



*LET'S
BUILD
TOMORROW
TODAY*

Cisco Nexus 5600/6000 Switch Architecture

Rohan Grover, Technical Marketing Manager

BRKARC-3452

Session Goal

- To provide a thorough understanding of the Nexus 5600/6000 switching architecture, fabric switching modes, packet flows, and key forwarding engine functions
- This session will also examine the Nexus 5600 Multicast Architecture, ACLs and QOS
- This session will not examine NX-OS software architecture or other Nexus platform architectures



Related Sessions at Cisco Live San Diego

Session Id	Session Name
BRKDCT-3100	Troubleshooting Nexus 5600/6000 Series switches
BRKARC-3470	Cisco Nexus 7000/7700 Switch Architecture
BRKARC-3454	In-depth and personal with the Cisco Nexus 2000 Fabric Extender Architectures, Features, and Topologies
BRKDCT-1980	Advanced Analytics in Nexus Switches
BRKDCT-3346	End-to-End QoS Implementation and Operation with Cisco Nexus

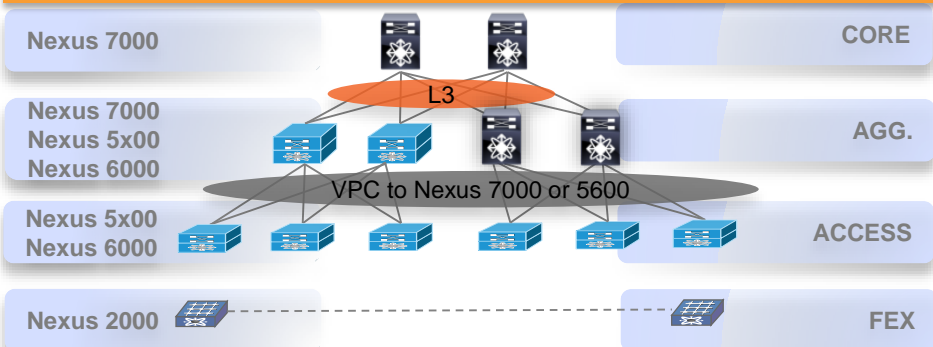
Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

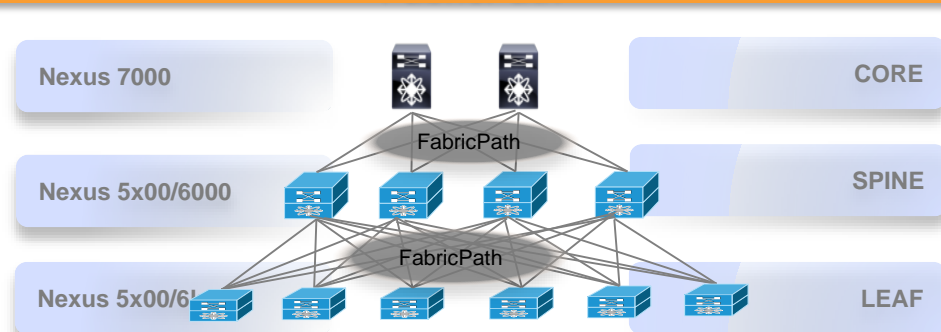
Introduction

Nexus 5500/5600/6000 Use-cases

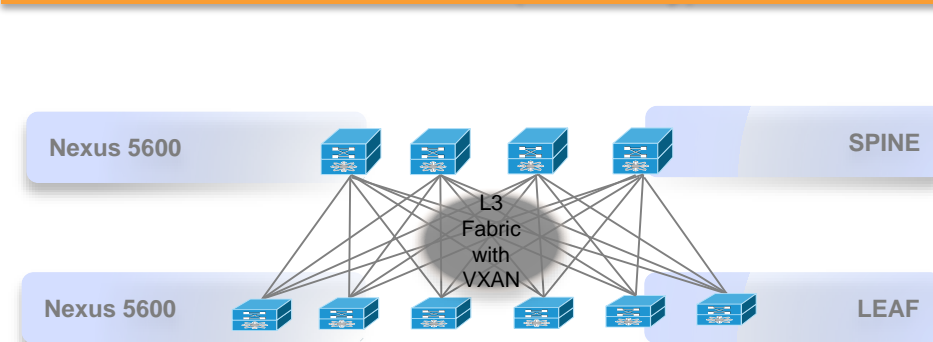
Classic 3-Tier with FEX



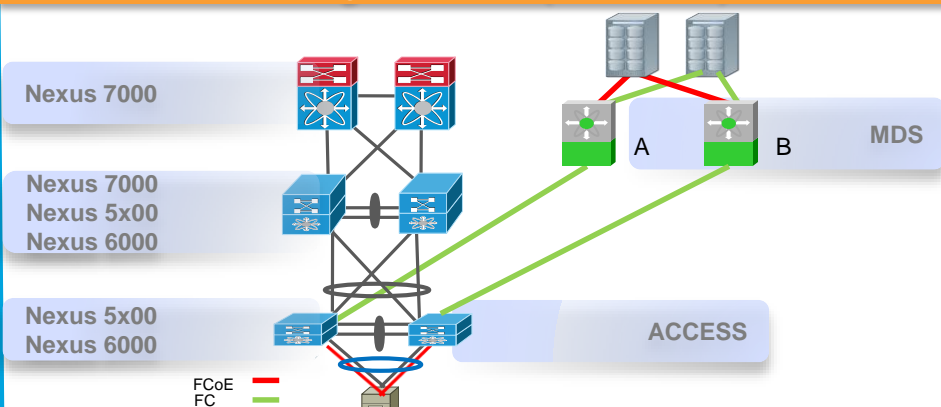
FabricPath



VXLAN Fabric (5600 only)

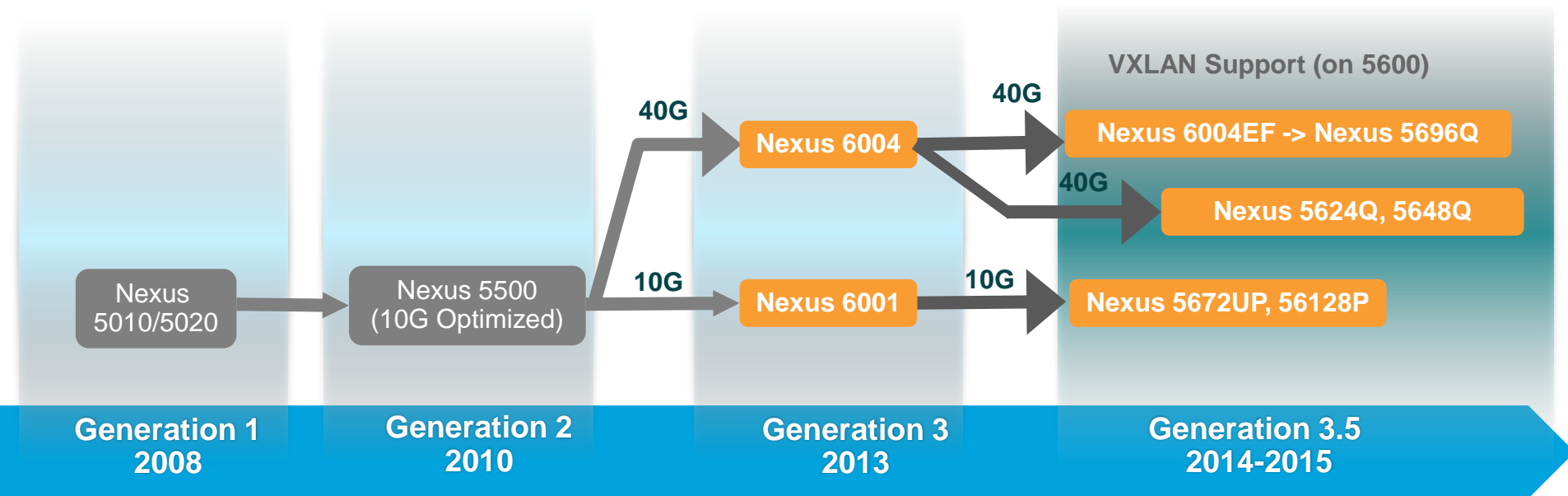


Converged Access (FC/FCOE)



CiscoLive!

Nexus 5000/6000 Evolution



Introducing the Nexus 5600 Series



Nexus 5600



Nexus 5500

VXLAN – Bridging and Routing

FabricPath with SegmentID

40G Uplinks

Programmability

Network Visibility

Lower End-to-end Latency

Line Rate L3

Bigger buffers and tables

Unified Ports

FEX Architecture

FabricPath

FCoE



75%
Market Share*



25 Million
Ports Shipped



20,000+
Customers



125K+
Chassis till date

Next Generation Nexus 5K

Nexus 5000 Series Positioning

10G/40G



Nexus 5600/6000

High 10G Density
40G Flexibility
100G Uplinks
Unified Ports
Buffers/Tables

Nexus 5500

Flexibility
10G Uplinks

Nexus 5010/5020

Large Buffers and Tables
FabricPath with SegID (DFA)
VXLAN (5600 only)

Density
Programmability
Network Visibility

LAN/SAN Convergence

FabricPath

FEX Architecture

- ✓ 20K+ Customers Over 5 Years
- ✓ 25M+ Ports Shipped
- ✓ 125K+ Chassis Shipped
- ✓ 75%+ Market Share*

CUSTOMER VALUE
CISCO INNOVATION

Nexus 5600, 6000 and 5500 Comparison



	Nexus 5600 Series	Nexus 6000 Series	Nexus 5500 Series
Switch Fabric Throughput	Upto 7.68 Tbps	Upto 7.68 Tbps	Upto 960 Gbps
Port-to-Port Latency	~ 1us	~ 1us	~ 1.8us
Layer 3 Capability	Integrated Line Rate	Integrated Line Rate	L3 Module
VXLAN	✓	✗	✗
Unified Ports	✓	✓	✓
MAC Table	256K MAC/ARP (flexible)	256K MAC/ARP (flexible)	32K MAC/ARP
IP Unicast Routes	32K (v4) or 8K (v6)	32K (v4) or 8K (v6)	8K (v4) or 4K (v6)
IP Hosts	128K (v4) or 64K (v6)	128K (v4) or 64K (v6)	16K (v4) or 8K (v6)
IP Multicast routes	32K (v4) or 16K (v6)	32K (v4) or 16K (v6)	8K (v4) or 4K (v6)
Buffers	25MB/12p X 10G)	25MB/12p X 10G)	640KB/10G port
IGMP Snooping Groups	32K*	32K*	8K
VRFs	4K	4K	1K
SPAN	31*, 16 Can Be ERSPAN	31*, 16 Can Be ERSPAN	4

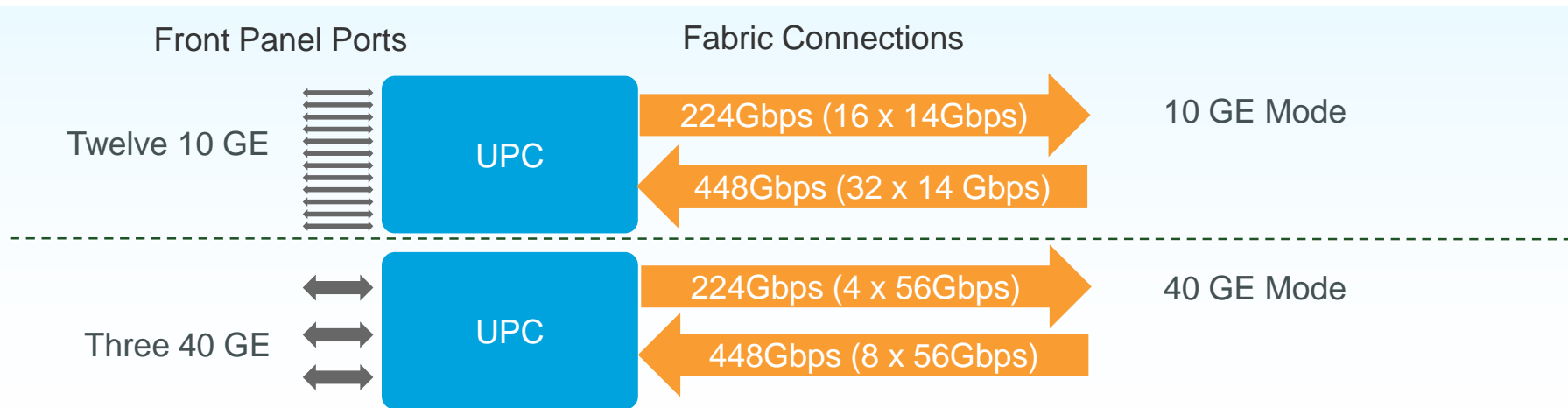
Agenda

- Introduction
- **Architecture**
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

Architecture

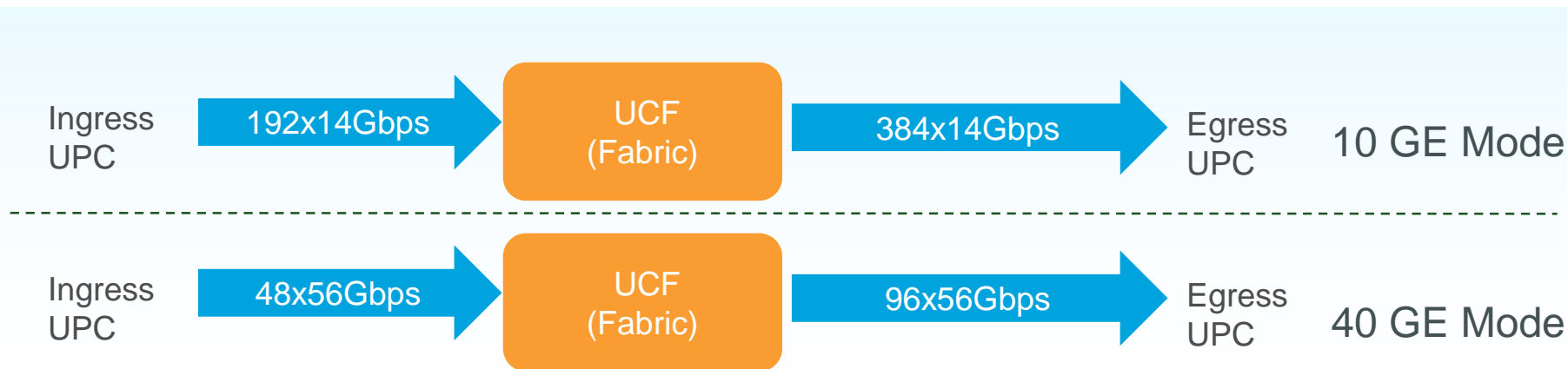
Unified Port Controller ASIC

- Multimode MAC; built-in PHY for 1, 10, and 40 GE
- Packet parsing and rewriting
- Lookup engine and access control: L2, L3, FabricPath, VxLAN, ACL, FCoE, and policing
- Buffer management, PFC for lossless traffic, queuing (Strict Priority Queuing and DWRR), and packet replication (SPAN and multicast)
- Extra fabric bandwidth for SPAN and multdestination traffic



Unified Crossbar Fabric ASIC

- Fabric: Can operate at 10 GE optimized mode or 40 GE optimized mode
- Lossless Fabric
- **In 10 GE mode:**
 - 192x384 X-BAR
 - 14Gbps per link
- **In 40GE Mode:**
 - 48x96 X-BAR
 - 56Gbps per link

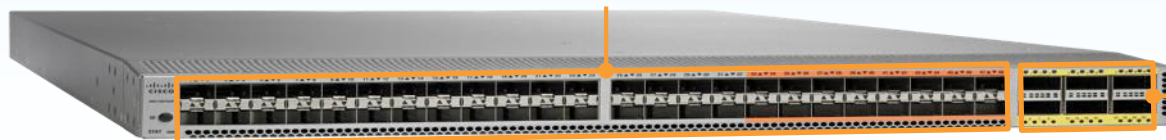


Nexus 5600 Series

Chassis Nexus 5672UP

Port Side of
Switch

48 Fixed 1/10G SFP+ Ports of which 16 Ports Unified
16 Unified Ports provide 2/4/8G FC, 10/40G Ethernet/FCoE
Traditional Ethernet or FCoE or FC



6x 40G
QSFP+ Ports
Flexibility to use
4x10G or 40G

Fan Side of
Switch

Redundant 1+1
AC/DC Power
Supplies



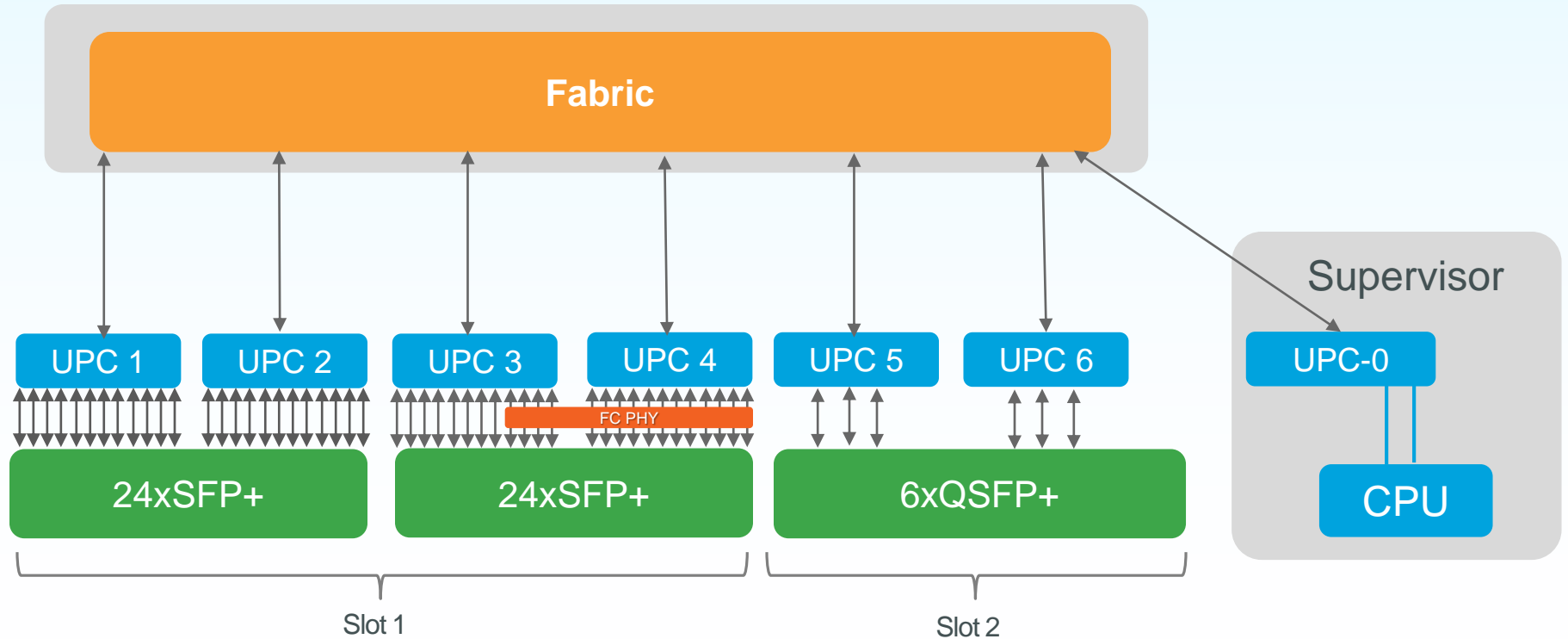
Redundant Fan
Modules

Mgmt 0,
Console, USB

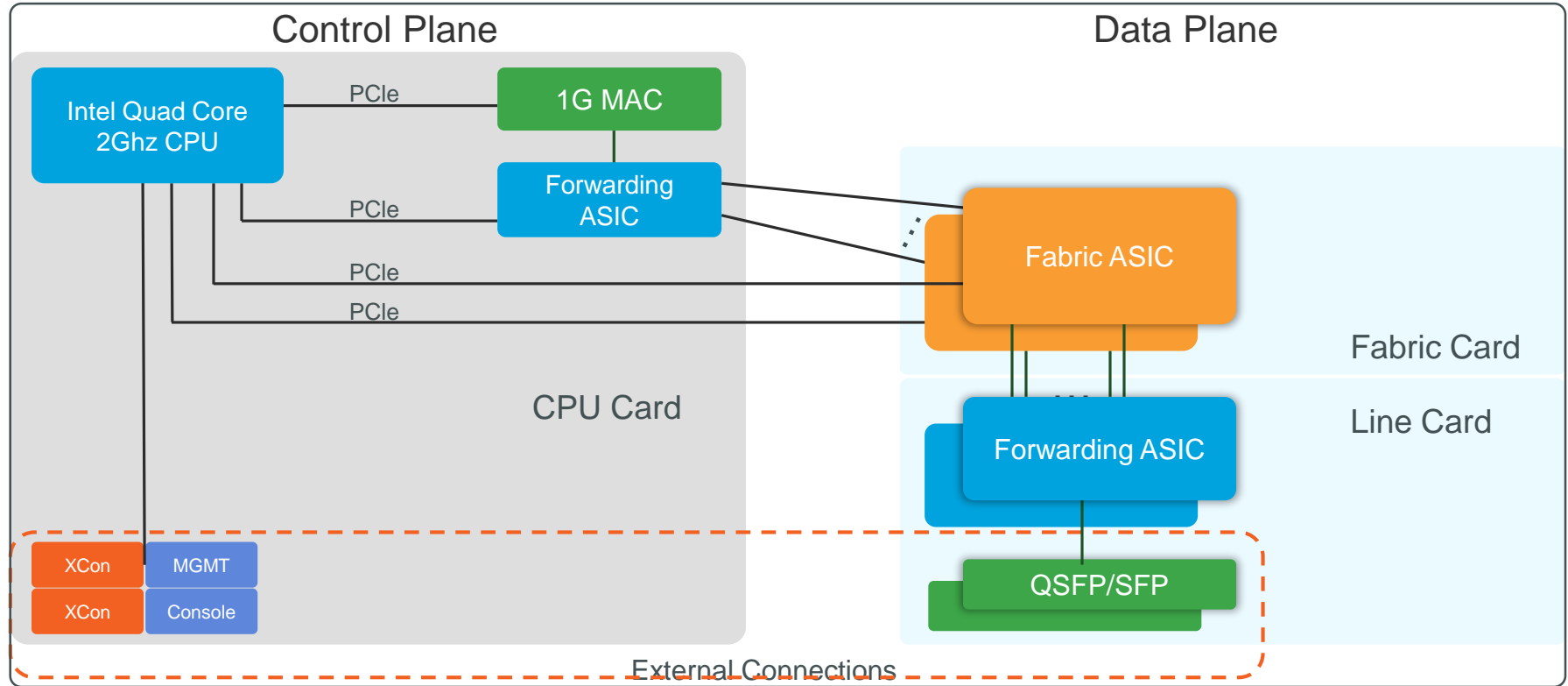
Redundant 1+1
AC/DC Power
Supplies

Compact 1RU Switch in the Nexus 5600 Platform

Nexus 5672UP Internal Architecture



Nexus 5600 Supervisor Block



Nexus 5600 Series

Chassis Nexus 56128P

Port Side of
Switch

48 Fixed 1/10G SFP+ Ports

2 Expansion Slots

Expansion Slots can be 24 port 10G Ethernet/FCOE or 2/4/8 FC with 2
40G QSFP+ Ethernet/FCOE



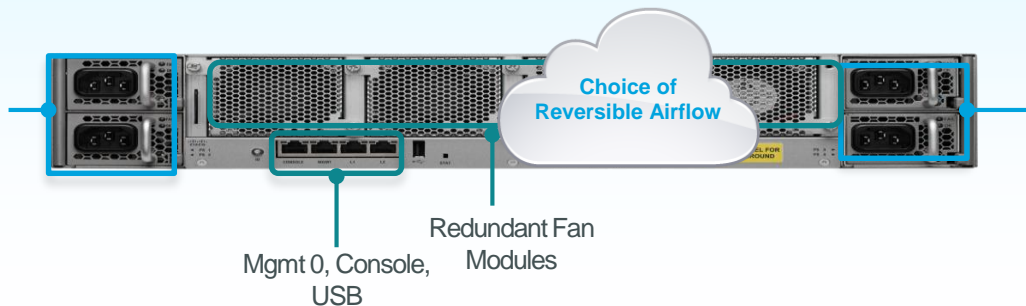
Base Chassis

4x 40G
QSFP+ Ports

Flexibility to use
4x10G or 40G

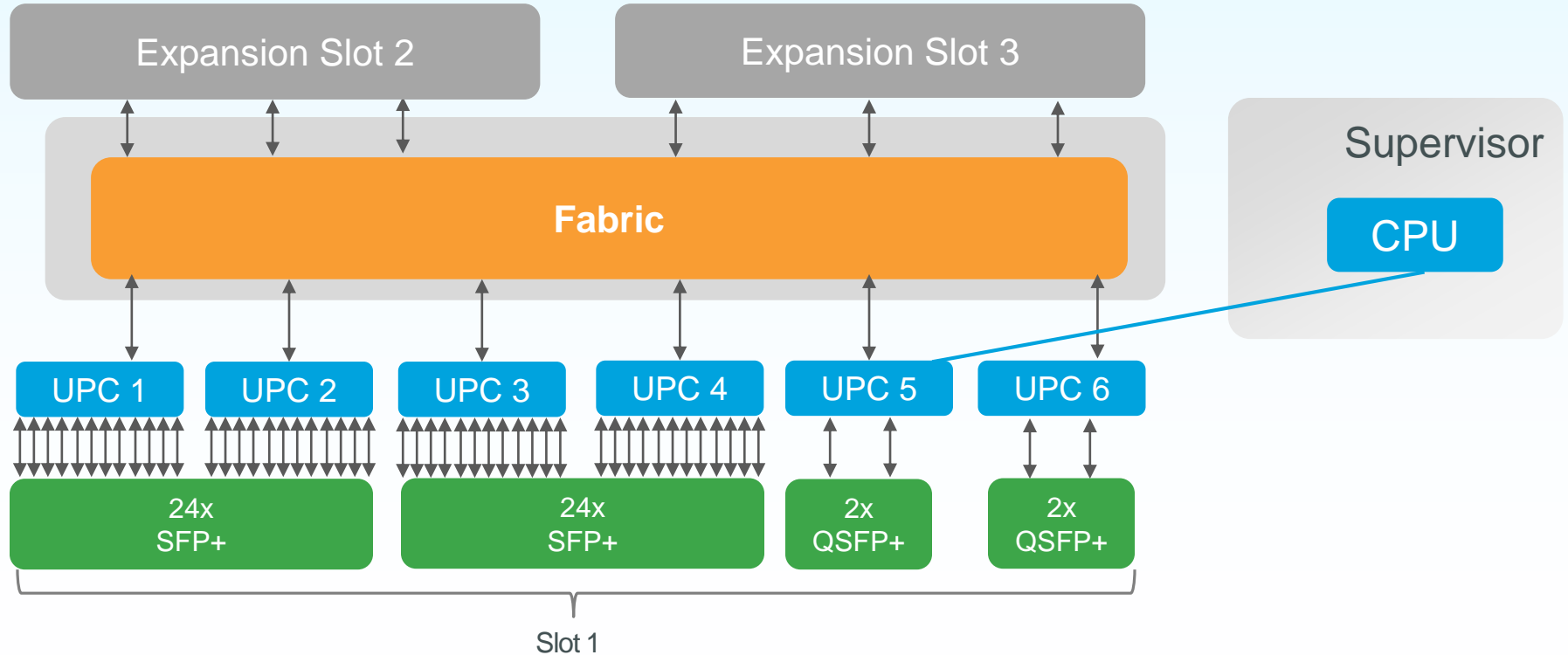
Fan Side of
Switch

Redundant 1+1
AC/DC Power
Supplies

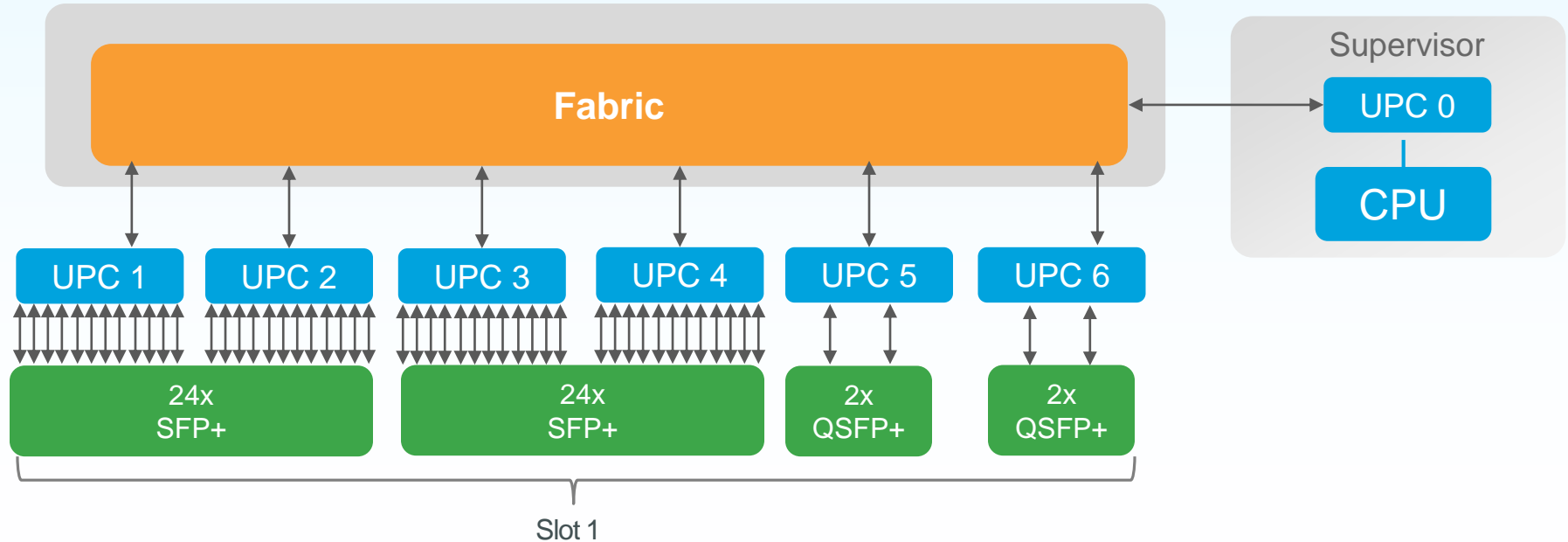


Redundant 1+1
AC/DC Power
Supplies

Nexus 56128P Internal Architecture

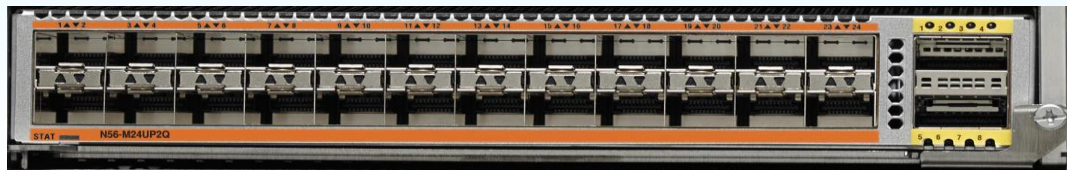


Nexus 6001 Internal Architecture



Nexus 5600 Series

Nexus 56128 Expansion Module



2 QSFP ports:

- Native 40G or 4x 10GE
- Ethernet and FCoE

Offers same performance, features, and scalability as fixed ports

24 UP ports

- 10G or 1G
- Ethernet, FCoE, and FC

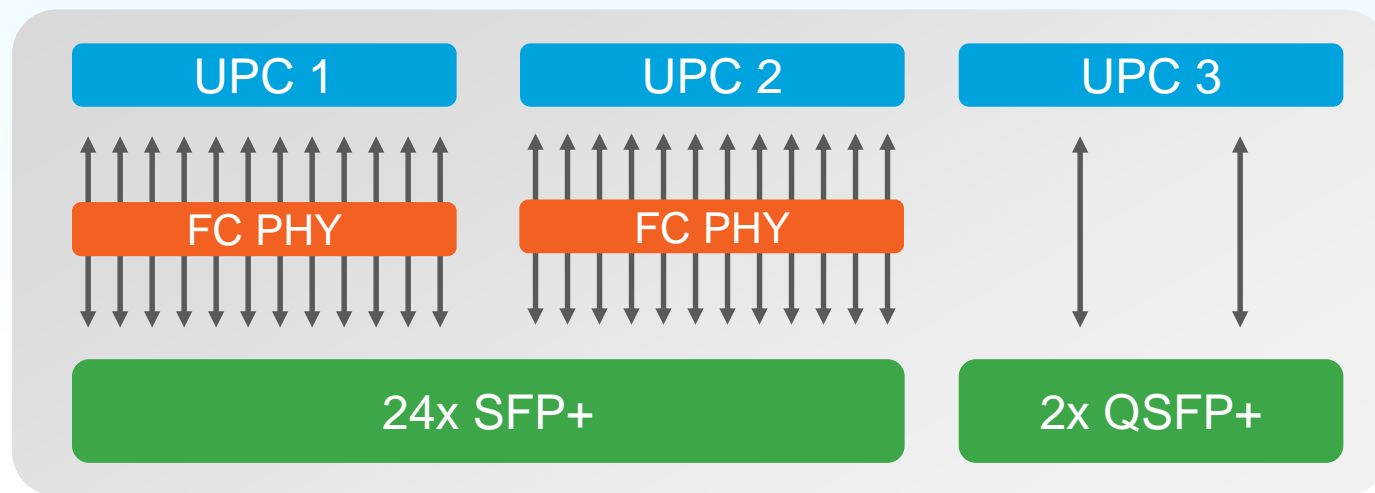
Supports OIR *

* shutdown the interfaces on the module and then “poweroff module <>” before swapping the module.

Cisco Nexus 56128P Expansion Module

M24UP2Q

To Mid-Plane / Switch Fabric



UPC Mapping

	Slot 1	Slot 2
UPC1	7	10
UPC2	8	11
UPC3	9	12

Nexus 5600 Series

Chassis Nexus 5624Q

Port Side of
Switch

Base Chassis

12 Fixed QSFP+ 40G Ports
Flexibility to use 4x10G or 40G
10/40G Ethernet/FCoE

Expansion Modules

12x 40G QSFP+ Ports
Flexibility to use 4x10G
or 40G
10/40G Ethernet/FCoE



Fan Side of
Switch

Redundant 1+1
AC/DC Power
Supplies

Choice of
Reversible Airflow

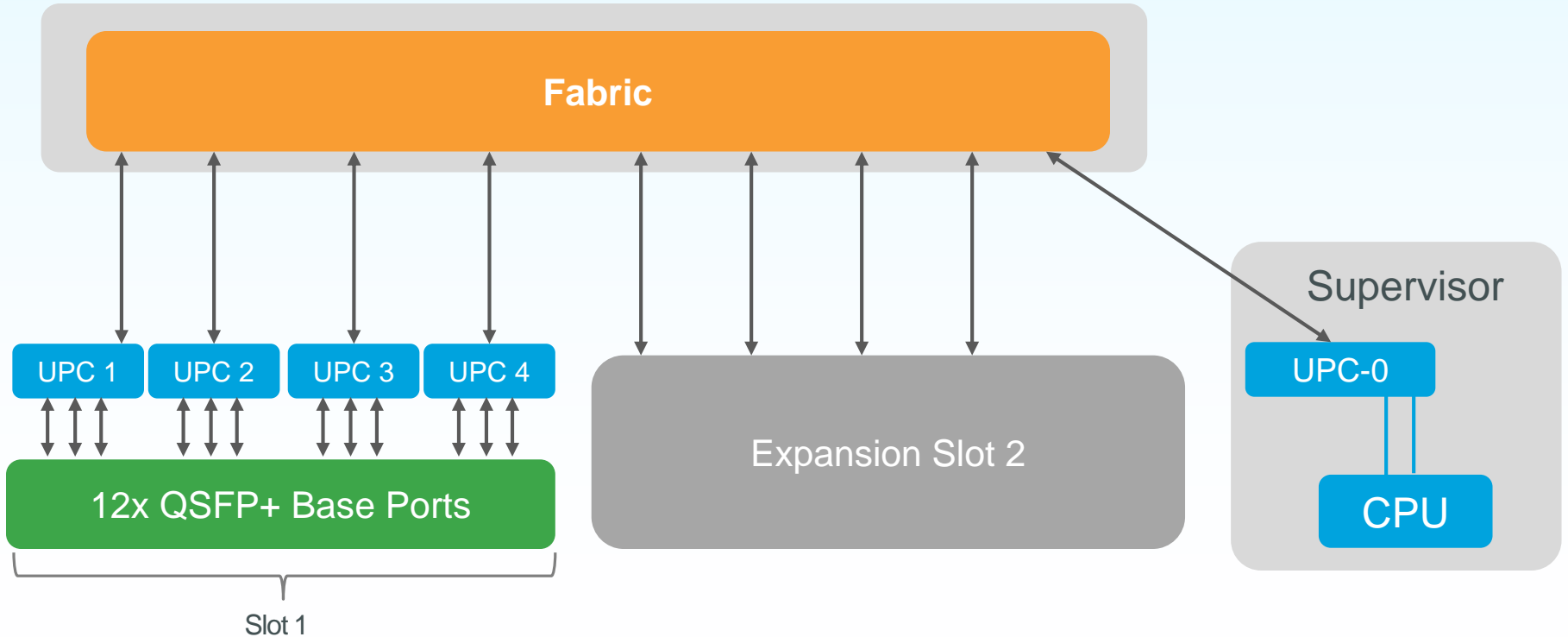
Redundant Fan
Modules

Mgmt 0, Console,
USB

Redundant 1+1
AC/DC Power
Supplies



Nexus 5624Q Internal Architecture



Nexus 5600 Series

Chassis Nexus 5648Q

Front of Switch

2 Expansion Slots

Flexibility to use 4x10G or 40G

10/40G Ethernet/FCoE



Base Chassis

24x 40G QSFP+ Ports

Flexibility to use 4x10G or 40G

10/40G Ethernet/FCoE

Back of Switch

Redundant n+1
AC/DC Power
Supplies

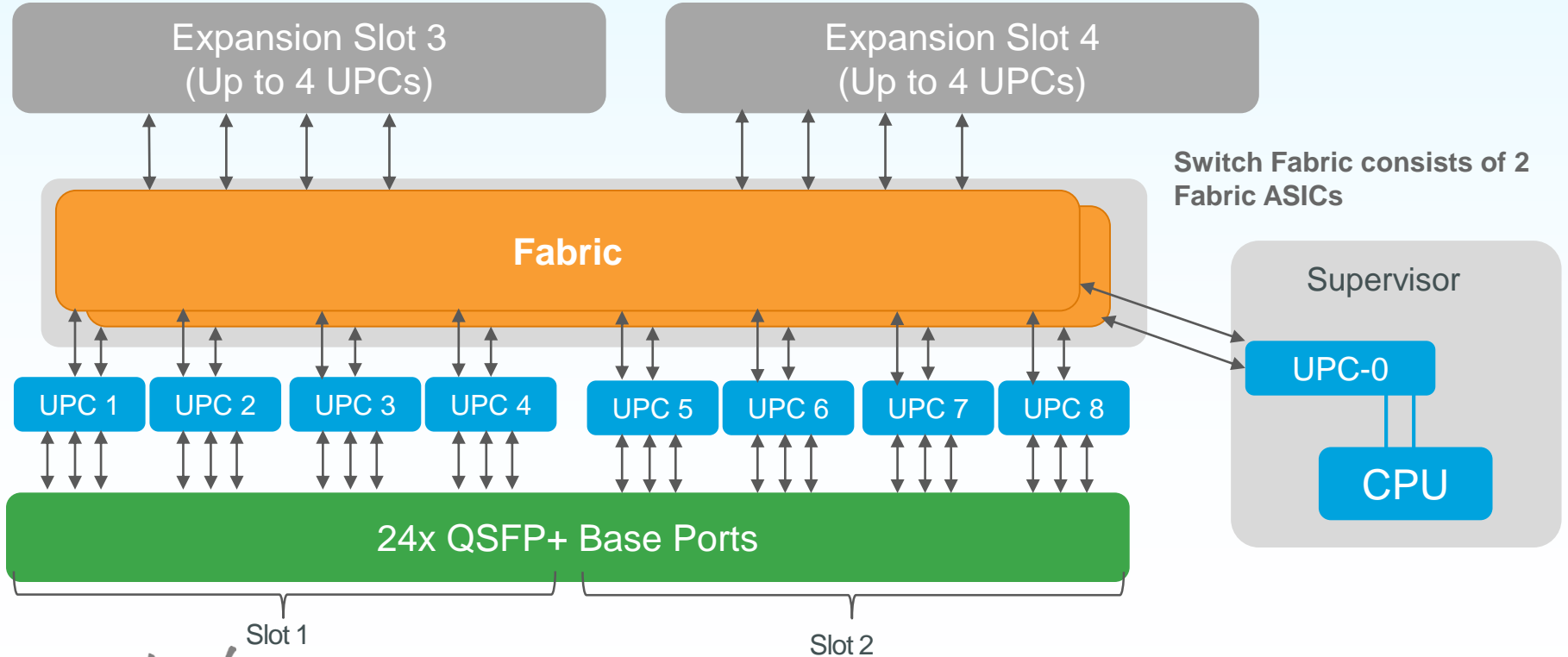


Redundant n+1
AC/DC Power
Supplies

Redundant Fan
Modules

Mgmt 0, Console,
USB

Nexus 5648Q Internal Architecture



Nexus 5624Q/5648Q Expansion Modules

N5600-M12Q



Provides 12 QSFP+ ports

Offer same performance, features, and scalability as fixed ports

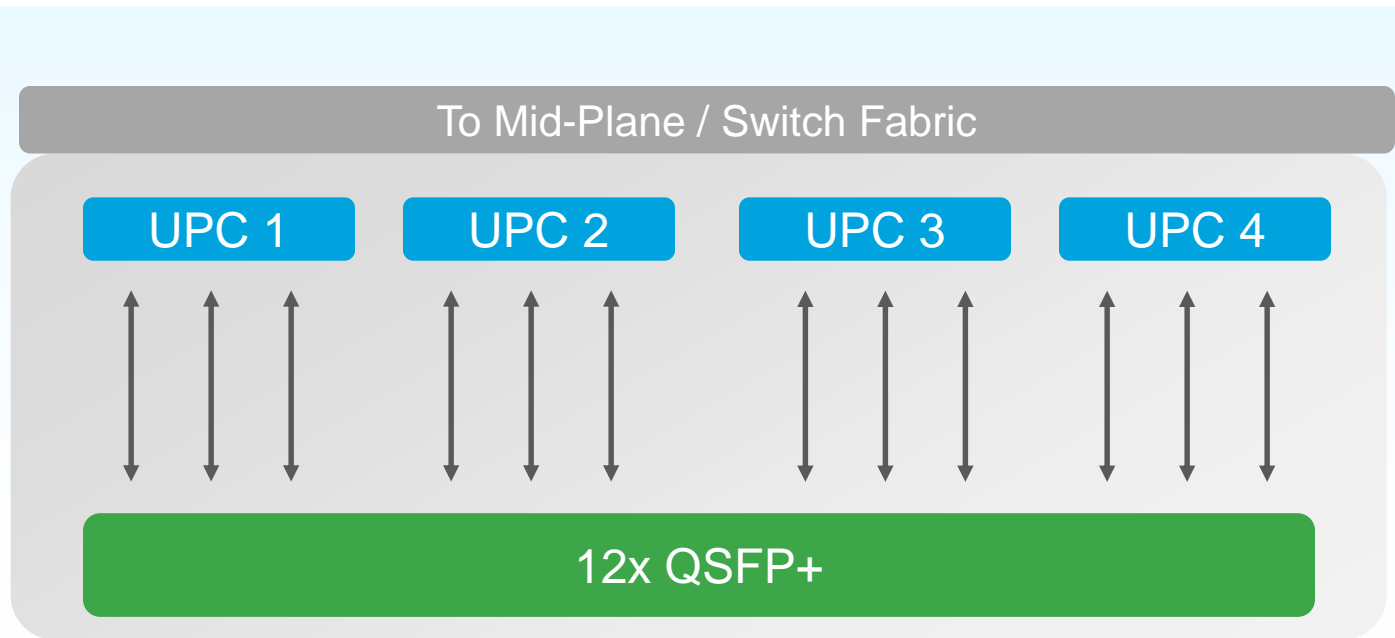
Support 40G, 10G, FCoE

Supports OIR *

* shutdown the interfaces on the module and then “poweroff module <>” before swapping the module.

Nexus 5624Q/5648Q Expansion Module

N5600-M12Q

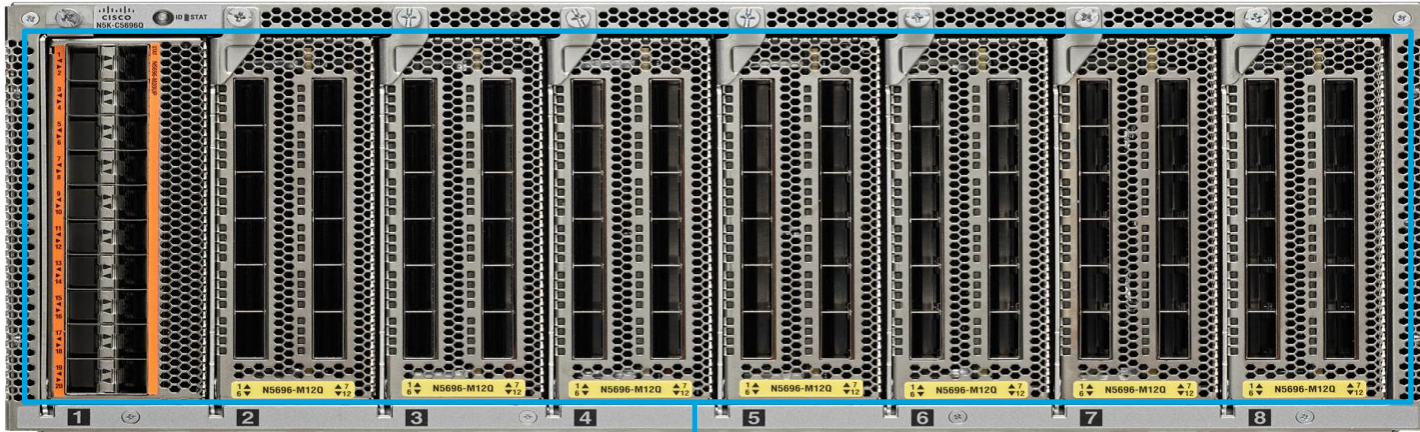


Nexus 5600 Series

Chassis Nexus 5696Q

- Upto 96 ports of 40G
- Upto 384 ports of 10G (with breakout)
- 4 RU compact Form factor

Port Side of
Switch

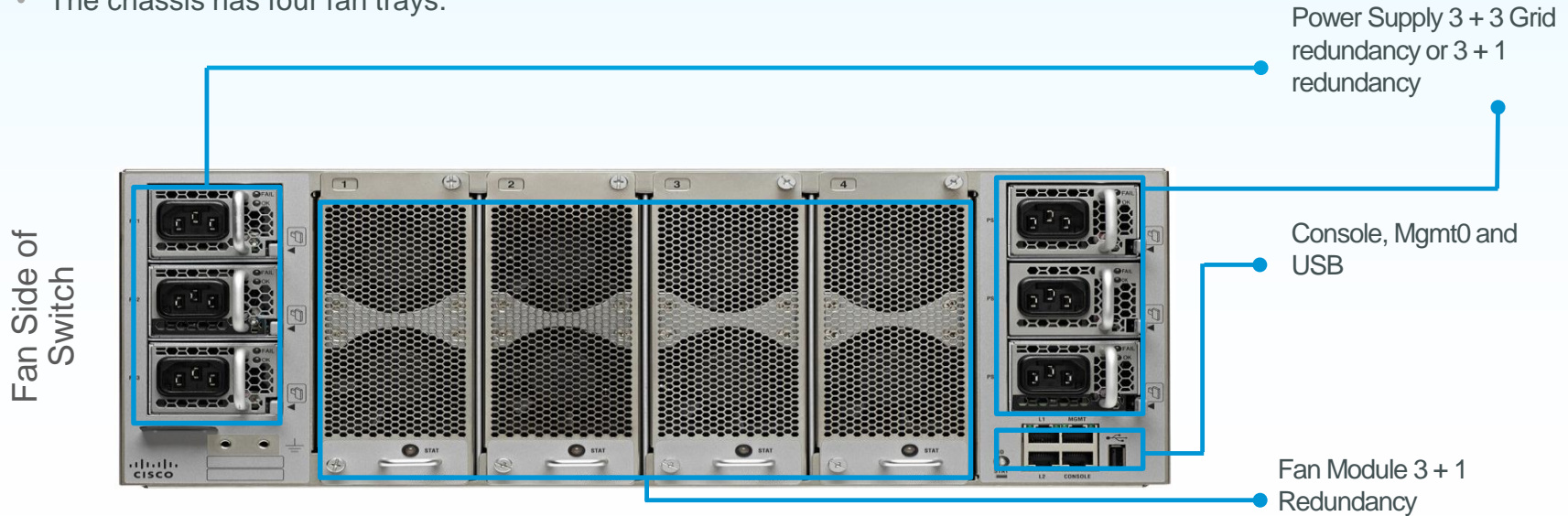


8 removable Expansion modules

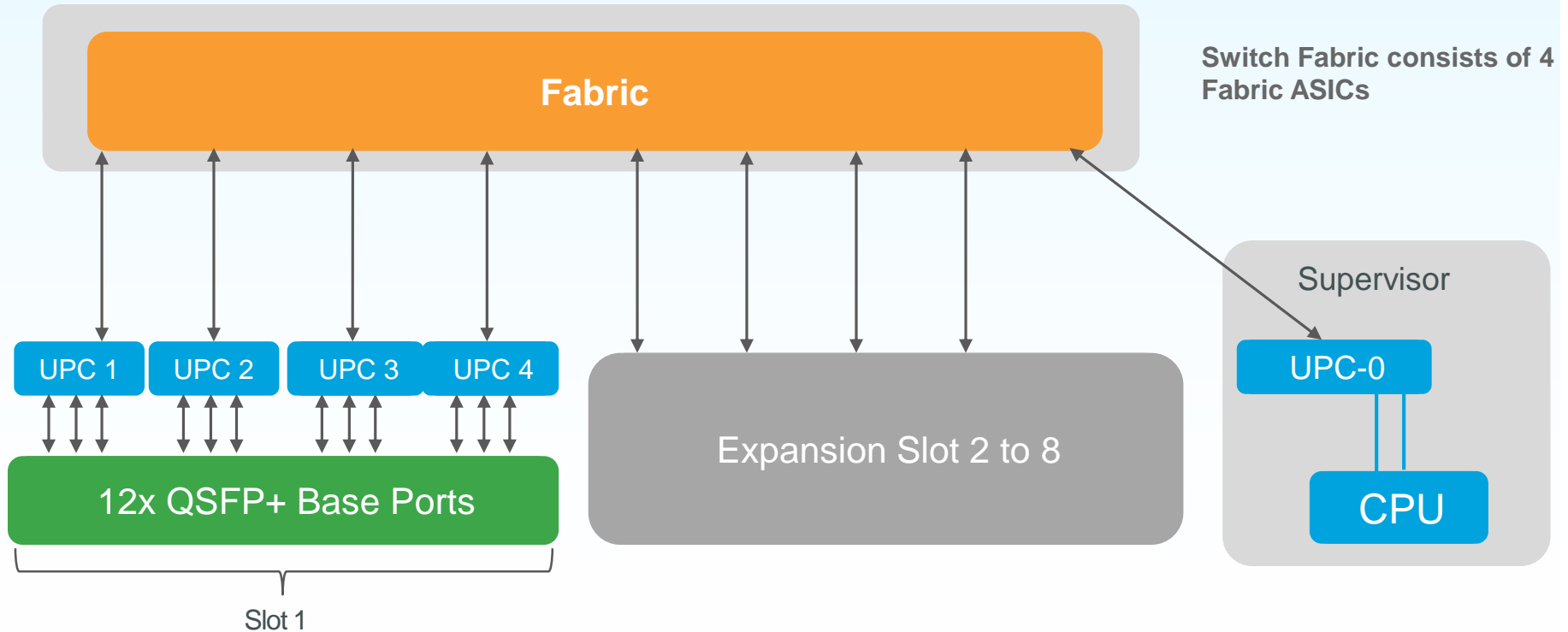
Nexus 5600 Series

Chassis Nexus 5696Q

- The chassis has six power supply slots; minimum of three is required for 40G/10G LEM and five for 100G LEM.
- Each power supply is rated 1100W, 90—240 VAC.
- The chassis has four fan trays.



Nexus 5696Q/6004 Internal Architecture



Nexus 5696Q/6004 Expansion Modules

N5696-M12Q / N6004-M12Q

Provides 12 QSFP interfaces

Supports 10/40GE Ethernet/FCoE

Offers same performance, features, and scalability as fixed ports

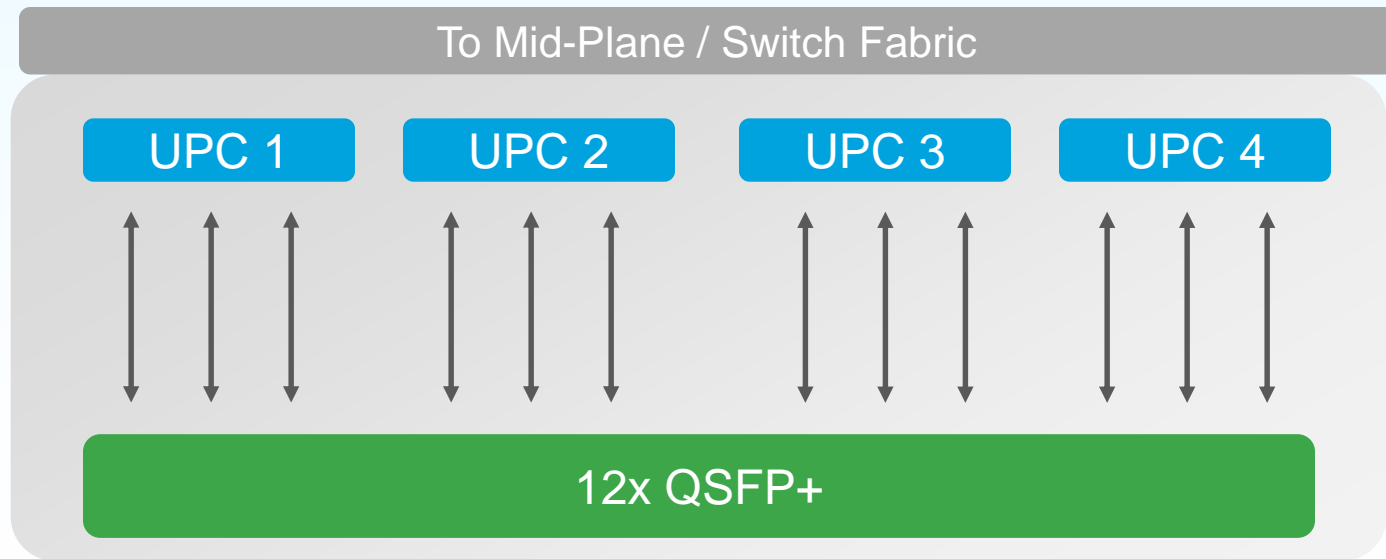
Supports OIR *



* shutdown the interfaces on the module and then “poweroff module <>” before swapping the module.

Nexus 5696Q/6004 Expansion Module

N5696-M12Q / N6004-M12Q



Nexus 5696Q/6004 Expansion Modules

N5696-M20UP / N6004X-M20UP

Provides 20 SFP/SFP+ interfaces

Supports 1/10GE Ethernet/FCoE

Supports 2/4/8G FC

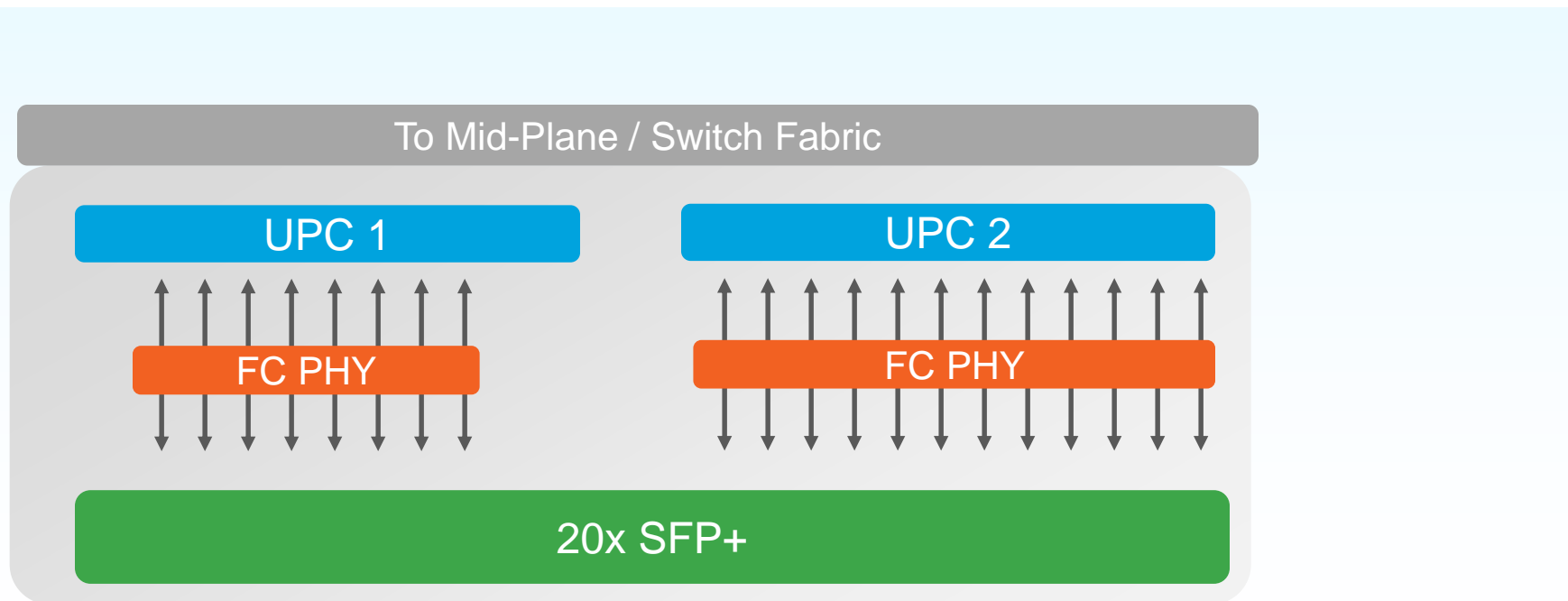
Supports OIR *



* shutdown the interfaces on the module and then “poweroff module <>” before swapping the module.

Nexus 5696Q/6004 Expansion Module

N5696-M20UP / N6004X-M20UP



Nexus 5696Q Expansion Modules

N5696-M4C

Provides 4 CXP interfaces

Supports 100GE Ethernet
(10GE/40GE Breakout on roadmap)

Supports FCOE

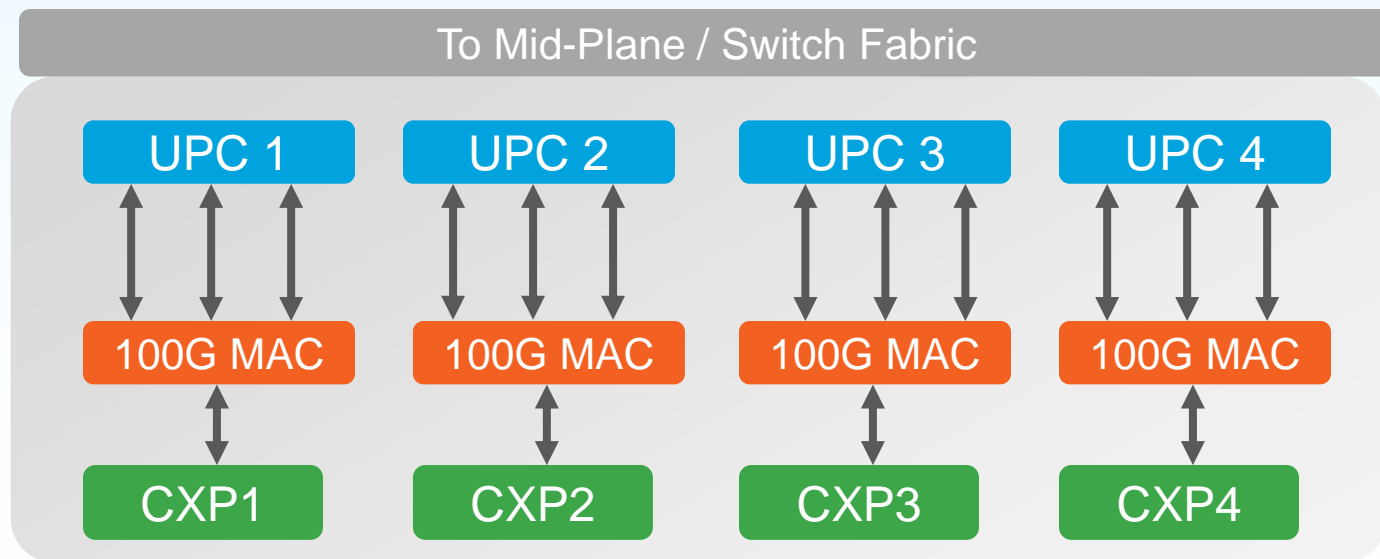
Supports OIR *



* shutdown the interfaces on the module and then “poweroff module <>” before swapping the module.

Nexus 5696Q Expansion Module

N5696-M4C



- Ingress hashing on 100G MAC is fixed with L2/L3/L4
- Subsequent release will provide configurable hashing

LEM Matrix with 5696Q and 6004s

Chassis	40G			10G		100G ¹
Chassis	6004 LEM (N6K-C6004-M12Q) V6.0(2)N1(1)	6004EF LEM (N6004-M12Q) V6.0(2)N2(2)	5696Q LEM (N5696-M12Q) V7.0(4)N1(1)	6004EF LEM (N6004X-M20UP) V7.0(1)N1(1)	5696Q LEM (N5696-M20UP) V7.0(4)N1(1)	N5696-M4C V7.1(0)N1(1a)
N6004 N6K-C6004-96Q	Yes	Yes	No	Yes	No	No
N6004EF N6K-C6004	No (Mechanical)	Yes	Yes	Yes	Yes	Yes
N5696Q	No (Mechanical)	No	Yes (VXLAN ³)	Yes ² (VXLAN ³)	Yes (VXLAN ³)	Yes (VXLAN ³)

1. 100G LEM requires minimum of 4 PS.
2. N5696Q requires a minimum SW version of 7.0(4)N1(1)
3. Minimum release to support VXLAN is 7.1(0)N1(1a)

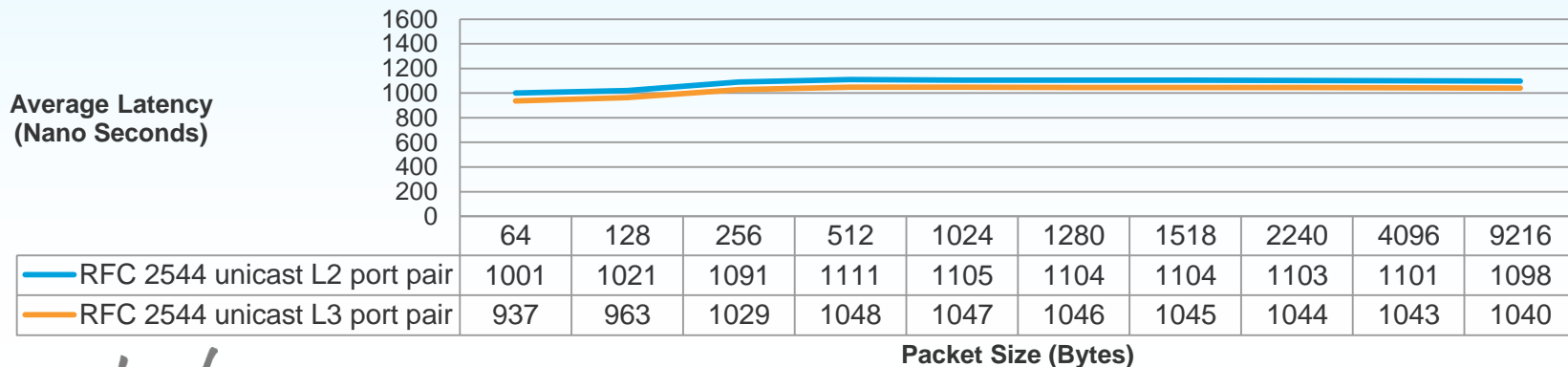
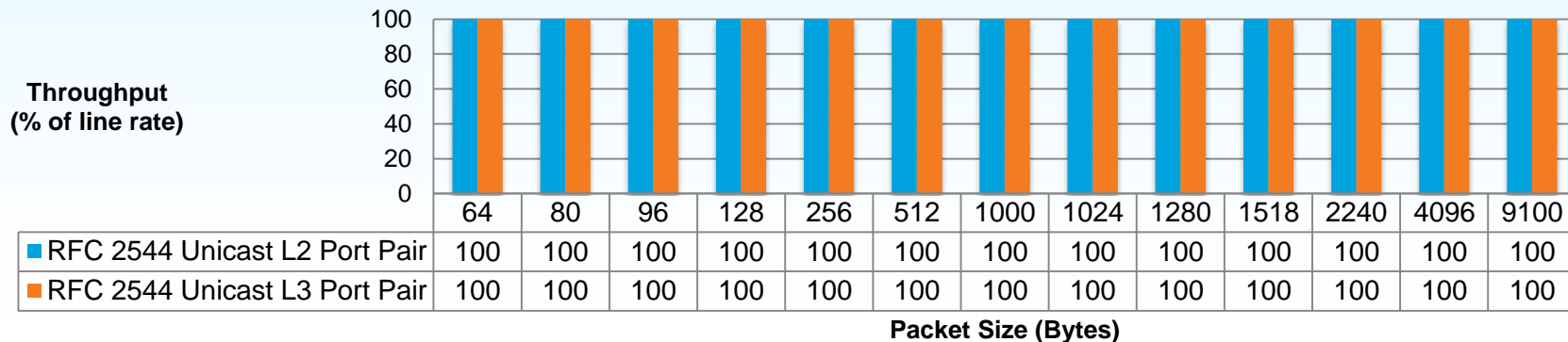
NEXUS 5600 SERIES



	Nexus 5672UP 7.0(1)N1(1)	Nexus 56128P 7.0(2)N1(1)	Nexus 5624Q 7.1(0)N1(1a)	Nexus 5648Q 7.1(1)N1(1)	Nexus 5696Q 7.0(4)N1(1)
PERFORMANCE	Switch Fabric Throughput	1.44 Tbps	2.56 Tbps	1.92 Tbps	3.84 Tbps
	Port-to-Port Latency	~1.0 us			
	Layer 3 Capability	Integrated Line-Rate			
DENSITY	Switch Footprint	1RU	2RU	1RU	2RU
	10 Gigabit Ethernet Port	72	128	96	192
	40 Gigabit Ethernet Port	6	8	24	48
SCALE	Unified Ports	16	48	N/A	N/A
	1/10 Gigabit Ethernet With FEX	1152	1152	1152	1152
	Packet Buffer	25MB per 3x40G (or 12x10G)			

* Current software supports 60 ports

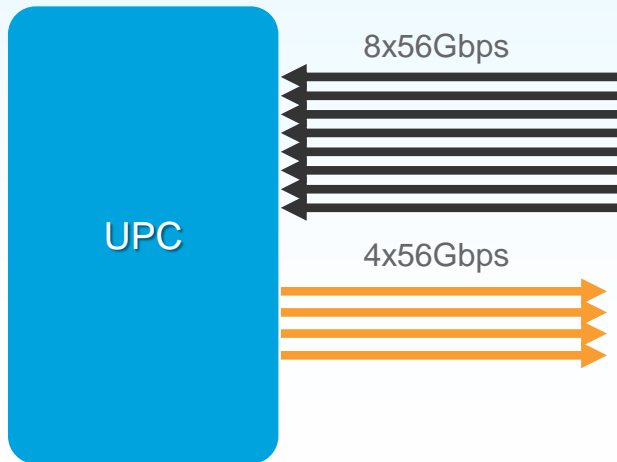
Nexus 5600/6000 - Throughput and Latency



Switch Fabric Mode

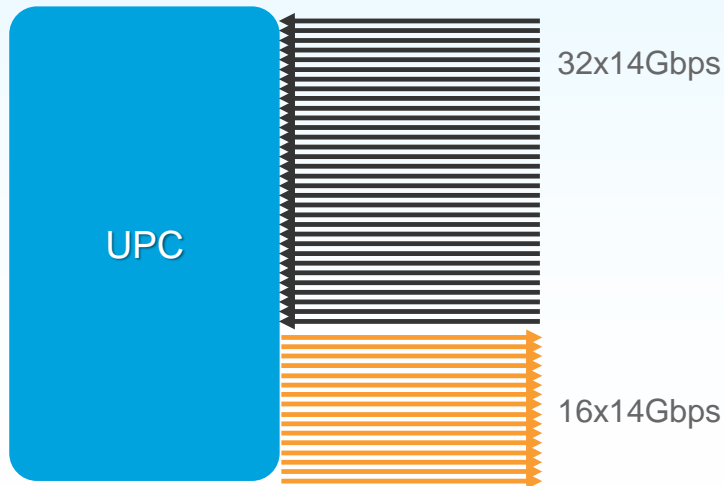
- Switch fabric mode determines the speed of a single fabric link between UPC and Fabric.
- Switch fabric mode is independent of front panel port speed
- Default switch fabric mode is 10 GE for 5672UP/56128P and 40 GE for 5624Q/5648Q/5696Q
- Fabric Mode Configuration is system wide

40 GE Switch Fabric Mode



Cisco *live!*

10 GE Switch Fabric Mode



Switch Fabric Mode and Throughput

- 40 GE switch fabric mode

- Support for 40 GE flow

- 10 GE switch fabric mode

- Cannot support >10-Gbps flow
- Line rate between 10 GE and 40 GE interfaces (no dependency on load balance)
- Throughput between 40 GE ports depends on hashing result

40 GE Fabric Mode			10 GE Fabric Mode		
Egress Ingress	10 GE	40 GE	Egress Ingress	10 GE	40 GE
10 GE	Line-Rate	Line-Rate	10 GE	Line-Rate	Line-Rate
40 GE	Line-Rate	Line-Rate	40 GE	Line-Rate	<10GE Flow Throughput depends on hashing

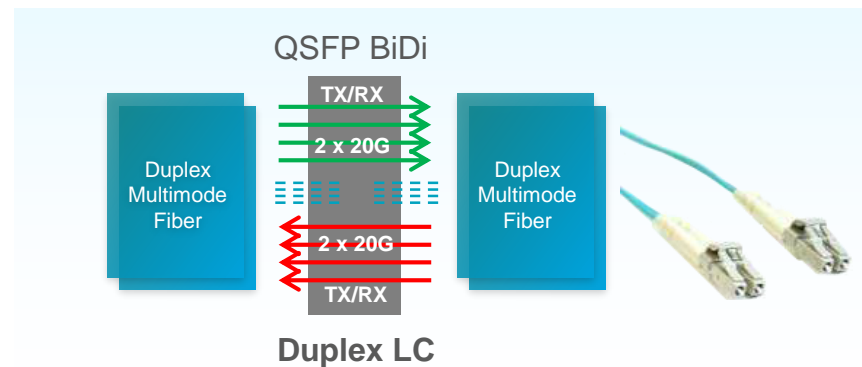
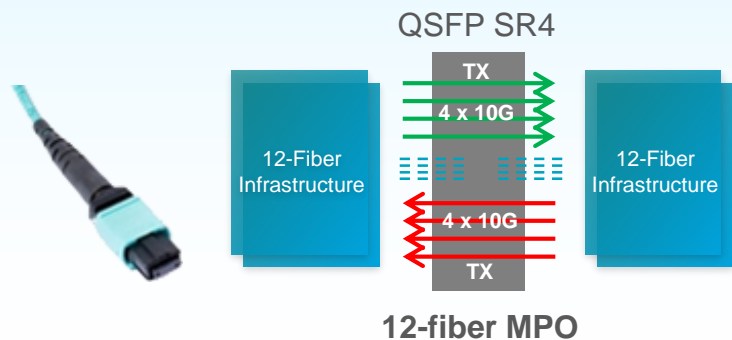
Switch Fabric Mode and Latency

- With 40 GE fabric, traffic from 10 GE to 10 GE may fall back to store-and-forward with a heavy traffic load.
- Recommendation is 10 GE fabric when most of the ports are 10 GE and latency is important.

40 GE Fabric Mode			10 GE Fabric Mode		
Ingress \ Egress	10 GE	40 GE	Ingress \ Egress	10 GE	40 GE
10 GE	Cut-through and store-and-forward	Store-and-forward	10 GE	Cut-through	Store-and-forward
40 GE	Cut-through	Cut-through	40 GE	Cut-through	Store-and-forward

Enabling 10 to 40GbE Transition

QSFP BiDi



- Utilize existing duplex fiber commonly deployed in 10G environment today
- 100m with OM3 grade fiber and 125m with OM4 fiber (850 nm)
- IEEE 802.3ba Compliant
- Power consumption < 3.5w
- 75% average savings over parallel fiber for new deployments



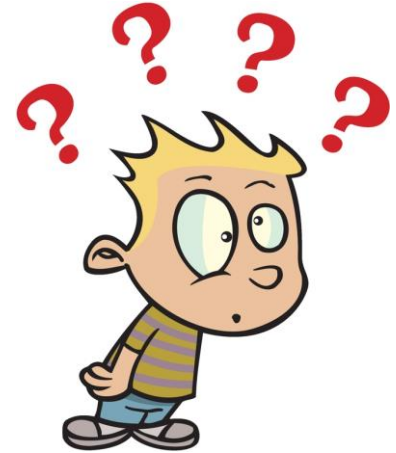
Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

Forwarding

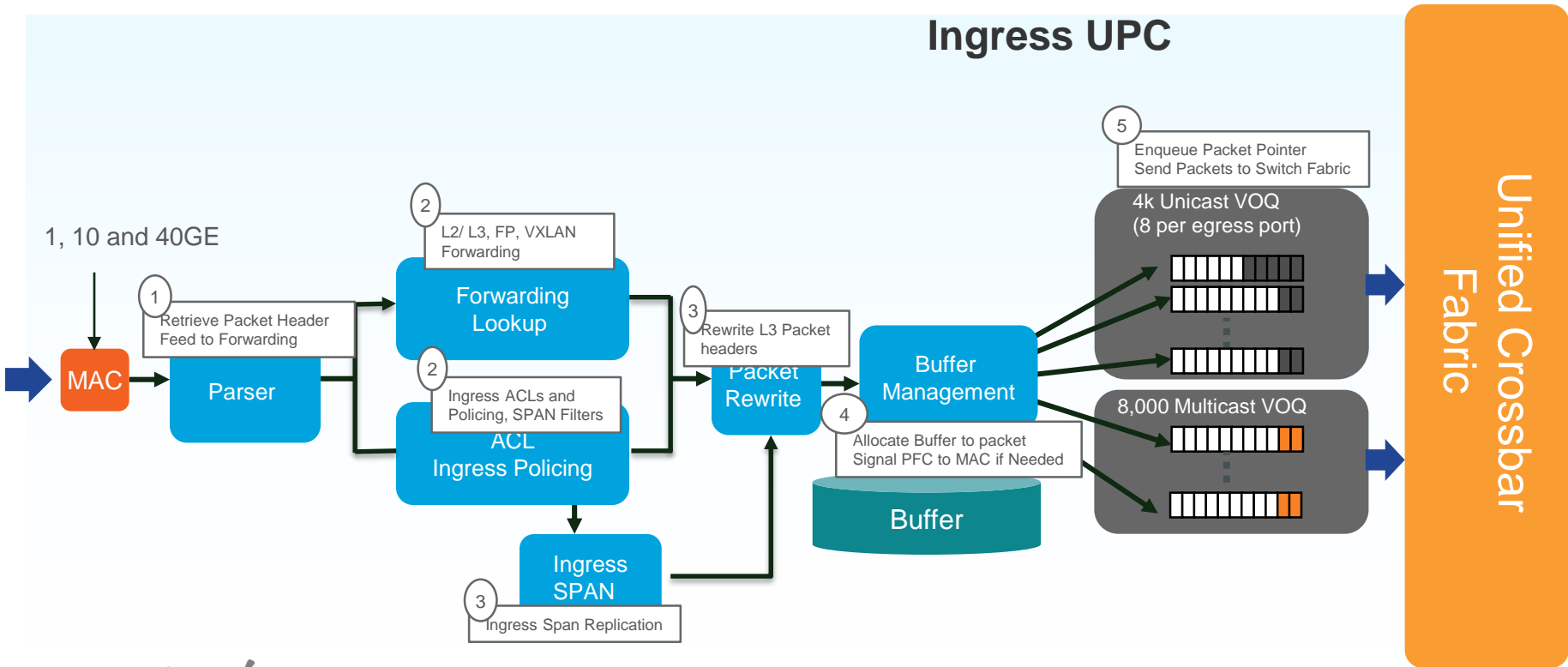
What is a VoQ?

- **Virtual Output Queues (VOQs)** - Provides buffering and queuing for ingress-buffered switch architecture
- Avoids Head of Line Blocking for unicast packets
- 576 ports with 8 class of service: 4608 queues per UPC (or)
- 1152 ports with 4 classes of service: 4608 queues per UPC



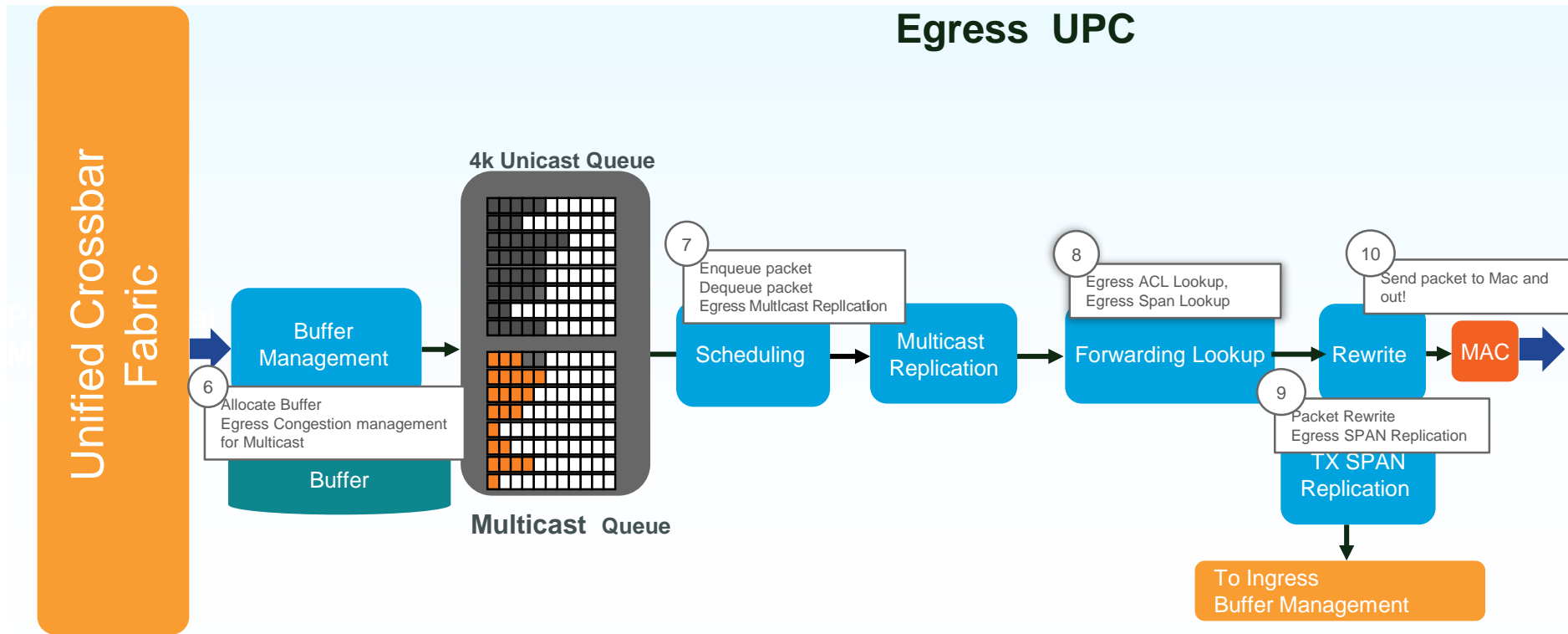
Nexus 5600 Packet Processing Flow

Ingress Pipeline

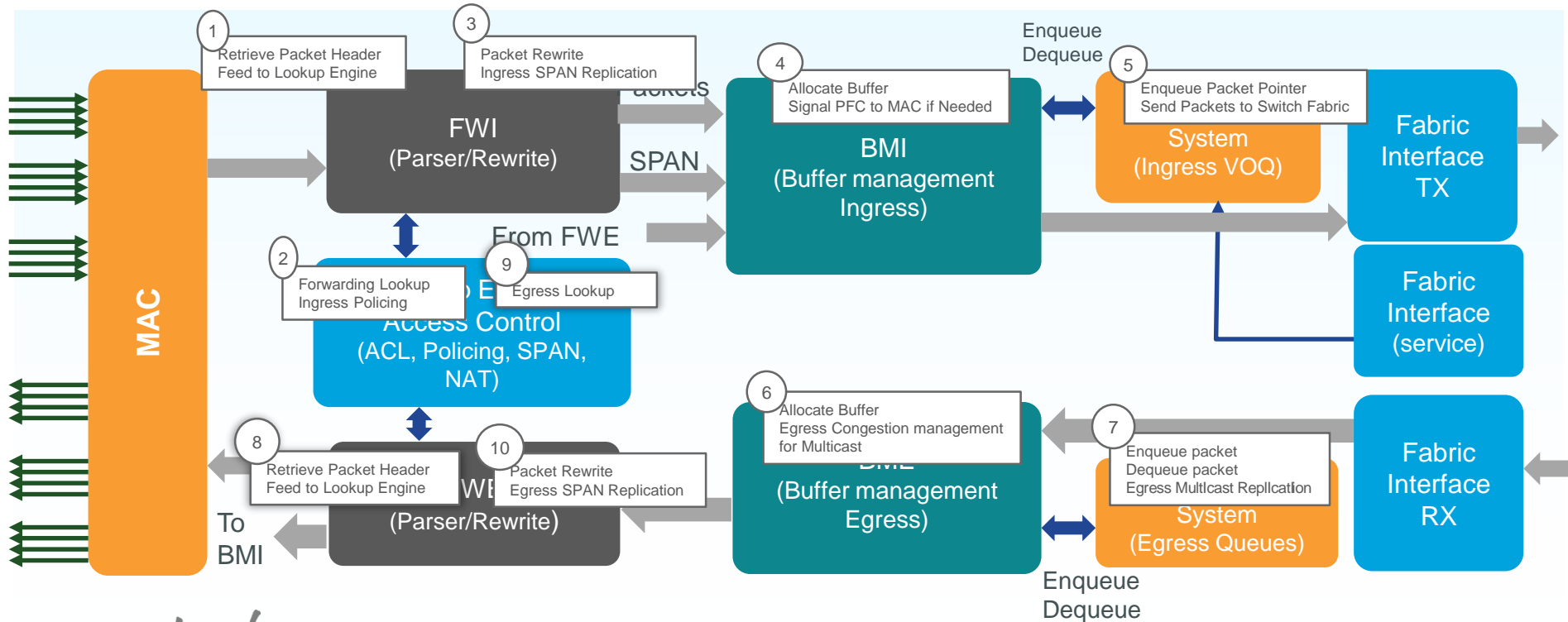


Nexus 5600 Packet Processing Flow

Egress Pipeline

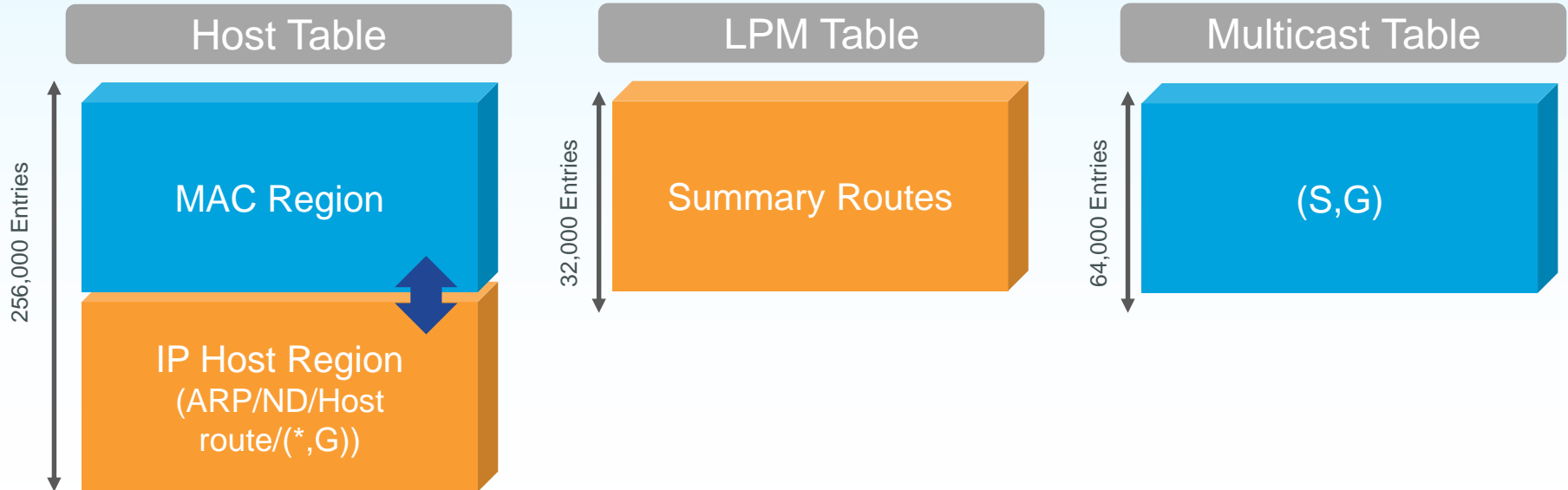


UPC Functional Blocks and Packet Walk



Nexus 5600 Key Forwarding Tables

- Host table: 256,000 -entry hashing table
- Host table: Shared between MAC, ARP, and ND and /32 host route
- Host table FCS carving: 128,000 MAC, 128,000 IP host
- LPM table: 32,000 entries. Also known as summary routes
- Mroute table: 64,000 entries



* Hardware table size. Please check configuration limit for software scaling

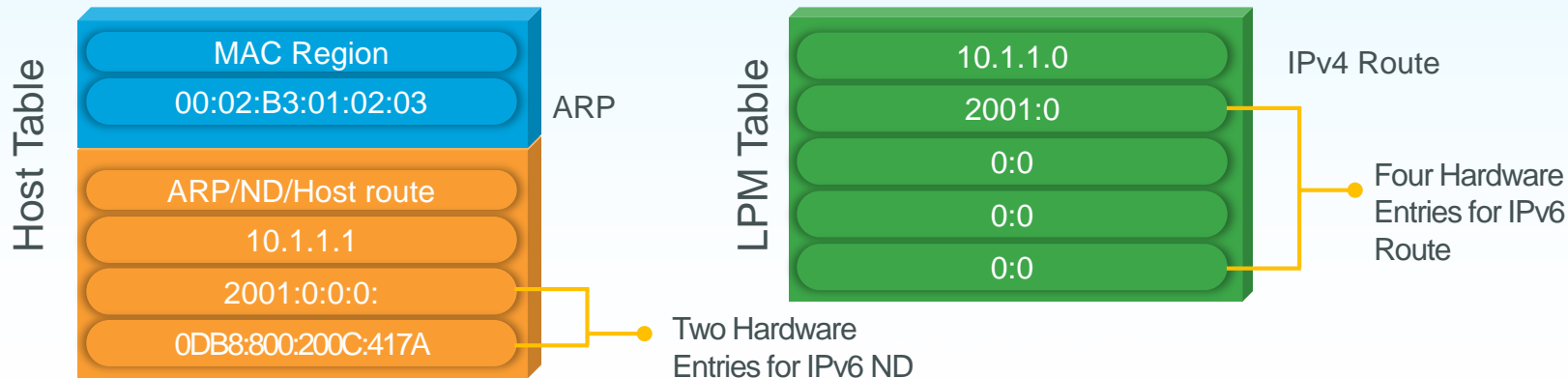
Nexus 5600 Unicast Table Scaling

- Each IPv6 ND(Neighbor Discovery) entry consumes two entries in host table.
- Each IPv6 route consumes four entries in LPM table.

10.1.1.0/24
2001::/64

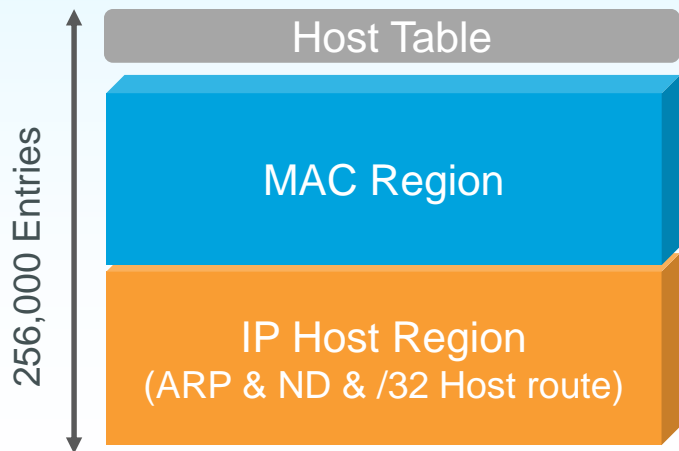


MAC: 00:02:B3:01:02:03
IPv4: 10.1.1.1
IPv6: 2001::0DB8:800:200C:417A



Nexus 5600 Host Table

- N5600 has a 256k host table divided in two parts:
 - MAC region
 - IP region: holds ARP, IPv6 ND, /32 Host Routes
- Default sizes are 128k for MAC region and 128k for IP region.



Deployment Scenario	Scalability
L2 switch	256,000 MAC
L2/L3 gateway with IPv4 only	128,000 virtual machines
L2/L3 gateway with IPv6 only	85,000 virtual machines
L2/L3 gateway with dual stack	64,000 virtual machines

Host Table Carving Profiles

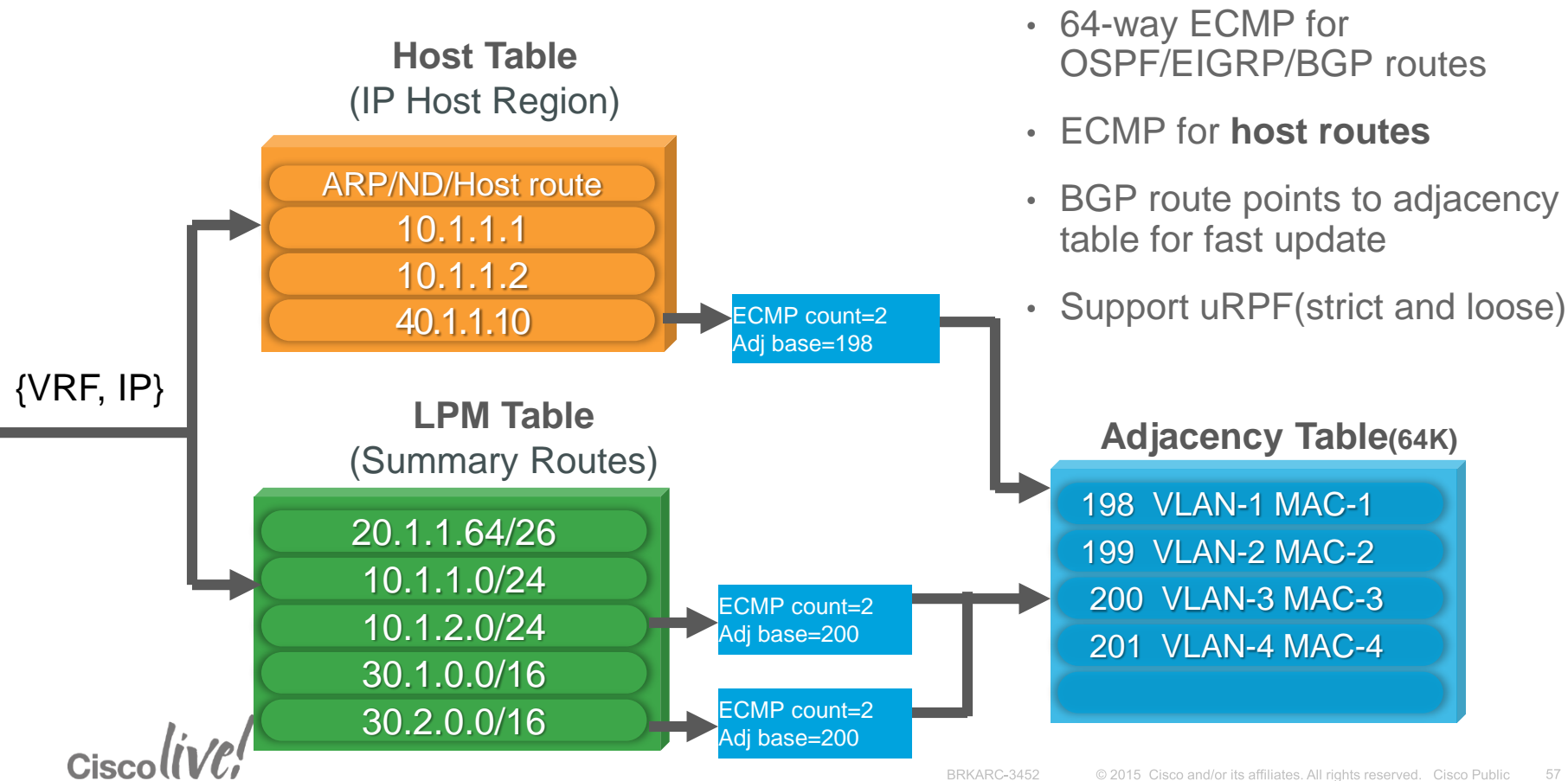
- Use of pre-defined carving profiles.
- HRT = Host Route Table = IP host region.
- STM = Station Table Management = MAC region.

Template Profile	Description
hrt-128-stm-128	HRT size: 128k, STM size: 128k (default profile)
hrt-96-stm-160	HRT size: 96k, STM size: 160k
hrt-64-stm-192	HRT size: 64k, STM size: 192k
hrt-32-stm-224	HRT size: 32k, STM size: 224k

MAC/ARP Resource Carving CLI

- Specify the resource template to use:
 - `switch(config)# hardware profile route resource service-template template-name`
 - Need to save the config and reload the switch to be applied.
- Show commands:
 - `show hardware profile route resource template`
 - `show hardware profile route resource template default`
 - `show running-config hardware profile route resource template`
 - `show startup-config hardware profile route resource template`

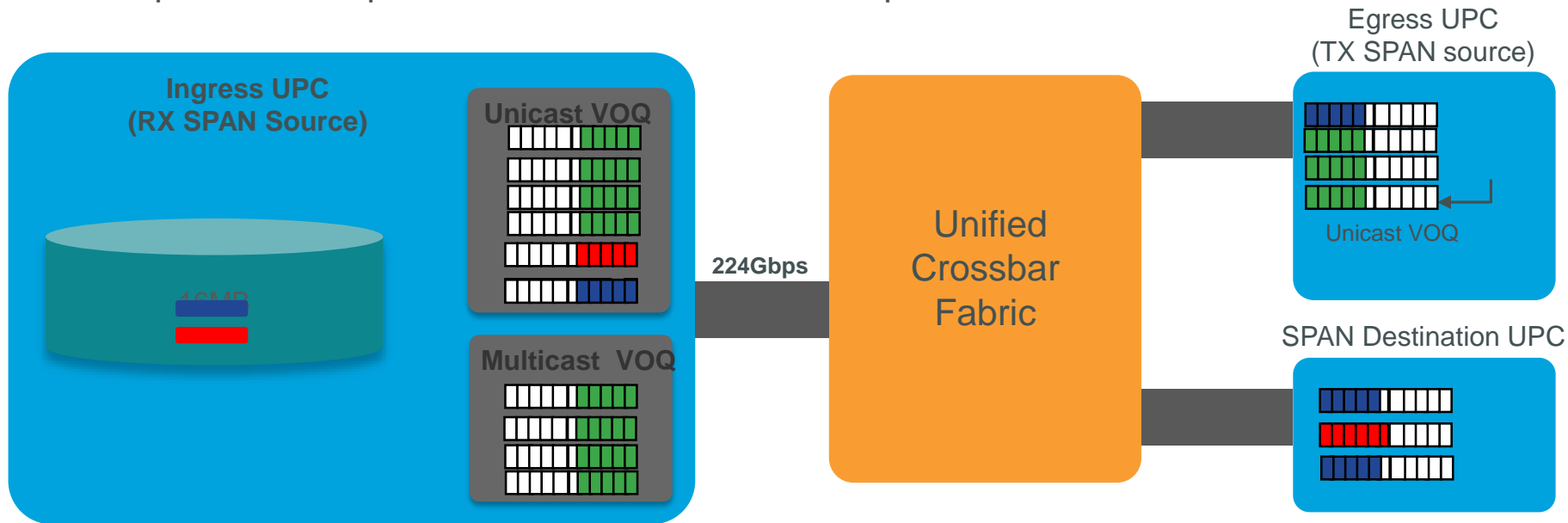
Unicast L3 Forwarding Lookup



- 64-way ECMP for OSPF/EIGRP/BGP routes
- ECMP for **host routes**
- BGP route points to adjacency table for fast update
- Support uRPF(strict and loose)

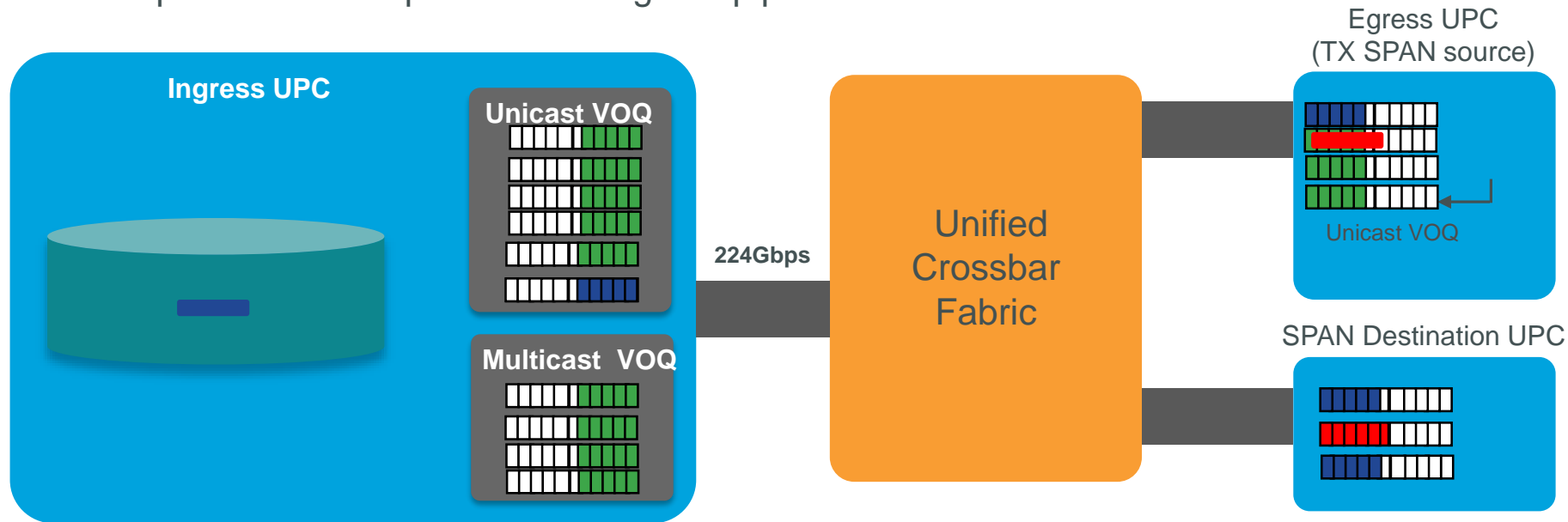
Ingress SPAN Packet Flow

- Data is replicated at ingress port ASIC-Unified Port Controller (UPC).
- SPAN packets are queued at the SPAN destination port VOQ.



Egress SPAN Packet Flow

- SPAN copy is made at egress pipe of the TX SPAN source port.
- SPAN packets are looped back to ingress pipe of UPC and sent to switch fabric.

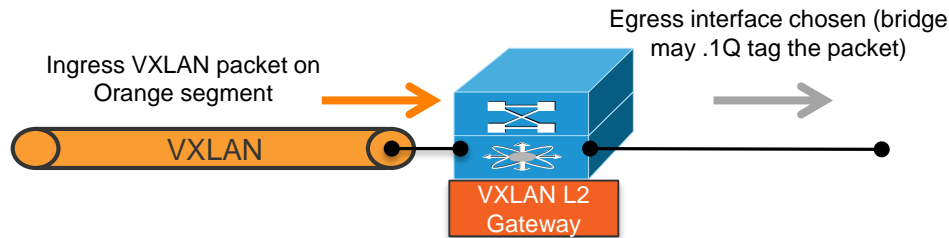


VxLAN on Nexus 5600

Line-Rate Bridging and Routing

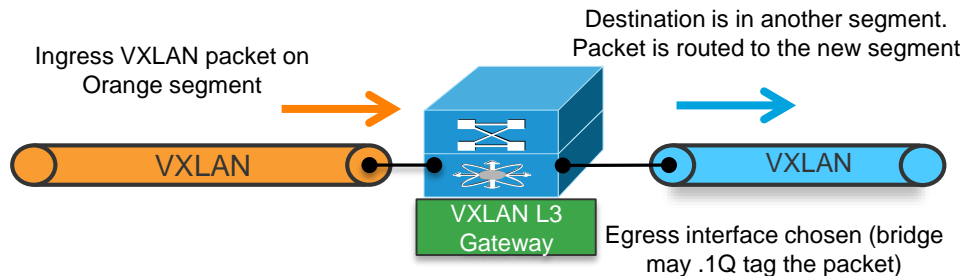
- VxLAN to VLAN Bridging (L2 Gateway)

- Single Pass Through UPC
- Line-Rate Bridging



- VxLAN to VxLAN Routing (L3 Gateway)

- Single Pass Through UPC
- Line-Rate Routing
- Key Differentiator



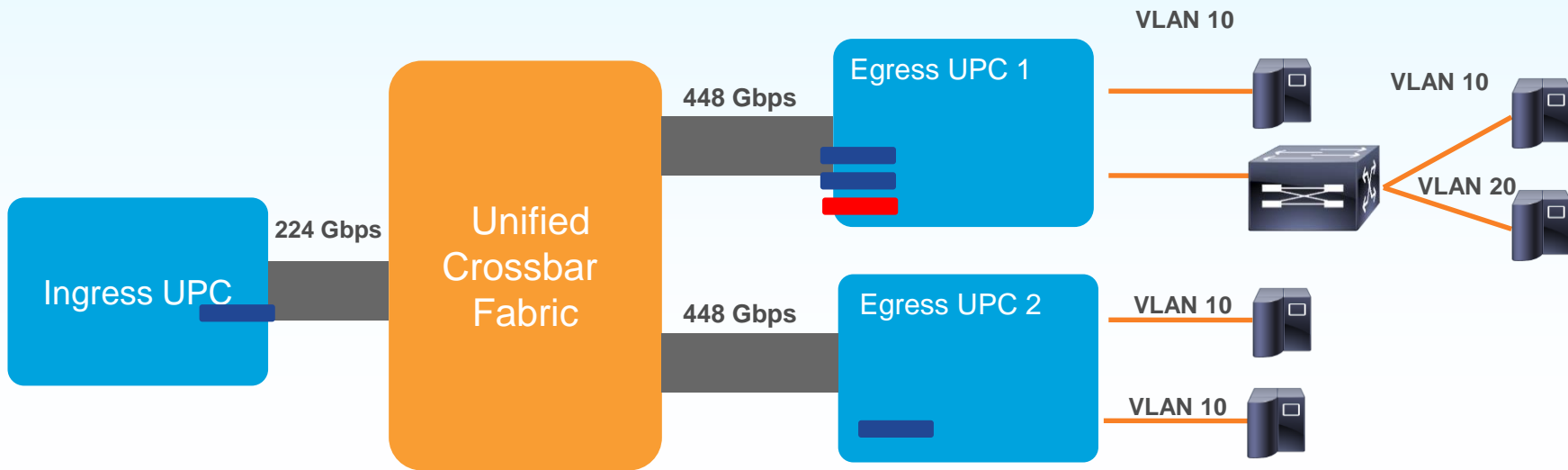
Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

Multicast

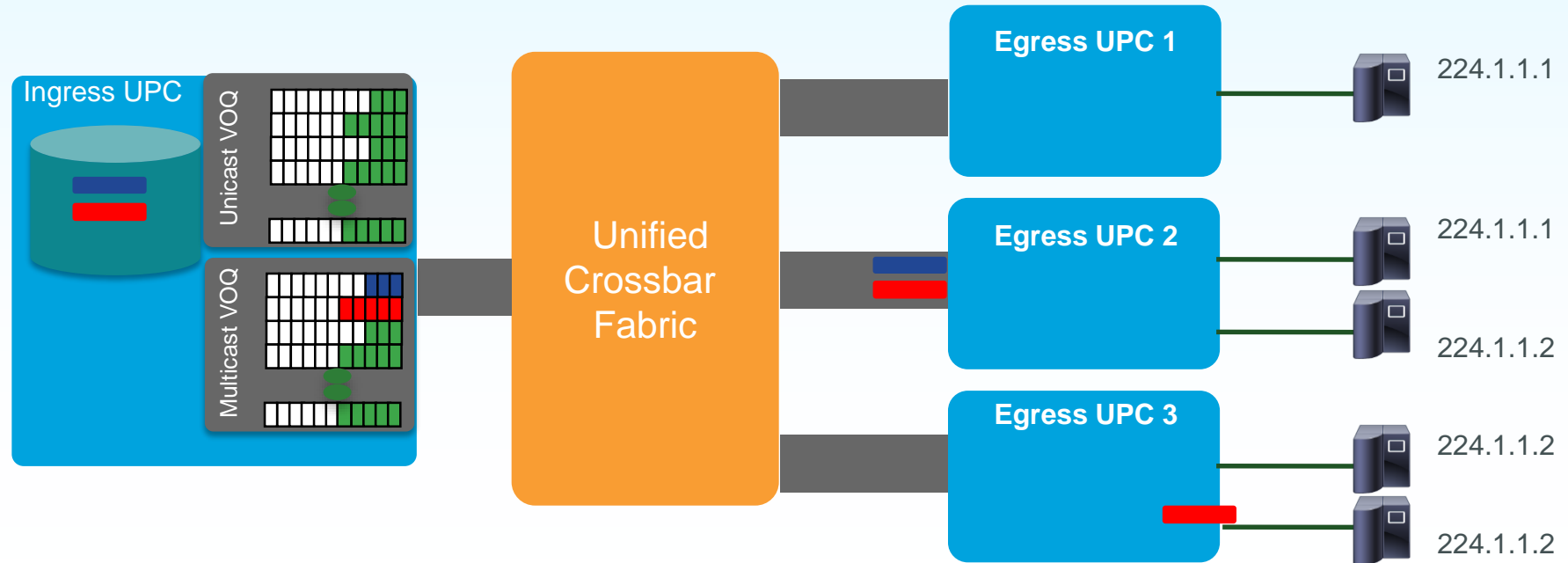
Efficient Multicast Replication

- Optimized multicast replication throughout the system
- Fabric replication and egress replication; one copy is replicated to egress UPC, where there is a receiver
- Line-rate multicast replication at fabric and egress UPC for all frame sizes



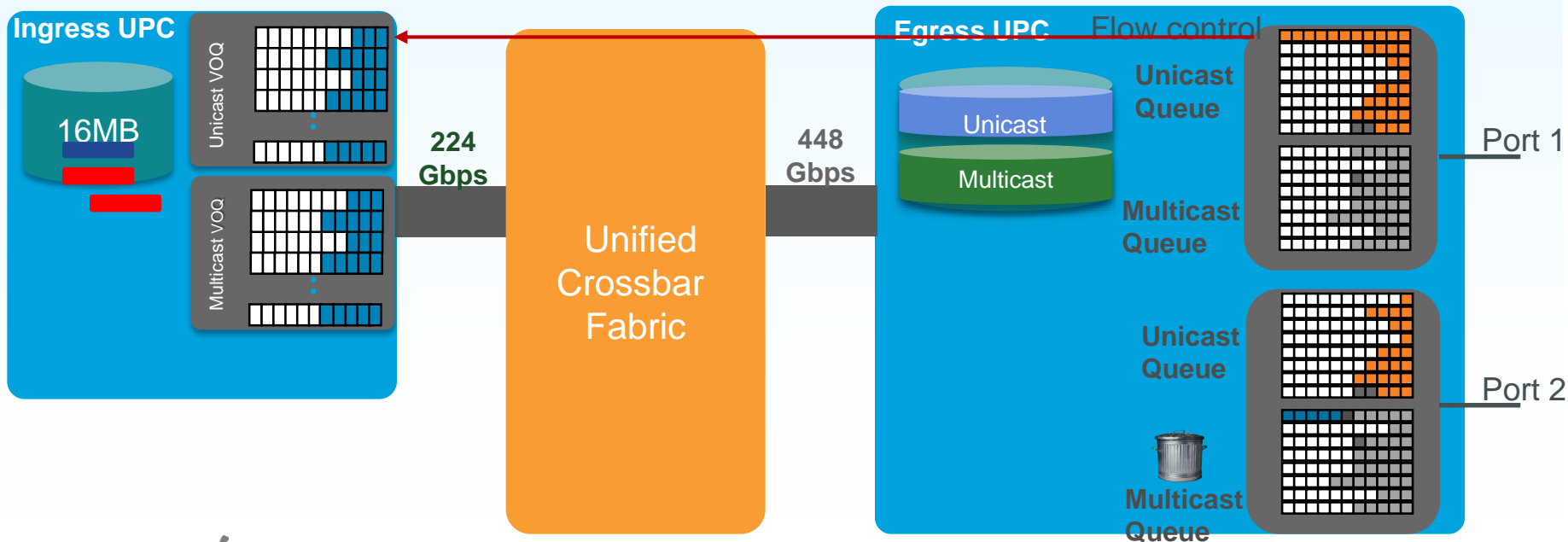
Multicast VoQ

- 8000 multicast VoQs to eliminate HOLB and help ensure high throughput
- Packets with different egress UPC fanouts are assigned to different VoQs



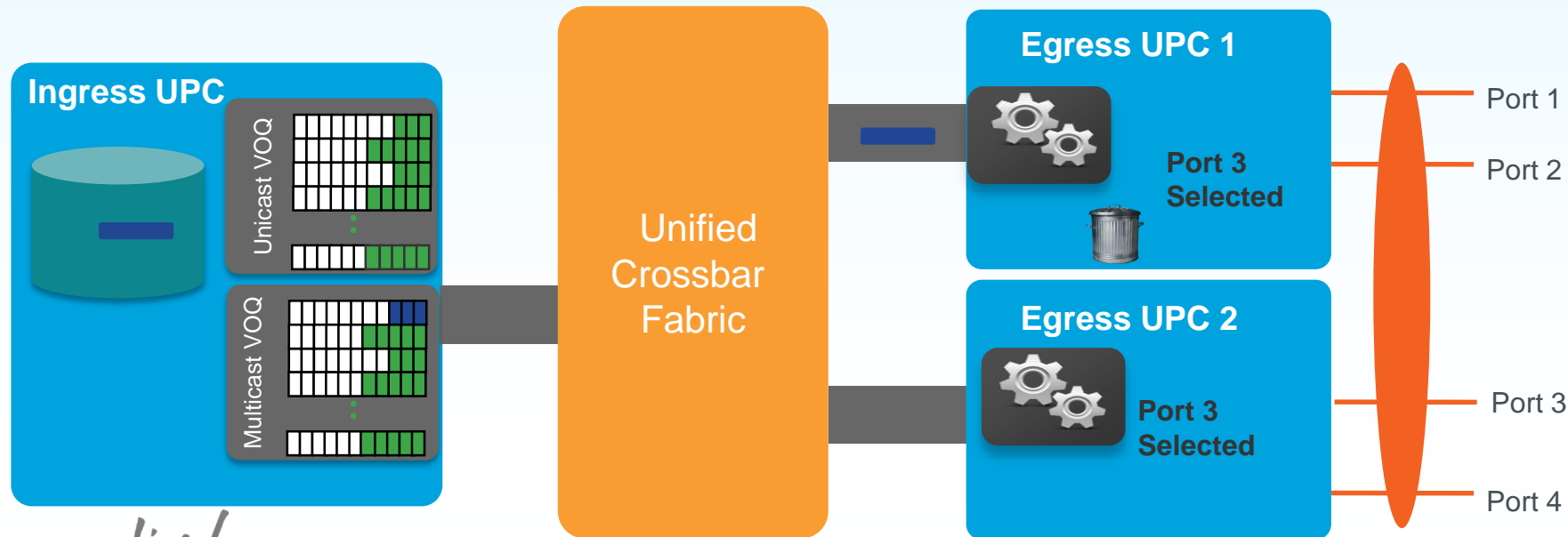
Multicast Buffering and Congestion Control

- Multicast packets are buffered and dropped at egress.
- Multicast packets are dropped when the multicast queue depth reaches threshold
- Maintain lossless fabric for unicast.



Multicast Hashing over Port Channel

- Flow based hashing for multi-destination traffic with 5-tuple packet header.
- Traffic replicated to all egress UPC where Port Channel member/s resides
- Egress UPC runs hash calculation and one egress port is chosen, others drop



RPF Check Failure

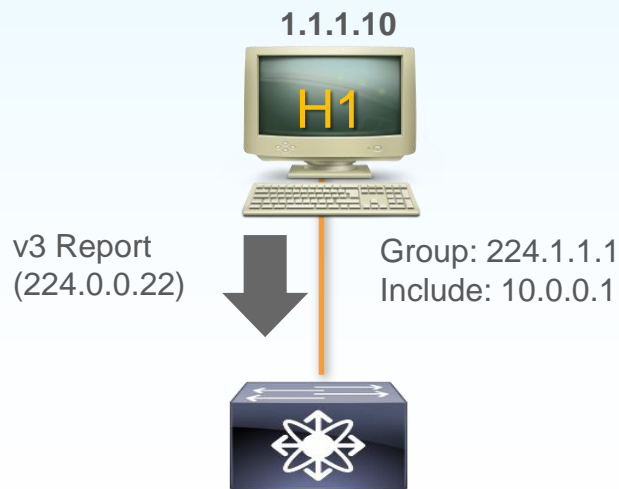
- Packets that fail RPF check are copied to CPU to generate PIM assert.
- Incoming Interface (IIF)
 - Layer 3: Redirect to CPU.
 - SVI: copy to CPU and flood to incoming VLAN ports.
- A filter is implemented so that only one packet from small flow is copied to CPU
 - Reduces CPU load.
 - Periodically allows more packets from single flow to CPU

CPU-Bound Multicast Data Packet Processing

- Multicast data packets are copied to CPU to create multicast forwarding state.
- First-hop router: Source registration
- Last-hop router: Share tree to source tree switchover
- RPF check failure packets: For PIM assert
- These three types of packets will be rate-limited by three different policers.

IP-Based Forwarding for IGMP Snooping

- Source IP and group address-based forwarding for IGMPv3 snooping
- Can filter traffic based on source IP
- No concern of overlapping multicast MAC addresses



Multicast MAC based forwarding

Vlan10, 0100.5E01.0101, eth1/1

IP-Based Forwarding

Vlan10, 10.0.0.1, 224.1.1.1, eth1/1

Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- **ACL**
- QOS
- Conclusion

Access Control Lists

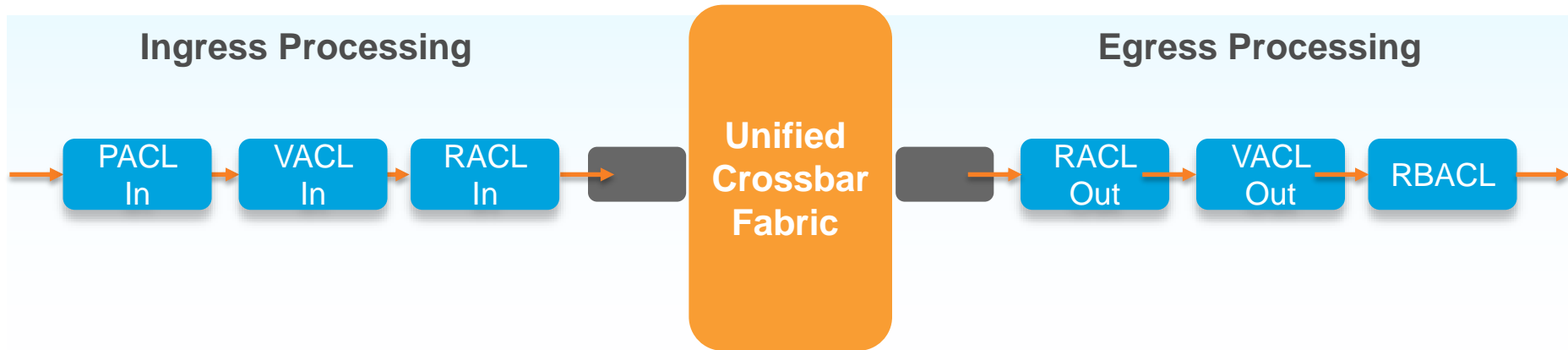
ACL Types and Features

- Security ACL
 - MAC, IPv4, and IPv6 ACLs
 - PACL: ACL enabled under L2 interface
 - VACL: ACL enabled for L2 VLAN traffic
 - RACL: ACL enabled for routed traffic
- ACL for QoS classification and Policing
- Policy Based Routing (PBR)
 - User ACL to redirect traffic
- ACL for SPAN
 - To identify, the traffic needs to be mirrored
- ACL for control traffic
 - To redirect control traffic to CPU; not user-configurable



Security ACL Processing Order and Priority

- Conceptual Illustration, lookup happens concurrently
- A packet is dropped if it hits the deny rule in any of these types of ACLs.
- RACL is applied only to traffic that is L3 forwarded.



Security ACLs configuration

```
interface Ethernet1/5
```

```
ip port access-group PACL-1 in
```

→ PACL: Port ACL Ingress Direction only

```
switchport access vlan 100
```

```
vlan access-map VACL-map
```

```
match ip address VACL-1
```

```
action forward
```

```
vlan filter VACL-map vlan-list 100
```

→ VACL: VLAN ACL for both Ingress and Egress

```
interface Vlan100
```

```
no shutdown
```

```
ip access-group RACL-1 in
```

```
ip access-group RACL-2 out
```

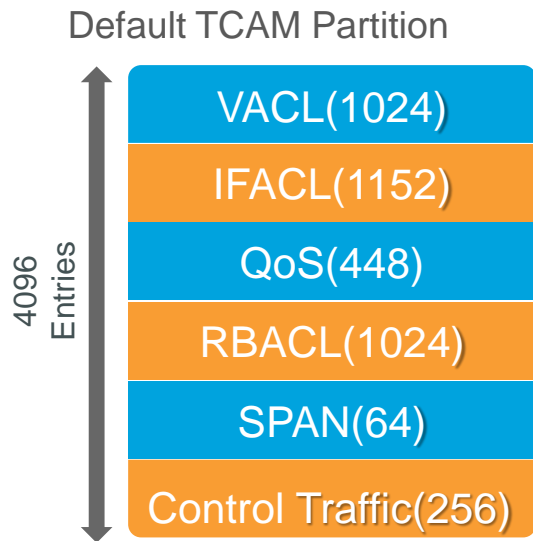
→ RACL: Router ACL only for routed traffic,
both ingress and egress

```
ip address 100.1.1.1/24
```

ACL Type	Attach Point	Direction
PACL	L2 interface, L2 PortChannel	IN
VACL	VLAN	IN OUT
RACL	L3 interface, L3 PortChannel, sub-interface SVI	IN OUT

ACL Scaling and TCAM Partition

- 4096 ACE entries per UPC are organized in blocks of 64 ACE entries.
- Software provides CLI to change the size of each region and create templates.
- Each IPv6 ACL (without port range) requires two TCAM entries.



```
N5600-TME1# sh platform afm info tcam 0 region ifacl
ifacl tcam TCAM configuration for ASIC ID 0:
[vacl tcam]: range      0 - 1023
[ifacl tcam]: range    1024 - 2175 *
[ qos tcam]: range    2176 - 2623
[rbacl tcam]: range    2624 - 3775
[ span tcam]: range    3776 - 3839
[ sup tcam]: range    3840 - 4095
```

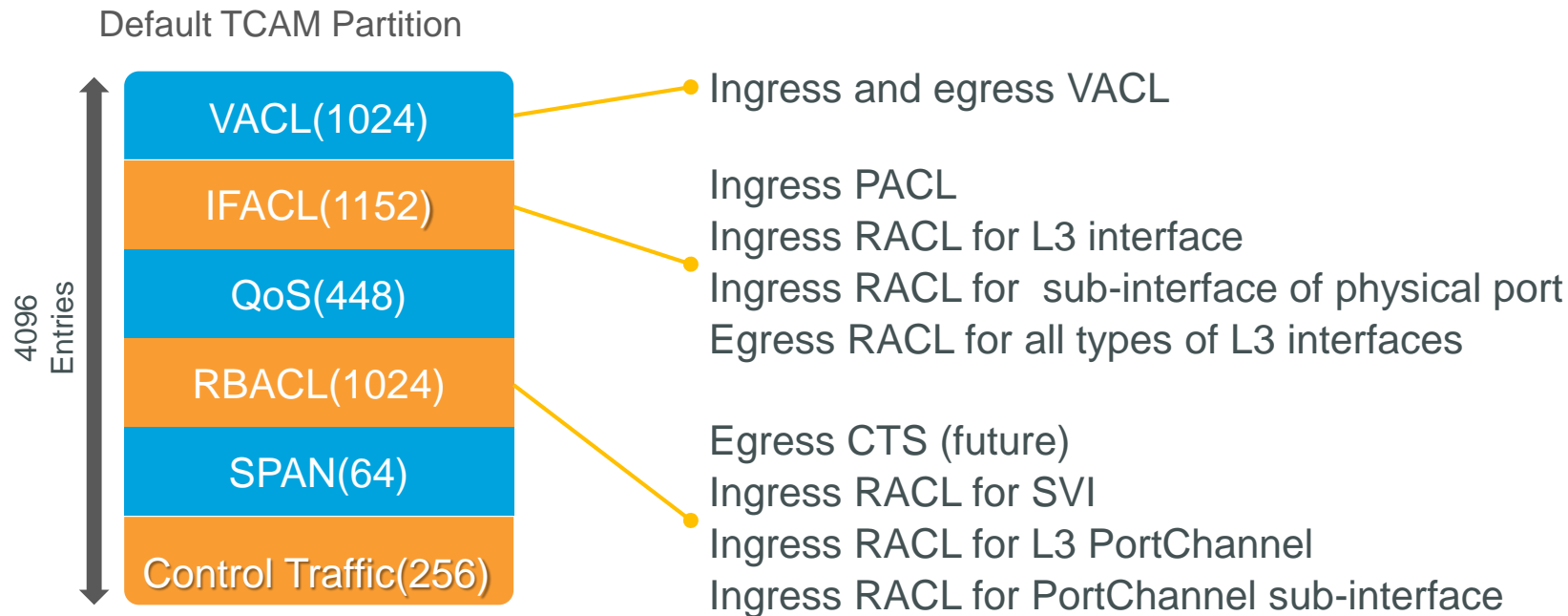
```
TCAM [ifacl tcam]: [v:1, size:1152, start:1024
end:2175]
```

```
In use tcam entries: 24
1024-1039, 2168-2175
```

TCAM Region Usage

```
N5600-TME1#
```

Security ACL-to-TCAM Region Mapping



TCAM Carving

- Creating a template:

```
switch(config)# hardware profile tcam resource template t1
switch(config-tcam-templ)#
```

- Setting region sizes

```
switch(config-tcam-templ)# ifacl ?
<320-3584> Enter size of ifacl region (in 64 entry increments)
```

- Displaying Template

```
switch(config)# sh hardware profile tcam resource template name ?
WORD Select name of a template
```

Sharing ACE (Label Sharing)

- When the same ACL policy is applied to multiple interfaces or VLAN, only one copy is stored in TCAM.
- Each ACL policy has a label. By assigning the same label to multiple interfaces and VLANs, you can apply the same TCAM rule to multiple interfaces or VLANs.

```
interface Ethernet1/10
  ip port access-group ip-list-1 in
interface Ethernet1/11
  ip port access-group ip-list-1 in
interface Ethernet1/12
  ip port access-group ip-list-1 in
```



Label	
xyz	eth1/10
xyz	eth1/11
xyz	eth1/12
xyz	IPV4 ACL ip-list-1

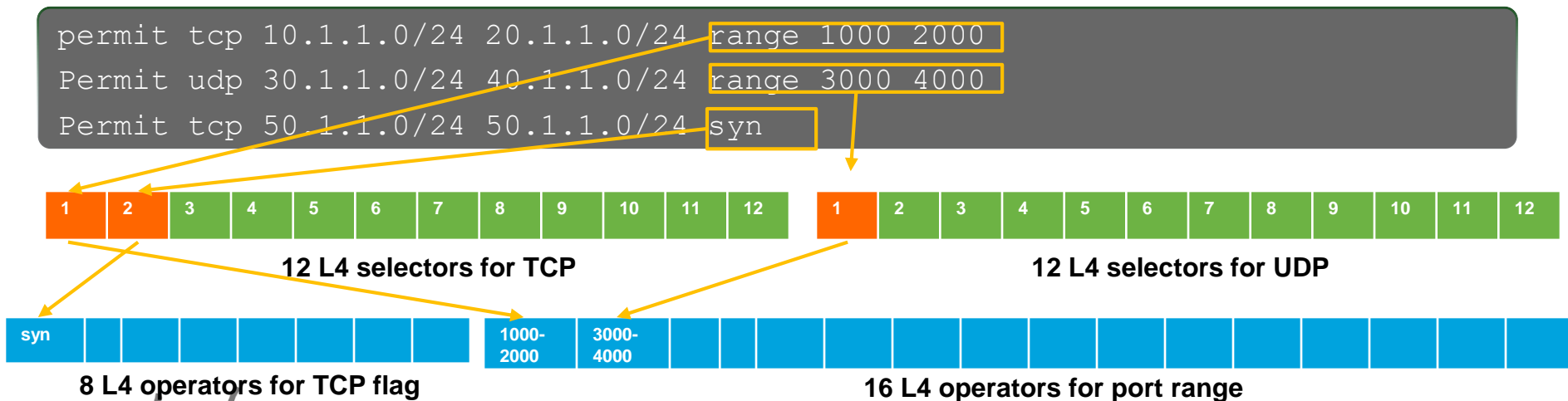
- 10 permit ip 100.1.1.0/24 200.1.1.0/24
- 20 permit ip 100.1.2.0/24 200.1.2.0/24
- 30 permit tcp 100.1.3.0/24 200.1.3.0/24 range 100 2000

ACL Logging

- Supported on all flavors of ACL: PACL, RACL, VACL & RBACL
- ACL logging feature allows the logging of the packets hitting IPv4/IPv6 ACL.
- Supported on all interfaces including FEX HIF interfaces and management.
- Logs only packets that hit the “deny” rule (For mgmt0, logs permit/deny)
- `switch(config)# ip access-list example`
`switch(config-acl)# permit tcp any any log`

ACL Layer 4 Operators

- L4OPs are hardware resources to support ACLs with special operators (range, gt, lt, neg etc)
- There are eight L4 operators for TCP flag per UPC
- 12 L4 selectors for UDP and 12 L4 selectors for TCP.
- Each unique source port range or destination port range burns one L4 operator.



ACE Expansion

FYI

- Software turns one ACL rule with UDP and TCP port range to multiple ACE entries in TCAM.
- Each TCAM field has an associated mask including the UDP or TCP port field.
- The mask can be used to support UDP and TCP port ranges.

permit tcp 10.1.1.0/24 20.1.1.0/24 range 0 100

User-Configured ACL Rule

Software Expands the Rule to Multiple ACEs

		Action	Src IP	Dst IP	Protocol	Src port	Dst port	
1	Content	Permit	10.1.1.0	20.1.1.0	TCP	0	0	Port Range 0-63
	Mask	0	0.0.0.255	0.0.0.255	0	255.255	0.63	
2	Content	Permit	10.1.1.0	20.1.1.0	TCP	0	64	Port Range 64-95
	Mask	0	0.0.0.255	0.0.0.255	0	255.255	0.31	
3	Content	Permit	10.1.1.0	20.1.1.0	TCP	0	96	Port Range 96-99
	Mask	0	0.0.0.255	0.0.0.255	0	255.255	0.3	
4	Content	Permit	10.1.1.0	20.1.1.0	TCP	0	100	Port Range 100
	Mask	0	0.0.0.255	0.0.0.255	0	255.255	0.0	

Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

Quality of Service

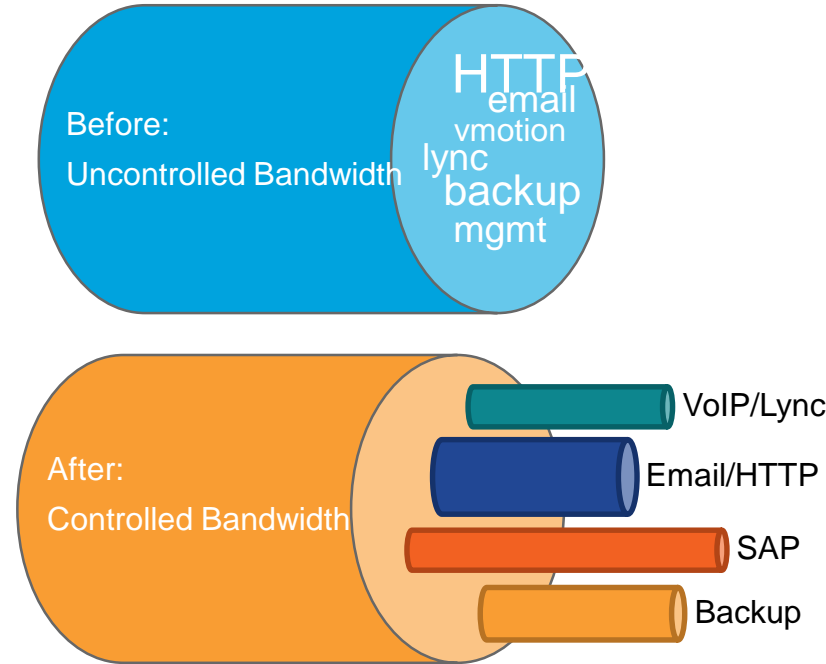
Nexus 5600/6000 Quality Of Service

- Traffic classification
 - DSCP, CoS, IP Precedence and ACL
- Packet marking
 - DSCP, CoS, and ECN
- Strict Priority Queuing and DWRR
 - Priority Flow Control
 - DCBX 802.1Qaz
- Ingress policing (No egress policing)
 - 4096 policers per ASIC
- Flexible buffer management



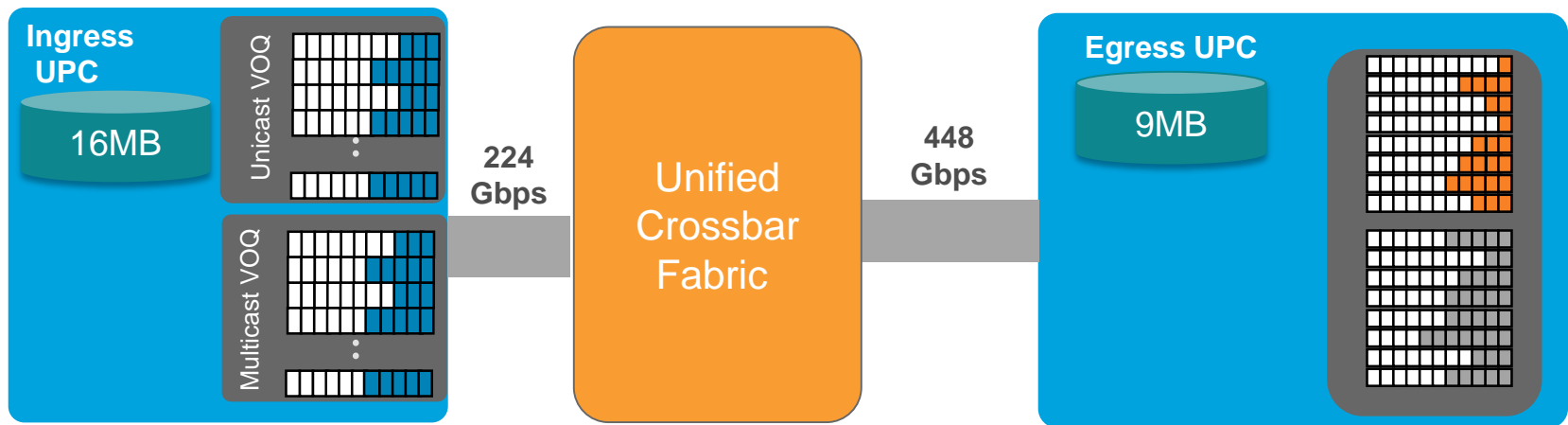
Packet Classification

- Classification based on the DSCP, IP Prec, COS or ACL.
- Default TCAM carving allocates 448 ACE entries for QoS
- Packet classification takes place at ingress.
- Packet marking is trusted by default.



Packet Buffering

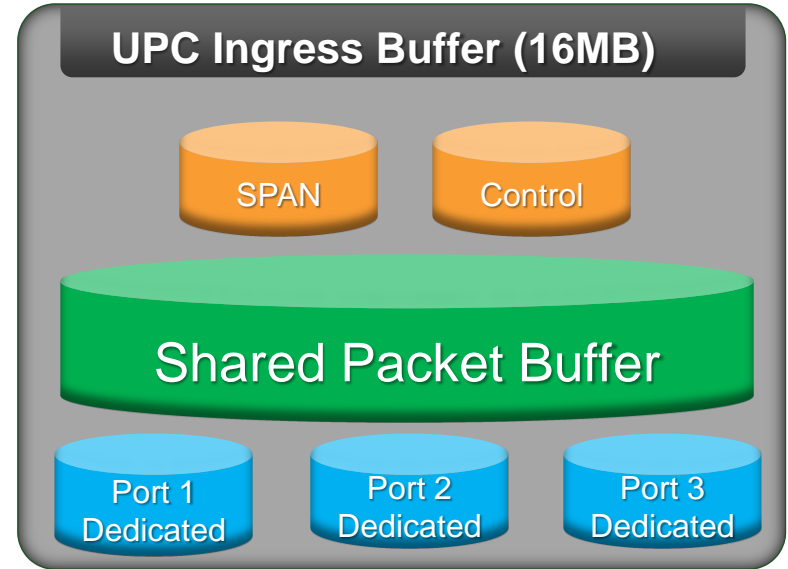
- 25MB packet buffer is shared by every three 40 GE ports or twelve 10 GE ports.
- Buffer is 16MB at ingress and 9MB at egress.
- Unicast packet can be buffered at both ingress and egress.
- Multicast Buffered at egress only



Flexible Buffer Management

Ingress Buffer

- Shared buffer is good for burst absorption.
- Dedicated buffer is good for predictable performance for each port.
- On by default, no configuration needed
- Long-distance FCoE, video editing (i.e., AVID), Big Data, and distributed storage



Default Ingress Buffer Allocation

- Each cell is 320 bytes.
- Total number of cells for ingress buffer is 48,840.

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

Tune Buffer Allocation at Ingress

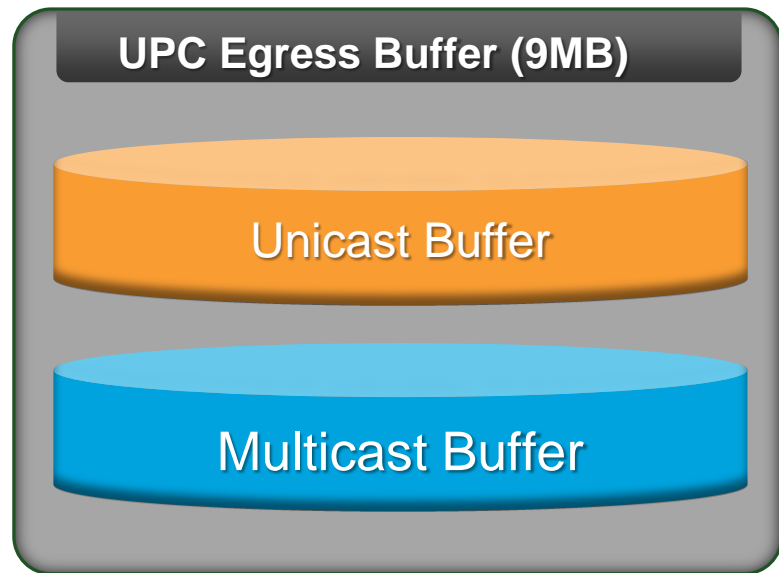
- “queue-limit” under “network-qos” policy specifies the dedicated buffer for each port and each class. The dedicated buffer can be used by the port for only that class of service.
- Without “queue-limit” each class of service will get 100 KB of dedicated buffer.
- The size of dedicated buffer can be different for different classes of service. The policy applies to all ports in the chassis.
- Total ingress buffer minus the dedicated buffer and buffer for control and SPAN will be in the shared buffer pool.
- The following example sets the dedicated buffer for “class-default” to be 400 KB for all ports.

```
switch(config)# policy-map type network-qos Policy-buffer
switch(config-pmap-nq)# class type network-qos class-default
switch(config-pmap-nq-c)# queue-limit 400000 bytes
switch(config-pmap-nq-c)# system qos
switch(config-sys-qos)# service-policy type network-qos Policy-buffer
```

Flexible Buffer Management

Egress Buffer

- 9-MB packet buffer is shared among three 40 GE or twelve 10 GE.
- CLI is provided to allocate buffer between unicast and multicast (future).
- Unicast traffic can be buffered at egress and ingress.
- Multicast is buffered at egress in case of interface oversubscription.



Default Egress Buffer Allocation

- By default the majority of egress buffer is allocated for multicast traffic.
- Future software will provide CLI to tune the egress buffer allocation.
- At egress, unicast buffer is allocated on a per-port basis. For multicast, the egress buffer is shared among all ports.

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

WRED/ ECN Configuration

- ECN parameters are configurable only at system level.
- ECN is enabled by default along with WRED
- Packet Threshold below minimum – **Transmit**
- Packet Threshold between minimum and maximum – **Mark ECN bits**
- Packet Threshold above maximum – **Drop**

```
switch(config)# hardware random-detect min-thresh 10g 10g-min-threshold 40g 40g-min-  
threshold max-thresh 10g 10g-max-threshold 40g 40g-max-threshold ecn qos-group qos-  
group-number
```

Nexus 5600 QoS Processing

Ingress UPC

Trust CoS/DSCP and
Match on COS/DSCP/ACL

If Buffer Usage Crosses Threshold:

- Tail drop for drop class
- Assert Pause for no-drop COS

MAC

Traffic
Classification

Ingress
Cos/DSCP
Marking

Ingress
Policing

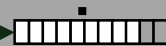
Forwarding

Per-class
Buffer Usage
Monitoring

Make a forwarding decision

Ingress Buffer

VoQs
(8 per egress port)



Crossbar
Fabric

Egress UPC

MAC

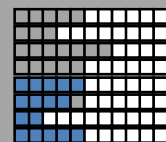
WRED/ECN
Marking

Egress
Scheduling

Strict Priority +
DWRR Scheduling

Egress Buffer

Egress Queues



Agenda

- Introduction
- Architecture
- Forwarding
- Multicast
- ACL
- QOS
- Conclusion

Conclusion

Nexus 5600/6000 Summary



High Performance

- Line rate L2 and L3 at all frame sizes
- 1-microsecond port-to-port latency
- True 40-Gbps flow
- 40-Gbps FCoE
- Cut-through switching for 40 and 10 GE
- 25-MB buffer per port ASIC
- 800-MB maximum (on 5696Q)

High Scalability

- Up to 224,000 MAC (Flexible Carving)
- Up to 128,000 ARP (Flexible Carving)
- 32,000 LPM
- 16,000 bridge domains

Feature-Rich

- Classic L2 and L3 features
- vPC and FabricPath
- FC and FCoE
- FabricPath with segment ID (DFA)
- VxLAN Bridging and Routing (5600 only)

Visibility and Analytics

- Line-rate SPAN, SPAN on drop, SPAN on Latency
- Sampled NetFlow
- Micro-burst and buffer monitoring
- Latency monitoring

BRKDCT-1980 -Advanced Analytics in Nexus Switches

Participate in the “My Favorite Speaker” Contest

Promote Your Favorite Speaker and You Could Be a Winner

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

Complete Your Online Session Evaluation

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



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

Continue Your Education

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



Thank you



TOMORROW starts here.