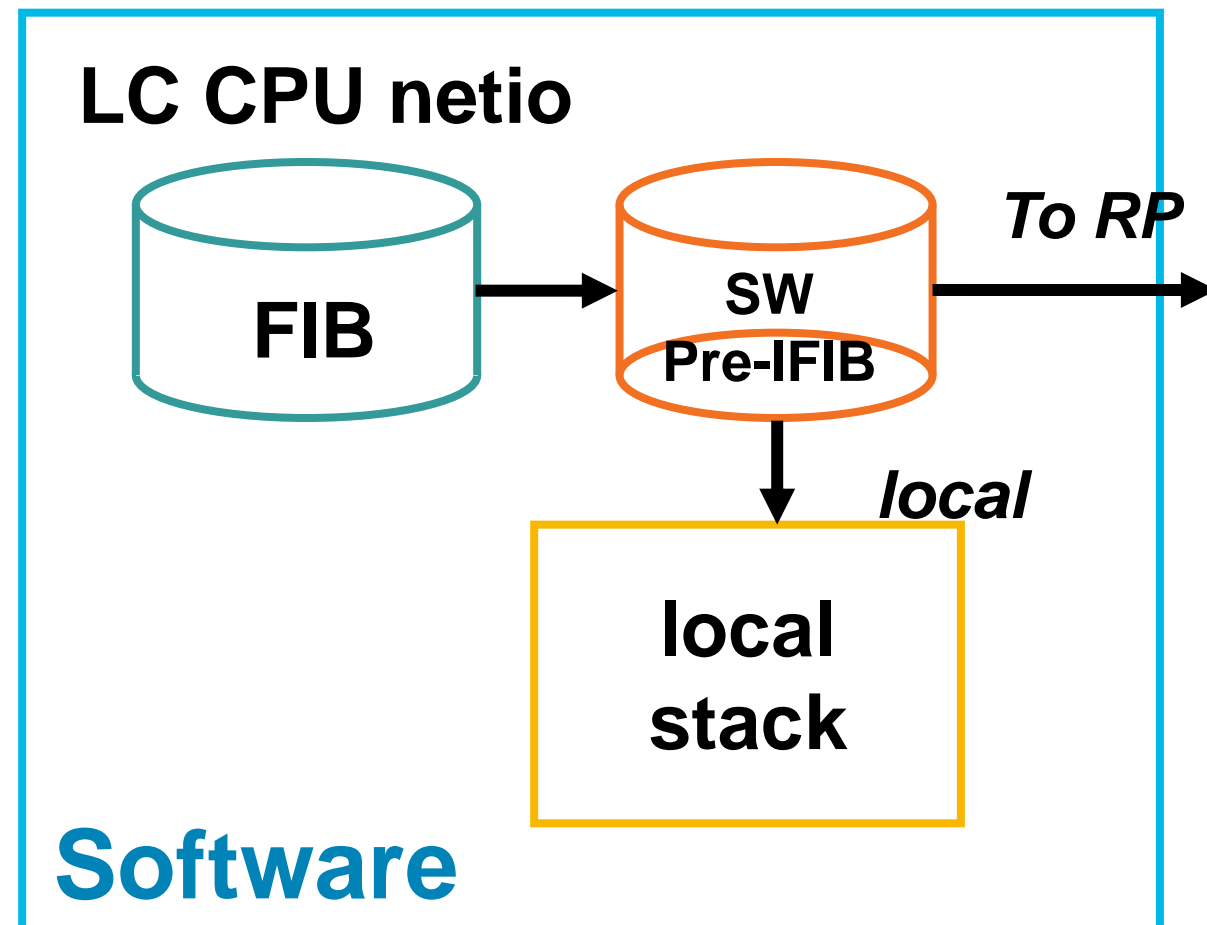
FlowType	Policer	Type	Cur. Rate	Accepted	Dropped
ICMP-local	112	Static	1500	100	0
ICMP-app	152	Static	1500	0	0
ICMP-control	140	Static	1000	3	0
ICMP-default	153	Static	1500	0	0



LPTS – Software Pre-IFIB

```
ASR2#show lpts pifib entry
      location 0/0/CPU0
```

```
-----
L3 Protocol      : IPV4
L4 Protocol      : ICMP
VRF-ID           : * (0x00000001)
Destination IP   : any
Source IP        : any
Port/Type        : ICMP:ECHO
Source Port      : any
Is Fragment      : 0
Is SYN           : 0
Interface        : any (0x0)
O/F/L/I/T       : DROP/ICMP-local/IPv4_STACK/1/0
Deliver List     : na
Accepts/Drops : 100/0
Is Stale         : 0
```



Q & A



Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2013 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 8 March 12:00pm-2:00pm



Cisco *live!* 365

Don't forget to activate your Cisco Live 365 account for access to all session material,

communities, and on-demand and live activities throughout the year. Log into your Cisco Live portal and click the "Enter Cisco Live 365" button.

www.ciscoliveaustralia.com/portal/login.wv

