Cisco Networkers 2011

November 7, 2011 - Hong Kong

Cisco ASR9000 Router Systems Architecture

Lim Fung, CTG Technical Marketing
limfung@cisco.com

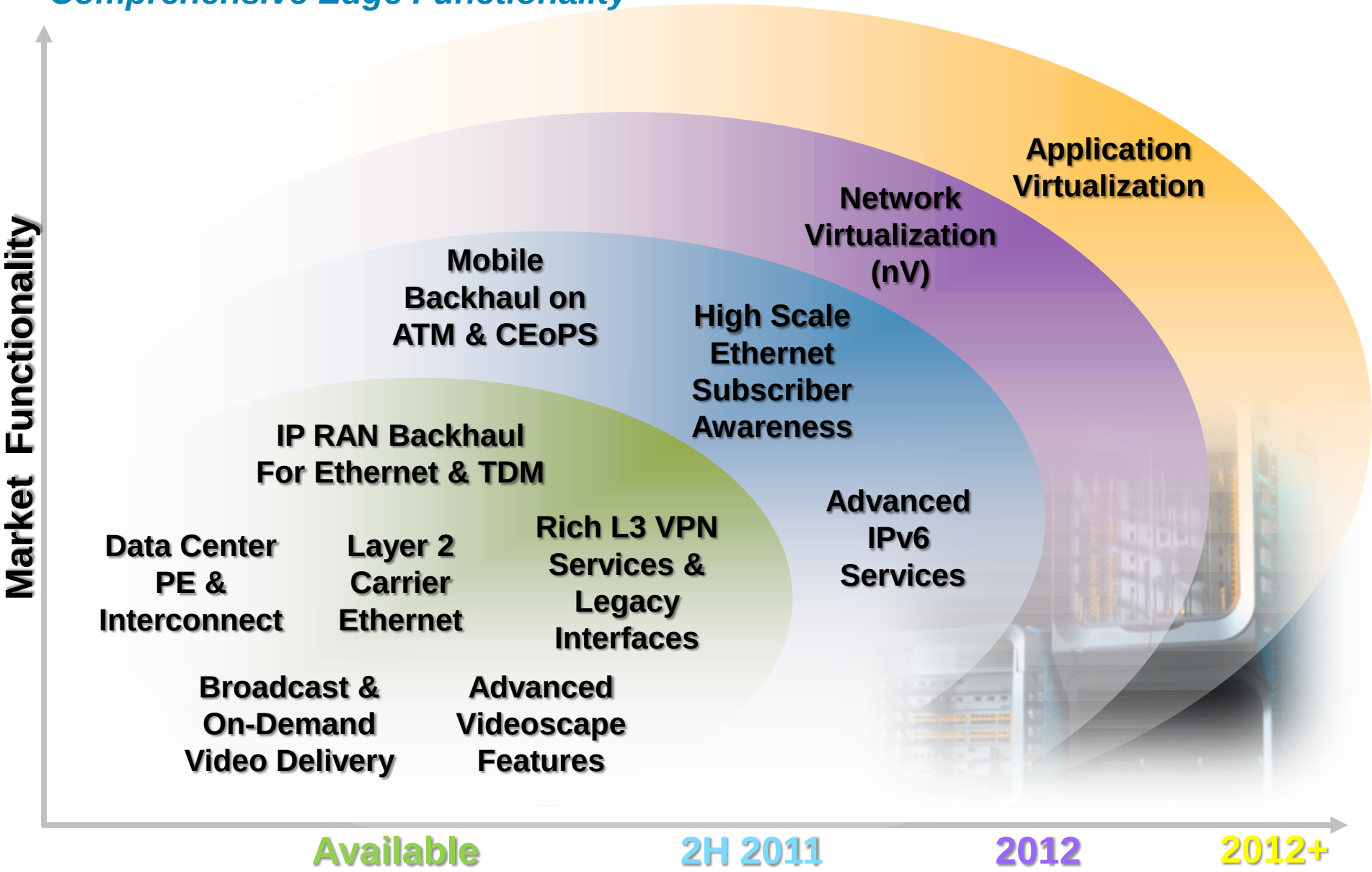
Service Edge Foundation: ASR 9000 System

- Optimized for Aggregation of **Dense 10GE & 100GE**
- Designed for Longevity & TCO: **Scalable up to 400 Gbps of Bandwidth per Slot**
- Based on IOS-XR & Cisco PRIME for **Nonstop Availability & Manageability**
- Enables **Network Convergence** of Business & Residential Services for both Fixed & Mobile Networks
- Advanced **Videoscape & Mobility DNA**
- Industry Leading **Operational Savings & Management** with Cisco nV Technology

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Cisco ASR 9000 Functional Evolution

Comprehensive Edge Functionality



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Cisco ASR 9000 Portfolio Evolution

Uncompromising Density Across a Scalable Portfolio

2012					
# of Slots	22 Slot	10 Slot	6 slots	Pizza Box	9000v Satellite
Max. Linecards per Chassis	20 LC	8 LC	4 LC	2 IO Slots	Fixed Ports
Max. Linecard Bandwidth	360 Gbps	360 Gbps	360 Gbps	Modular + 4x10GE	44xGE + 4x10GE
Max. Slot Bandwidth	1.2 Tbps	440 Gbps	440 Gbps		
Chassis Bandwidth	48 Tbps	7.0 Tbps	3.5 Tbps	120 Gbs	80 Gbs

ASR 9922

ASR 9010

ASR 9006

ASR 9000v

ASR 9001

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Hardware Architecture

Chassis, RSP, Switch Fabric and Linecards

and/or its affiliates. All r eserved.
Cisco Confidential

ASR 9000 System Architecture

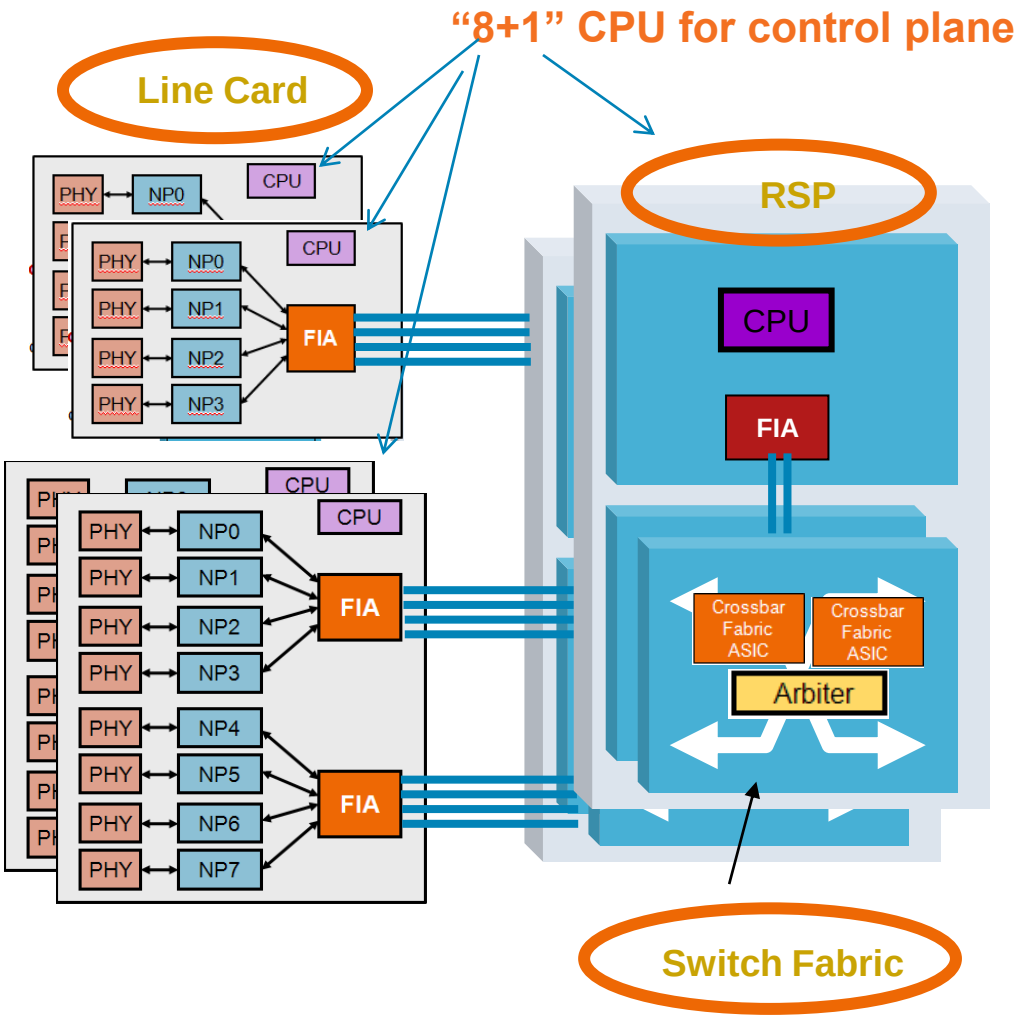
"At-a-Glance"

Fully Distributed Architecture for High Performance and High Multi-dimensional Control Plane Scale

- Data forwarding is fully distributed on the line cards
- Control plane split among RSP and LC CPU (same type of CPU as RSP)
- L2 protocols, BFD, CFM, Netflow runs on the LC CPU for high scale

IOS-XR: True Modular OS for HA and Operational Simplicity

- Micro-kernel based, true modular OS
- High availability and System stability
- SW patch granularity for operational simplicity



Active-Active Intelligent Switch Fabric

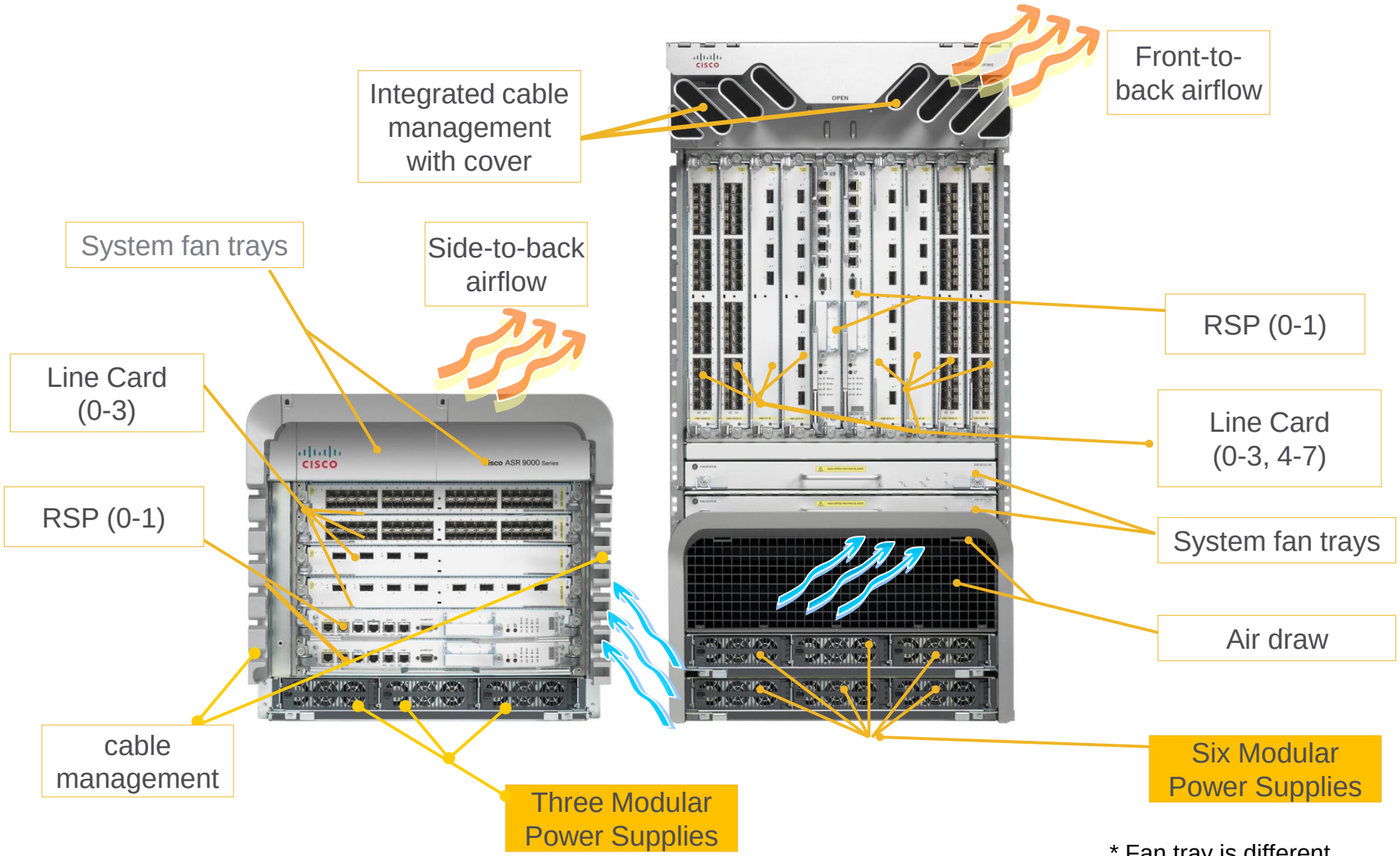
Guarantee "0" packet loss during RSP failover

Advanced internal system QoS, Intelligent multicast replication

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

ASR 9010 and ASR 9006 Chassis

Identical HW components across two chassis*



* Fan tray is different

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Power and Cooling

Existing Power Supply and Fan are ready for 400G/slot

ASR 9010 Fan Tray

ASR 9006 Fan Tray

- Fans unique to chassis
- Variable speed for ambient temperature variation
- Redundant fan-tray
- Low noise, NEBS and OSHA compliant

Power Supply

DC Supplies

A
B
2.1/1.5 kW

A
B
2.1 kW

AC Supplies

A
3 kW

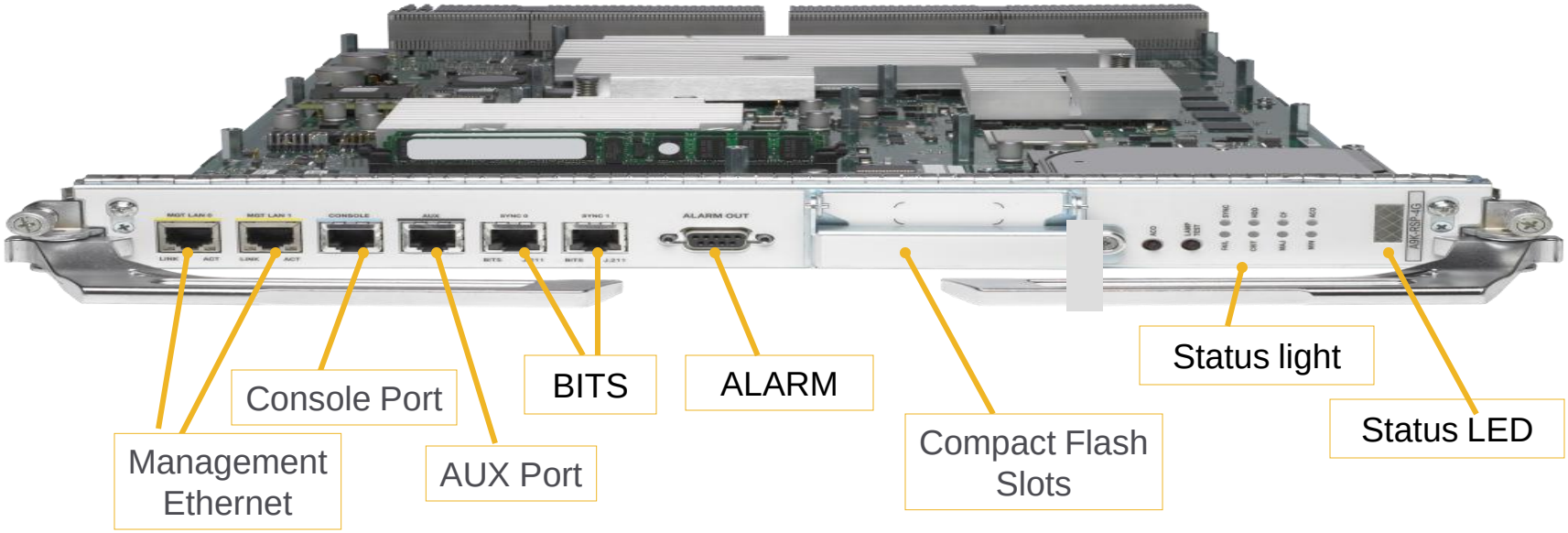
B
3 kW

- 6 & 10 slot use same power supplies
- Single power zone
- All power supplies run in active mode
- Power draw shared evenly
- 50 Amp DC Input or 16 Amp AC for Easy CO Install

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

RSP Engine

- Performs control plane and management functions
- Dual Core CPU processor with 4GB or 8GB DRAM
  (4GB→8GB is not field upgradable)
- 2MB NVRAM, 4GB internal bootdisk, 2 external compact flash slots
- Dual Out-of-band 10/100/1000 management interface
- Console & auxiliary serial ports
- Hard Drive: 70G HDD

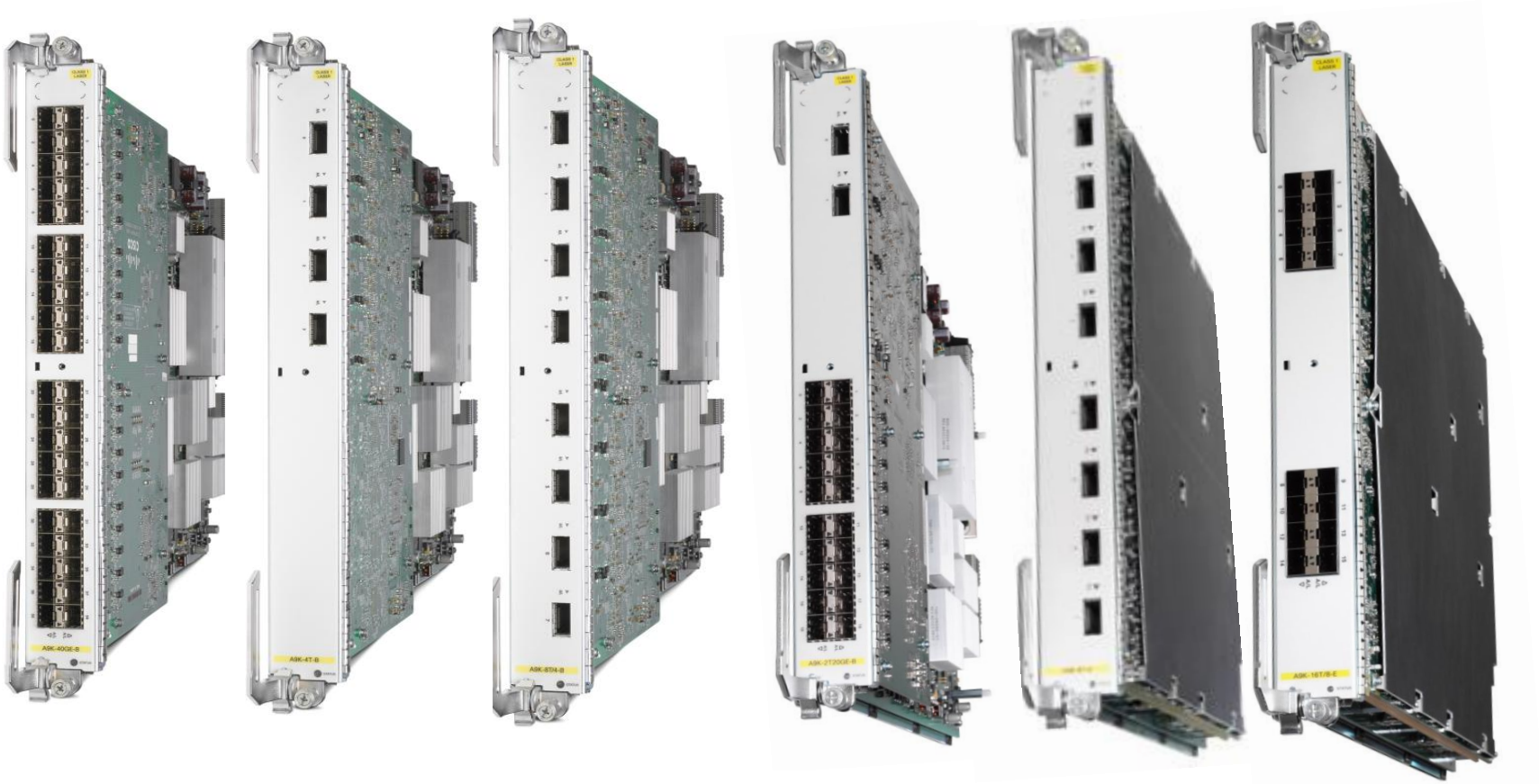


© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Ethernet Line Card Family

Shipping

- Common HW ASIC and HW architecture
- Identical SW features
- Line card Memory options for QoS scale



A9K-40G | A9K-4T | A9K-8T/4 | A9K-2T20G | A9K-8T | A9K-16T/8

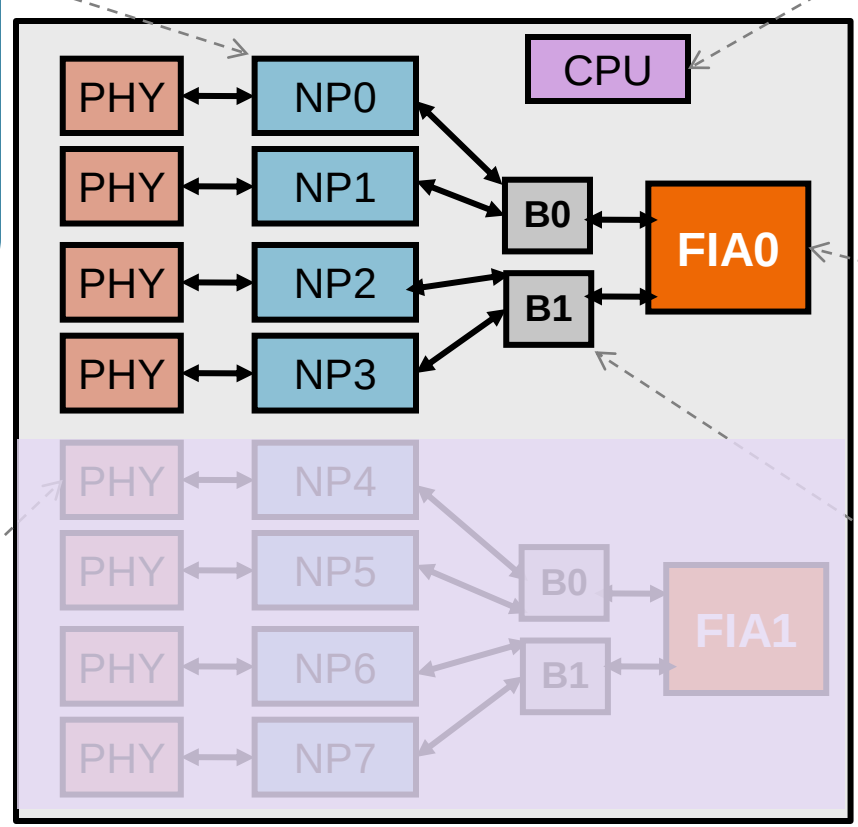
© 2010 Cisco and/or its affiliates. All rights reserved.
Cisco Confidential

Line Card Architecture – Hardware Components

NP: Network Processor
Main forwarding ASIC
L2 & L3 forwarding, features (QoS, ACL, etc), control plane policing, mcast replication, etc

10Gbps bi-directional with features applied

CPU (same as RSP)
Program HW forwarding tables
Distributed Control planes
SW switched packets
Inline Netflow



FIA: Fabric Interface ASIC
Provide non-blocking data connection to switch fabric
Internal system queues/VoQ
Intelligent mcast replication

10G PHY for one 10G port, or 10x1G port

B: Bridge FPGA
Provide non-blocking data connection between NP and FIA
Internal System queues
Intelligent mcast replication

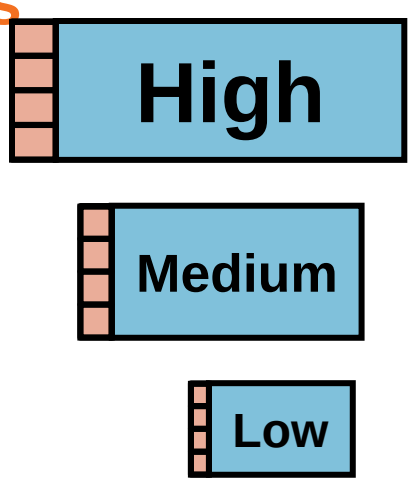
Example: A9K-8T

Note, Bridge FPGA provide non-blocking connection between NP and the FIA. Functionally it does the HW conversion due to different interface format on NP and FIA. It's part of the switch fabric connection. To make it logically simple, it will be removed from the remaining slides.

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Line Card Memory Options

Different Queue Scale for Different Deployments



- 3 memory options for each line card: **E**xtended (or high queue), **B**ase (medium queue), **L**ow (low queue)*
- Different memory option has different QoS queue/policer scale, L2 sub-interface scale and ACL scale
- All other system wide scale is the same across different type of the line cards, including FIB, MAC address, Bridge-domain, L3 sub-interface, VRF, etc
- All line cards have the same HW → Identical features
- Mixed different type of line cards are supported on the same chassis with same system wide scale and identical features

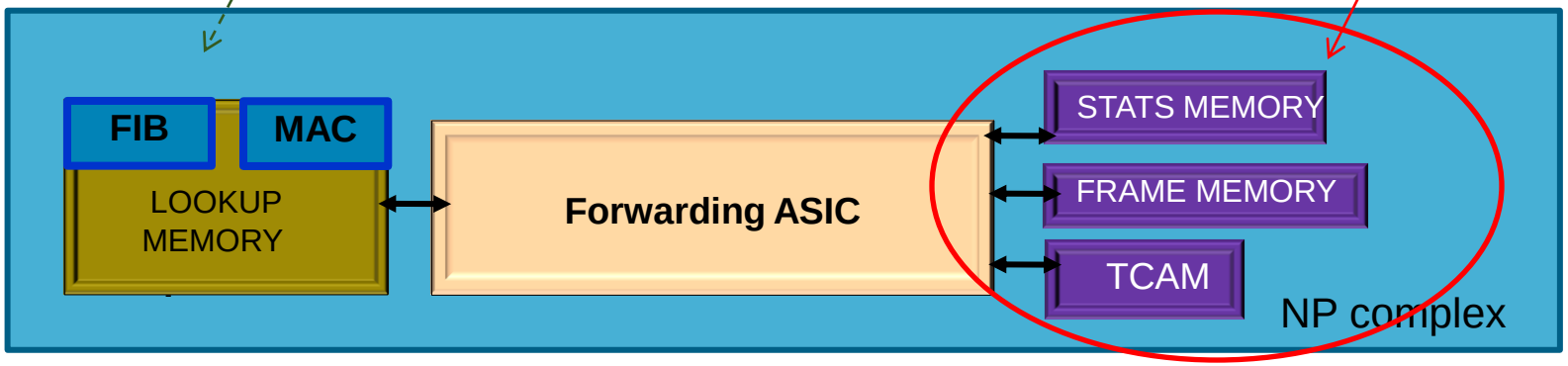
* A9K-16T/8 only have “B” option. It doesn’t have “E” or “L” option as of the 4.1.0 release

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L/B/E Line Cards – What's the Difference?

Same lookup memory size across different type of line cards → mix-and-match LCs, same system level scale

Different memory size for different QoS queuing/policing, L2 EFP and ACL scale



- Each NPU has Four Main Associated memories: TCAM , Search/Lookup memory , Frame/buffer memory and statistics memory
  - **TCAM** is used for VLAN tag, QoS and ACL classification
  - **Lookup Memory** is used for storing FIB tables, Mac address table and Adjacencies
  - **Stats memory** is used for all interface statistics, forwarding statistics etc
  - **Frame memory** is buffer memory for Queues
- E/B/L line card have different TCAM , Stats and Frame Memory size, which give different scale number of the QoS queues and L2 sub-interfaces per line card
- Lookup Memory is the same across line cards → why?
  - To support mix of the line cards without impacting the system wide scale including routing, multicast, MAC address, L3 interface, MPLS label space scale

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

SIP/SPA – Non-Ethernet Support

SPA support list:
http://www.cisco.com/en/US/partner/prod/collateral/routers/ps9853/data_sheet_c78-573452.html

Strong QoS
- 128+K Queues,
- 128+K Policers
- H-QoS
- Color Policing

Scalability
- Distributed control and data plane; 20G; 4 Bays
- L3 i/f, route, session, protocol – scaled up for MSE needs

High Availability
- IC-Stateful Switch Over capability,
- MR-APS
- IOS-XR base for high scale and reliability

Powerful and Flexible QFP Processor
- Flexible ucode Architecture for Feature Richness
- L2 + L3 Services: FR, PPP, HDLC, MLPPP, LFI, L3VPN, MPLS, Netflow, 6PE/VPE

A9K-SIP-700

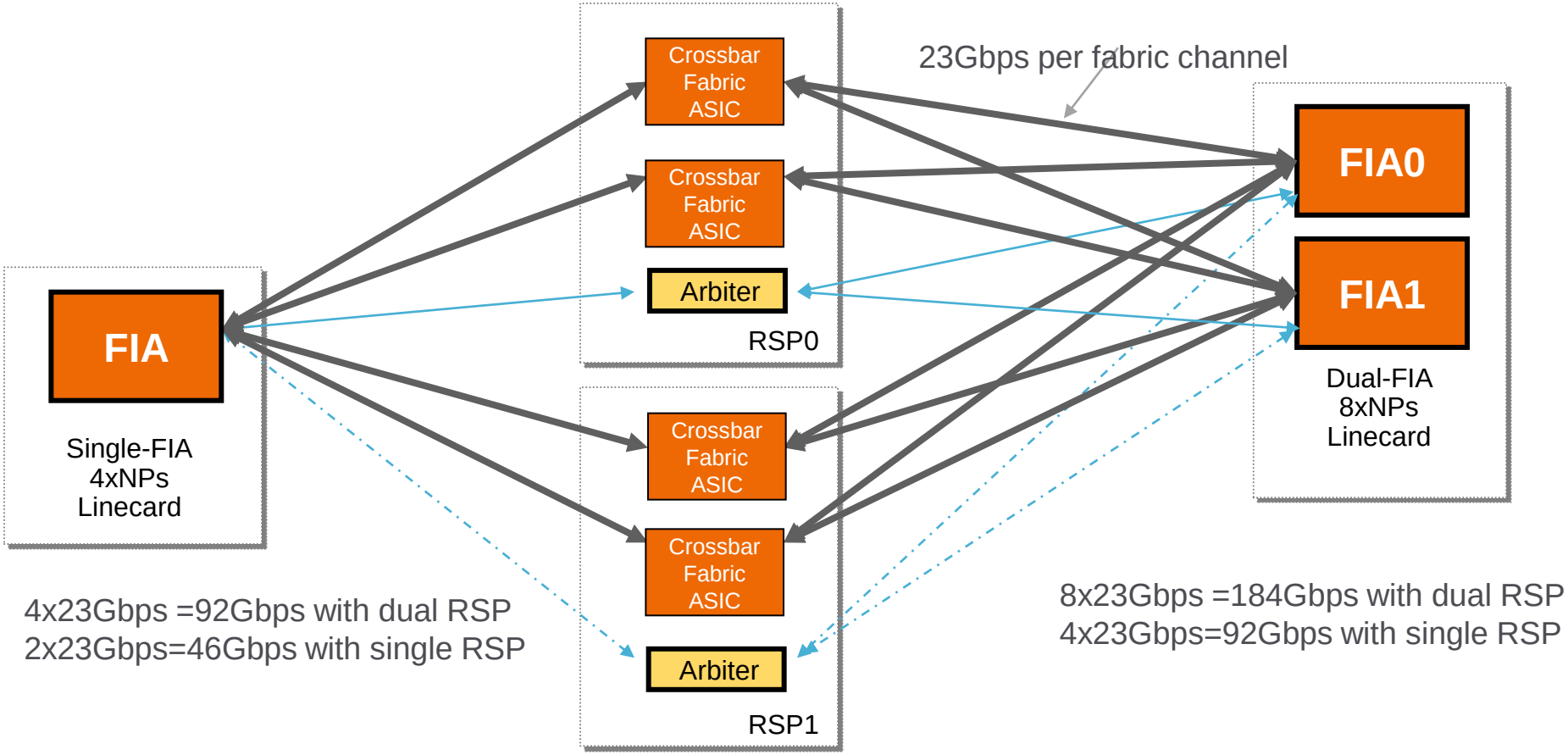
SPA Support:
–ChOC-12: T1 / T3, MLPPP, LFI, IC-SSO, MR-APS
–ChOC-3/STM1 + ChOC-12: DS0 / STM1 + ChOC-48: DS3 / E3 / STM4
–POS: OC3/STM1, OC-12/STM4, OC-48/STM16, OC-192/STM64

SPA Roadmap (2011):
Ch T1/E1, Ch T3/E3, CEoPs, ATM

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Switch Fabric Overview

- Active-active load balancing: Unicast: per-packet load balancing, Multicast: per (S,G) load balancing
- Arbiter for fabric access control. Arbiter is in active/standby mode, which is controlled by low level hardware signalling
- Frame format over fabric: super-frame, it can aggregate multiple small packet into a big sup-frame to improve the fabric throughput



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

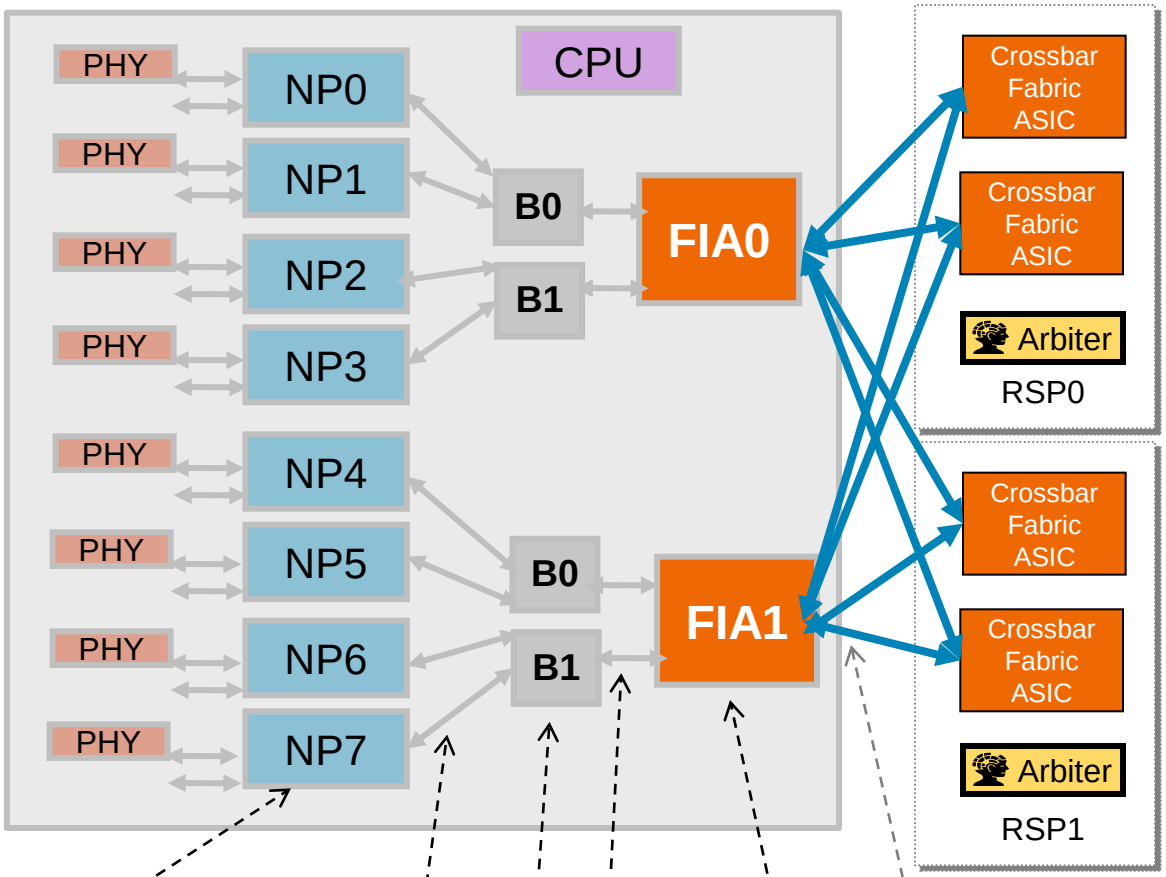
System Internal Bandwidth

NO bottleneck inside the system

A9K-8T/4
~60Gbps raw bandwidth

A9K-16T/8-B
~120Gbps raw bandwidth

Use A9K-8T line card as example



30Gbps and 25M pps
(combined ingress and egress)

15Gbps bi-directional

30Gbps bi-directional

30Gbps bi-directional

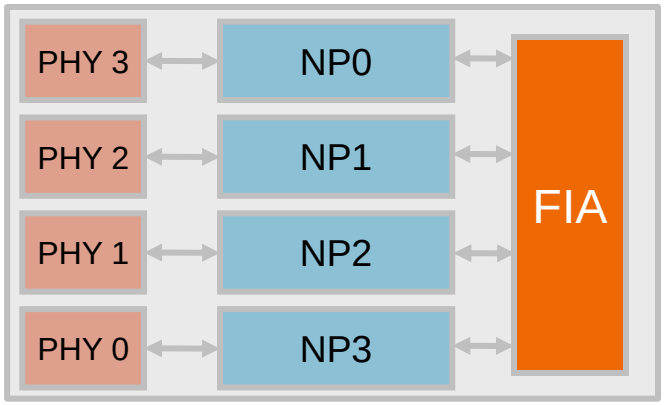
60Gbps bi-directional

Each FIA has one fabric channel which is 23 Gbps bi-directional, to each of the switch fabric ASIC

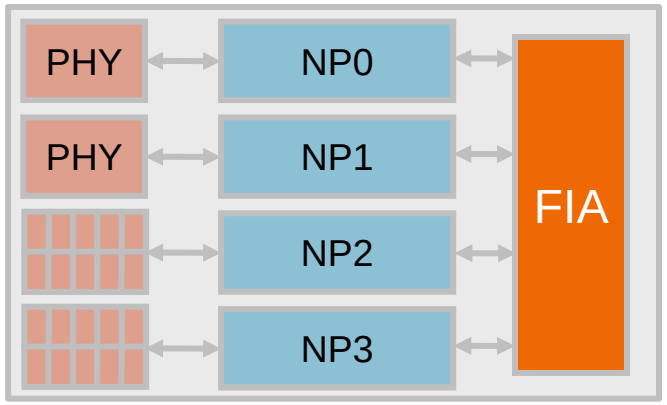
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

4xNPs Line Card Family

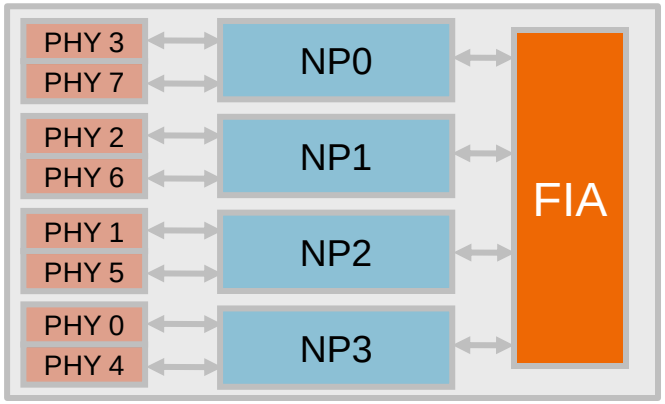
A9K-4T-E/B/L



A9K-2T20G-E/B/L

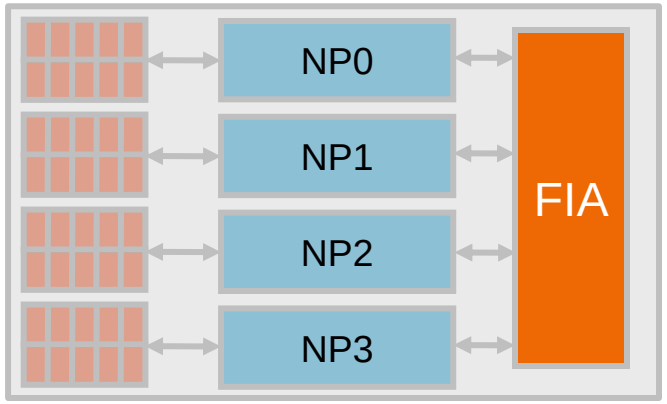


A9K-8T/4-E/B/L



Oversubscribed line card Up to 60Gbps (~57Gbps) bandwidth

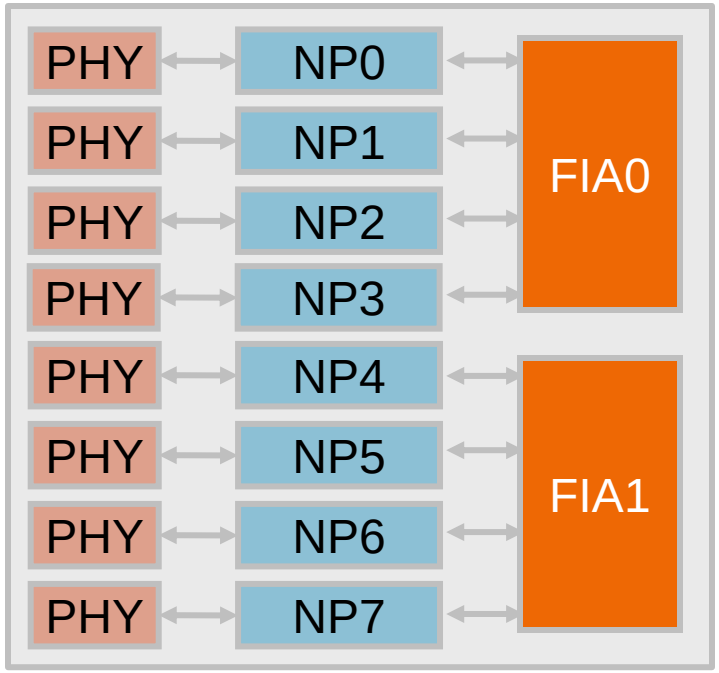
A9K-40G-E/B/L



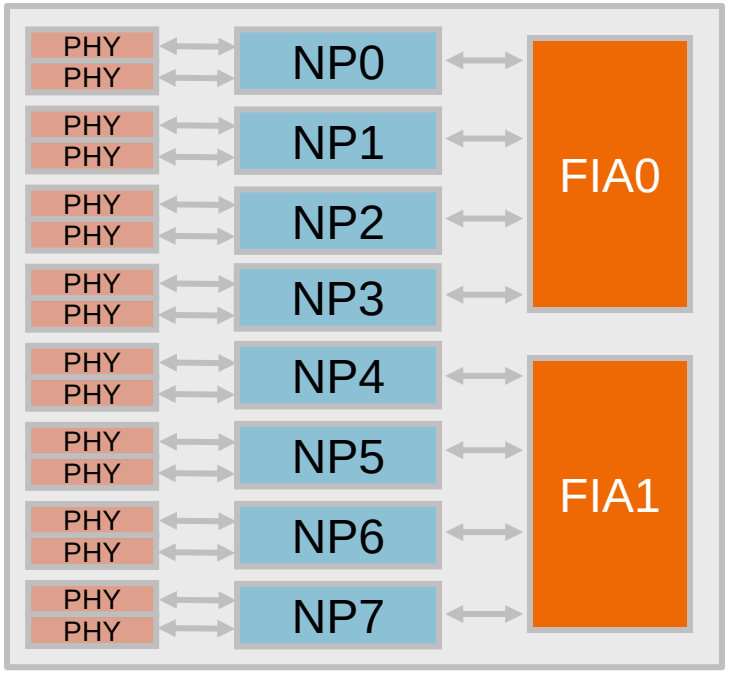
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

8xNPs Line Card Family

A9K-8T-E/B/L



A9K-16T/8-B



Oversubscribed line card
Up to 120Gbps (~117Gbps) bandwidth

© 2010 Cisco and/or its affiliates. All rights reserved.
Cisco Confidential

Packet Forwarding Architecture

Distributed Forwarding, Multicast Forwarding, EVC Architecture

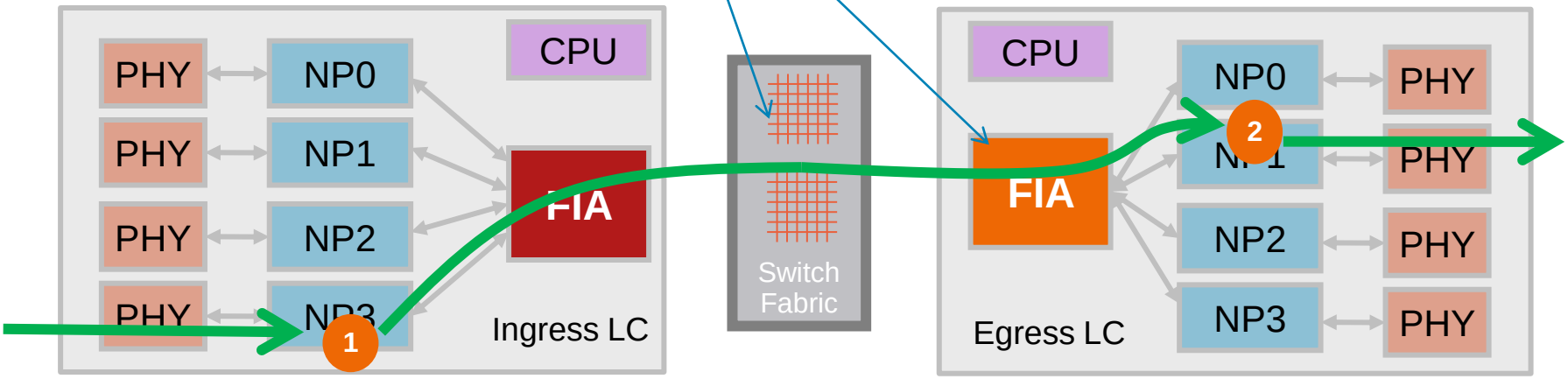
and/or its affiliates. All r eserved.
Cisco Confidential

IOS-XR Two-Stage Packet Forwarding

Fully Distributed Forwarding on Line Cards

- Each line card has independent AIB only for local interfaces
- Each line card has independent Interface DB for local interfaces
- Both Ingress and Egress FIB – allows forwarding features to be independently applied on LCs

Packet is forwarded to the egress NP based on the information in the NP/fabric header



Ingress NP look up →
Get egress NP information (added into the NP/fabric header), apply ingress features

Egress NP look up →
Get egress logical port, VLAN, MAC, ADJ information, etc for packet rewrite, apply egress features

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Multicast Packet Forwarding

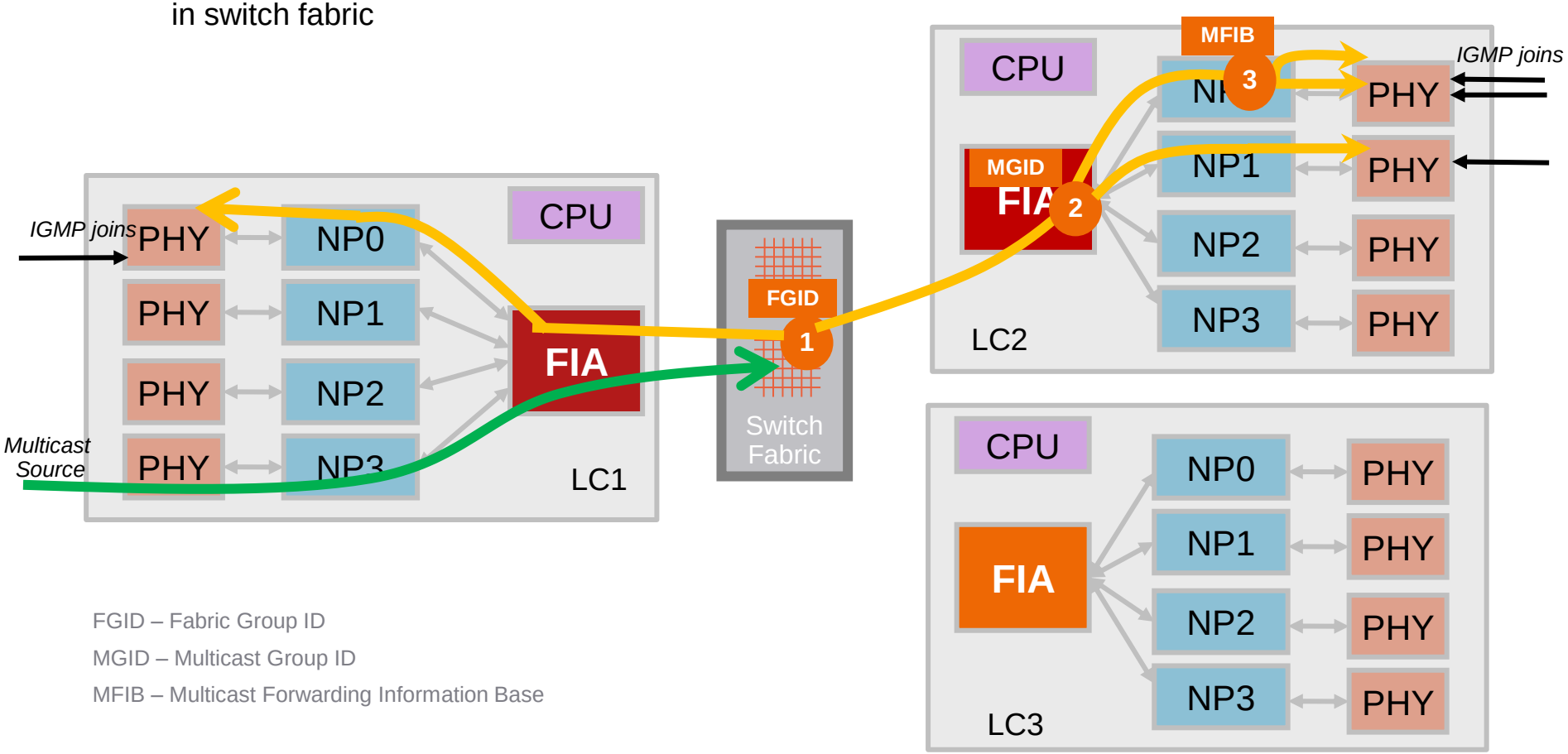
Intelligent Multicast Replication

- Efficiency
- Line rate
- Scalable
- Simple

1. **Fabric Replication →** replicate single copy to each LC which has IGMP join, based on FGID table in switch fabric

2. **FIA Replication →** replicate single copy to each NP which has IGMP join, based on MGID table in FIA

3. **NP Replication →** replicate single copy per each receiver based on multicast FIB table



FGID – Fabric Group ID

MGID – Multicast Group ID

MFIB – Multicast Forwarding Information Base

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Intelligent Multicast Replication

Switch Fabric and Egress LC Replication

1 **Fabric Replication** → replicate single copy to LCs which receive IGMP join, based on FGID table in switch fabric

2 **FIA Replication** → replicate single copy to NPs which receive IGMP join, based on MGID table in FIA

3 **Bridge Replication** → similar as FIA replication, single copy to NP

4 **NP Replication** → replicate copy per receiver based on multicast FIB table

FGID=3

1	0000000000
2	0000000001
3	hit 0000000011
	…
N	1111001111

FPOE Table

1Kx10 bitmap representing all combination of the egress linecards

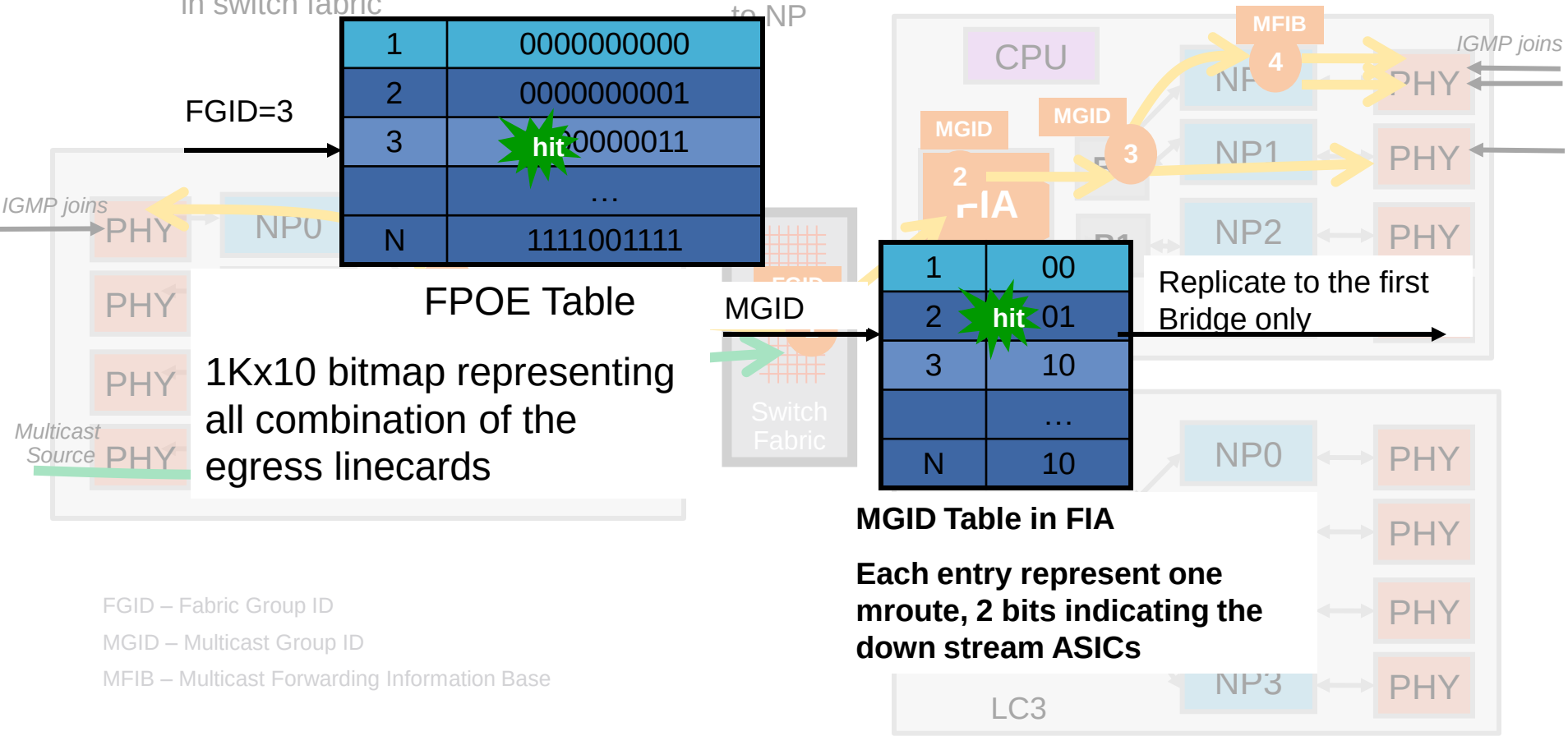
MGID

1	00
2	hit 01
3	10
	…
N	10

Replicate to the first Bridge only

MGID Table in FIA

Each entry represent one mroute, 2 bits indicating the down stream ASICs



FGID – Fabric Group ID

MGID – Multicast Group ID

MFIB – Multicast Forwarding Information Base

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

ASR 9000 EVC Infrastructure

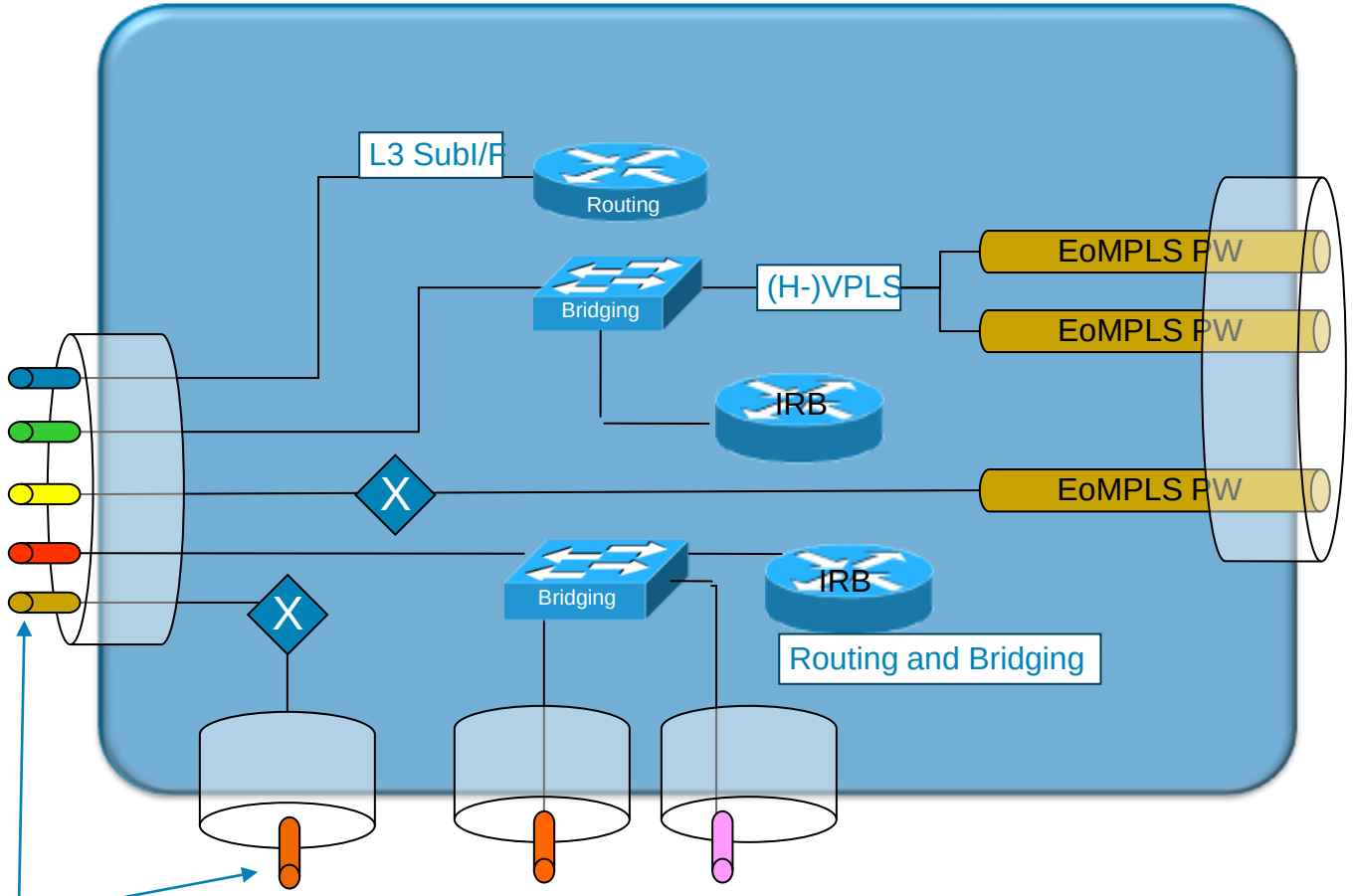
VLAN tag local significant

Flexible VLAN tag classification

Flexible VLAN tag rewrite

Flexible Ethertype (.1Q, QinQ, .1ad)

1



L2 or L3 sub-interfaces (802.1q/qinq/.1ad)

2

Flexible service mapping and multiplexing. Support all standard based services concurrently on the same port:

Regular L3, L2 interface/sub-interface

Integrated L2 and L3 – IRB/BVI

Mixed L2 and L3 sub-interfaces on the same port

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Flexible VLAN Tag Classification

```
RP/0/RSP0/CPU0:PE2-asr(config)#int gig 0/0/0/4.100 l2transport
RP/0/RSP0/CPU0:PE2-asr(config-subif)#encapsulation ?
  default   Packets unmatched by other service instances
  dot1ad    IEEE 802.1ad VLAN-tagged packets
  dot1q     IEEE 802.1Q VLAN-tagged packets
  untagged  Packets with no explicit VLAN tag

RP/0/RSP0/CPU0:PE2-asr(config-subif)#encapsulation dot1q 100-200,205 ?
  comma  comma
  exact  Do not allow further inner tags

RP/0/RSP0/CPU0:PE2-asr(config-subif)#encapsulation dot1q 100 second-dot1q 10-20,35 ?
  comma  comma
  exact  Do not allow further inner tags

RP/0/RSP0/CPU0:PE2-asr(config-subif)#encapsulation dot1ad 20 dot1q 10-20 ?
  comma  comma
  exact  Do not allow further inner tags
```

- Double tag
- Single tag
- Multiple tag
- Range of tag
- .1q and/or .1ad
- Loose or exact match
- Untagged
- Default tag

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Flexible VLAN Tag Rewrite

```
RP/0/RSP0/CPU0:PE2-asr(config)#int gig 0/0/0/4.100 l2transport

RP/0/RSP0/CPU0:PE2-asr(config-subif)#rewrite ingress tag ?
  pop        Remove one or more tags
  push       Push one or more tags
  translate  Replace tags with other tags

RP/0/RSP0/CPU0:PE2-asr(config-subif)#rewrite ingress tag pop ?
  1  Remove outer tag only
  2  Remove two outermost tags

RP/0/RSP0/CPU0:PE2-asr(config-subif)#rewrite ingress tag push ?
  dot1ad  Push a Dot1ad tag
  dot1q   Push a Dot1Q tag

RP/0/RSP0/CPU0:PE2-asr(config-subif)#rewrite ingress tag push dot1q 100 ?
  second-dot1q  Push another Dot1Q tag
  symmetric     All rewrites must be symmetric

RP/0/RSP0/CPU0:PE2-asr(config-subif)#rewrite ingress tag translate ?
  1-to-1  Replace the outermost tag with another tag
  1-to-2  Replace the outermost tag with two tags
  2-to-1  Replace the outermost two tags with one tag
  2-to-2  Replace the outermost two tags with two other tags
```

Pop tag 1 or 2

Push tag 1 or 2

Tag translation

- 1-1
- 1-2
- 2-1
- 2-2

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L2VPN P2P

EFP configuration example

```
Interface gig 0/0/0/1.101 l2transport
encapsulation dot1q 101 second 10
rewrite ingress pop 2 Symmetric

Interface gig 0/0/0/2.101 l2transport
encapsulation dot1q 101
rewrite ingress pop 1 Symmetric

Interface gig 0/0/0/3.101 l2transport
encapsulation dot1q 102-105
rewrite ingress push dot1q 100 Symmetric
```

L2VPN P2P service configuration example

```
l2vpn
 xconnect group cisco
   p2p service1 ← local connect
     interface gig 0/0/0/1.101
     interface gig 0/0/0/2.101
   p2p service2 ← VPWS
     interface gig 0/0/0/3.101
     neighbor 1.1.1.1 pw-id 22
  p2p service3 ← PW stitching
     neighbor 2.2.2.2 pw-id 100
    neighbor 3.3.3.3 pw-id 101
```


© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L2VPN Multi-Point

EFP configuration example

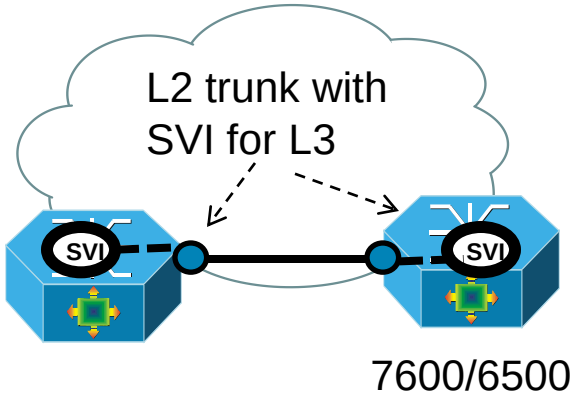
```
Interface gig 0/0/0/1.101 l2transport
encapsulation dot1q 101
rewrite ingress pop 1 Symmetric

Interface gig 0/0/0/2.101 l2transport
encapsulation dot1q 101
rewrite ingress pop 1 Symmetric

Interface gig 0/0/0/3.101 l2transport
encapsulation dot1q 102
rewrite ingress push dot1q 100 Symmetric
```

L2VPN MP service configuration example

```
l2vpn
 bridge group cisco
  bridge-domain domain1 ← local bridging
   Interface gig 0/0/0/1.101
     split-horizon group ← no bridging among same SHG
   Interface gig 0/0/0/2.101
     split-horizon group

  bridge-domain domain2 ← VPLS
   Interface gig 0/0/0/1.101
   Interface gig 0/0/0/2.101
   vfi cisco
    neighbor 192.0.0.1 pw-id 100
    neighbor 192.0.0.2 pw-id 100

  bridge-domain domain3 ← H-VPLS
   Interface gig 0/0/0/1.101
   neighbor 192.0.0.3 pw-id 100 ← spoke PW
   vfi cisco ← for core PWs
    neighbor 192.0.0.1 pw-id 100 ← core PW
    neighbor 192.0.0.2 pw-id 100
```

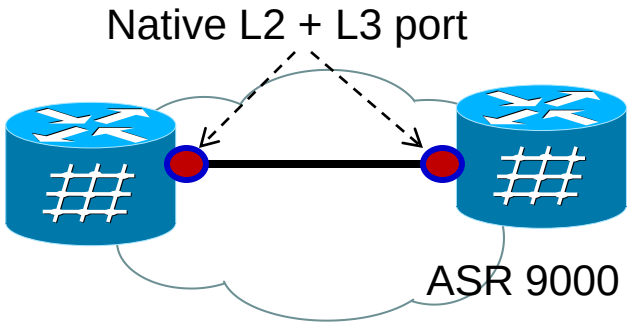

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Mixed L2 and L3 on the Same Port



L2 trunk port, use SVI for L3

```
interface gig 1/2
  switchport
  switchport mode trunk
  switchport trunk allow vlan 50-1000

Interface vlan 50
  ip address 1.1.1.1 255.255.255.0
```



Native L2 and L3 sub-interface on the same port

```
interface gig 0/0/0/1.50
   encapsulation dot1q 50
   ip address 1.1.1.1 255.255.255.0

Interface gig 0/0/0/0/1.51 l2transport
   encapsulation dot1q 51
   rewrite ingress tag pop 1 Symmetric

l2vpn ← define L2 service
  bridge group test
  <snip>
```

Native L2 and L3 sub-interface support on the same physical port
No such limitations of the L3 features on the SVI

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

ASR 9000 L3 "At-a-Glance"

Extend IOS-XR L3/MPLS to the ASR9000

Mature IOS-XR L3 Routing & MPLS SW Infrastructure

IOS-XR routing and MPLS → Widely deployed, proven stability and feature rich

Built-in HA → NSF/GR & NSR & SSO

Distributed control plan → High Multi-dimensional control plane scale, for example 5000 BGPs + 10,000 BFDs + etc

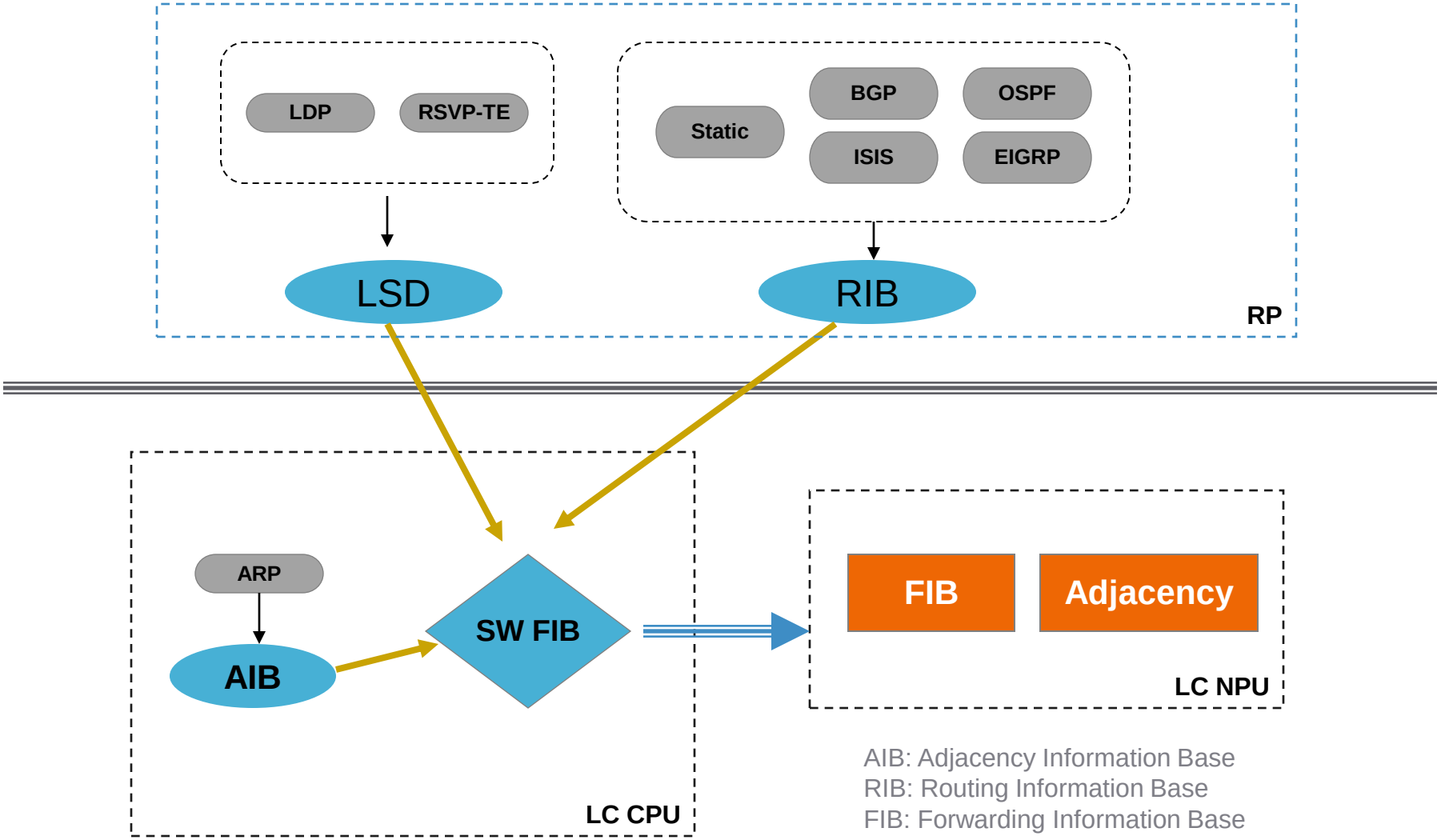
Hierarchical FIB structure and Fast FIB programming → Prefix independent fast network convergence

IPv6 ready → Dual stack, 6PE/6VPE, high scale, line-rate IPv6 forwarding performance

Distributed inline Netflow → 1M entries/LC , 100Kpps /LC, flexible sampling and IPv4/IPv6/MPLS feature support

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L3 Control Plane Overview

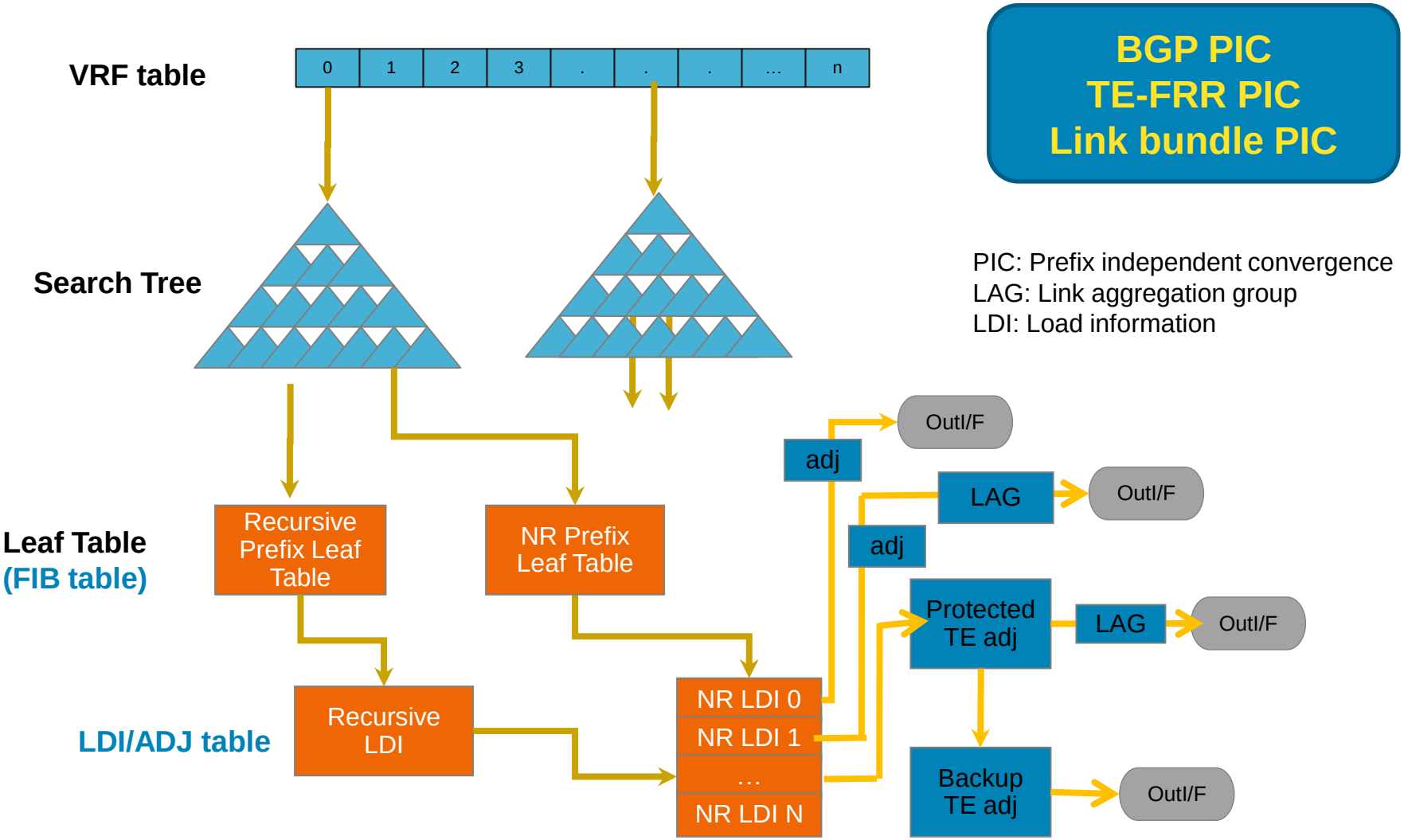


AIB: Adjacency Information Base
RIB: Routing Information Base
FIB: Forwarding Information Base
LSD: Label Switch Database

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L3 Forwarding – Hierarchical Data Structure

FIB and ADJ Tables Overview



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Cisco ASR 9000 L2 Behavior

It does FULL L2, but not a classic switch

ASR 9000 support full L2 service (multi-point, point-to-point, E-Tree native L2 or L2 over MPLS). It's fully MEF certified.

However, the behavior and configuration is quite different than traditional L2 switch

- **By default, it's Router**, it doesn't do any bridging function unless you configure it specifically
- There is **no "switch-port"** concept. All L2 features are done with "EVC" SW infrastructure
- There is **no global VLAN** concept. All VLAN is local port significant. By default VLAN 10 on port A doesn't talk to VLAN 10 on port B. It require "bridge-domain" configuration to bridge port/VLAN together
- By default, **STP is not enabled** on any L2 port unless you "explicitly" configure it under STP process

NO global VLAN (no 4K limit)
Per-VLAN features, more flexible, more secure, more scale

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

ASR 9000 L2 “At-a-Glance”

Built on New Cisco EVC SW Foundation

Cisco “EVC” Carrier Ethernet Foundation

Flexible Cisco “EVC” Carrier Ethernet SW infrastructure → VLAN tag matching, manipulation, service mapping

Distributed control plan → High Multi-dimensional L2 service scale, VLAN, PW, Bridge-domain, VPLS, MAC address scale, etc

Distributed HW based MAC learning, aging/flushing, 4-5Mpps

Rich L2 fast convergence features

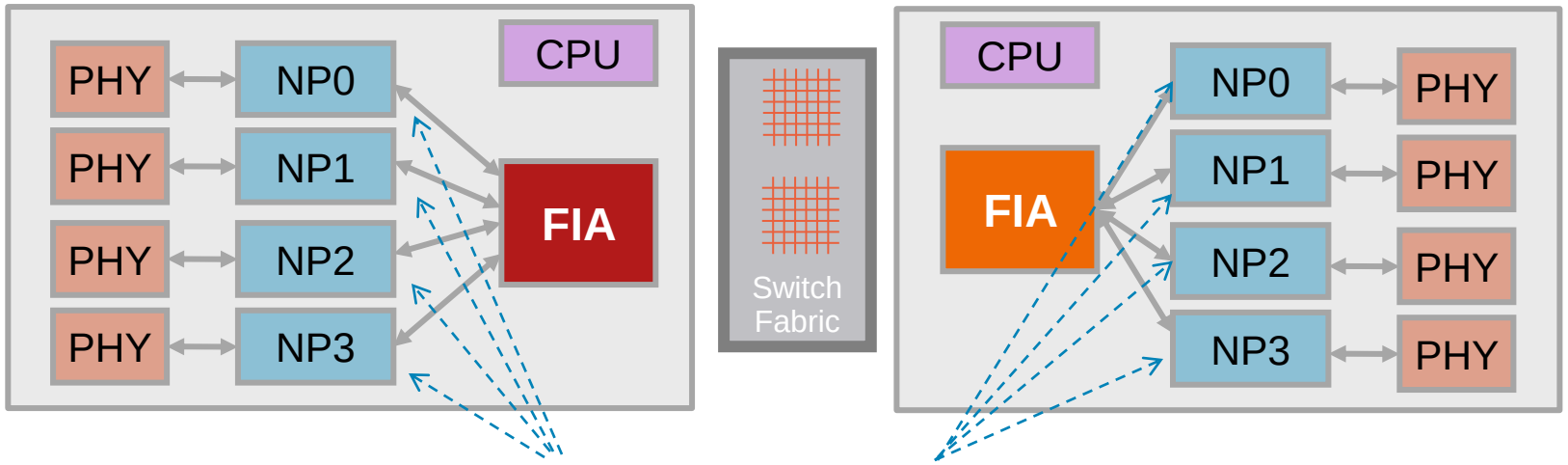
Standard compliant, MEF certified, 802.1q/qinq/.1ad/.1ah EoMPLS, H-VPLS/VPLS supported, L2 feature rich

E-OAM: CFM, Y.1731, Link OAM

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

ASR 9000 MAC Learning Overview

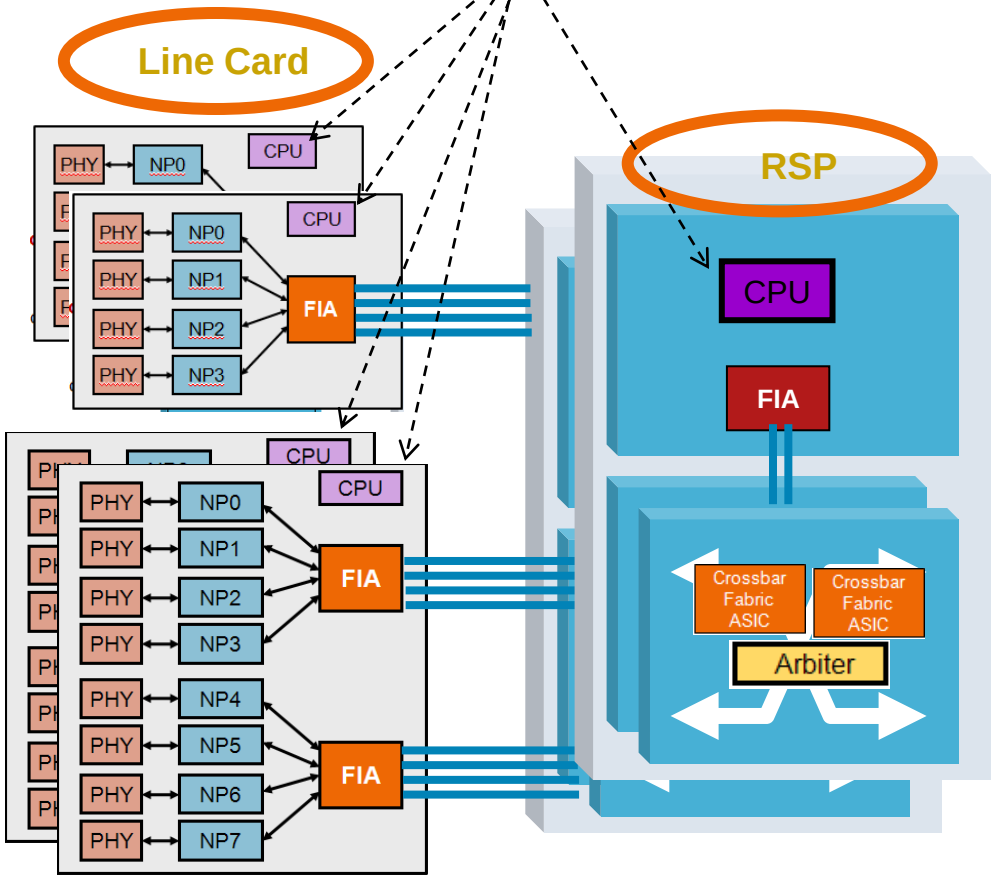
- **Distributed HW based MAC Learning**
  - –MAC learning /aging/flushing is fully distributed to the NP, independently
  - –Each NP can learn around 4-5 Millions MAC addresses per second in hardware
- **Data-plane MAC table synchronization**
  - –MAC address are synchronized across all NPs in the entire system, regardless of the bridge-domain or bridge port configuration.
  - –MAC address synchronization is achieved by special MAC notification messages which is implemented in the data-plane microcode



MAC learning/aging/flushing is done by hardware and fully distributed on each NP independently

ASR 9000 Scalable System Architecture

- Control plane split among RSP and LC CPU (same type of CPU as RSP)
- L2 protocols, BFD, CFM, Netflow runs on the LC CPU for high scale



"8+1" CPU for Ultra-High multi-dimensional control plane scale

Scale examples

Multi-dimensional scale	
FIB	1.3M (v6 use two entries)
MAC address	512K
L2 interfaces	64K
Pseudowires	64K
Bridge-domain/VFI	8K
EFPs per BD	16K
L3 interfaces, VRF	8K
CFM MEP/MIP	32K/LC
Netflow	1M entries/LC, 100K pps/LC
BFD	1250/LC (with 150msec hello timer), minimal hello timer is 15msec

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

HA and Fast Convergence

Business L3 VPN
Business L2 VPN (E-Line/E-LAN/E-Tree)
Residential Data/Voice
Residential IPTV/VoD

Service layer resiliency, service termination point (node or link) fail., service is down and need to re-route

- L2: PW redundancy, L3: BGP PIC, Multicast: MoFRR
- Access dual home: REP-AG/MST-AG, Rapid-PVST/PVST-AG, MC-LAG, G.8032, Flexible Link (not supported), HSRP/VRRP
- State sync between primary and backup service node: ANCP, IGMP snooping, DHCP/ARP/IGMP (in future), etc

L2 and L3 transport layer resiliency, transit link or transit node fail, L2 or L3 transport protocol re-converge. Service layer is not aware

- L2: MST, G.8032, REP(not supported), PVST(not supported)
- L3: IGP/BGP, TE/FRR(link, node, path protection), IP/FRR (ISIS per-link and per-prefix , OSPF per-link, OSPF per-prefix IP/FRR
- Multicast: PIM, P2MP-TE/FRR

Link layer resiliency: transport and service layer is not aware, for link protection only
Link Bundle (active-active, 1:1 backup, m:n backup)

Resiliency infrastructure: distributed BFD, distributed EFD, HW based LoS detection, IPoDWDM FEC/EFC & proactive protection, Hierarchical Forwarding table structure and in-place modification, Fast programming FIB table

System HA: IOS-XR Modular OS, Micro-kernel, RSP failover, SMU, ISSU(4.2.1), Process restart. All protocols are SSO, all protocol support NSF/GR, OSPF/ISIS/LDP/BGP support NSR. PIM support NSF and is NSR compliant

Feature Architecture

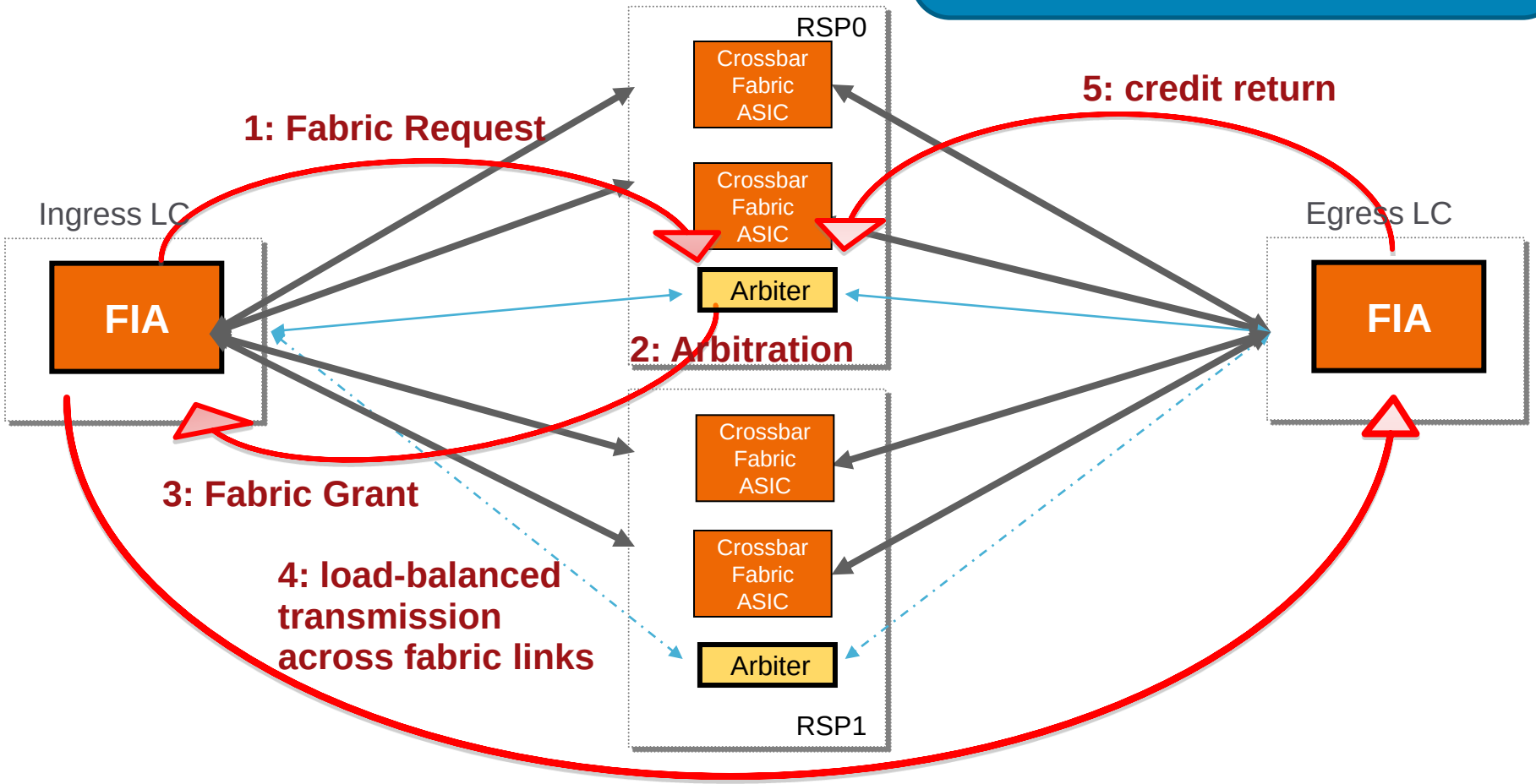
Quality of Service, MoFRR, Vidmon, New Features

and/or its affiliates. All r eserved.
Cisco Confidential

Switch Fabric Bandwidth Access Overview

Intelligent Fabric and Internal System QoS

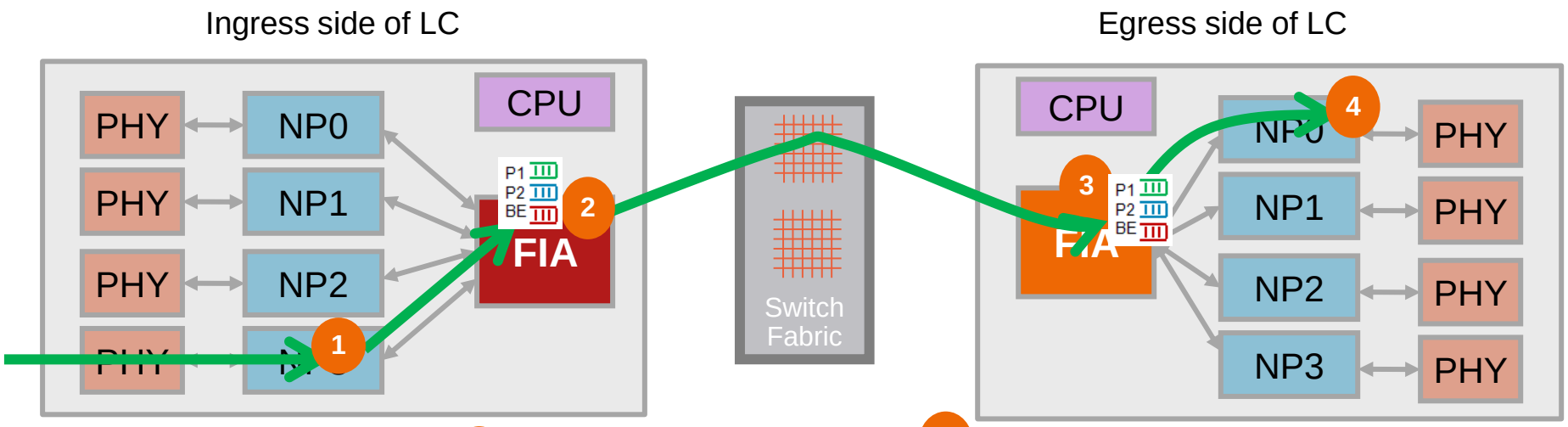
- 3 strict priority scheduling/queueing
- Back pressure and virtual output queue
- Multicast and Unicast separation (separated queues and fabric plane)



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

End-to-End System QoS Queuing

End-to-End priority (P1,P2, Best-effort) propagation → Guarantee bandwidth, low latency for high priority traffic at any congestion point
3 strict priority level across all internal HW components



1 Ingress (sub-)interface QoS Queues

2 Virtual Output Queues

3 Egress FIA Queues

4 Egress (sub-)interface QoS Queues

Configure with
Ingress MQC
4-layer hierarchy
Two strict high priority +
Normal priority

Implicit Configuration
Two strict high priority +
Normal priority

Configure with
Egress MQC
4-layer hierarchy
Two strict high priority +
Normal priority

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Backpressure and VoQ Mechanism

VoQ Scale: Each FIA has P1/P2/BE queue set for every NP and RSPs in the entire system

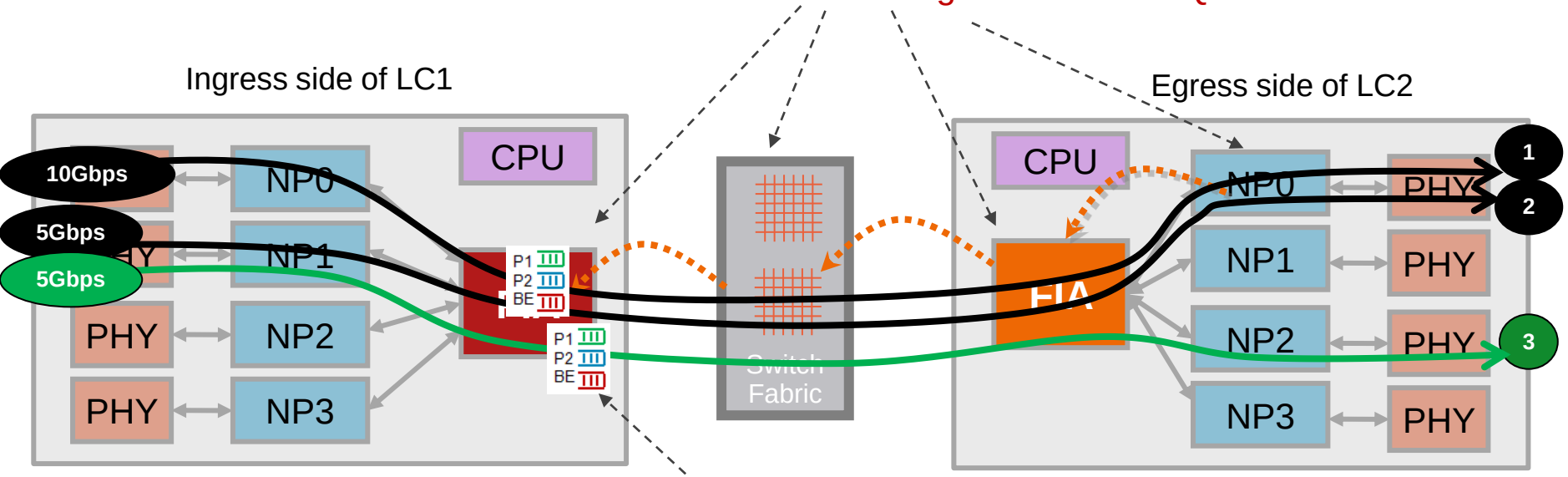
Egress NP congestion → → backpressure to ingress FIA →

Packet is en-queued in the dedicated VoQ →

No impact of the packet going to different egress NP →

No head-of-line-block issue

Backpressure: egress NP → egress FIA → fabric Arbiter → ingress FIA → VoQ

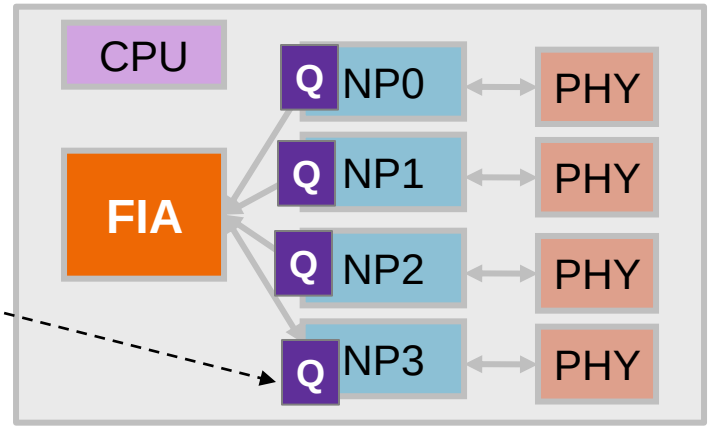


Packet going to different egress NP put into different VoQ set →
Congestion on one NP won't block the packet going to different NP

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Port QoS Overview

User configure port level QoS via MQC CLIs
QoS function is done by special Queuing HW chip within the NP complex
Different memory options have different scale



- High scale
  - –Up to 3 Million queues per system (with E linecard)
  - –Up to 2 Million policers per system (with E linecard)
- Highly flexible: 4 layer hierarchy queuing/scheduling support
  - –Four layer scheduling hierarchy →Port, Subscriber Group, Subscriber, Class
  - –Egress & Ingress, shaping and policing
- Three strict priority scheduling with priority propagation
- Flexible & granular classification, and marking
  - –Full Layer 2, Full Layer 3/4 IPv4, IPv6

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

QOS Configuration Basics

- IOS XR MQC CLI
- Attachment Points
  - –Main/physical Interface: MQC applied to a physical port will take effect for traffic that flows across all sub-interfaces on that physical port
  - –L3 sub-interface
  - –L2 sub-interface (EFP)
  - –You can have either port-based or sub-interface based policy on a given physical port, but not both at same time

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

QoS Classification Criteria

- Very flexible L2/L3 field classification on L2 interfaces
  - Inner/outer cos
  - Inner/Outer vlan
  - DEI
  - Outer EXP
  - DSCP/TOS
  - TTL, TCP flags, source/destination L4 ports
  - Protocol
  - Source/Destination IPv4
  - Source/Destination MAC address
  - Discard-class
  - Qos-group
  - match all/match any
- Max. 8 match statements per class

Note: a given match statement can have multiple entries (i.e COS 1,2,3)

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Marking Details

- "settable" packet fields:
  - DSCP/Precedence
  - EXP imposition
  - EXP topmost
  - COS inner/outer
  - qos-group
  - discard-class

Sample configuration for the marking

```
class CLASS-UNTAGGED-PD
 set cos 2
 set dscp 21

police rate 10240000 bps burst 320000 bytes peak-rate 20480000
bps peak-burst
  conform-action set cos 2
  conform-action set prec 2
  exceed-action set cos 1
  exceed-action set dscp 8
  violate-action set cos 0
  violate-action dscp 0
```

- ASR9K supports maximum of 2 fields per class-map. The same 2 fields can be placed in any combination below
  - 2 sets per police-conform/exceed/violate
  - 2 sets without policing

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

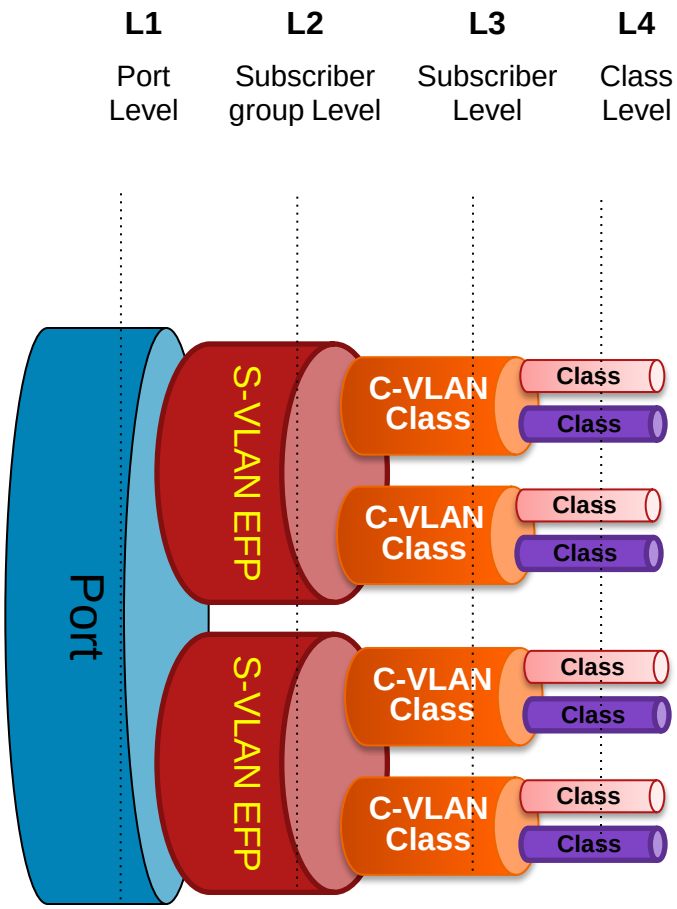
Policing Details

- RFC 2698 supported (2r3c) and 1r2c
  - ✓color blind mode
- Ingress & egress policing supported
- General Rule: Policing required on priority queues.
  - Priority level 2 classes can also accept shaping instead of policing.
- Granularity of 64Kbps supported.
- 2-level nested policy maps supported
  - Note: policers at parent and child work independently
- 64k policers per NPU (shared for ingress/egress) on extended linecards
- Policer actions supported:
  - –transmit
  - –drop
  - –set (implicitly behaves like set and transmit)
  - –each color can have **two** set actions:

```
Policy-map parent
   Class class-default
         Police rate 10 Mbps  peak-rate 20 mbps
               conform-action set dscp af12
               conform-action set cos 2
               exceed-action set dscp af13
               exceed-action set cos 3
```


© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

4 Layer Hierarchy QoS Queuing Overview



L1 level is implicitly assumed

Hierarchy levels used are determined by how many nested levels a policy-map is configured for and applied to a given sub-interface

Max 8 classes per subscriber level

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

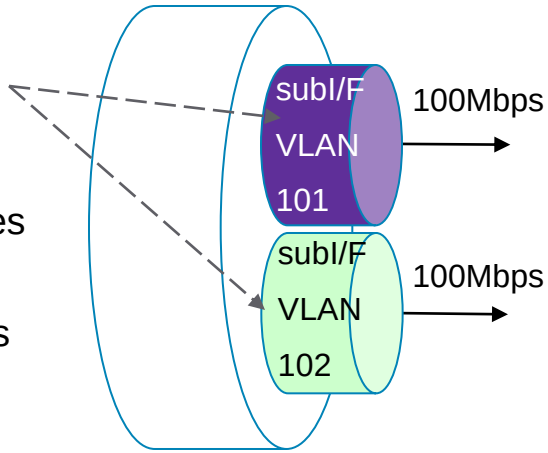
Shared Policy Instances (SPI)

Policy-map X

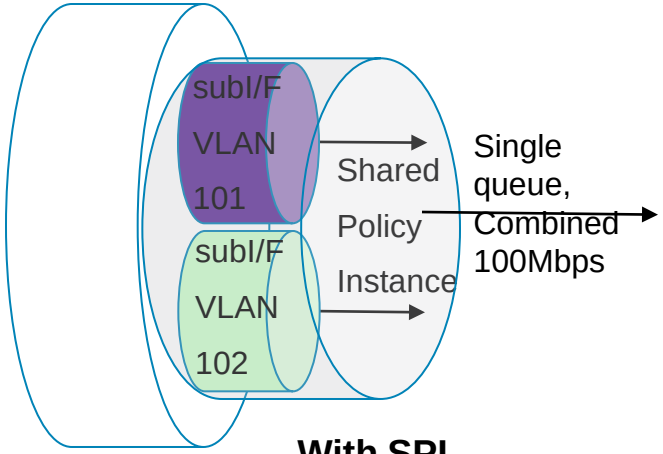
shaping 100Mb

Applied to both sub-interfaces

Creating two queues, each queue is shaped to 100Mbps



Without SPI



With SPI

- Enable distinct forwarding entities on the same port to use the exact same policy instance in order to share a common SLA enforcement
- Support 3 and 4 layer hierarchies
- Lightweight implementation with extension to existing MQC
- Require a new key-word to the service-policy comand:

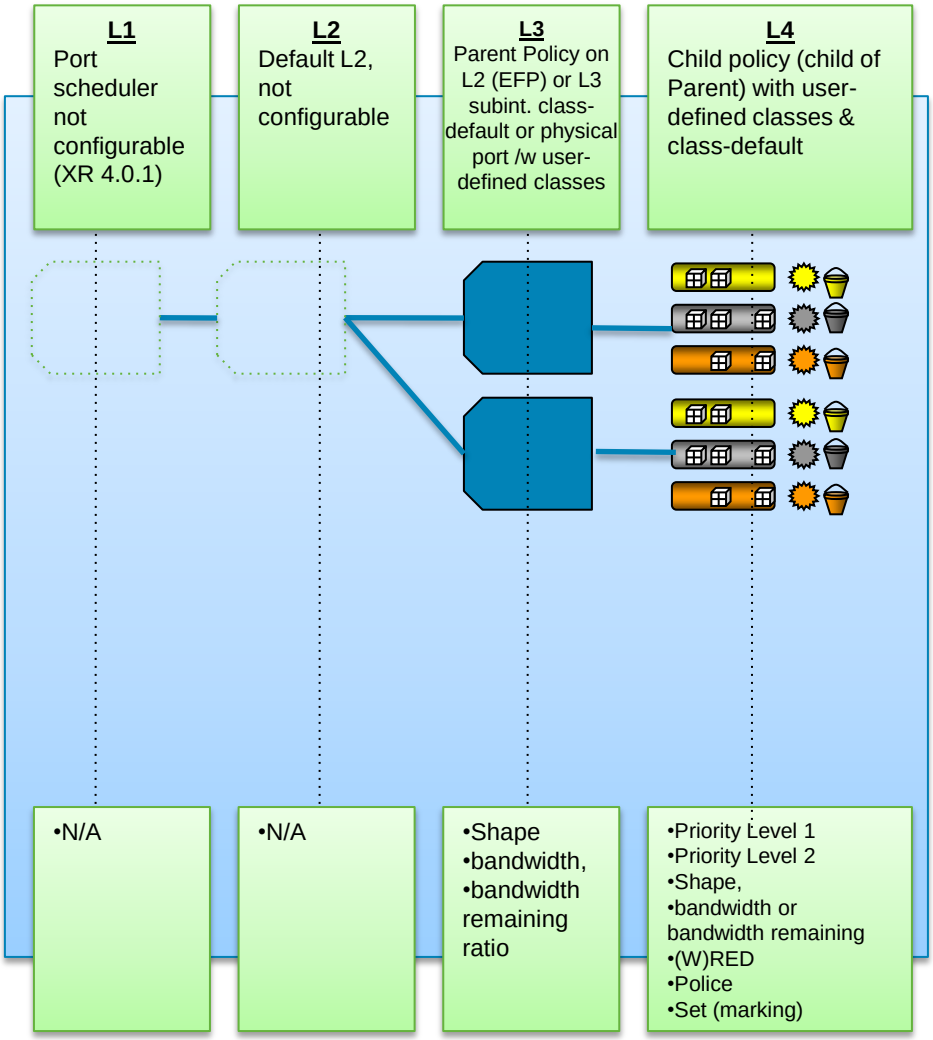
```
service-policy output/input <name> shared-policy-instance <name>
```

- Require that both EFPs/SubI/Fs have to reside on the same port

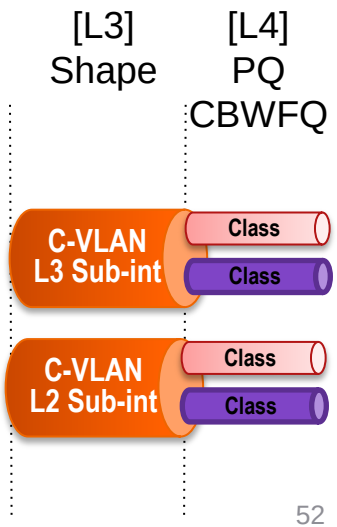
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

3-Layer H-QoS – CLI Examples (1)

Example 1 – policy-map on subinterface




```
policy-map parent (configures L3 layer)
  class-default
    shape average x
    …
    service-policy child

  policy-map child (configures L4 layer)
   class-1
     priority level x
   class-2
     bandwidth remaining percent x
   …

int GigE 0/1/2/3.10 l2transport
encapsulation dot1q 10
service-policy output parent

int GigE 0/1/2/3.20
Ipv4 address …
encapsulation dot1q 20
service-policy output parent
```

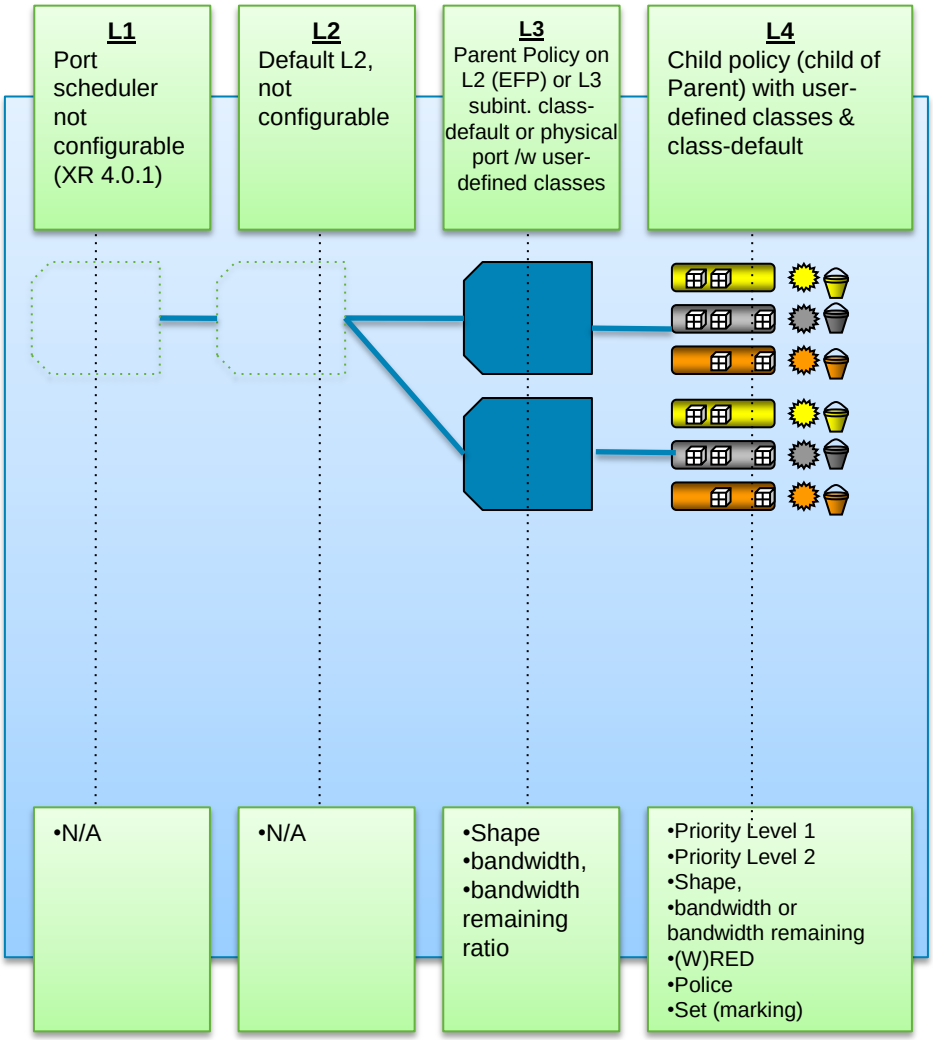


Note: Example applies to L3 and L2 VLAN subinterfaces

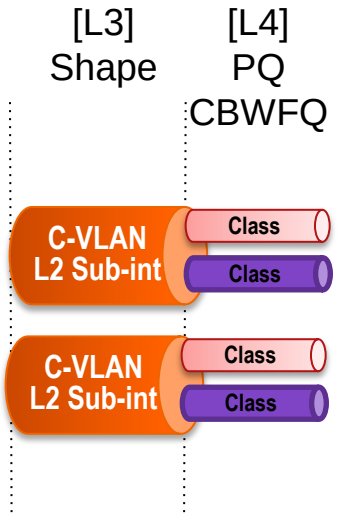
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

3-Layer H-QoS – CLI Examples (2)

Example 2 – policy-map on physical interface




```
policy-map parent (configures L3 layer)
 class VLAN10 (class-map match on dot1q tag 10)
    shape average x
    …
    service-policy child

 class VLAN20 (class-map match on dot1q tag 20)
    shape average x
    …
    service-policy child


  policy-map child (configures L4 layer)
   class-1
     priority level x
   class-2
     bandwidth remaining percent x
   …

int GigE 0/1/2/3
service-policy output parent

int GigE 0/1/2/3.10 l2transport
encapsulation dot1q 10

int GigE 0/1/2/3.20 l2transport
encapsulation dot1q 20
```



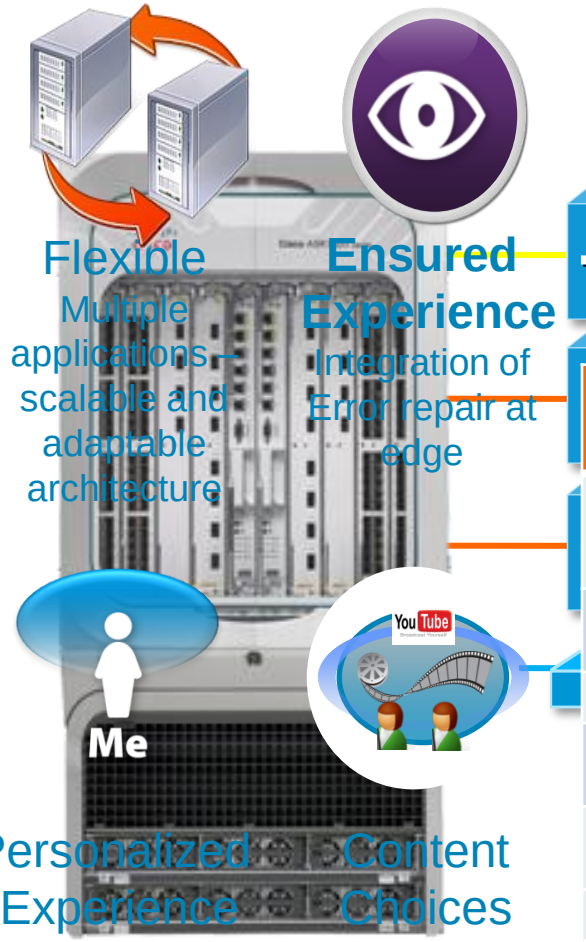
Note: Example applies to L2 VLAN subinterfaces only, since match VLAN is not supported for L3 interfaces

© 2010 Cisco and/or its affiliates. All rights reserved.
Cisco Confidential

“Video Aware” Edge Router

High Bandwidth Scale + High Quality

- Transparent Caching
- VOD, Internet Caching & Streaming
- Inline Video Monitoring
- High Density for VoD Bandwidth Requirements
- Best-in-Class Multicast Replication and scale
- Best-in-Class High quality: System QoS, HA, Fast convergence (MoFRR)



Flexible
Multiple applications – scalable and adaptable architecture

Ensured Experience
Integration of Error repair at edge

Personalized Experience
Virtual personal channels
Personalized ads

Content Choices
VoD, Online, User generated content

Feature	ASR 9000 CDS-TV Capabilities
Applications	Ultra-Dense VoD, TV, Internet Streaming, Error Repair, NPS*, CGN*, SBC*
Bandwidth	30-40 Gbps streaming capacity ~3 Gbps cache fill rate
Compatibility	Works with all CDS appliances
Concurrent Streams	Up to 10,000 SD equivalent
Content Cache	3.2 TBytes at FCS - Modular Design
Environmental	NEBS / ETSI compliant

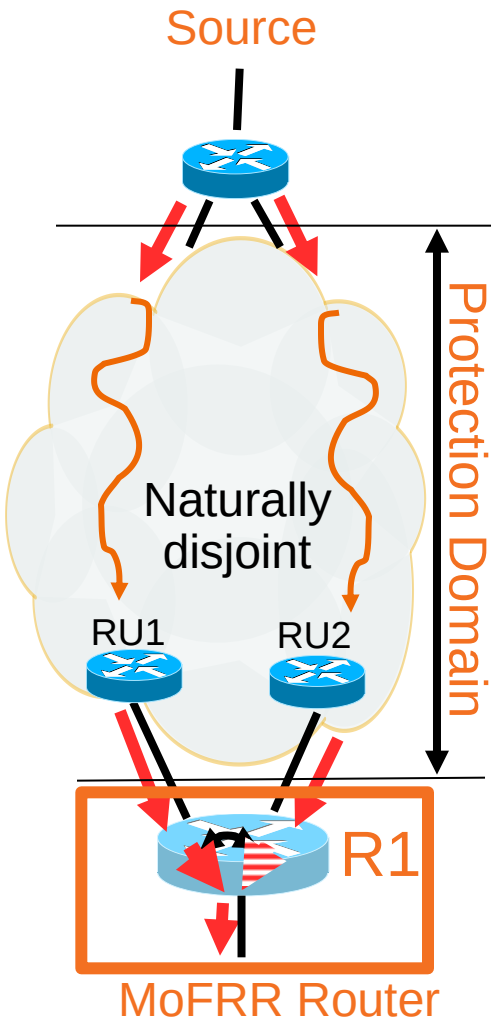
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Multicast Network HA

Truly Resilient Multicast by ASR 9000 MoFRR

Multicast Only Fast Re-Route

- **ASR 9000 MoFRR Overview**
  - –Tail end (Receiver) PE send PIM join on both of its redundant uplinks to receive identical streams over disjointed network paths: primary and backup
  - – HW monitor the video flows on the primary path
  - – Automatic Switch over to back up path based on configurable criteria (hard coded 30msec in phase 1)
- **Key Differentiators**
  - – Only need to be enabled on the last hop Router for operational simplicity
  - – Fast Convergence: <50msec (1000 channels tested) convergence time for source-to-receiver path (link and node) protection
  - – Native IP multicast solution without the complexity of MPLS P2MP TE/FRR (but can be integrated with TE)
  - – Pro-active protection by using Vidmon metric as trigger
  - –Towards lossless Video transport solution



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

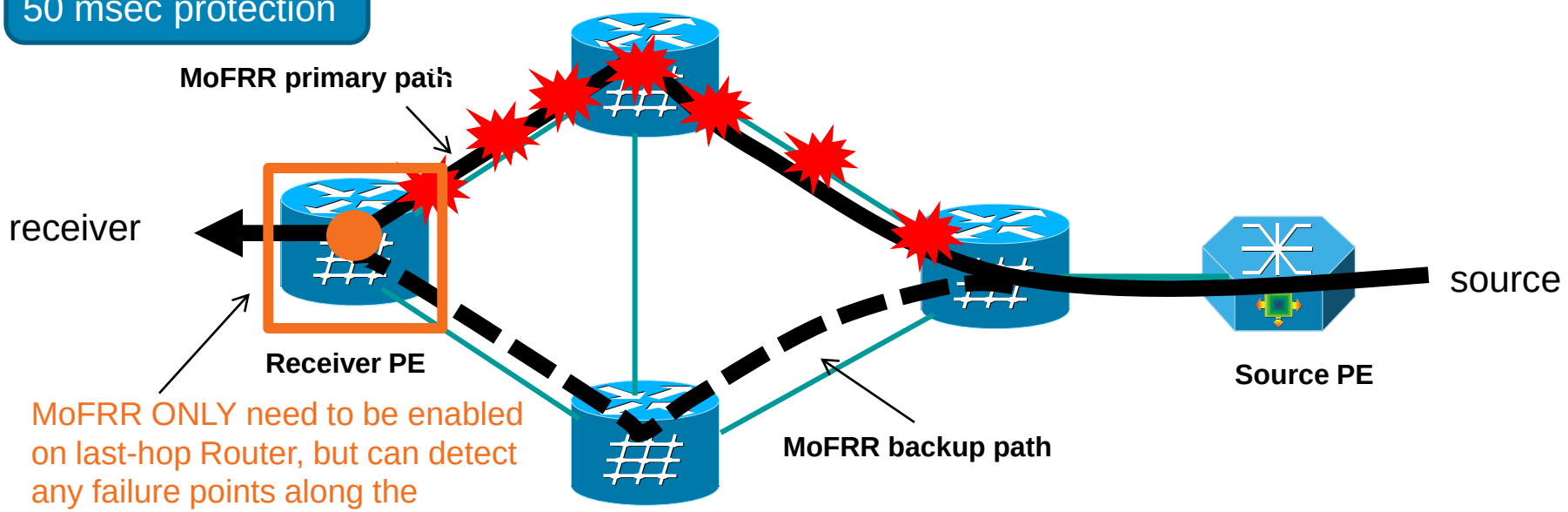
MoFRR – A Simple CLI Example

Configuration on receiver PE

```
ipv4 access-list mofrr-acl-group1
 20 permit ipv4 any 231.1.0.0 0.0.255.255
 30 permit ipv4 any 232.1.0.0 0.0.255.255

router pim
 address-family ipv4
 mofrr mofrr-acl-group1
```

```
RP/0/RSP0/CPU0:receiver#show mrib route 232.1.1.1 1.3.13.100 detail
<snip>

(1.3.13.100,232.1.1.1) Ver: 0x383a RPF nbr: 10.0.102.2 Flags: MoFE MoFS, FMA: 0x50130f98
  Up: 04:55:57
  MOFRR State: Inactive Sequence No 1
  Incoming Interface List
    TenGigE0/0/0/0.1 Flags: A2, Up: 00:03:11  ← MoFRR backup interface
    TenGigE0/0/0/1.1 Flags: A, Up: 04:00:09  ← MoFRR primary interface
  Outgoing Interface List
    TenGigE0/0/0/7.13 Flags: F NS LI, Up: 04:55:57

 #show mfib hardware route mofrr location <loc>
```

End-to-End path
50 msec protection



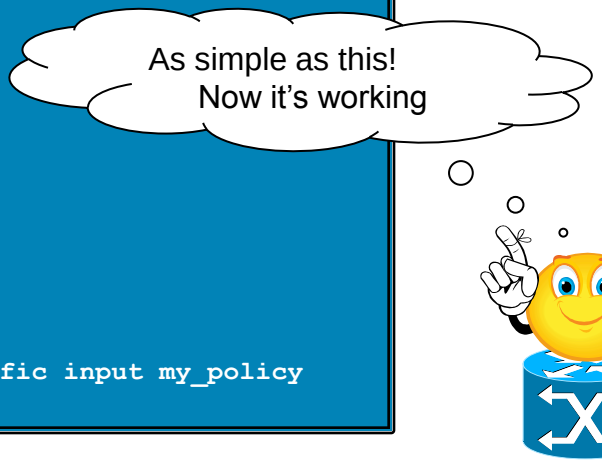
MoFRR ONLY need to be enabled on last-hop Router, but can detect any failure points along the forwarding path

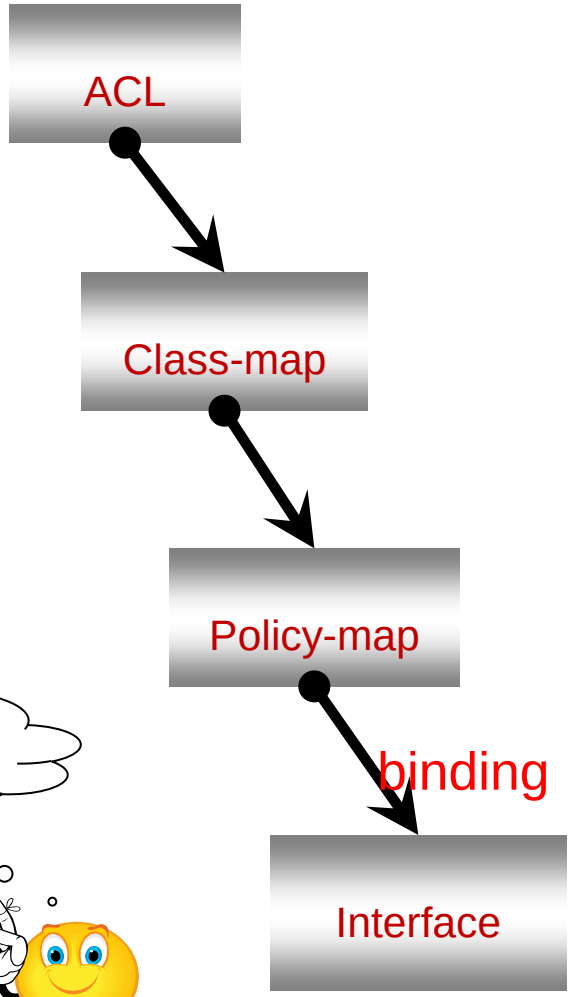
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

What's "Inline" Video Monitoring?

Monitor the Video flow quality on the transport port. It's just another feature like QoS/ACL which you apply to the normal interface

```
ipv4 access-list my_acl:
 10 permit udp any eq 4000 any eq 4000
!
class-map type traffic match-any my_class
 match access-group ipv4 my_acl
 end-class-map
!
policy-map type performance-traffic my_policy
 class type traffic my_class
  monitor parameters
   interval duration 10
   history 10
   timeout 2
  !
  monitor metric ip-cbr
   rate layer3 packet 100 pps
  !
  react 1 packet-rate
   threshold type immediate
   threshold value gt 200.0
   action syslog
   alarm severity warning
   alarm type discrete
 !
 end-policy-map
!
interface GigabitEthernet0/0/0/8
 ipv4 address 10.1.1.1 255.255.255.0
 service-policy type performance-traffic input my_policy
!
```





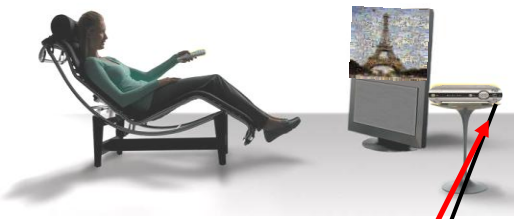
© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Video Trap and Clone

For Deep Analysis

CAPEX: No need for DPI probe deployment all over network

OPEX: Leverage the shared, "expert" analysis equipment in the NOC



Video Ops

Video Flow trapped and cloned

Regional Network

Regional Headend

Backbone

ASR9000

ASR9000

Hub

ASR9000

CRS-1

CRS-1

- In-line monitoring Trap or Customer calls in and opens trouble ticket
- Service personnel trap the Video flow and clone it
- Video Flow transported to Video Operations Site for detailed codec level Analysis / Troubleshooting

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

SW/HW Architecture Overview

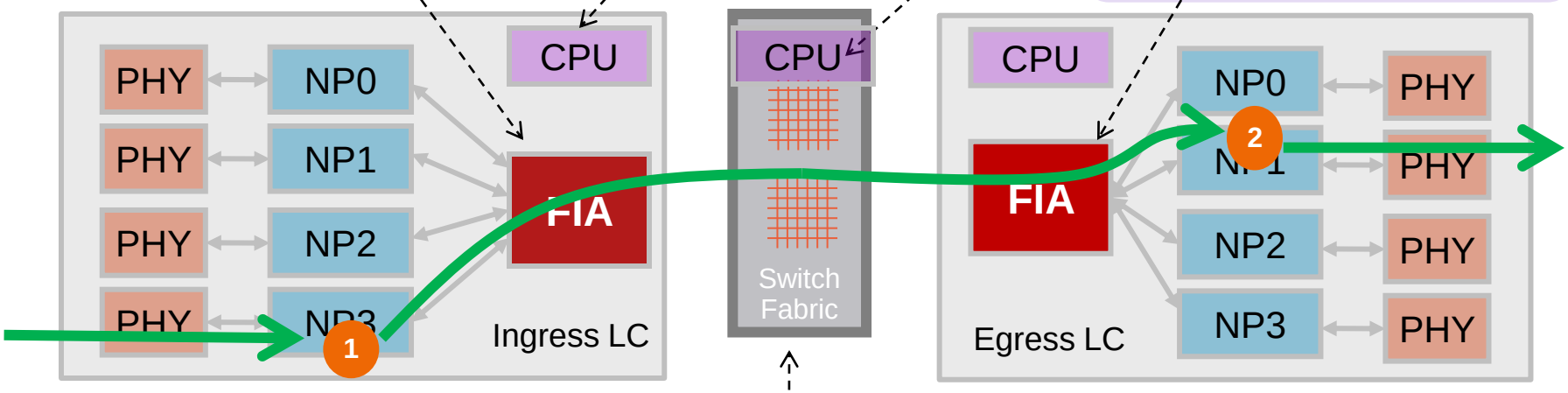
Where to Implement the Features

Features on RP CPU: Routing, MPLS, Multicast control plane, HSRP, VRRP, link bundle, etc

Features on LC CPU: BFD, OAM, STP, Netflow, ARP, ICMP, etc

Ingress Features on FIA: System QoS - VoQ

Egress Features on FIA: System QoS queuing, Multicast replication



Ingress NP look up →
Get egress NP information for packet forwarding

Apply ingress port features: ACL, QoS, uRPF, SPAN, Vidmon, etc

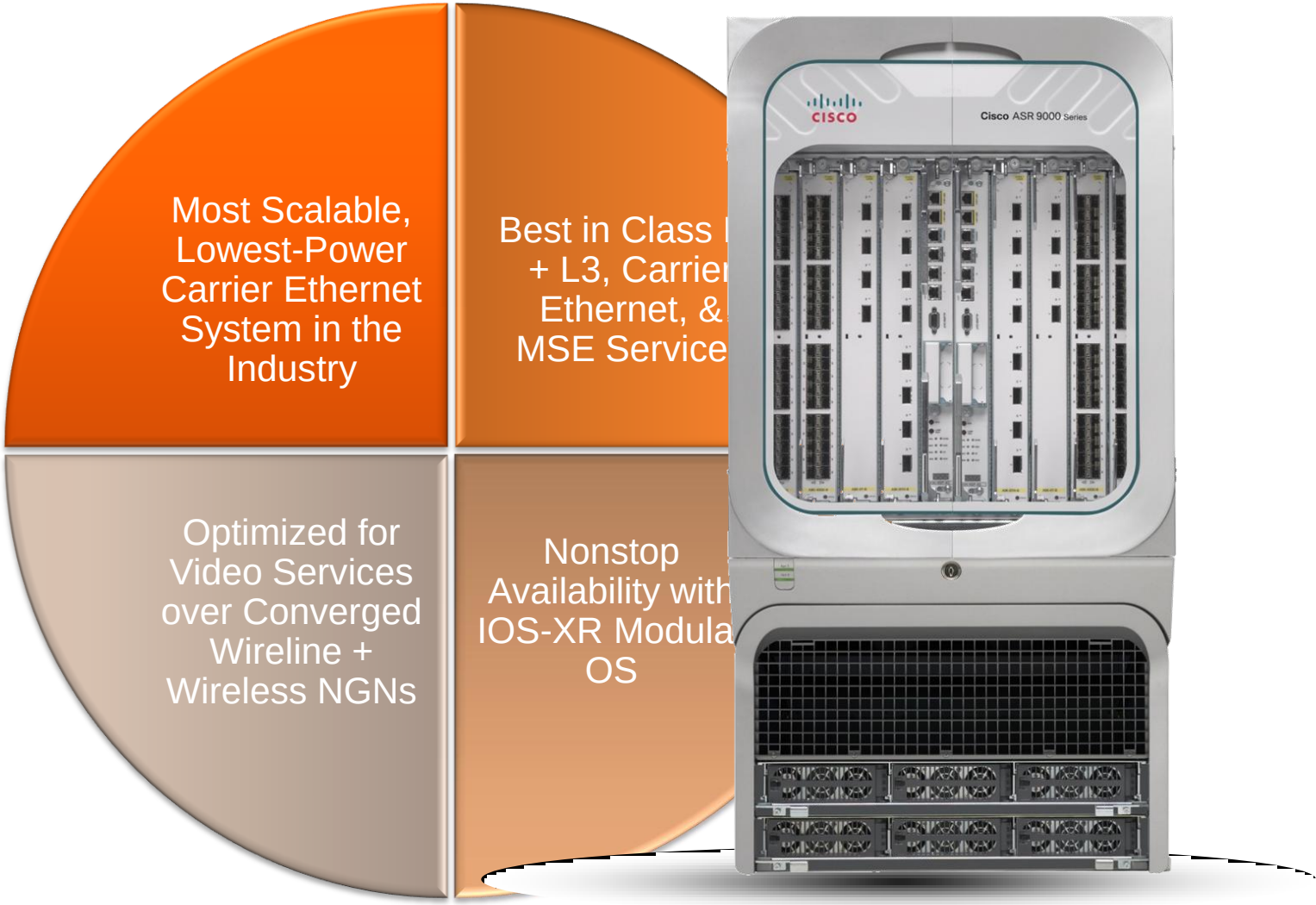
Switch Fabric:
Multicast replication
QoS scheduling and Fabric access Arbitration

Egress NP look up →
Get egress logical port, VLAN, MAC, ADJ information, etc for packet rewrite

Apply egress port features: ACL, QoS, Multicast replication, etc

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Cisco ASR 9000 Summary



© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Complete Your Session Evaluation

- Please give us your feedback!!

  Complete the evaluation form you were given when you entered the room
- This is session 5.2

Don't forget to complete the overall event evaluation form included in your registration kit

YOUR FEEDBACK IS VERY IMPORTANT FOR US!!! THANKS

© 2010 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

© 2010 Cisco and/or its affiliates. All rights reserved.
Cisco Confidential