
cisco
You make possible

ASR 1006-X and ASR1009-X Overview and Architecture

David Roten

Session ID : BRKARC-2013

Barcelona | January 27-31, 2020

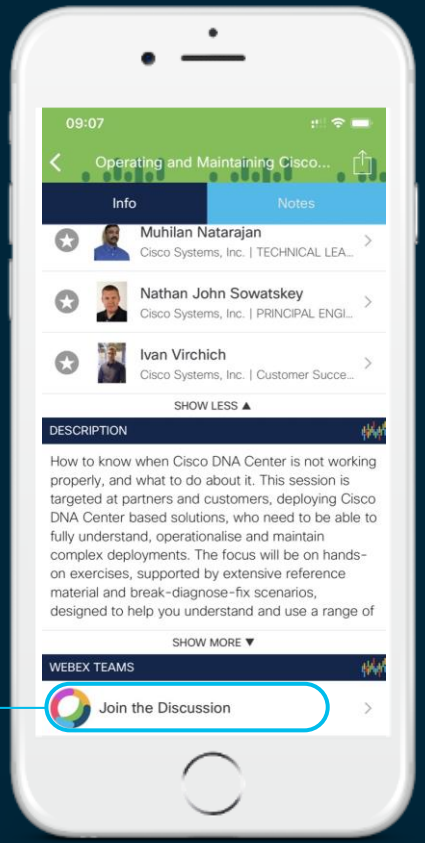
Cisco Webex Teams

Questions?

Use Cisco Webex Teams to chat with the speaker after the session

How

1. Find this session in the Cisco Events Mobile App
2. Click “Join the Discussion”
3. Install Webex Teams or go directly to the team space
4. Enter messages/questions in the team space

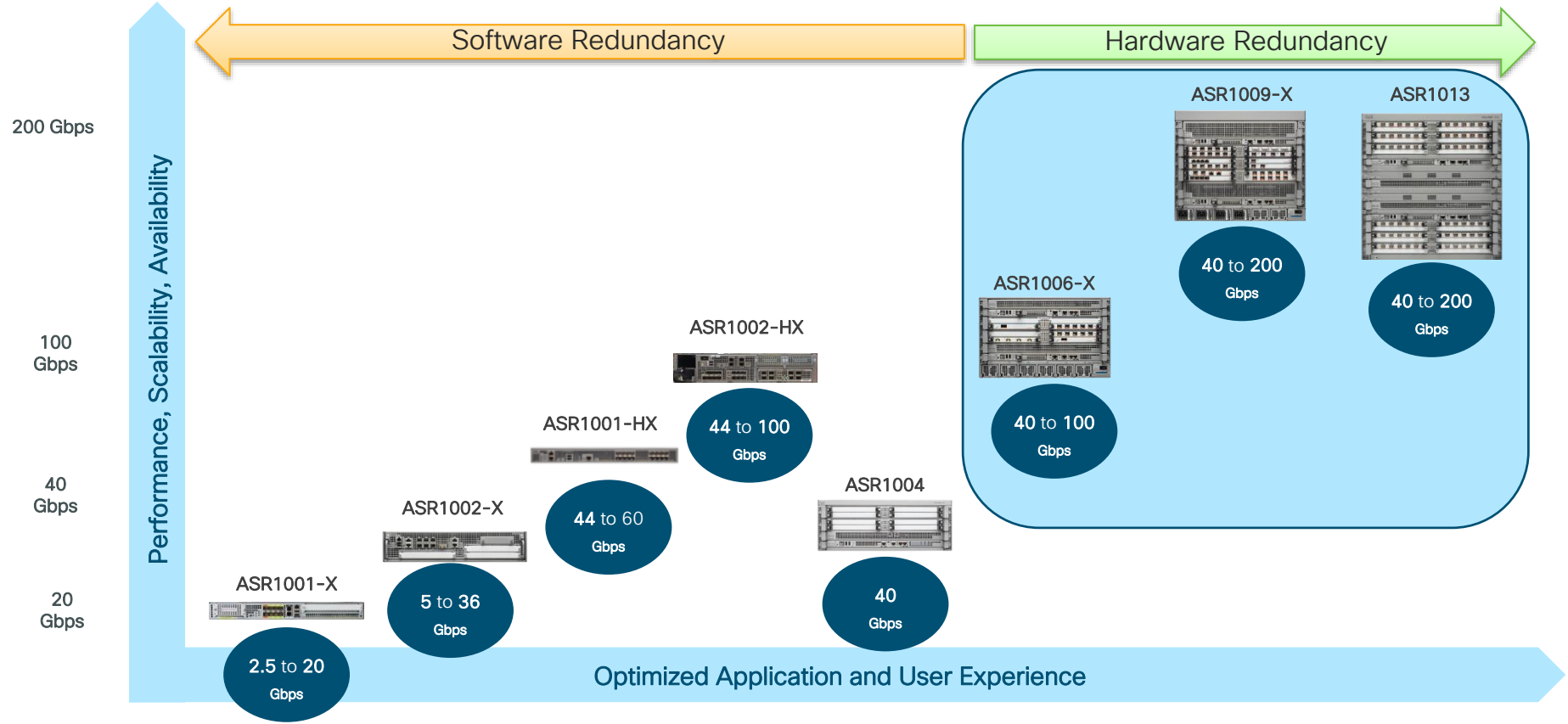


Agenda

- Introduction
- Chassis Overview
- Control Plane Hardware
- Data Plane Hardware
- New 3rd Generation QFP ASIC Hardware
- Linecards for Input / Output
- Redundancy
- Monitoring
- Key Takeaways
- Q & A

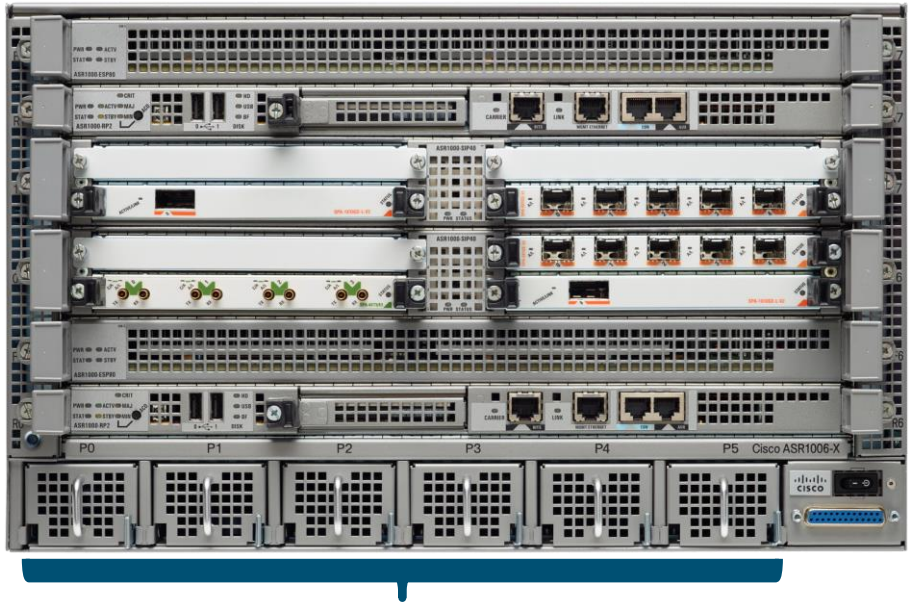
ASR1000 Platform Introduction

Cisco ASR1000 Series Routers



Hardware Overview

Chassis ASR1006-X



ESP slots
(ESP40 / 100 / 100X / 200X)

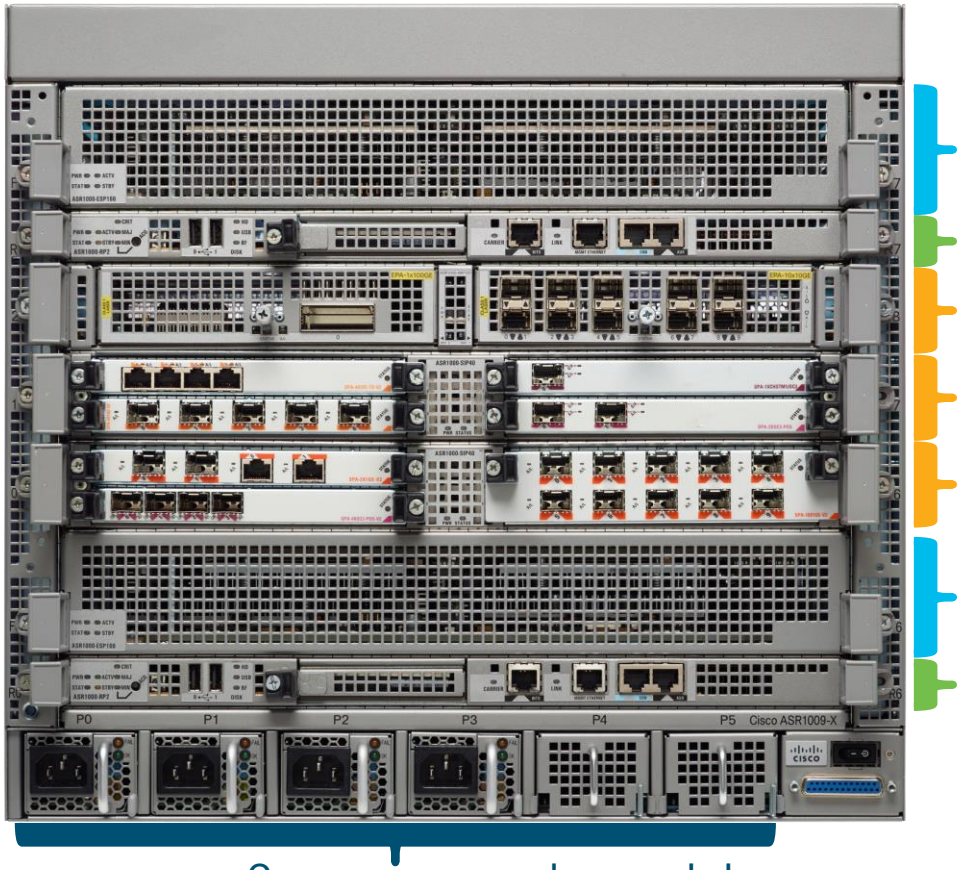
RP slots

Linecard slots
(SIP40 / MIP100 / Fixed Ethernet)

6 power supply modules

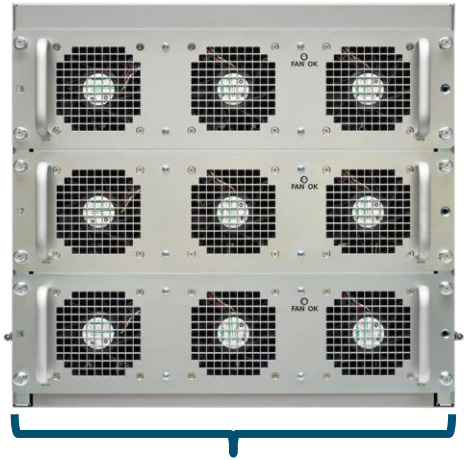
2 fan trays

Chassis ASR1009-X



6 power supply modules

ESP slots
(ESP40 / 100 / 100X / 200 / 200X)
RP slots
Linecard slots
(SIP40 / MIP100 / Fixed Ethernet)



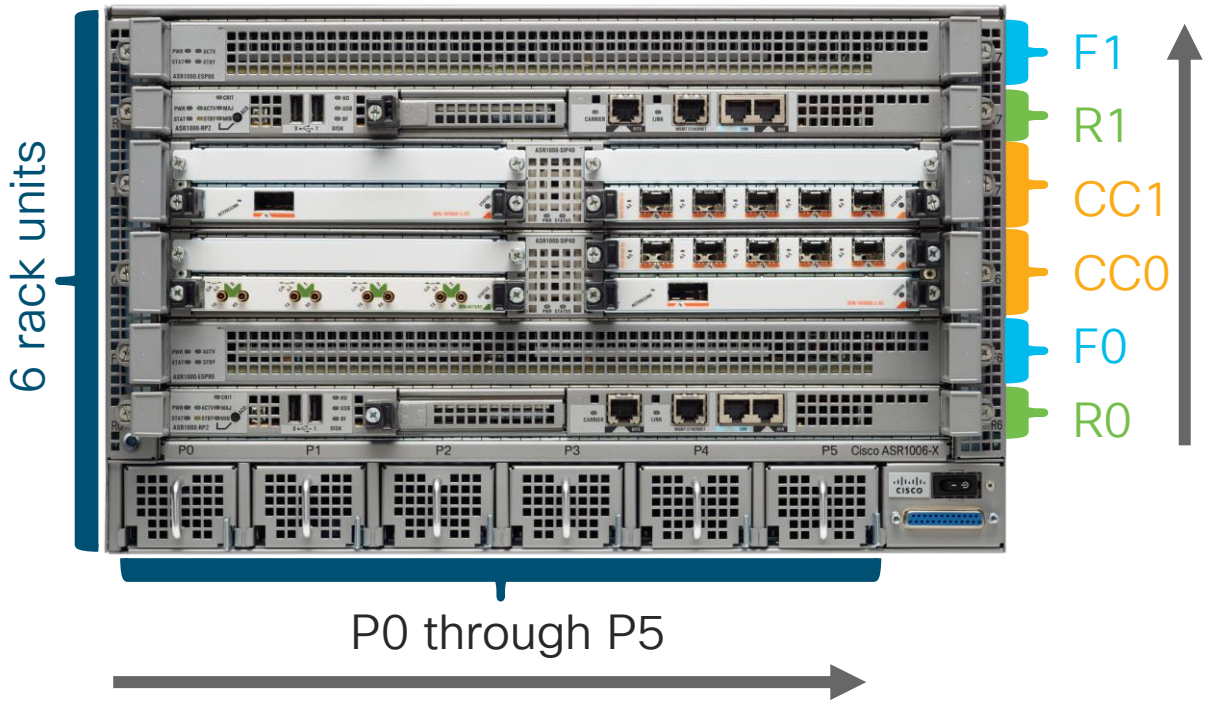
3 fan trays

ASR1000-X chassis linecard slots

- Each of the linecards slots are “superslots” supporting up to 100 Gb/sec full duplex traffic with currently available hardware
- Chassis is capable of delivering 200 Gb/sec full duplex per slot when paired with potential future ESP hardware supporting that functionality
- Additional support for linecards running at 40 Gb/sec full duplex
  - SIP-40
  - ASR1000-2T+20X1GE
  - ASR1000-6TGE

ASR1000 numbering conventions

- ASR1000 part numbers indicate the number of rack units
- Items are numbered from bottom to top, left to right



Chassis comparison

Legacy chassis

1004

RP2
RP3

ELC
SIP40 & SPAs
~~MIP100~~ & ~~EPAs~~

1006

ESP40
ESP100* & ~~ESP200~~

~~200Gbps slot ready~~

~~ESP100X~~
~~ESP200X~~

* Supported in ASR1006 only, no ASR1004 support.

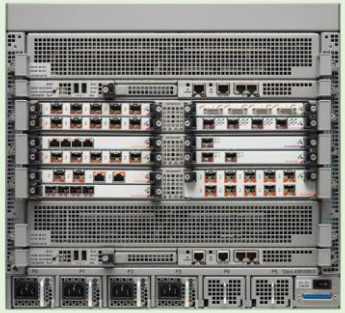
ASR1000X chassis

1006-X

RP2
RP3

ELC
SIP40 & SPAs
MIP100 & EPAs

1009-X



ESP40
ESP100* & ESP200

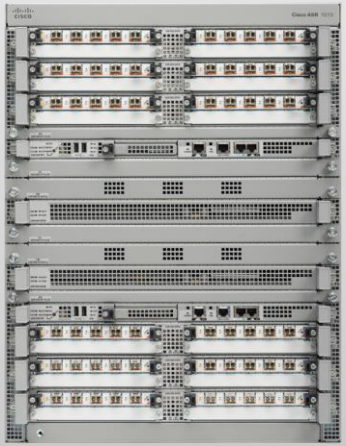
200Gbps slot ready

ESP100X
ESP200X

Chassis comparison

Legacy chassis

1013



RP2
RP3

ELC
SIP40 & SPAs
MIP100 & EPAs

ESP40
ESP100* & ESP200

~~200Gbps slot ready~~

ESP100X
~~ESP200X~~

* Slots 2 & 3 support 100Gbps backplane access, slots 0,1,4, and 5 are limited to 40 Gbps backplane access

ASR1000X chassis

1006-X

RP2
RP3

ELC
SIP40 & SPAs
MIP100 & EPAs

1009-X

ESP40
ESP100* & ESP200

200Gbps slot ready

ESP100X
ESP200X

ASR 1006 / ASR1009-X Power Supply

- AC and DC power supply modules
- Fault tolerance
  - Detects short circuits and component failures within the PS, if a failure is found, the unit is shut down
- High efficiency
  - More than 85% efficient to reduce power waste even at low loads
- Hot-swappable
- Always load sharing and redundant (N+1, N+M)
  - software configurable for redundancy level for operation
- Minimum power number of power supplies is 2 for both ASR1006-X and ASR1009-X
- Maximum of 6 power supplies
  - Provides both chassis level and facility level power fault tolerance.

ASR 1006 / ASR1009-X Power Supplies

```
Router# show platform power
Chassis type: ASR1009-X
Slot      Type                State                Allocation (W)
--------- ------------------- -------------- ---------
0         ASR1000-SIP40       ok                    64
0/0       SPA-5X1GE-V2        ok                    18
1         ASR1000-SIP40       ok                    64
1/0       SPA-8X1GE-V2        ok                    20
1/3       SPA-4XOC3-POS       ok                    14
2         ASR1000-SIP40       ok                    64
R0        ASR1000-RP2         ok, active           105
F0        ASR1000-ESP100      ok, standby          310
F1        ASR1000-ESP100      ok, active           350
P6        ASR1000X-FAN        ok                   125
P7        ASR1000X-FAN        ok                   125
P8        ASR1000X-FAN        ok                   125

Slot      Type                State    Capacity (W) Load (W)
--------- ------------------- -------- ------------ --------
P0        ASR1000X-AC-1100W   ok       1100         228
P1        ASR1000X-AC-1100W   ok       1100         216
P3        ASR1000X-AC-1100W   ok       1100         204

Total load: 648 W, total capacity: 3300 W. Load / Capacity is 19%
Power capacity:       3300 W
Redundant allocation:    0 W
PS/Fan allocation:     375 W
FRU allocation:       1009 W
-----------------------------------------
Excessive Power in Reserve:   1916 W
Excessive / (Capacity - Redundant) is 58%
Power Redundancy Mode: none
Power Allocation Status: Sufficient
```

```
Router# show platform power
Chassis type: ASR1006-X
Slot      Type                State                Allocation (W)
--------- ------------------- -------------- ---------
1         ASR1000-SIP40       ok                    64
R0        ASR1000-RP2         ok, active           105
F0        ASR1000-ESP100      ok, active           350
P6        ASR1000X-FAN        ok                   125
P7        ASR1000X-FAN        ok                   125
Slot      Type                State    Capacity (W) Load (W)
--------- ------------------- -------- ------------ --------
P0        ASR1000X-AC-1100W   ok       1100         132
P1        ASR1000X-AC-1100W   ok       1100         144
P2        ASR1000X-AC-1100W   ok       1100         144

Total load: 420 W, total capacity: 3300 W. Load/Capacity is 12%
Power capacity:       3300 W
Redundant allocation: 1100 W
PS/Fan allocation:     250 W
FRU allocation:        519 W
-----------------------------------------
Excessive Power in Reserve:   1431 W
Excessive / (Capacity - Redundant) is 65%
Power Redundancy Mode: nplus1
Power Allocation Status: Sufficient
```

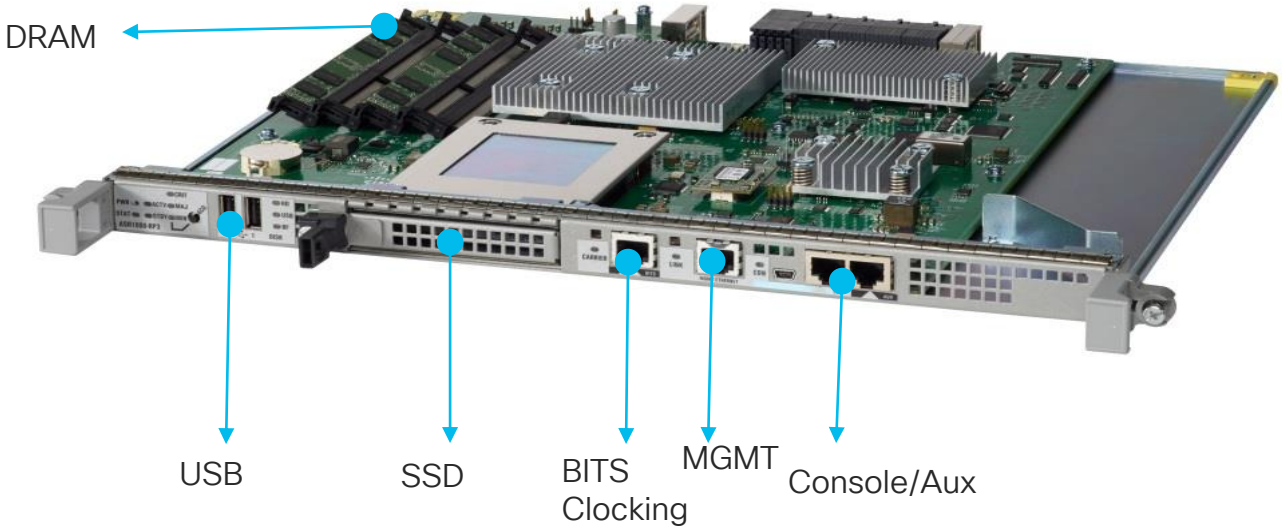
```
platform power redundancy-mode nplus1
```

Control Plane

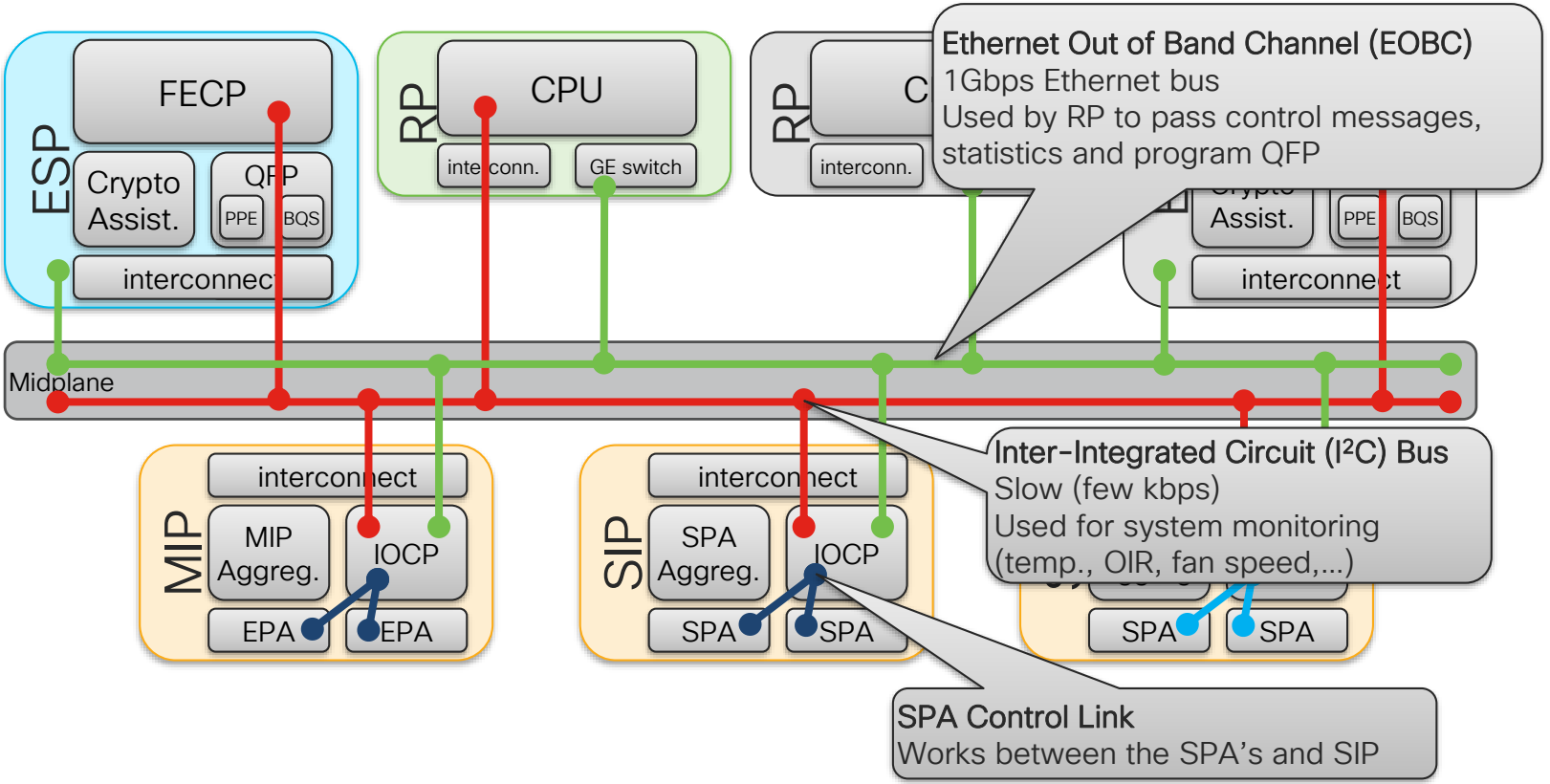
ASR1000 Route Processors

	RP2	RP3
CPU	Intel Dual-core Wolfdale 2.66GHz	Intel Quad-core Broadwell 2.2GHz
Memory	8, 16GB	8, 16, 32, 64 GB
Built-in Boot flash	2GB	8GB
Storage	80GB HDD External USB	100 – 400 GB SSD External USB
Chassis Support	ASR1004 ASR1006 ASR1006-X ASR1009-X ASR1013	ASR1006-X ASR1009-X ASR1013

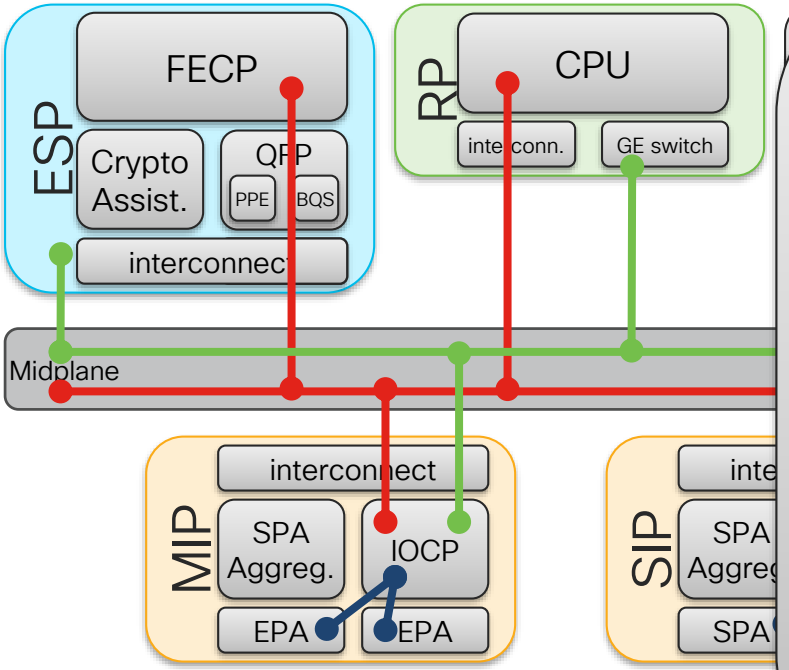
ASR1000 RP3



ASR1000 control plane architecture



ASR1000 control plane architecture



Ethernet out-of-band channel (EOBC)

- indication if cards are installed and ready loading images, stats collection
- messages to program QFP

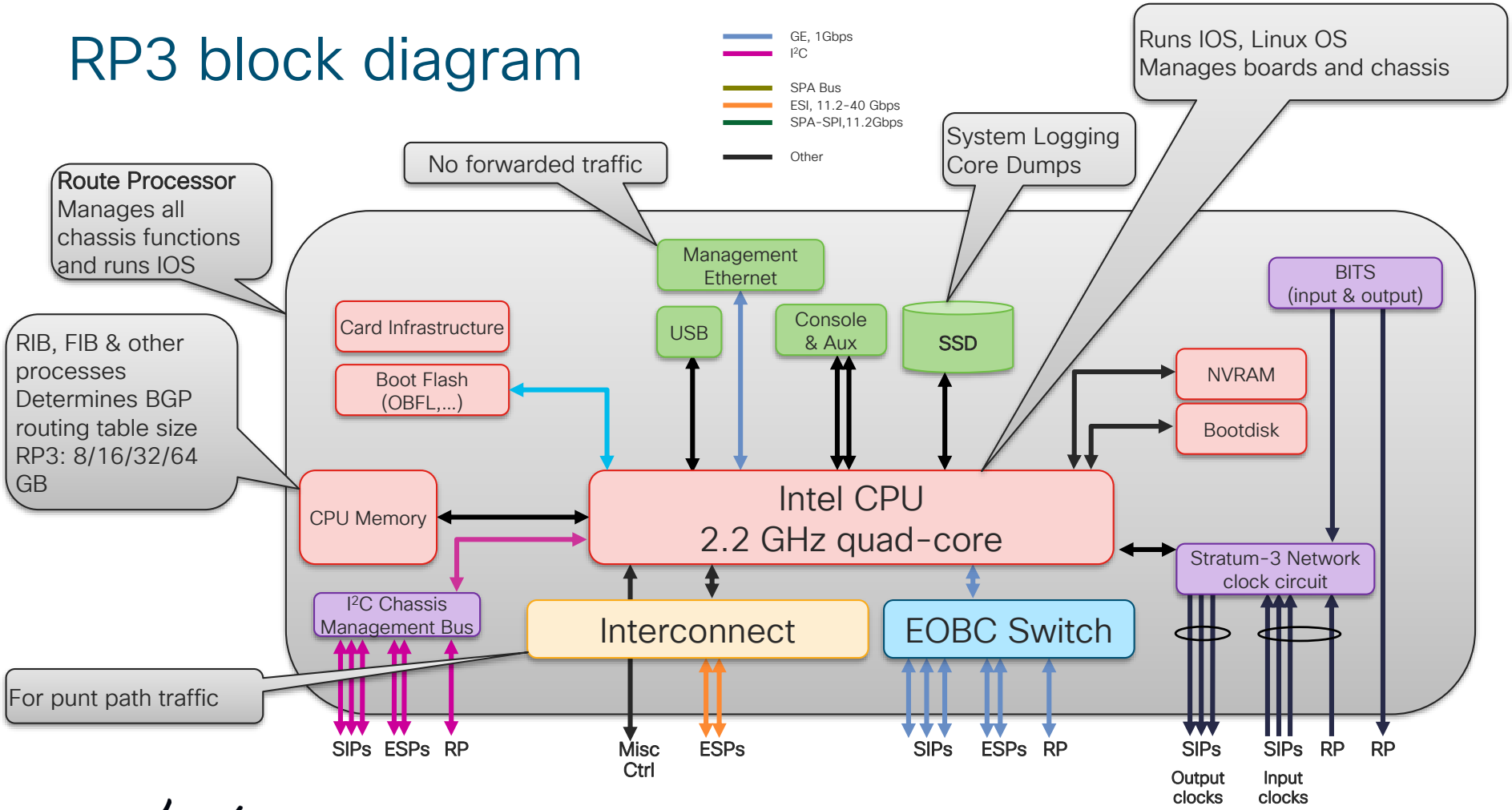
Inter-Integrated Circuit (I²C)

- monitor health of hardware components
- control resets
- communicate active/standby
- real time presence and ready indicators
- control the other RP (reset, power-down,etc.)
- report power-supply status
- EEPROM access

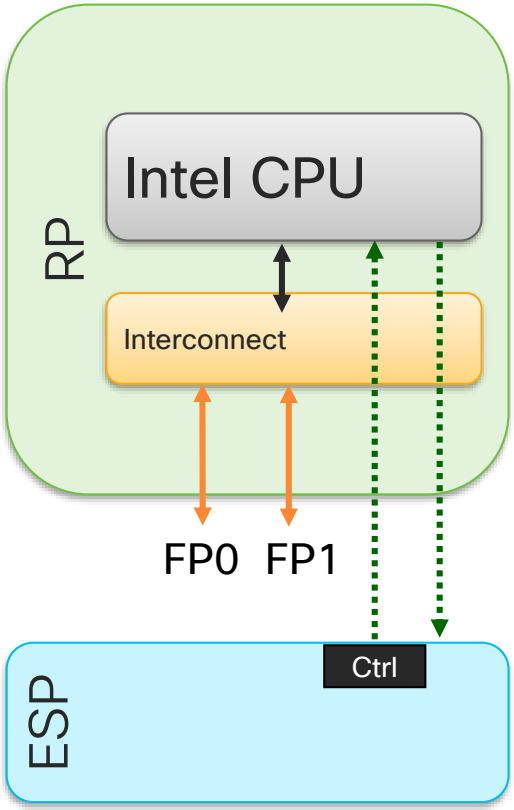
SPA control links

- detect SPA OIR
- reset SPAs (via I²C)
- power-control SPAs (via I²C)
- read EEPROMs

RP3 block diagram



Control Packet Flow: Punt to RP



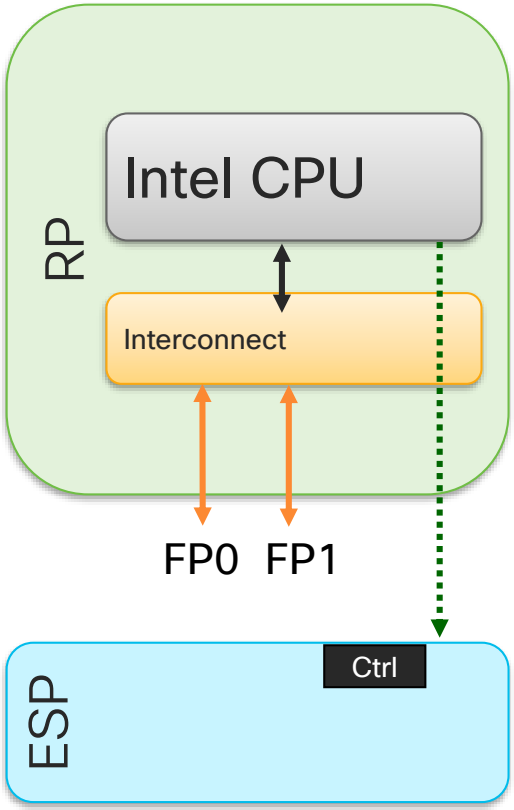

```
ASR1000#show platform software infrastructure punt

LSMPI interface internal stats:
enabled=0, disabled=0, throttled=0, unthrottled=0, state is ready
<snip>
IOSXE-RP Punt packet causes:
     8409150 Layer2 control and legacy packets
      142957 ARP request or response packets
      153783 Incomplete adjacency packets
     2159290 For-us data packets
     2927128 RP<->QFP keepalive packets
          13 Glean adjacency packets
           5 RP handled ICMP packets
          50 RP injected For-us data packets
     3284335 For-us control packets
           8 IP subnet or broadcast packet packets

   FOR_US Control IPv4 protcol stats:
        3284335 GRE packets
Packet histogram(500 bytes/bin), avg size in 139, out 121:
 Pak-Size      In-Count        Out-Count
      0+:      17075341         13995189
    500+:             3           204377
   1000+:             1                0
   4000+:          1374                0
```

ESI, 11.2Gbps

Control Packet Flow: Inject by RP

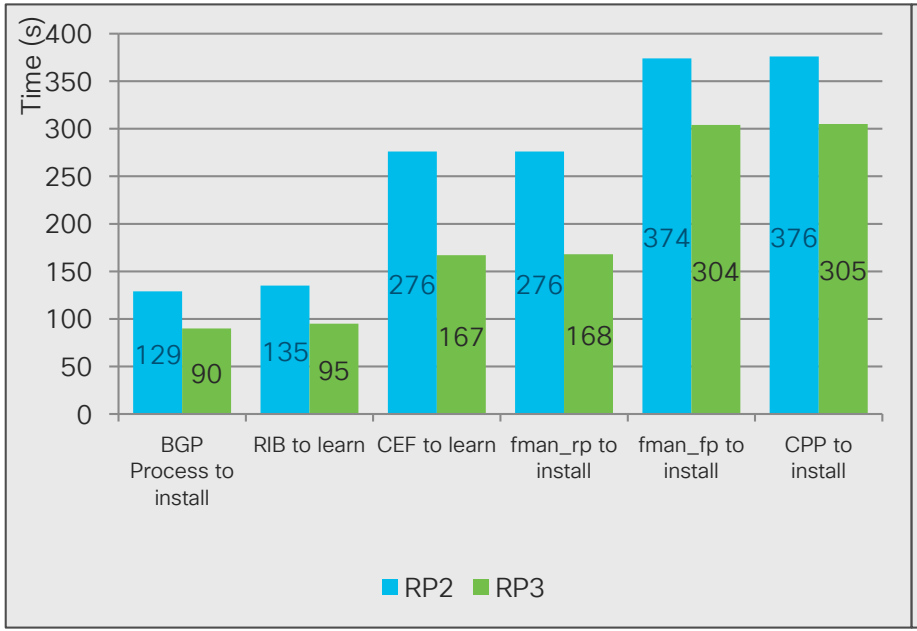



```
Sydney#show plat software infrastructure inject
Statistics for L3 injected packets:
 11342475 total inject pak, 0 failed
 0 sent, 6161717 prerouted
 0 non-CEF capable, 888119 non-unicast
 6393836 IP, 15   IPv6
 0 MPLS, 0 Non-IP Tunnel
  <snip>
 231032 normal, 60 nexthop
 4173263 adjacency, 0 feature
 0 undefined
 1989486 pak find no adj, 0 no adj-id
 1170 sb alloc, 6393791 sb local
 0 p2mcast failed count 0 p2mcast enqueue fail
  <snip>
 per feature packet inject statistics
 0 Feature multicast
 0 Feature Edge Switching Service
  <snip>
Statistics for L2 injected packets:
 0 total L2 inject pak, 0 failed
 0 total BD  inject pak, 0 failed
 0 total EFP inject pak, 0 failed
 0 total VLAN inject pak, 0 failed
```

ESI, 11.2Gbps

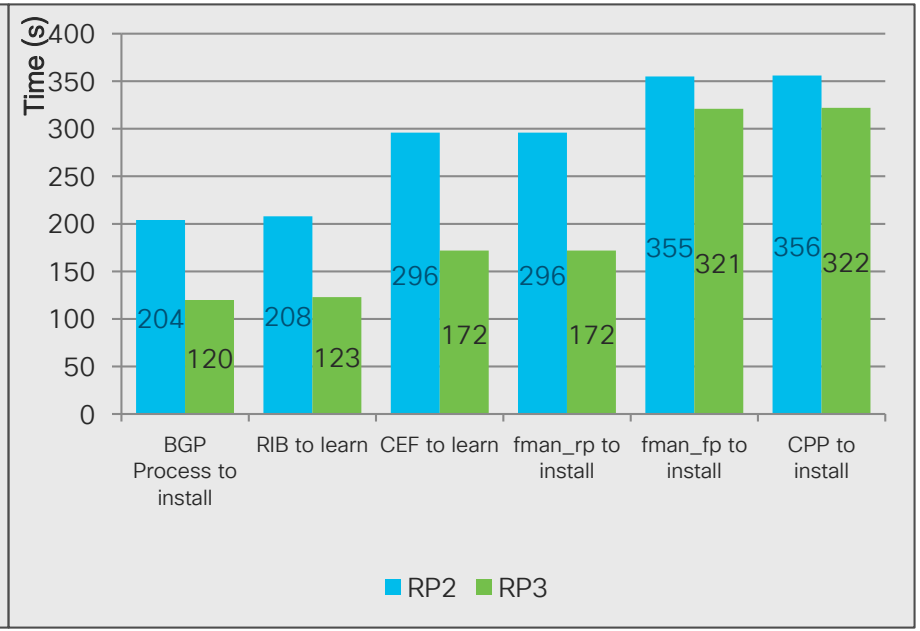
RP2 vs RP3 Performance

BGPv4 Performance



~20-30% better BGPv4 performance

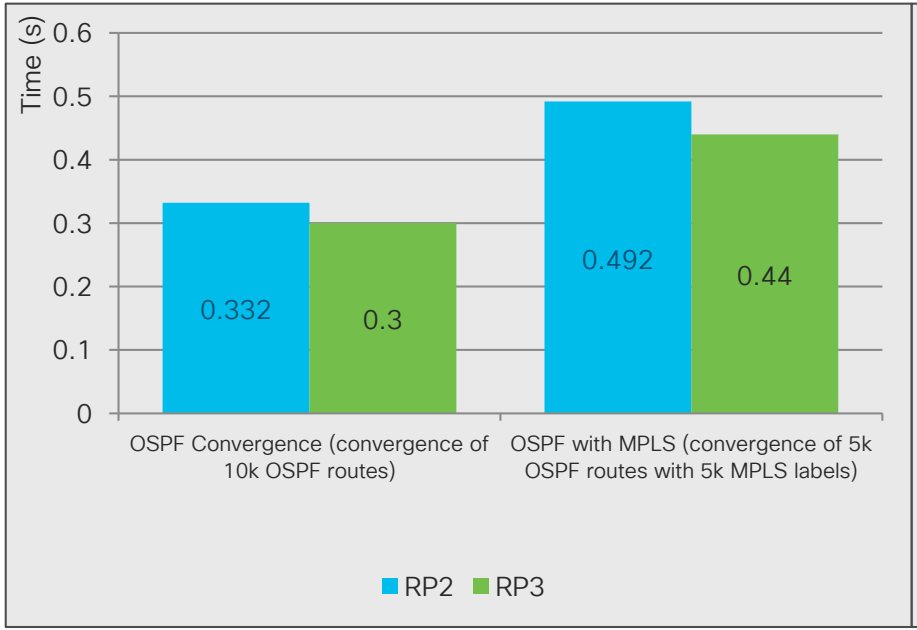
BGPv6 Performance



~40% better BGPv6 performance

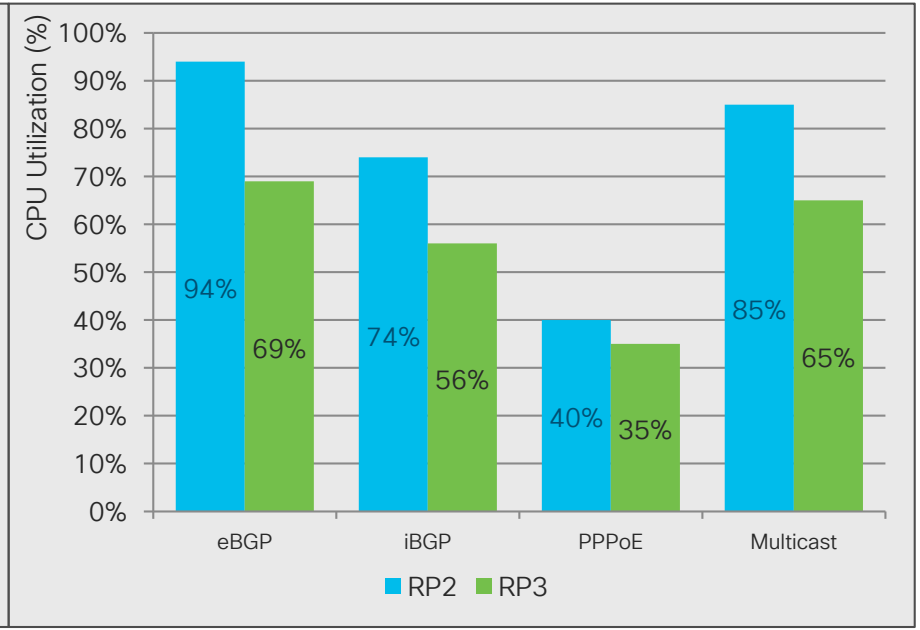
RP2 vs RP3 Performance

Route Convergence Time



~10% faster OSPF convergence

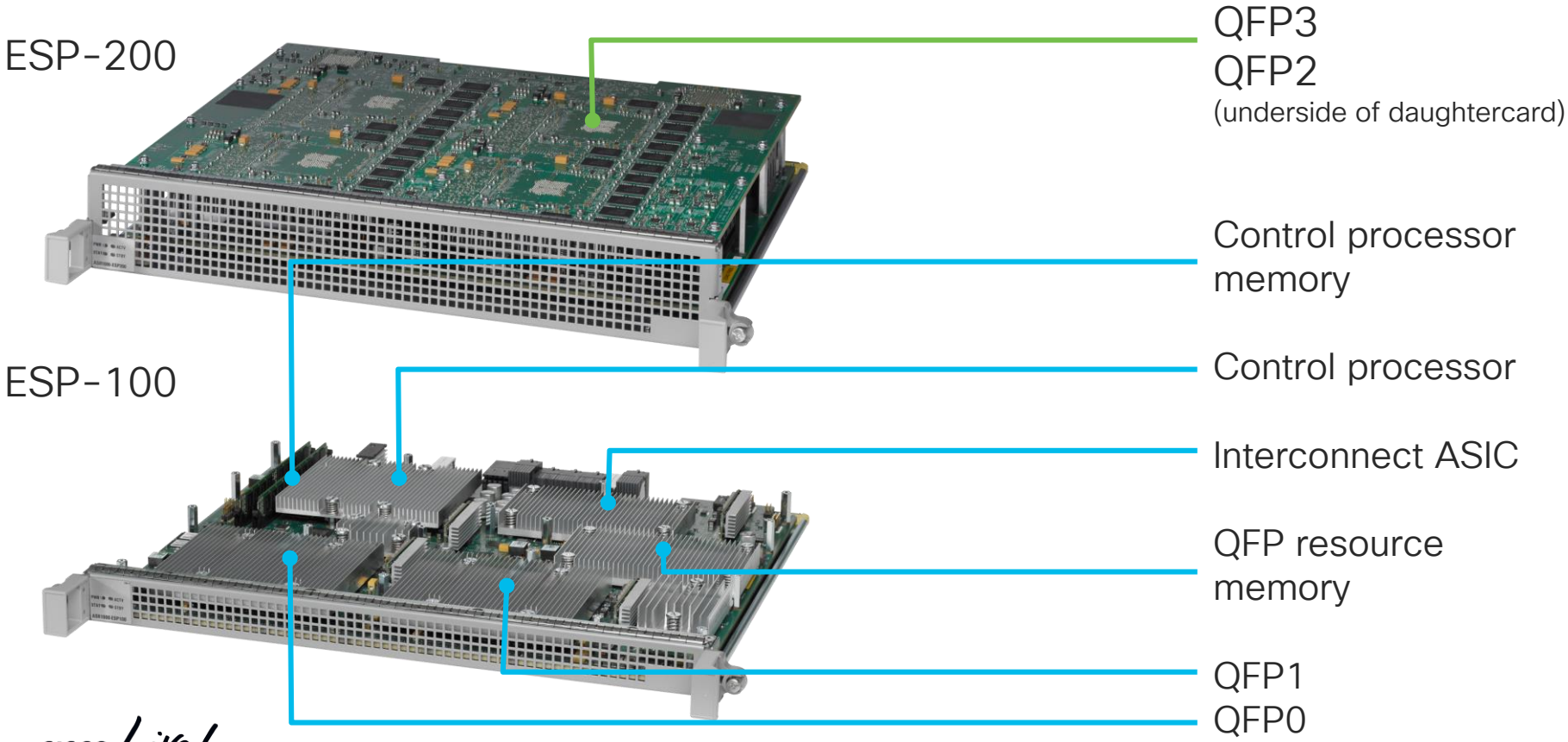
CPU Utilization (%)



Less CPU utilization on RP3

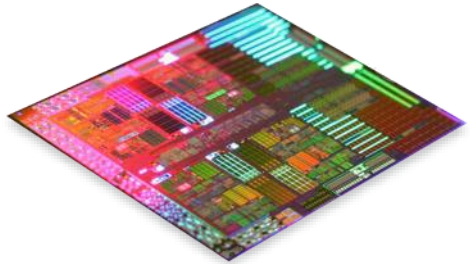
Data Plane Hardware

Frontside view of new ESP100 and ESP200

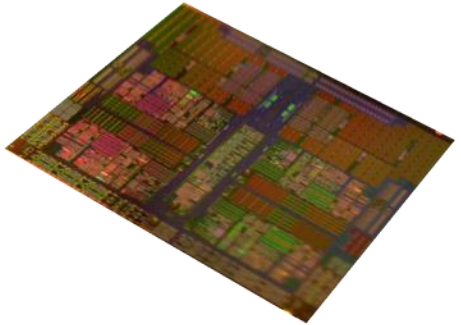


ASR 1000 Series Innovations
Cisco QuantumFlow Processor

- Multiple generations of ASIC hardware
- Massively parallel, 2nd generation ASIC has 64 multi-threaded cores
  - 4 threads per core
  - totaling 256 simultaneous processes available to handle traffic
- High-priority traffic has independent path through hardware
- Packet replication capabilities for Lawful Intercept
- Full visibility of entire L2 frame
- Latency: tens of microseconds with features enabled
- Interfaces on-chip for external cryptographic engine
- Dedicated hardware for traffic management
- Low power per core / thread versus generic chips

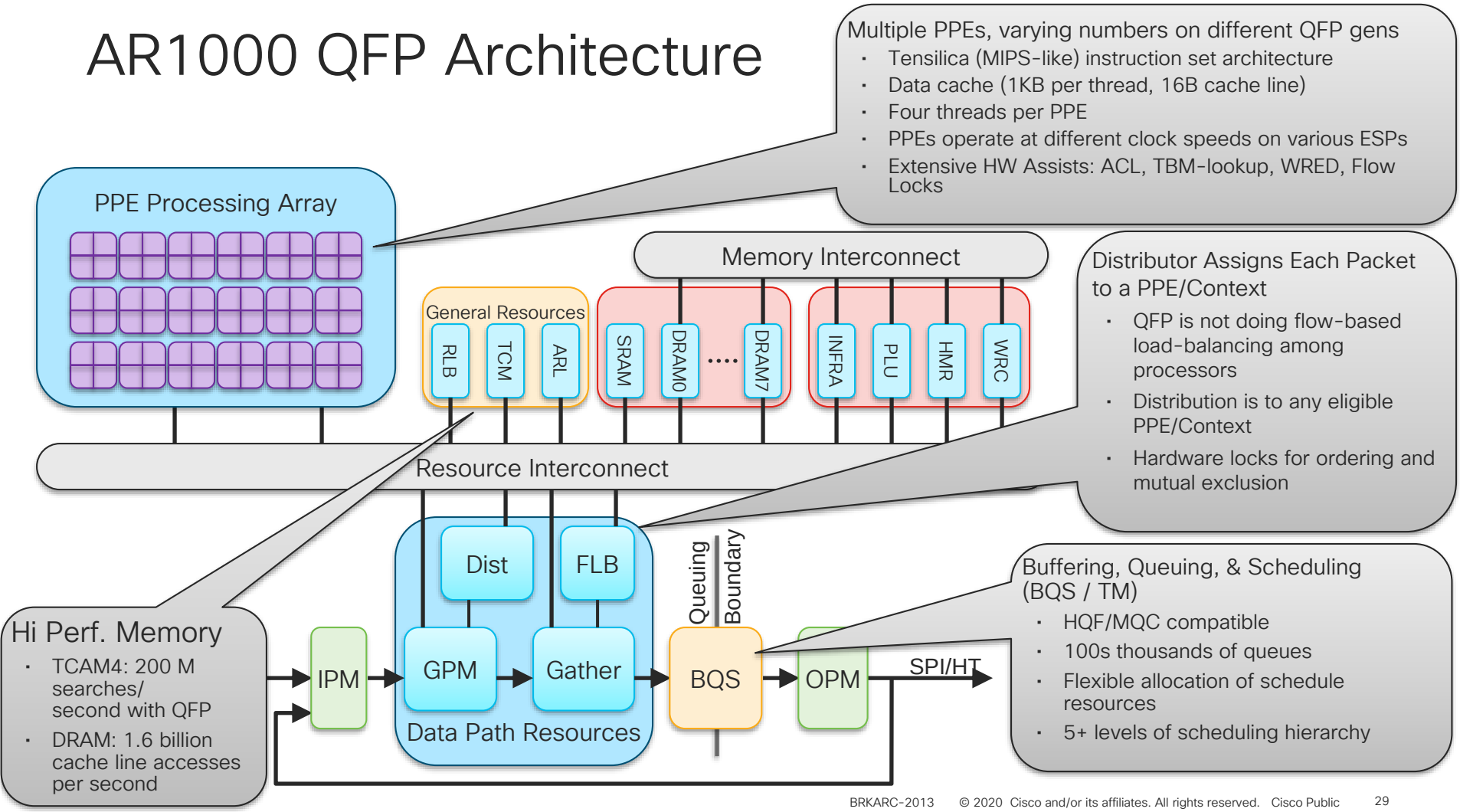


Cisco QuantumFlow Processor
(QFP)



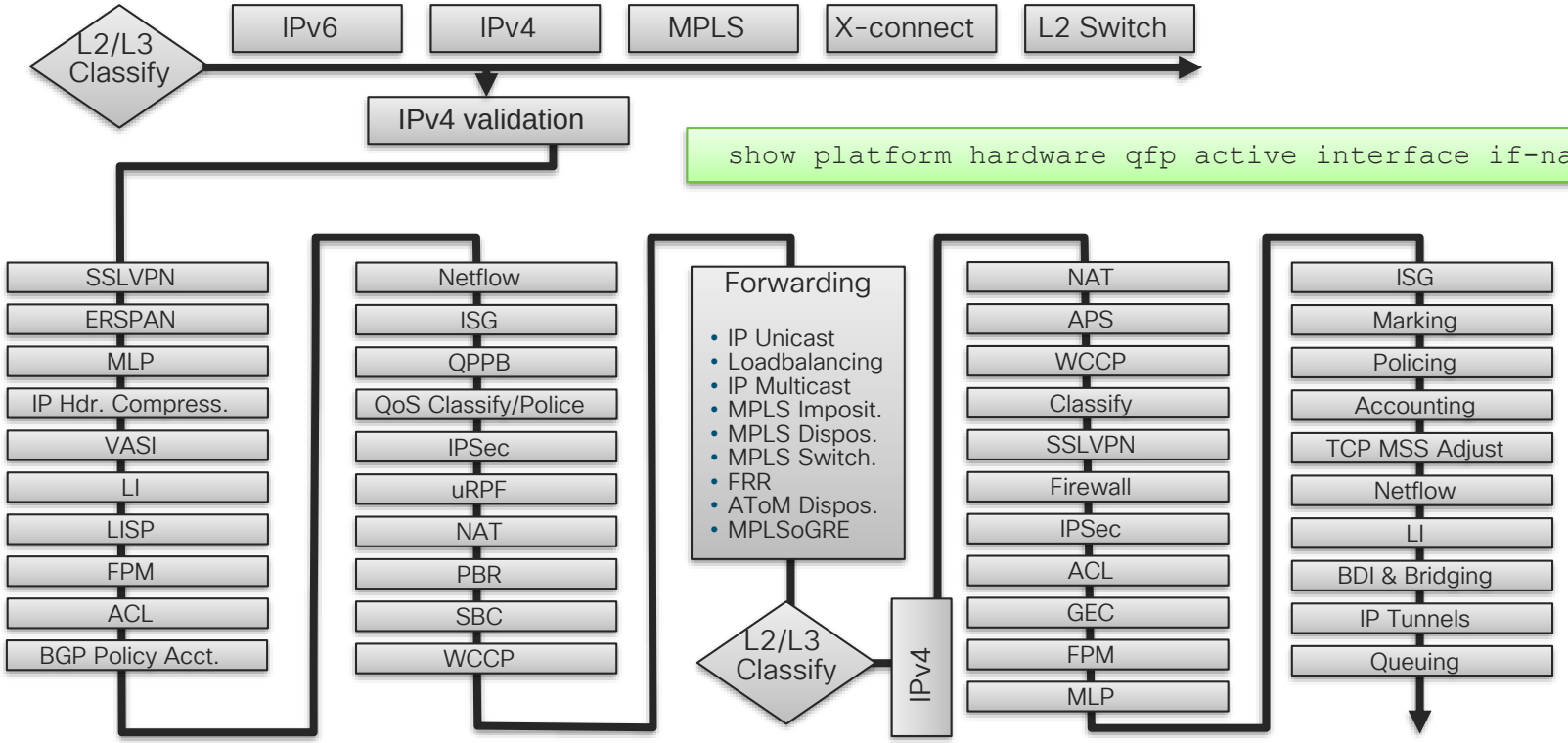
Cisco QuantumFlow Processor
Traffic Manager
(QFP TM)

AR1000 QFP Architecture



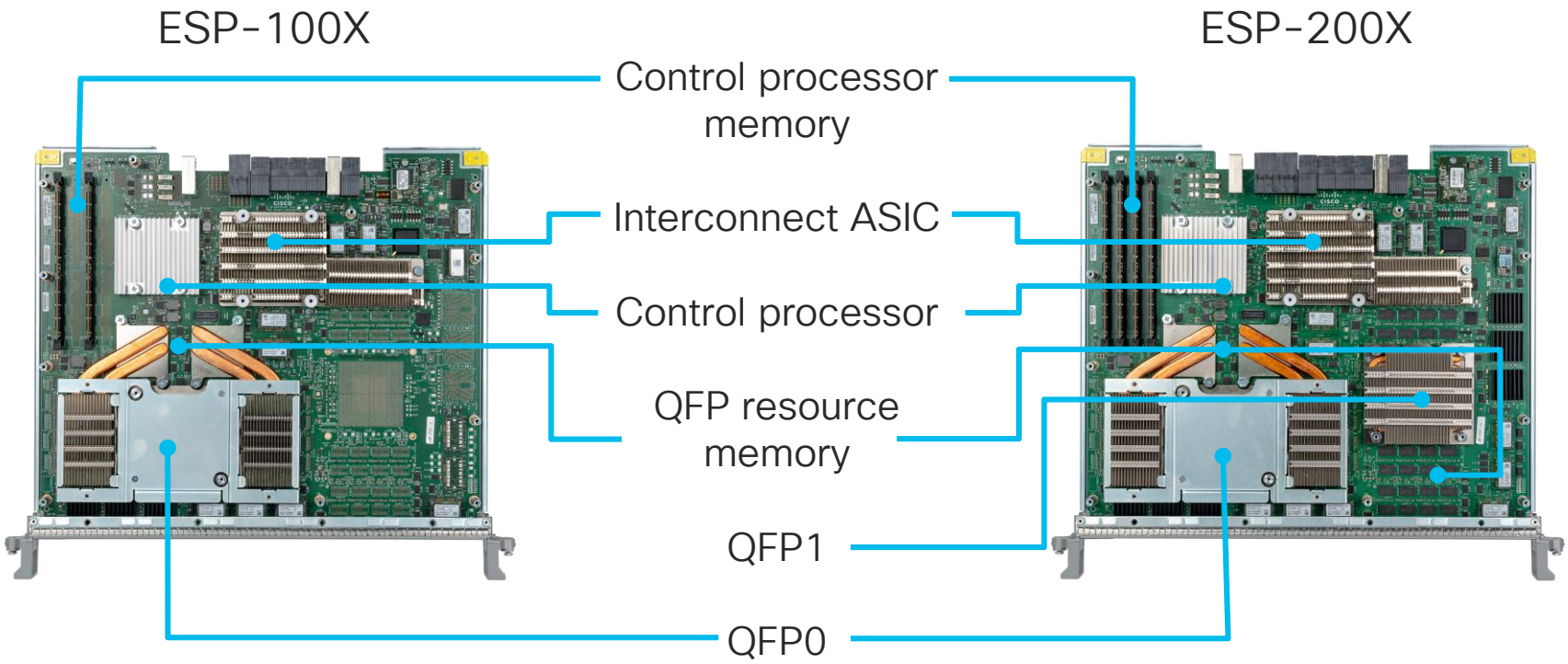
Feature Invocation Array – FIA

Animation



New 3rd Generation QFP ASIC Hardware

Overhead view of new ESP100X and ESP200X



Data plane hardware options

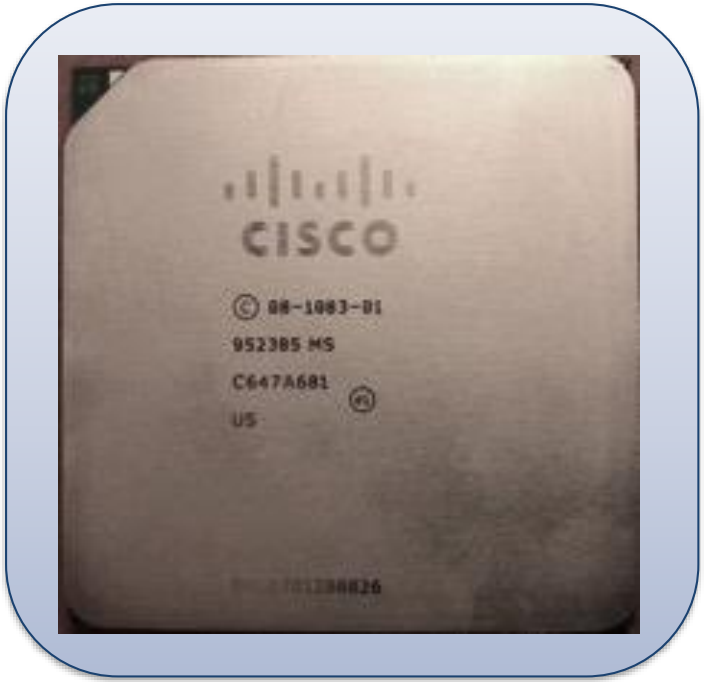
	ESP-40	ESP-100	ESP-100X *	ESP-200	ESP-200X *
Marketed throughput	Up to 40 Gb/s	Up to 100 Gb/s	Up to 100 Gb/s	Up to 200 Gb/s	Up to 200 Gb/s
Raw CEF throughput	Up to 40 Gb/s	Up to 138 Gb/s	Up to 130 Gb/s	Up to 260 Gb/s	Up to 260 Gb/s
Minimum software	IOS XE 3.2	IOS XE 3.7.1	IOS XE 17.2	IOS XE 3.10	IOS XE 17.2
Supported chassis	1004, 1006, 1006-X, 1009-X, 1013	1006, 1006-X, 1009-X, 1013	1006-X 1009-X 1013	1006-X, 1009-X, 1013	1006-X, 1009-X

Data plane hardware options

	ESP-40	ESP-100	ESP-100X	ESP-200	ESP-200X
Packet buffer memory	0.25 GB	0.5 GB	1.5 GB	1 GB	3.0 GB
QFP resource memory	1 GB	2 x 4 G	32 G	4 x 4 G	2 x 32 G
TCAM memory	40 Mb	2 x 80Mb	80 Mb	2 x 80 Mb	2 x 80 Mb
Max ACEv4 entries	100 k	320 k	380 k	320 k	380 k
Max ACEv6 entries	30 k	200 k	200 k	200 k	200 k
Max IPv4 routes	4 M	4 M	4 M	8.5 M	8.5 M
Max NAT44 sessions	2 M	8 M	12 M	8 M	32 M
Max NAT44 (CGN)	4M	12 M	32 M	12 M	52 M
Max firewall sessions	2 M	6 M	16 M	6 M	16 M

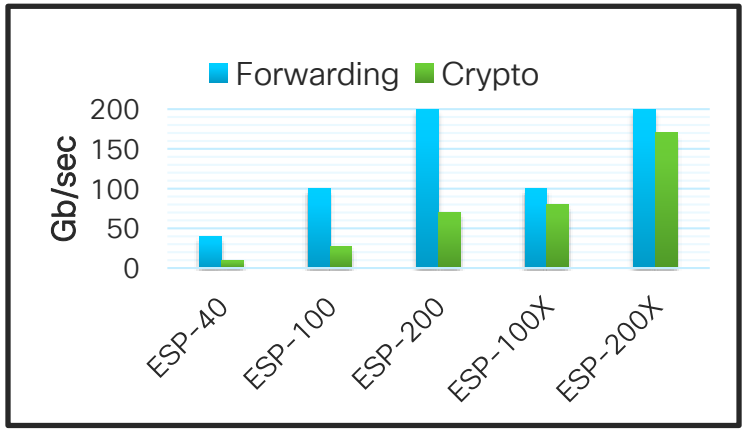
3rd generation QFP ASIC

- Multicore architecture
  - 28 clusters of 8 PPEs @ 1GHz
  - 224 PPEs with 4 threads each = 896 threads
- Specialized Hardware Assist
  - Integrated support for IPSEC and MACSEC cryptography
  - Flow queues for complex stateful features
  - Integrated Ethernet infrastructure
- Memory
  - 32GB DDR4 Resource memory
  - 3 to 4 Billion accesses per second
  - 8 times increase in memory for feature scale
  - 50% more packet buffer – 1.5GB packet buffer



Next Gen Data Plane – ESP-X

- Powered by 3rd generation QFP ASIC
- Complex services support without performance degradation.
- Enhanced services include:
  - AVC
  - IPSEC Crypto
  - MACSEC
  - NAT
  - Firewall
  - QoS
  - AppNav



- IPSEC is now done on ASIC with much higher performance

ESP-200X

ESP-100X

3rd generation QFP ASIC

- Used in the following platforms:
  - 1x 3rd generation QFP ASIC
    - ESP-100X, FP for modular chassis
    - Future platforms
  - 2x 3rd generation QFP ASIC
    - ESP-200X , FP for modular chassis
    - Future platforms

ESP-200X

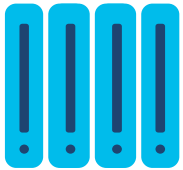
ESP-100X

3rd generation QFP ASIC Technology

- Unified architecture NPU supporting all existing feature code
- Processing Power per packet
  - one 3rd gen QFP ASICs = two 2nd gen QFP ASICs

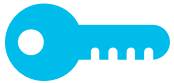
- System on a Chip (SoC)
  - Approximately 2x improvements in Watts/MPPS

- Integrated classification based L2 subsystem
  - Integrated 1/10/40/100 GE MACs
  - Native Support for ASR1000 / ISR4000 modular IO
  - Support for WAN MACSEC



- Virtual single NPU complex via mesh interconnect
  - Support for single use as well as 2x and 4x ASIC meshes

3rd generation QFP ASIC Technology

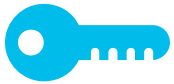


- Embedded security / crypto
  - Crypto engines embedded in the ASIC, with dedicated cores for IPSEC
- Acceleration infra for complex features and flow handling
  - 50% more instructions available per packet for feature processing
  - High performance DDR4 resource memory subsystem (16 channels)
- Improved Classification and QoS
  - MQC compliant hardware traffic management with 256K queues
  - Ether-channel support (200G max single bundle bandwidth)
  - Distributed traffic management with improved granularity

3rd generation QFP ASIC – L2 Subsystem

- Integrated support for 240G of aggregate ethernet ports (2x120G)
- Supports full per port, L2/L3, TCAM based classification
- Supports Rx sub-intf / MAC classification & accounting
  - 4K Rx VLAN
  - 8K HSRP-DA
  - 512 Ethertype
- Supports Tx sub-intf / MAC accounting
- Line-rate WAN MACSEC for 240G of Ethernet interfaces
- MACsec XPN extension for high speed interfaces

3rd generation QFP ASIC – Crypto



- ASIC targets much higher throughput of combined Cipher + Digest @ IMIX packet sizes

- 16 crypto engines
- Each Crypto Engine contains
  - packet input buffer, cipher, digest, and checksum engines
  - Each cipher and digest engine has all the logic necessary to execute any of the supported underlying cipher algorithms

QFP Comparison

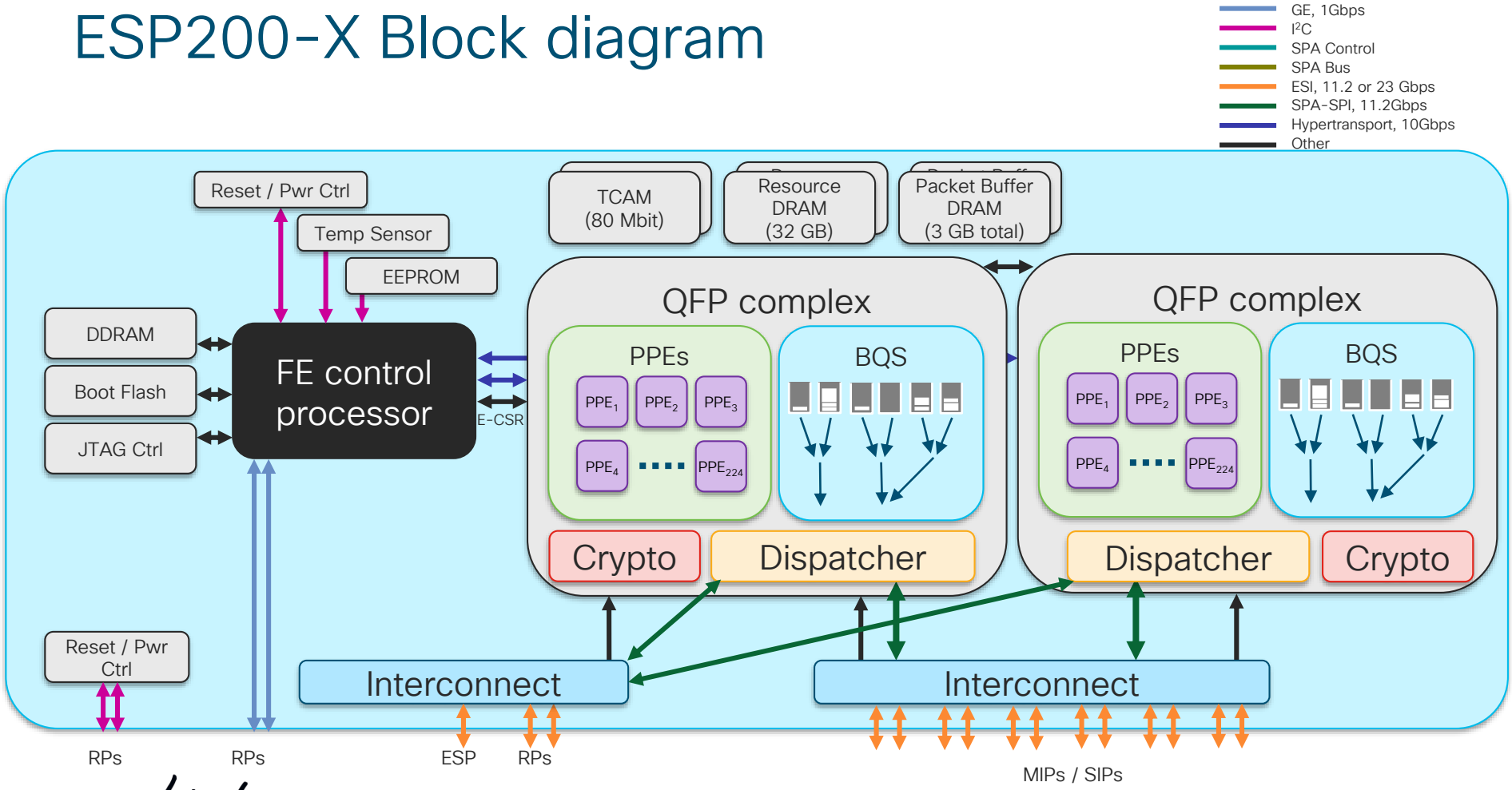
	Cores	Total threads	Crypto cores	QoS Queues	Ethernet sub-system
1st gen QFP	40	160	0	128 K	None
2nd gen QFP lite	32	128	0	16 K	Yes
2nd gen QFP	64	256	0	118 K	None
3rd gen QFP	224	896	16	256 K	Yes, 240 G

	MACSEC	Memory	Packet buffer memory	Platforms using
1st gen QFP	None	4 G	256 MB	ESP 2.5, 5, 10, 20, 40 ASR1001, ASR1002, ASR1002F
2nd gen QFP cost reduced	None	4 G	512 MB	ASR1001-X
2nd gen QFP	Integrated in some platforms	4 G	512 MB	ESP100, ESP200, ASR1002-X, ASR1001-HX, ASR1002-HX
3rd gen QFP	Integrated in all platforms	32 G	1.5 G	ESP100-X, ESP200-X, future platforms

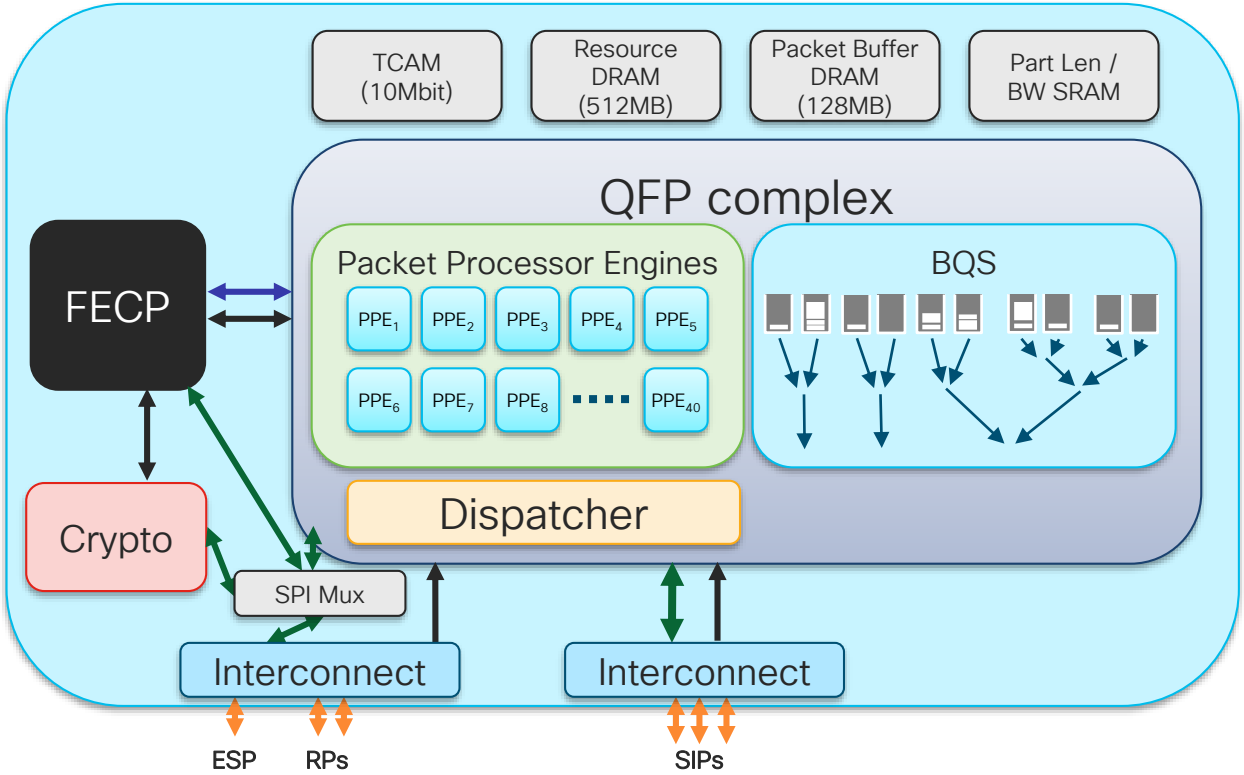
Platform QFP specification comparison

Product	ESP-40	1001-X	1002-X	ESP-100	ESP-200	1001-HX	1002-HX	ESP100-X	ESP-200X
ASICs:	QFP 1st gen	lite QFP 2nd gen	QFP 2nd gen	2x QFP 2nd gen	4x QFP 2nd gen	QFP 2nd gen	2x QFP 2nd gen	QFP 3rd gen	QFP 3rd gen
Performance (IPv4 @64B)	24 Mpps	19 Mpps	34 Mpps	58 Mpps	130 Mpps	43 Mpps	58 Mpps		
# of Processors	40	32	64	128	256	64	128	224	448
# of Threads	160	128	256	512	1024	256	512	896	1792
Clock Rate	1.2 GHz	1.5 GHz	1.2 GHz	1.5 GHz	1.5 GHz	1.5 GHz	1.5 GHz	1.0 GHz	1.0 GHz
Crypto Engine BW @ 1400B	10 Gbps	8 Gbps	4 Gbps	27 Gbps	70 Gbps	19 Gbps	39 Gbps	141 Gbps	197 Gbps
QFP Resource Memory	1 GB	4 GB / 0.5 GB (unified)	1 GB	4 GB	8 GB	1 GB	4 GB	32 GB	32 GB x2
Packet Buffer	256 MB		0.5 GB	1 GB	2 GB	0.5 GB	1 GB	1.5 GB	3 GB
TCAM	40 Mb	10 Mb	40 Mb	80 Mb	2 x 80 Mb	40 Mb	80 Mb	80 Mb	2 x 80 Mb

ESP200-X Block diagram



ESP data processing path



1. Packet arrives at ESP via interconnect
2. Packet assigned to an available PPE a by dispatcher
3. Input FIA invoked
4. Potentially forward through BQS to crypto
5. Forwarding decision is made
6. Egress FIA invoked
7. Packet forwarded through BQS for scheduling based on QoS and interface bandwidth
8. Packet leaves ESP via interconnect

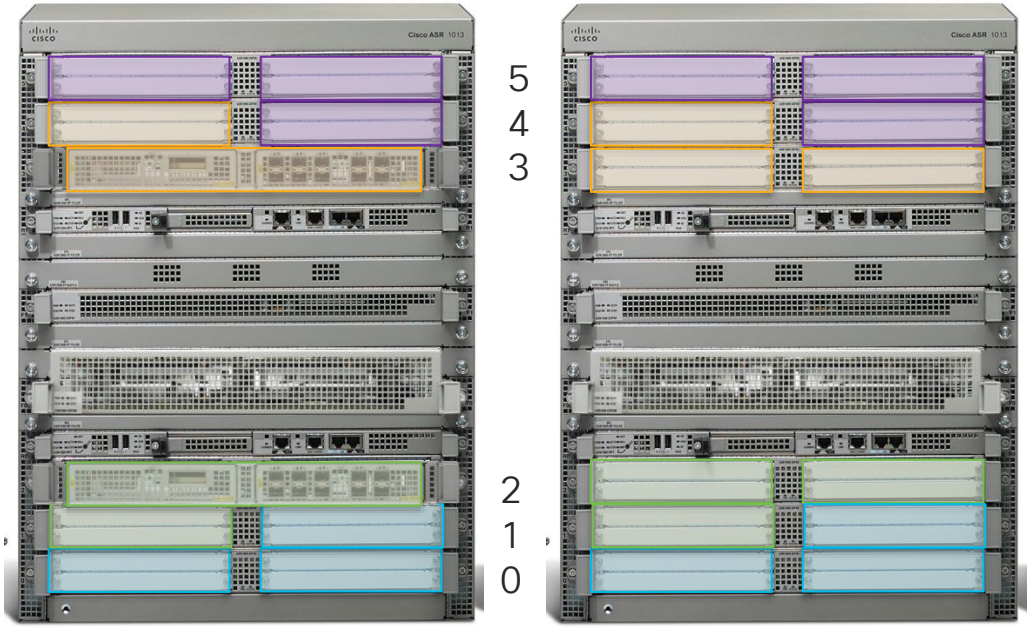
QFP queue distribution for 4 QFP systems

ESP200

- QFP 3
- QFP 2
- QFP 1
- QFP 0

Feature processing and packet buffering are spread across all QFPs. Queuing and scheduling is divided amongst the QFPs based on location of egress interface in the chassis.

For example all packets egressing interfaces in slot 5 with ESP200 in the ASR1013 chassis would be processed by QFP3.

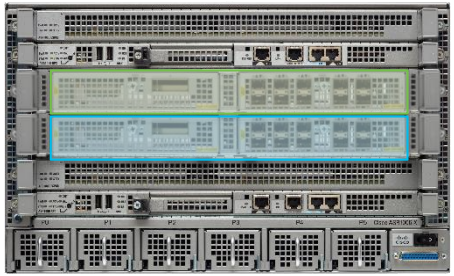


ASR1013 with ESP200 and MIP100

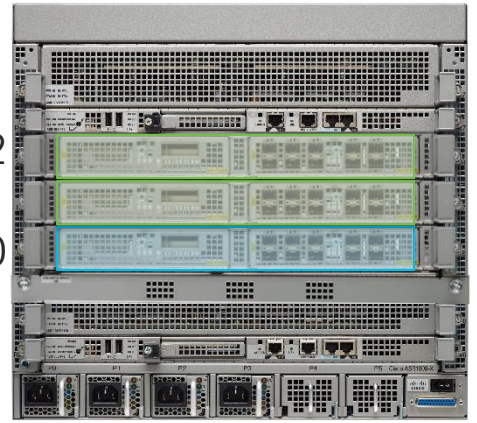
ASR1013 with ESP200 and SIP40

QFP queue distro for 2 QFP systems

ESP100 and ESP200X



ASR1006-X with MIP100



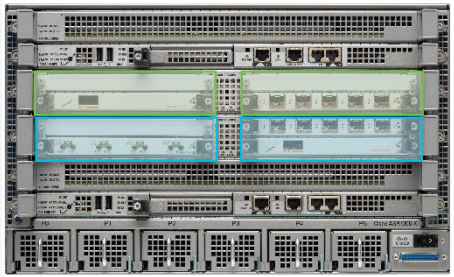
ASR1009-X with MIP100

QFP 1

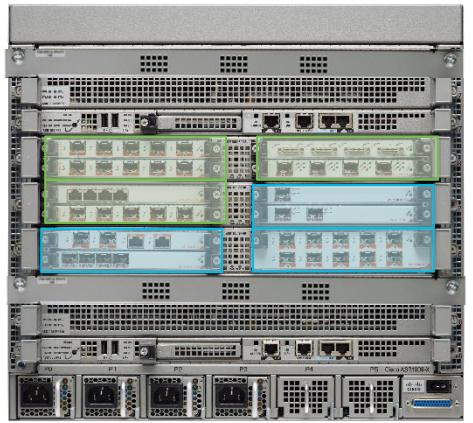
QFP 0

All interfaces on a MIP-100 linecard are handled by a single QFP since there is one channel between the MIP and the ESP.

SIP-40 cards have two channels (each at 20Gb/s) so they can be split between QFPs for a more equal distribution.



ASR1006-X with SIP40



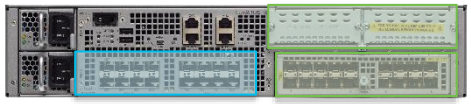
ASR1009-X with SIP40

QFP queue distribution for 2 QFP systems

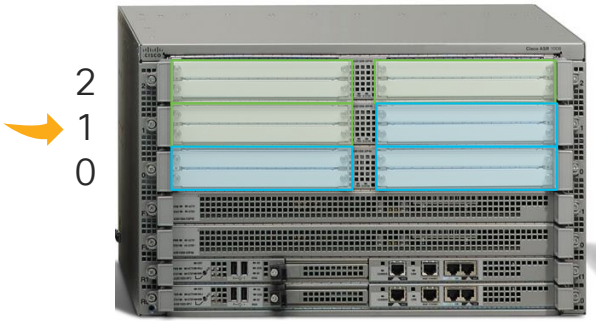
ESP100

QFP 1

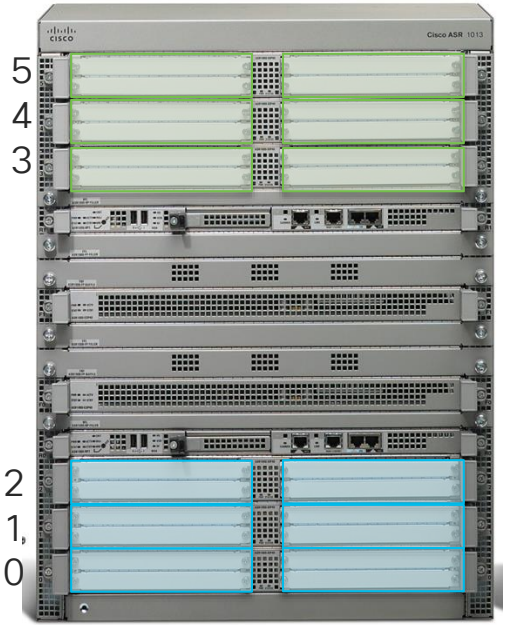
QFP 0



ASR1002-HX



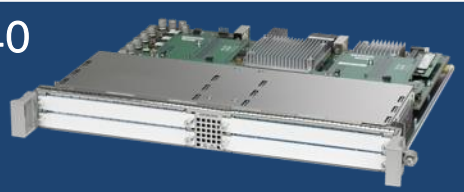
ASR1006 with
ESP100 and SIP40



ASR1013 with
ESP100 and SIP40

Linecards for Input / Output (SIP / MIP)

Modular line cards

	SIP40 	MIP100 
Bandwidth to backplane (full duplex)	40 G	100 G *
Installable cards	4 half height SPAs, ethernet and / or WAN interfaces	2 EPAs, ethernet interfaces only
Buffering	128 MB	Approximately 128 MB
Egress Buffering	8 MB	8 MB
Chassis support	ASR1004, ASR1006, ASR1006-X, ASR1009-X, ASR1013	ASR1006-X, ASR1009-X, ASR1013 *
Features	Basic ingress classification for high and low priority forwarding path to ESP. Egress high and low priority paths. All other features implemented on ESP.	
Backplane channels	Two 20 Gb/s channels	One 100 Gb/s channel

* When installed in ASR1013 chassis, 40 Gbps throughput in slots 0, 1, 4, and 5

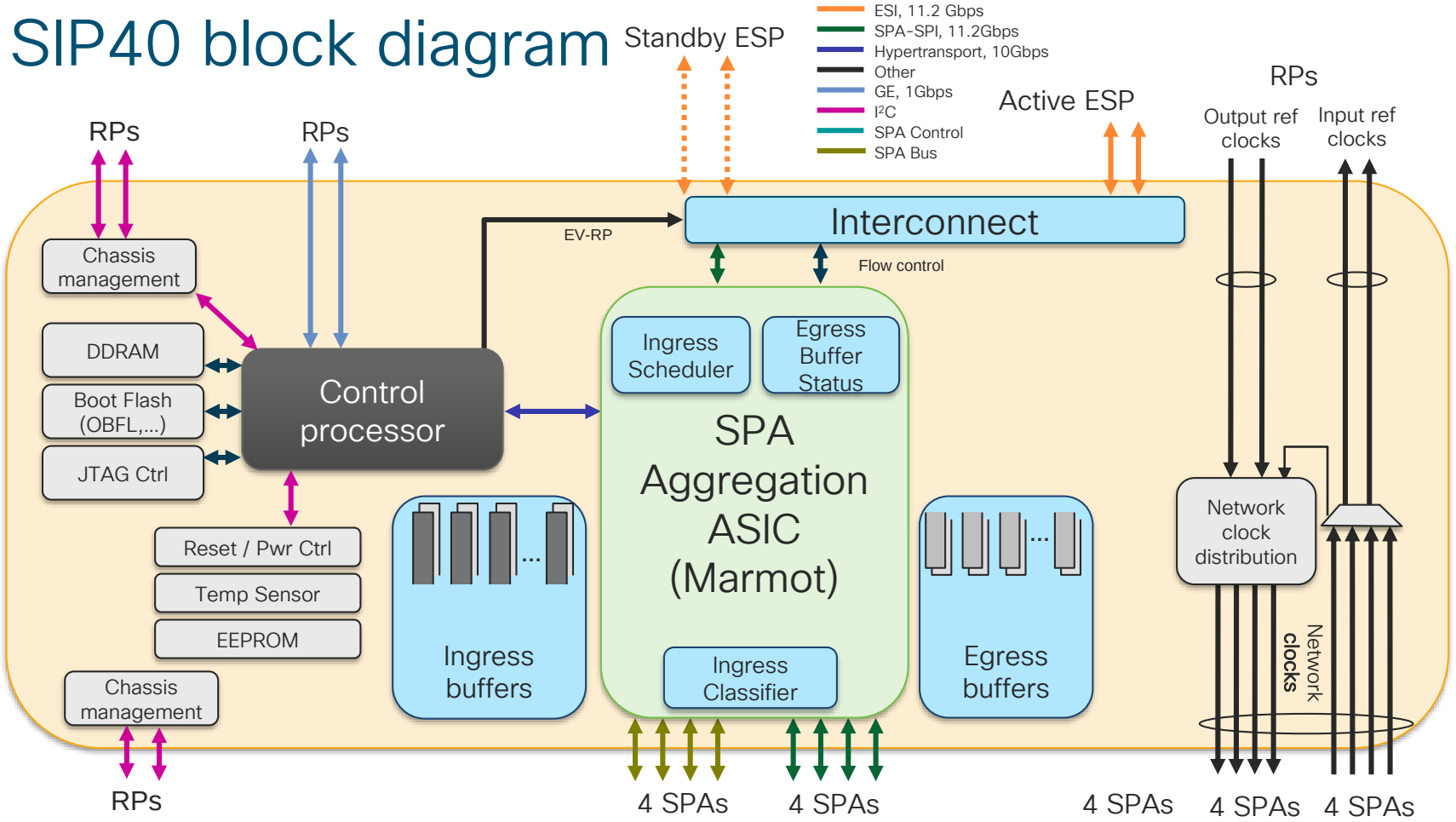
MIP-100: High Density Modular Ethernet

100G Carrier Card + 2xEthernet Port Adapters	
EPA options	• 1x100GE with CPAK (No MACSEC) • 1x100GE with QSFP (MACSEC with XFP) • 2x40GE via CPAK breakout cable (No MACSEC) • 2x40GE and 1x40GE with QSFP (MACSEC with XFP) • 10x10GE with SFP+ (MACSEC) • 18X1GE with SFP (MACSEC)
Throughput	• 200G I/O with up to 100G[1] throughput per line card
Key Features	• Feature Parity to 2x10GE+20xGE Plus • 256-bit WAN MACSEC with VLAN tags in the clear
RP	• RP2 + RP3
ESP	• ESP100, ESP200, ESP100X, ESP200X

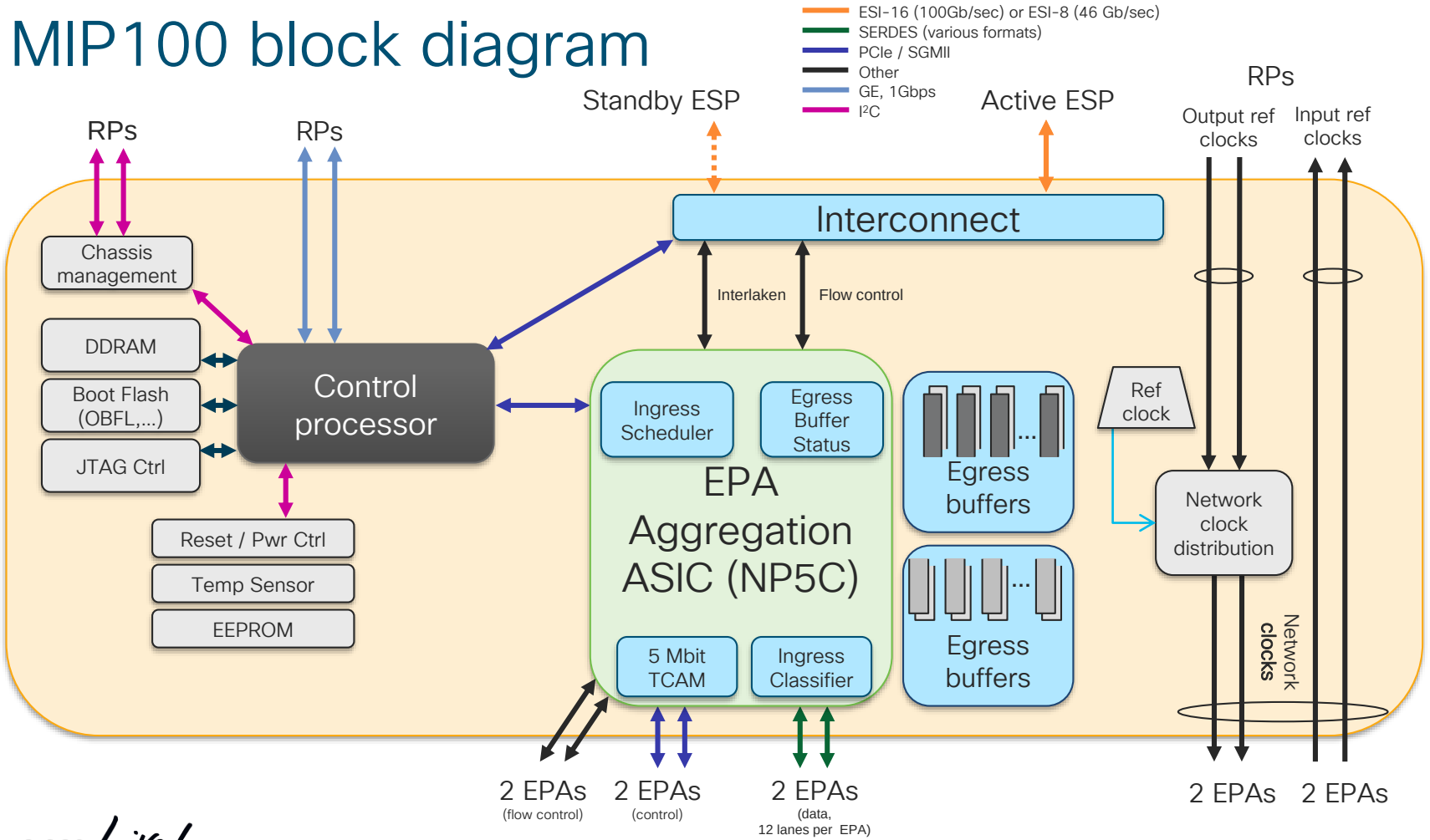
Chassis	Slots	BW
1013	Slots 2 & 3	100G
1013	Slots 0,1,4&5	40G
1006-X	All Slots	100G
1009-X	All Slots	100G
ASR1002-HX	Integrated CC	100G

[1] Max Bandwidth per slot for EPAs (ESP100 and ESP200)

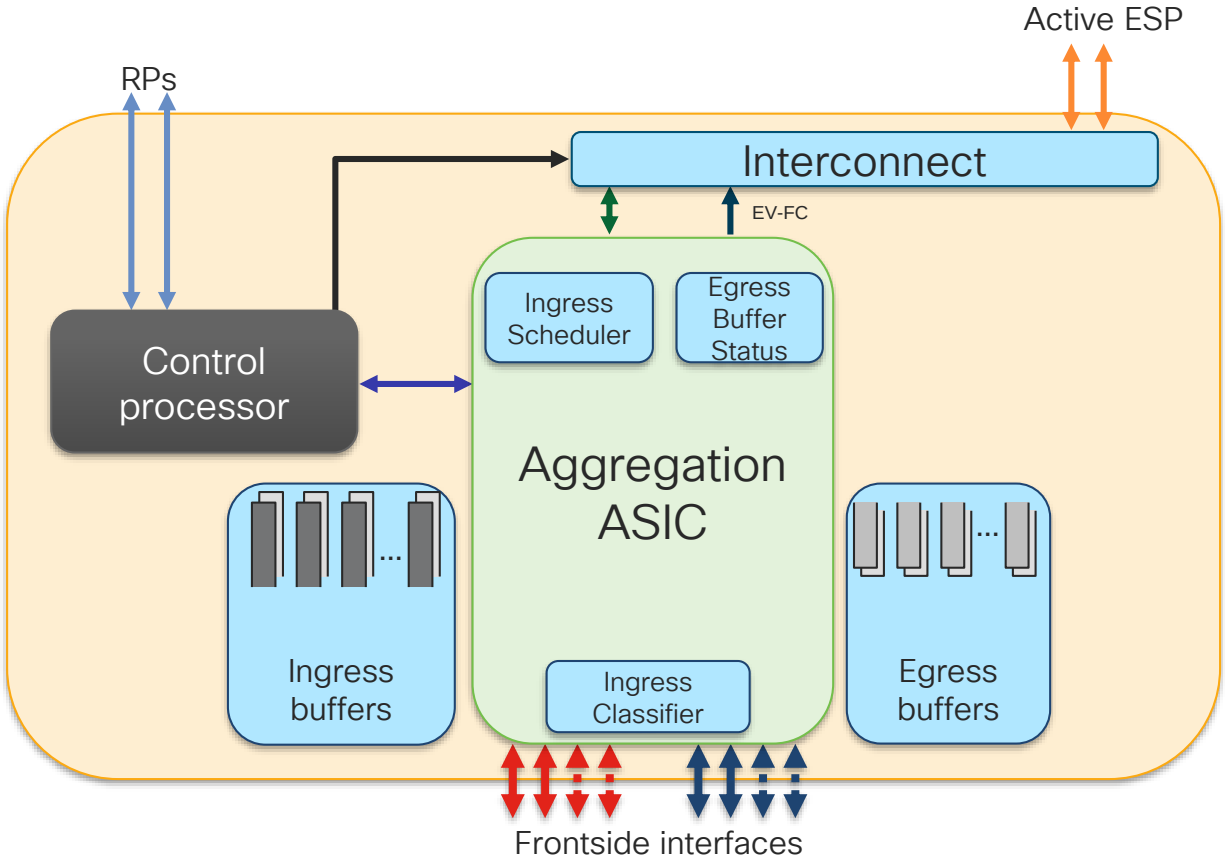
SIP40 block diagram



MIP100 block diagram



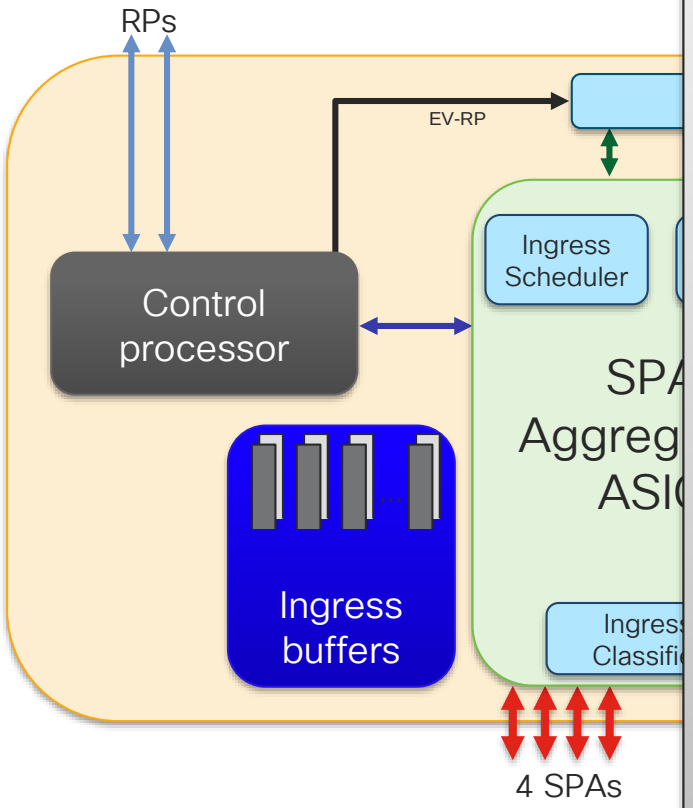
SIP40 / MIP 100 ingress data path



1. SPA receives packet data from its network interfaces and transfers the packet to the SIP
2. SPA Aggregation ASIC classifies the packet into H/L priority
3. SIP writes packet data to external 128MB of ingress buffers

   Ingress buffer memory is carved into 64 queues. The queues are arranged by SPA-SPI channel and optionally H/L. Channels on “channelized” SPAs share the same queue.
4. SPA ASIC selects among ingress queues for next pkt to send to ESP over ESI. It prepares the packet for internal transmission
5. The interconnect transmits packet data of selected packet over ESI to active ESP at up to 23 Gbps

SIP40 ingress data path



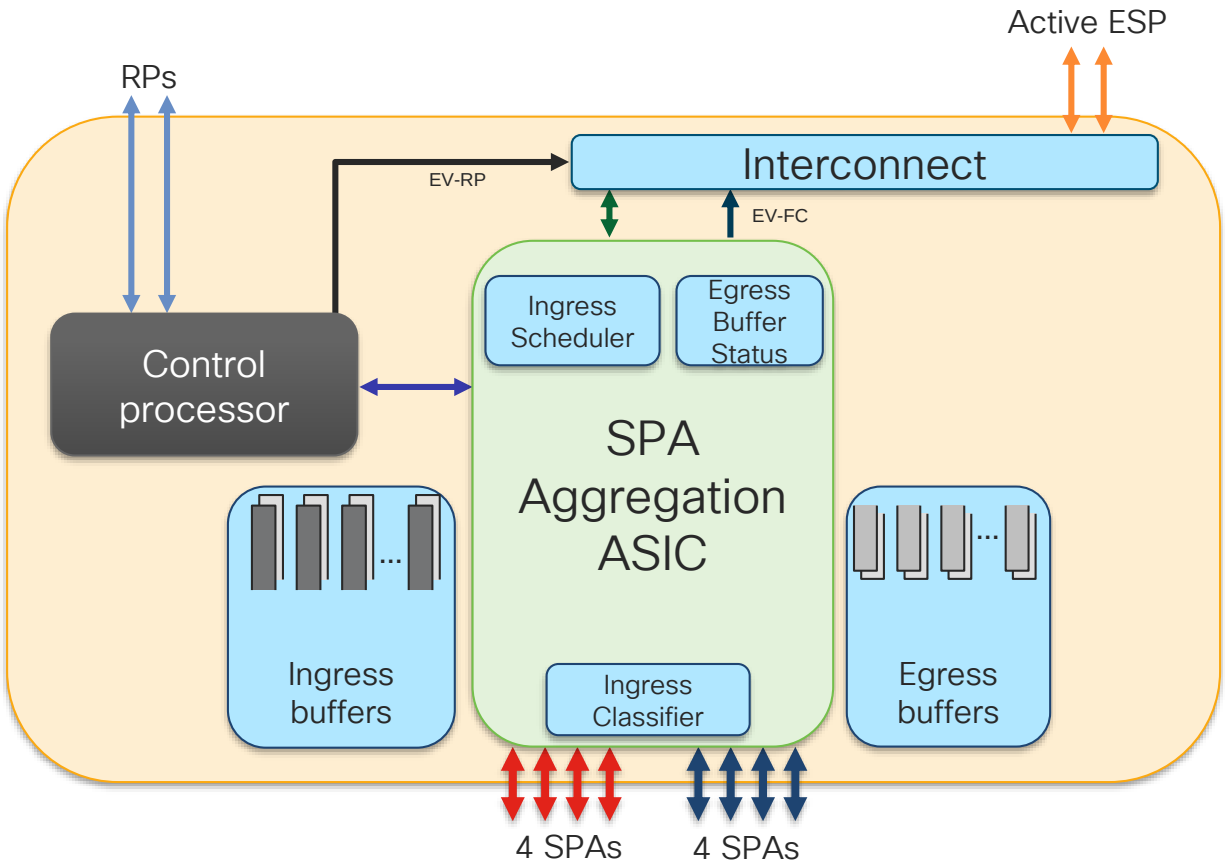

```
ASR1000#show platform hardware int gig0/2/0 plim qos input map

Interface GigabitEthernet0/2/0
    Low Latency Queue (High Priority):
        IP PREC, 6, 7
        IPv6 TC, 46
        MPLS EXP, 6, 7

ASR1000#show platform hardware port 0/2/0 plim buffer settings

Interface 0/2/0
RX Low
    Buffer Size 2064384 Bytes
    Drop Threshold 1020864 Bytes
    Fill Status Curr/Max 0 Bytes / 0 Bytes
TX Low
    Interim FIFO Size 48 Cache line
    Drop Threshold 35136 Bytes
    Fill Status Curr/Max 0 Bytes / 3072 Bytes
RX High
    Buffer Size 2064384 Bytes
    Drop Threshold 402624 Bytes
    Fill Status Curr/Max 0 Bytes / 0 Bytes
  TX High
    Interim FIFO Size 48 Cache line
    Drop Threshold 35136 Bytes
    Fill Status Curr/Max 0 Bytes / 5120 Bytes
```

SIP40 egress data path



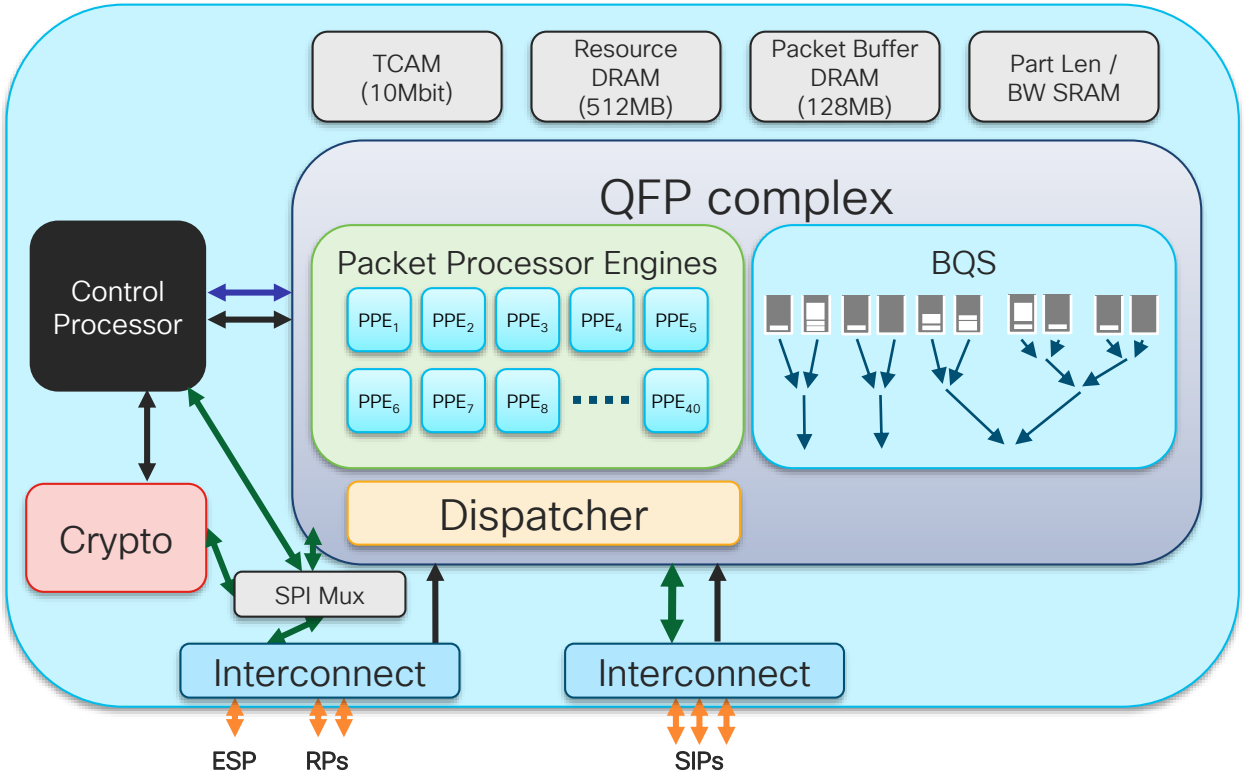
1. Interconnect receives packet data over ESI from the active ESP
2. SPA Aggregation ASIC receives the packet and writes it to external egress buffer memory

   Egress buffer memory is carved into 64 queues. The queues are arranged by egress SPA-SPI channel and optionally H/L. Channels on “channelized” SPAs share the same queue.
3. SPA Aggregation ASIC selects and transfers packet data from eligible queues to SPA-SPI channel (Hi queue are selected before Low)
4. SPA transmits packet data on network interface

   SPA can backpressure transfer of packet data burst independently for each SPA-SPI channel using SPI FIFO status

Animation

ESP data processing path

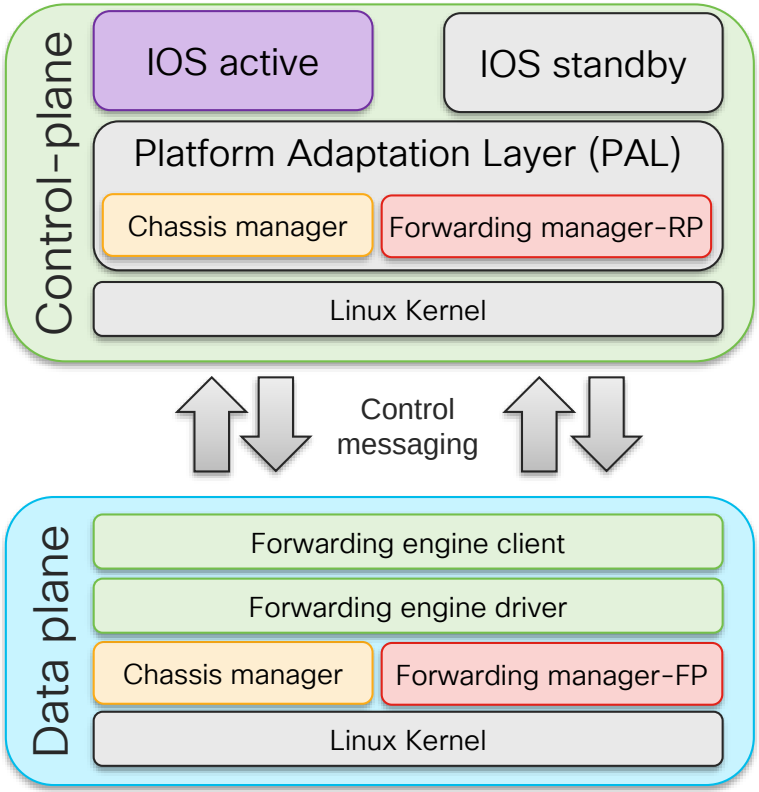


1. Packet arrives at ESP via interconnect
2. Packet assigned to an available PPE a by dispatcher
3. Input FIA invoked
4. Potentially forward through BQS to crypto
5. Forwarding decision is made
6. Egress FIA invoked
7. Packet forwarded through BQS for scheduling based on QoS and interface bandwidth
8. Packet leaves ESP via interconnect

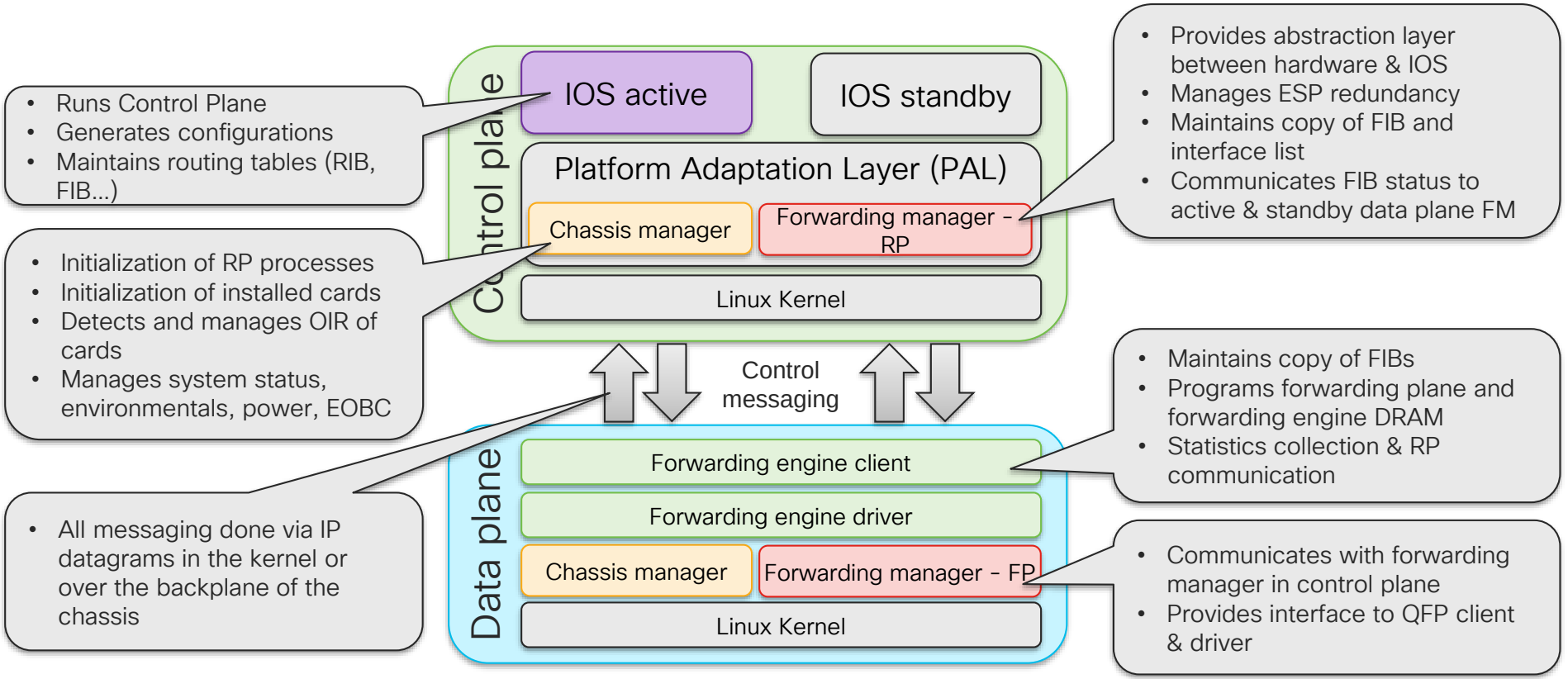
Animation

IOS XE software architecture

- IOS + IOS XE Middleware + Platform Software
- Operational Consistency
  - same look and feel as classic IOS Router
- IOS runs as its own Linux 64 bit process for control plane
- Linux kernel with multiple processes running in protected memory
- Fault containment, re-startability
- ISSU of individual SW packages
- With redundant data plane hardware packet loss is less than 50 ms at failover



IOS XE architecture building blocks



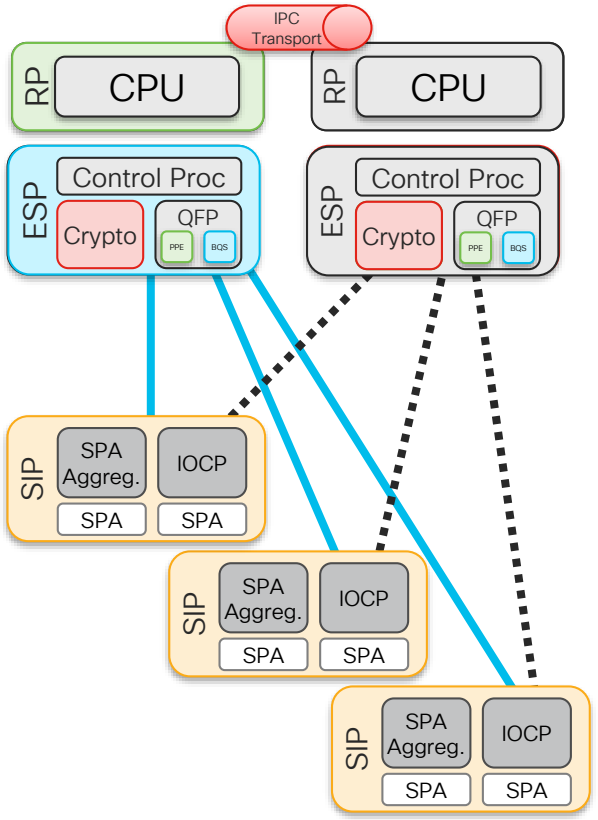
Redundancy

Flavors of High Availability

- Hardware redundancy (intrachassis redundancy)
- Redundancy groups between chassis (interchassis redundancy)
- IOS Software redundancy
- In Service Software Upgrade (ISSU)
- SMU (Service Maintenance Upgrade)

Intrachassis redundancy

- Redundant ESP / RP on ASR 1006 & ASR 1013
- Software Redundancy on ASR 1001, 1002 & 1004
- Max 50ms loss for ESP fail-over
- Zero packet loss on RP fail-over
- Intra-chassis Stateful Switchover (SSO)
  Stateful features: PPPoX, AAA, DHCP, IPSec, NAT, Firewall
- IOS XE also provides full support for Network Resiliency
  NSF/GR for BGP, OSPFv2/v3, IS-IS, EIGRP, LDP
  IP Event Dampening; BFD (BGP, IS-IS, OSPF)
  first hop redundancy protocols: GLBP, HSRP, VRRP
- Support for ISSU super and sub-package upgrades



Interchassis redundancy

Animation

2 ASR 1000 chassis with single RP / single ESP

- Inter-chassis plus intra-chassis redundancy not supported
- Maximum of 2 cluster members

Physical connectivity to both member systems from adjacent routers or switches

- Need a mechanism to direct traffic to either member system in case of failover

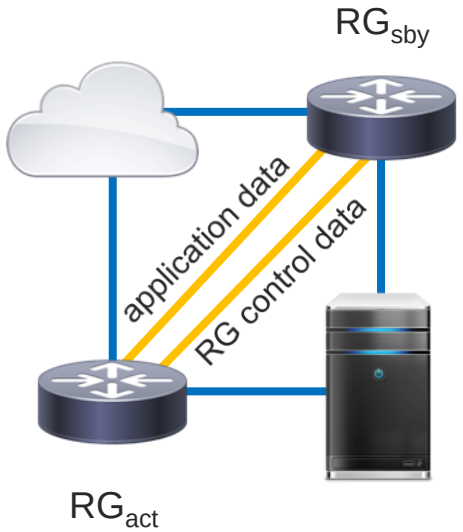
L2 connection between both member systems for RG control traffic

- Used to exchange control traffic (RG hellos, RG state, fail-over signaling etc.)
- Communication required between the two member systems to avoid split-brain condition

L2 connection between both member systems for application state data

- Synchronization of NAT/Firewall/SBC state tables
- FIBs are NOT synchronized by RG infrastructure

Possible a user data cross-connect for asymmetric routing cases



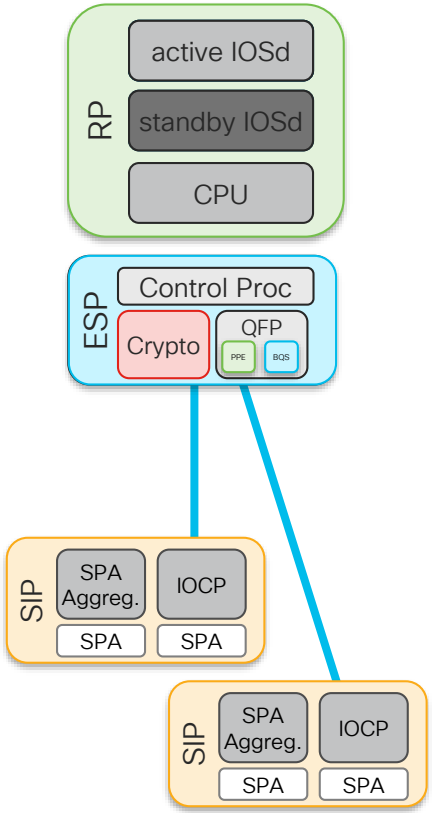
IOS Software redundancy

Animation

- IOS runs as its own Linux process for control plane (Routing, SNMP, CLI etc.)
- Linux kernel runs IOS process in protected memory for:

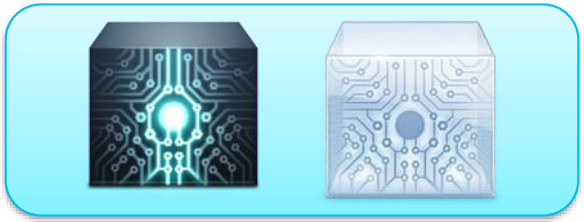
  Fault containment

  Restart-ability of individual SW processes
- Software redundancy helps when there is a RP-IOS failure/crash
- Active process will switchover to the standby, while forwarding continues with zero packet loss
- Can be used for ISSU of RP-IOS package for control-plane bug fixes and PSIRTs
- Other software crashes (example: SIP or ESP) do not benefit from Software redundancy



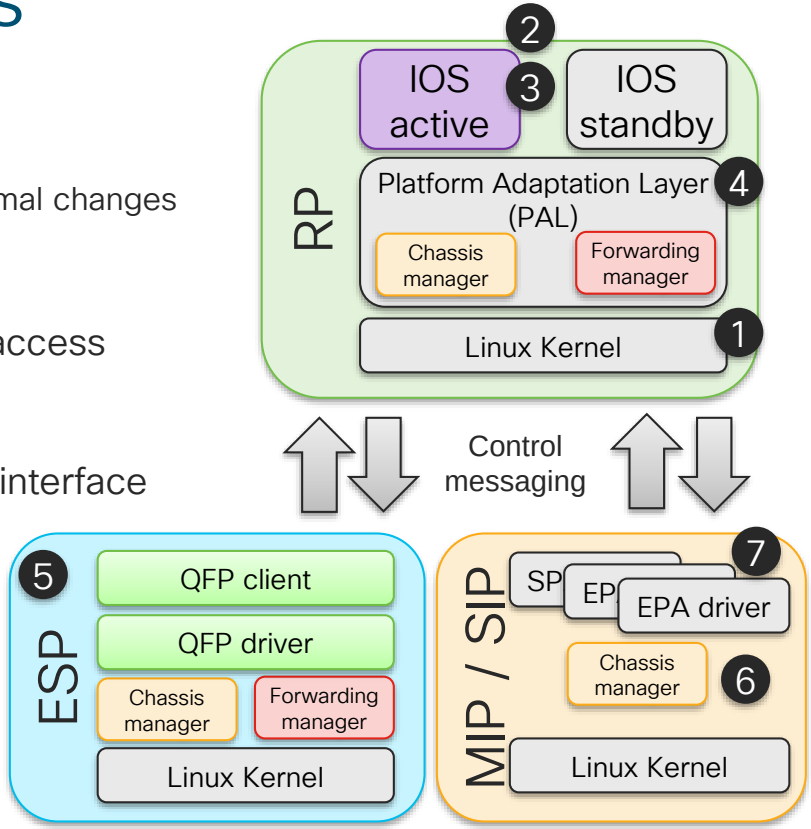
ASR1000 High Availability – ISSU

- Ability to perform upgrade of the IOS image on the single-engine systems
- Support for software downgrade
- “In Service” component upgrades (SIP-Base, SIP-SPA, ESP-Base) without requiring reboot to the system
- Hitless upgrade of some software packages
- RP Portability - installing & configuring hardware that are physically not present in the chassis
- This allows the user to configure an RP in one system i.e. a 4RU and then move it to another system i.e. a fully populated 6RU
- One-shot ISSU procedure available for H/W redundant platforms



ISSU software sub-packages

1. RPBase: RP Linux operating system
   Upgrading of the OS will require reload to the RP and expect minimal changes
2. RPIOS: IOS executable
   facilitates Software Redundancy feature
3. RPAccess (K9 & non-K9): Software required for Router access
   Two versions available (with and without open SSH & SSL)
   facilitates software packaging for export-restricted countries
4. RPControl : control plane processes for IOS / hardware interface
   IOS XE Middleware
5. ESPBase: All ESP code
   Any software upgrade of the ESP requires reload of the ESP
6. SIPBase: SIP OS & control processes
   OS upgrade requires reload of the SIP
7. SIPSPA: SPA drivers and SPA FPD
   Facilitates SPA driver upgrade of specific SPA slots



Software maintenance upgrade (SMU)

- Realtime patches for software that do not require reboot or interruption in forwarding of traffic
- Primary use case is for PSIRT fixes or critical updates that do not allow time for full image certification
- Minimal change footprint versus other upgrade methods such as sub-package upgrade
- Reduces overall maintenance and upgrade cost
- Not as granular as SMU updates from IOS XR. Individual processes for BGP, OSPF, etc. are not restartable on IOS XE.

ASR1000 – Hitless SMU

- Software maintenance update – runtime update without downtime

IOS XE 16.9.1 Hitless SMU Support Matrix		ASR1001-X ASR1002-X ASR1001-HX ASR1002-HX	ASR1004 ASR1006 ASR1006-X ASR1009-X ASR1013	CSR1000v
RP *		Yes	Yes	Yes
ESP *	client / driver	No	No	No
	microcode	No	No	No
SIP40 / MIP100 **		Yes	No	n/a

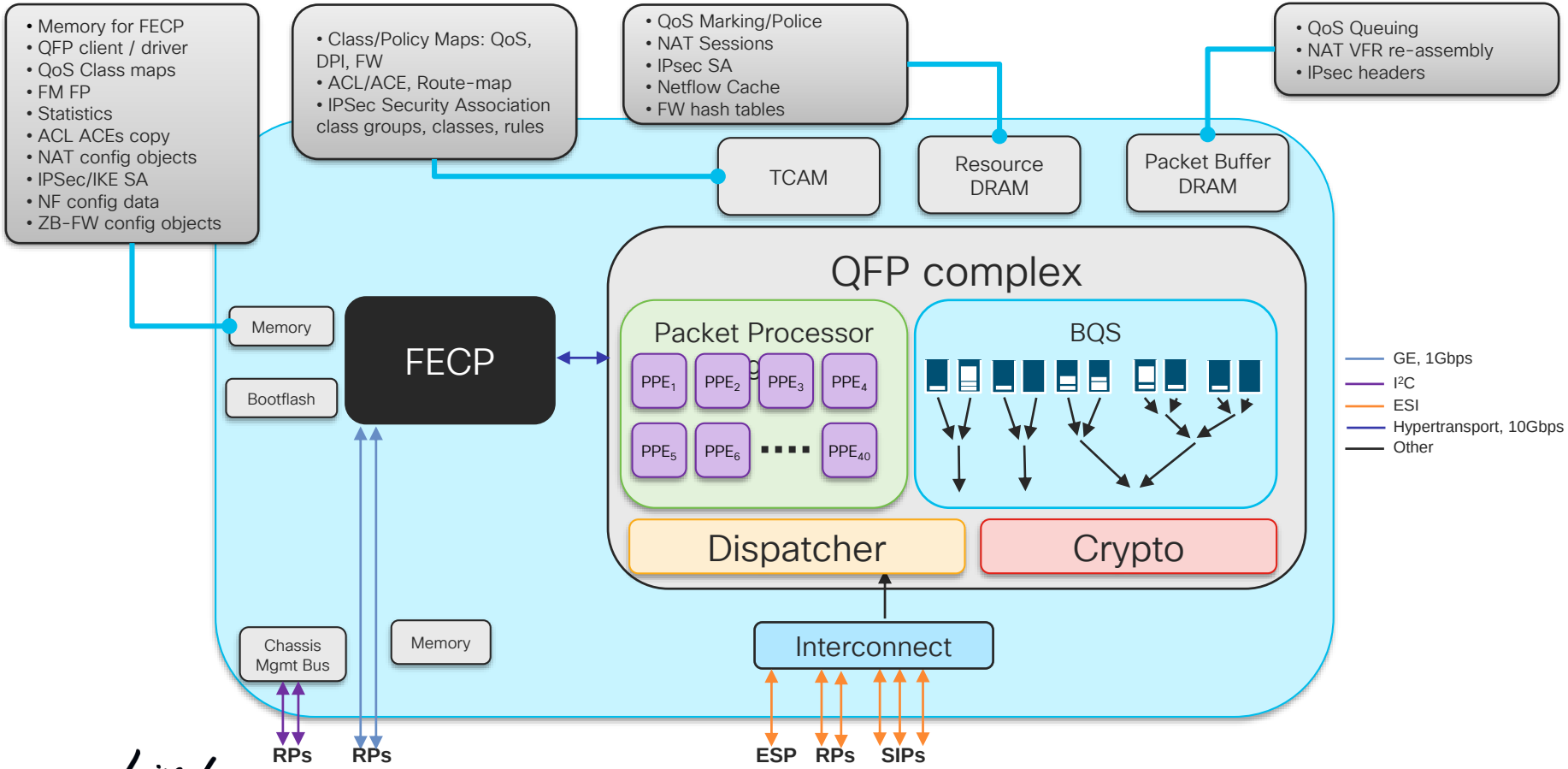
* Not every bug fix is 'hot patchable'
** Internal components for fixed chassis

SMU Features Supported in 16.6.1

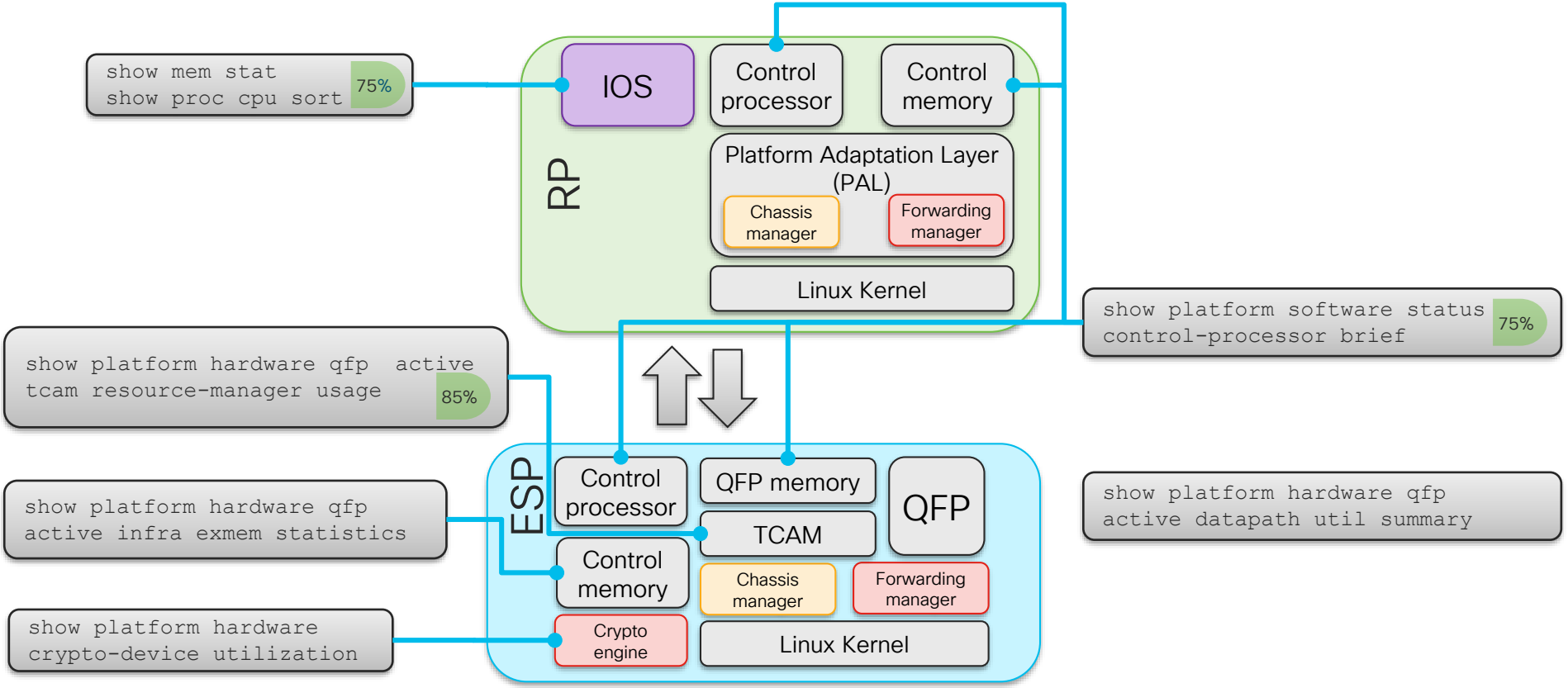
Components			
NAT	Multicast/PIM/MVPN	AAA	IPSec / IKEv2 / VPN
FW	Trustsec	DHCP	LISP (VxLAN)
ALG	RBAC	QoS	L2VPN
MACSec	ISIS	SNMP	MPLS (TE/RSVP/OAM/LDP)
Crypto	BGP	CDP	LLDP
CGN/NPTv6	RIB	ACL	
ALG	OSPF	SSH	

Monitoring

Feature to ESP resources dependency



Component utilization information



Mitigation plan when running out of resources

Before upgrading RP, Memory or ESP, immediate action can be taken to reduce system utilization:

IOS/RP Memory

- Reduce prefixes received from a peer

```
neighbor { ip-address} maximum-prefix <number of prefixes>
```

- Turn off Software Redundancy

```
redundancy mode none
```

QFP Resources DRAM

- Reduce NAT max-entries:

```
ip nat translation max-entries <number of entries>
nat64 translation max-entries <number of entries>
```

- Reduce FW session limit:

```
parameter-map type inspect <X>
  session total <count>
```

- Reduce FNF cache limit:

```
flow monitor <X>
  cache entries <number of entries>
```

IOSd CPU & Memory Utilization

- CPU Load in IOSd process

  ```
  show processes cpu
  ```

- In IOSd, to investigate the memory is occupied by which process use the traditional command:

  ```
  show memory
  show memory allocating-process totals
  ```

Control CPU & Memory Utilization (1)

For an overview of each Module CPU load on the ASR 1000, use the following command:

Sample EEM script to trigger the Load monitoring on the next slide

```
ASR1000# show platform software status control-processor brief
Load Average
 Slot   Status  1-Min  5-Min 15-Min
  RP0 Healthy   0.06   0.06   0.01
  RP1 Healthy   0.06   0.04   0.01
 ESP0 Healthy   0.01   0.00   0.00
 ESP1 Healthy   0.00   0.00   0.00
 SIP1 Healthy   0.04   0.03   0.01
 SIP2 Healthy   0.00   0.00   0.00
```

Load Average represents the process queue or process contention for CPU resources.

1. On a single core processor, an instantaneous load of “7” would mean that seven processes were “ready to run”, one of which is currently running.
2. On a dual core processor, a load of “7” would represent seven processes were ready to run, two of which are currently running.

Triggered EEM Script to monitor system load

This is a sample EEM script that monitors RP0 one minute load.

- A load of 5 triggers actions 1 through 5.
- Action 1 generates a log message when the script triggers.
- Actions 2 through 5 run CLI, outputs them to the bootflash, and appends the cpuinfo file

```
 event manager applet capture_cpu_spike
   event snmp oid 1.3.6.1.4.1.9.9.109.1.1.1.1.24.2 get-type exact entry-op ge entry-val 500 exit-time
180 poll-interval 2
   action 1.0 syslog msg ”Load is high.  Check bootflash:cpuinfo for details."
   action 2.0 cli command "en"
   action 3.0 cli command "show clock | append bootflash:cpuinfo"
   action 4.0 cli command "show platform software status control-processor br | append
bootflash:cpuinfo"
   action 5.0 cli command "show platform software process slot rp active monitor | append
bootflash:cpuinfo"
```

Control CPU & Memory Utilization (2)

Status: Critical, Warning, Healthy.
Definition in reference slide at section end

```
ASR1000# show platform software status control-processor brief
  <snip>
Memory (kB)
 Slot  Status    Total     Used (Pct)     Free (Pct) Committed (Pct)
  RP0 Critical  3919788  3891940 (95%)    27848 (0%)   2005100 (98%)
  RP1 Healthy   3919788  1164924 (28%)  2754864 (66%)   1994212 (48%)
 ESP0 Healthy   2030288   520744 (24%)  1509544 (71%)   2816620 (134%)
 ESP1 Healthy   2030288   514972 (24%)  1515316 (72%)   2816356 (134%)
 SIP1 Healthy    484332   311868 (59%)   172464 (32%)    262472 (50%)
 SIP2 Healthy    484332   332252 (63%)   152080 (29%)    317648 (60%)

CPU Utilization
 Slot  CPU   User System   Nice   Idle    IRQ   SIRQ IOwait
  RP0    0   1.28   1.15   0.00  97.25   0.01   0.10   0.20
  RP1    0   0.94   1.23   0.00  97.48   0.00   0.02   0.30
 ESP0    0   0.56   0.66   0.00  98.76   0.00   0.00   0.00
 ESP1    0   0.52   0.64   0.00  98.82   0.00   0.00   0.00
 SIP1    0   0.47   0.45   0.00  99.04   0.00   0.01   0.00
 SIP2    0   0.58   0.53   0.00  98.85   0.00   0.01   0.00
```

Memory utilization is represented by the following:

- Total – Total card memory
- Used – Consumed memory
- Free – Available memory
- Committed – Virtual memory committed to processes

Control CPU & Memory Utilization (3)

CPU utilization is a two second relative percentage average of the number of processes requesting CPU resources at a given time and is represented by the following fields:

- `CPU` – The allocated processor
- `User` – Non-Linux kernel processes
- `System` – Linux kernel process
- `Nice` – Low priority processes
- `Idle` – Percentage of time the CPU was inactive
- `IRQ` – Interrupts
- `SIRQ` – System Interrupts
- `Iowait` – Percentage of time CPU was waiting for IO

This command must be executed two times as the first sample will not report accurate CPU utilization information.

```
ASR1000# show platform software process slot RP active monitor cycles 2 | inc Cpu|Mem
Cpu(s):  1.1%us,  1.0%sy,  0.0%ni, 97.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:  16343244k total,  3988416k used, 12354828k free,   202964k buffers
Swap:        0k total,        0k used,        0k free,  1414668k cached
Cpu(s):  3.8%us,  0.3%sy,  0.0%ni, 95.8%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:  16343244k total,  3988788k used, 12354456k free,   202964k buffers
Swap:        0k total,        0k used,        0k free,  1414796k cached
```

Control CPU & Memory Utilization (4)

- To check process in each Module, use following command to check in VTY
- Enter “m” to sort by memory usage

```
ASR1000# monitor platform software process fp active
Tasks:  80 total,   4 running,  76 sleeping,   0 stopped,   0 zombie
Cpu(s):  1.0% us,  0.3% sy,  0.0% ni, 98.7% id,  0.0% wa,  0.0% hi,  0.0% si
Mem:   2030288k total,   525260k used,  1505028k free,    21228k buffers
Swap:        0k total,        0k used,        0k free,   192024k cached

  PID USER      PR  NI  VIRT  RES  SHR S %CPU %MEM    TIME+  COMMAND
 4750 root      20   0  645m  92m  31m S  0.7  4.6  26:36.97 cpp_cp_svr
 5597 root      20   0  502m  45m  24m S  0.3  2.3   6:00.44 fman_fp_image
 5737 root      20   0 16108 5732 4104 R  0.3  0.3  12:39.08 hman
 7321 root      20   0  8876 2200 1712 R  0.3  0.1   0:00.03 in.telnetd
 7392 binos     20   0  2496 1212  976 R  0.3  0.1   0:00.10 top
    1 root      20   0  2132  632  544 S  0.0  0.0   0:10.63 init
```

The "monitor" command does not work with serial console. VTY sessions work by default.

Do not use the first iteration. Allow for at least 2 refreshes to get the most accurate data.

Control CPU & Memory Utilization (5)

CISCO-PROCESS-MIB

- support 64 bits architecture which runs on IOS XE
- monitor CPUs on RP, ESP and SIP. Only Active RP/ESP can be monitored, not standby.

1) Find out the index for the RP's `cpmCPUTotal1min`

```
host> getmany -v2c 9.0.0.52 cpmCPUTotalPhysicalIndex
cpmCPUTotalPhysicalIndex.2 = 7031
```

7031 is RP CPU physical index in entity mib, so use 2 as index for RP `cpmCPUTotal1min`

2) The OID used to retrieve instance for the RP's `cpmCPUTotal1min`

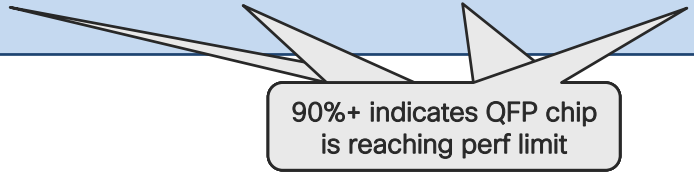
```
host> getone -v2c 9.0.0.52 cpmCPUTotal1min.2
cpmCPUTotal1min.2 = 58
```

Please note that "`cpmCPUTotal1min.2`" is same as OID "`1.3.6.1.4.1.9.9.109.1.1.1.1.4.2`"

QFP & Resource DRAM Utilization (1)

To display the QFP utilization, use the following command:

```
ASR1000# show platform hardware qfp active datapath utilization summary
 CPP 0:                  5 secs       1 min       5 min      60 min
Input:  Total (pps)     1625349     1625340     1625345     1625345
              (bps)  1708810504  1708399184  1708085344  1708039368
Output: Total (pps)     1625333     1625338     1625344     1625344
              (bps)  1786828168  1786418448  1786105008  1786059008
Processing: Load (pct)        2           2           2           2
```



QFP & Resource DRAM Utilization (2)

DRAM on QFP usage can be found on the following command

% Utilization = InUse/Total

124180480 / 1073741824 = 11.56%

```
ASR1000# show platform hardware qfp active
infrastructure exmem statistics
QFP exmem statistics
Type: Name: DRAM, QFP: 0
  Total: 1073741824
  InUse: 124180480
  Free: 949561344
  Lowest free water mark: 949561344
Type: Name: IRAM, QFP: 0
  Total: 134217728
  InUse: 8134656
  Free: 126083072
  Lowest free water mark: 126083072
Type: Name: SRAM, QFP: 0
  Total: 32768
  InUse: 15088
  Free: 17680
  Lowest free water mark: 17680
```

QFP & Resource DRAM Utilization (3)

Syslog when throughput exceeds BW license (ASR1001-X, ASR1002-X)

```
Exceeding 95% threshold:
*Sep 24 10:15:14.249: %BW_LICENSE-5-THROUGHPUT_THRESHOLD_LEVEL: F0: cpp_ha:  Average
throughput rate
had exceeded 95 percent of licensed bandwidth 10000000000 bps 1 times, sample period
300 seconds, in last 24 hours

Exceeding total bw:
Sep 24 10:42:28.450: %BW_LICENSE-4-THROUGHPUT_MAX_LEVEL: F0: cpp_ha:  Average
throughput rate had
exceeded the total licensed bandwidth 10000000000 bps and dropped 1 times, sample
period 300 seconds, in last 24 hours.
```

Upgrade throughput via (licensing implications apply):

```
platform hardware throughput level <x>
reload
```

TCAM

QFP TCAM usage can be found in following command:

```
ASR1000# show platform hardware qfp active tcam
resource-manager usage
QFP TCAM Usage Information

80 Bit Region Information
-------------------------
Name                          : Leaf Region #0
Number of cells per entry     : 1
Current 80 bit entries used   : 0
Current used cell entries     : 0
Current free cell entries     : 0

160 Bit Region Information
-------------------------
Name                          : Leaf Region #1
Number of cells per entry     : 2
Current 160 bits entries used : 6
Current used cell entries     : 12
Current free cell entries     : 4084
```

```
320 Bit Region Information
-------------------------
Name                          : Leaf Region #2
Number of cells per entry     : 4
Current 320 bits entries used : 0
Current used cell entries     : 0
Current free cell entries     : 0


Total TCAM Cell Usage Information
---------------------------------
Name                          : TCAM #0 on CPP #0
Total number of regions       : 3
Total tcam used cell entries  : 12
Total tcam free cell entries  : 524276
Threshold status              : below critical limit
```

Control-Process Health Definition (1)

Board	WARNING	CRITICAL	WARNING	CRITICAL	WARNING	CRITICAL
	1 minute		5 minute		15 minutes	
SIP10	5	8	5	8	5	8
SIP40	5	8	5	8	5	8
ESP5	5	8	5	8	5	8
ESP10	5	8	5	8	5	8
ESP20	5	8	5	8	5	8
ESP40	5	8	5	8	5	8
ESP100	5	8	5	8	5	8
ESP200	5	8	5	8	5	8
RP1	5	8	5	8	5	8
RP2	5	8	5	8	5	8
ASR1001-X	8	12	8	12	10	15
ASR1002-X	8	12	8	12	10	15

"`show platform software status control-processor brief`" output in slide 30, the Load Average Status can be Healthy, Warning and Critical, this table provides the Warning and Critical status threshold for each field

Control-Process Health Definition (2)

Board	FIELD	WARNING	CRITICAL	FIELD	WARNING	CRITICAL	FIELD	WARNING	CRITICAL
SIP10	Committed	95%	100%	MemFree	10%	5%	MEMUSED	90%	95%
SIP40	Committed	95%	100%	MemFree	10%	5%	MEMUSED	90%	95%
ESP5	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ESP10	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ESP20	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ESP40	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ESP100	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ESP200	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
RP1	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
RP2	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ASR1001-X	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%
ASR1002-X	Committed	90%	95%	MemFree	10%	5%	MEMUSED	90%	95%

"`show platform software status control-processor brief`" output in previous slide, the Memory Status can be Healthy, Warning and Critical, this table provides the Warning and Critical status threshold for each field

BQS memory utilization

```
ASR1006-X #show platform hardware qfp active bqs 0 packet-
buffer utilization

Packet buffer memory utilization details:
  Yoda: 0
    Total:      512.00 MB
    Used :       66.50 KB
    Free :      511.94 MB

    Utilization:    0 %

    Threshold Values:
      Vital              :    511.95 MB, Status: False
      Packet Priority    :    507.12 MB, Status: False
      Priority           :    487.59 MB, Status: False
      Non-Priority       :    438.76 MB, Status: False
  Yoda: 1
    Total:      512.00 MB
    Used :       66.50 KB
    Free :      511.94 MB

    Utilization:    0 %

    Threshold Values:
      Vital              :    511.95 MB, Status: False
      Packet Priority    :    507.12 MB, Status: False
      Priority           :    487.59 MB, Status: False
      Non-Priority       :    438.76 MB, Status: False
```

Total amount of memory per ASIC

Total amount of memory free on ASIC 0

Current utilization on ASIC 0

50 kBytes reserved for vital packets

4% (95 thru 99) for PAK_PRI packets

10% (85 thru 95) for user priority packets

Lower 85% available for generic packets

If shows true, the no memory available for queueing this type of packet for the specific ASIC

TCAM Exhaustion

```
ASR1006-X #show platform hardware qfp active tcam resource-manager usage

QFP TCAM Usage Information

80 Bit Region Information
-------------------------
Name                                : Leaf Region #0
Number of cells per entry           : 1
Current 80 bit entries used         : 0
Current used cell entries           : 0
Current free cell entries           : 0

160 Bit Region Information
--------------------------
Name                                : Leaf Region #1
Number of cells per entry           : 2
Current 160 bits entries used       : 6
Current used cell entries           : 12
Current free cell entries           : 4084

320 Bit Region Information
--------------------------
Name                                : Leaf Region #2
Number of cells per entry           : 4
Current 320 bits entries used       : 0
Current used cell entries           : 0
Current free cell entries           : 0


Total TCAM Cell Usage Information
---------------------------------
Name                                : TCAM #0 on CPP #0
Total number of regions             : 3
Total tcam used cell entries        : 12
Total tcam free cell entries        : 1048564
Threshold status                    : below critical limit
```

Once TCAM starts to become fragmented, these individual counters for region size will start to change.

There are no deny statements in TCAM hardware. It is constructed only with allow rules. Therefore, for every deny statement encountered the search pointer jumps to the next class to find a permit statement. This phenomenon is called deny-jump.

For each deny statement the algorithm builds a list of entries which are derived with the product of deny and the subsequent permit statements from other classes. This increases the number of TCAM entry exponentially which leads to TCAM exhaustion.

This is the overall free TCAM memory counter/

Key Takeaways

- ESP100X and ESP200X for ASR1000 deliver
  - significant crypto performance improvements
  - significant growth for stateful feature scale
- ASR1000 platforms are a combination of multiple systems
  - Independent data plane cores and memory
  - Independent control plane cores and memory
  - Ingress / egress linecard classification and buffering
  - Other hardware resources – TCAM, hardware QoS scheduling

Understand the problem and know where to look!

Continue your education

Demos in the Cisco campus

Walk-in labs

Meet the engineer 1:1 meetings

Related sessions

Complete your online session survey

- Please complete your session survey after each session. Your feedback is very important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (starting on Thursday) to receive your Cisco Live t-shirt.
- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Content Catalog on ciscolive.com/emea.

Cisco Live sessions will be available for viewing on demand after the event at ciscolive.com.

cisco Live!

Thank you

cisco Live!

You make **possible**