You make **possible**

End-to-End QoS Implementation and Operation with Nexus

Nemanja Kamenica
Technical Marketing Engineer

BRKDCN-3346

Cisco live!
June 9-13, 2019 • San Diego, CA
#CLUS

Session Objectives

- Provide a refresh of QoS Basics
- Understand the basic switch architecture for the Nexus switch family
- Provide a detailed understanding of QoS on Nexus platforms
- Learn how to configure QOS on Nexus devices through real-world configuration examples

Session Non-Objectives

- Data Centre QoS Methodology
- Nexus hardware architecture deep-dive
- Application Centric Infrastructure (ACI) QOS

Cisco Webex Teams

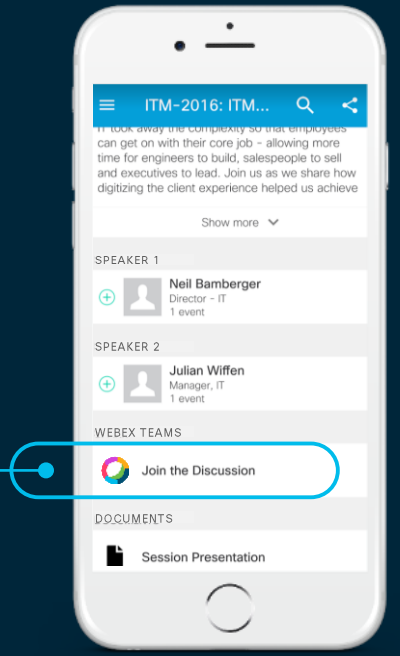
Questions?

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

How

1. Find this session in the Cisco Live 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

Webex Teams will be moderated by the speaker until June 16, 2019.



cs.co/ciscolivebot#BRKDCN-3346

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

Congestion Happens Everyday!

Why QoS in the Data Centre?

Assign Colour to Traffic

Manage Congestion

Maximise Throughput

Maximise Throughput and Manage Congestion!

Can Traffic Control help ...

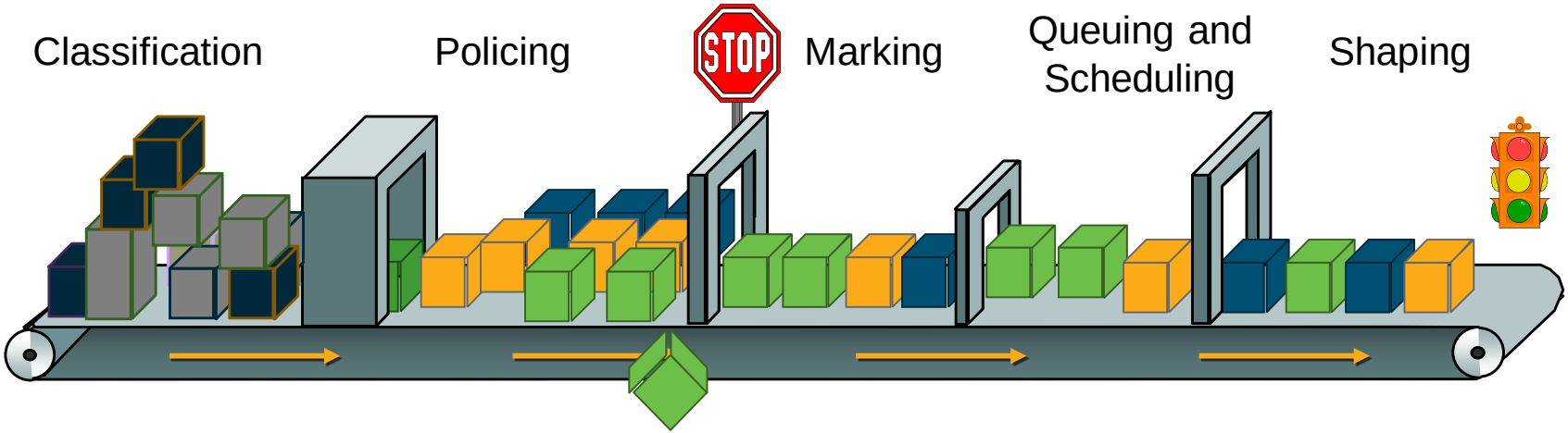
… or confuse

… or hurt

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

The QoS Toolset



Identify and Split Traffic into Different Classes

Discard Misbehaving Traffic to Maintain Network Integrity

Mark Traffic According to Behaviour and Business Policies

Prioritise, Protect and Isolate Traffic Based on Markings

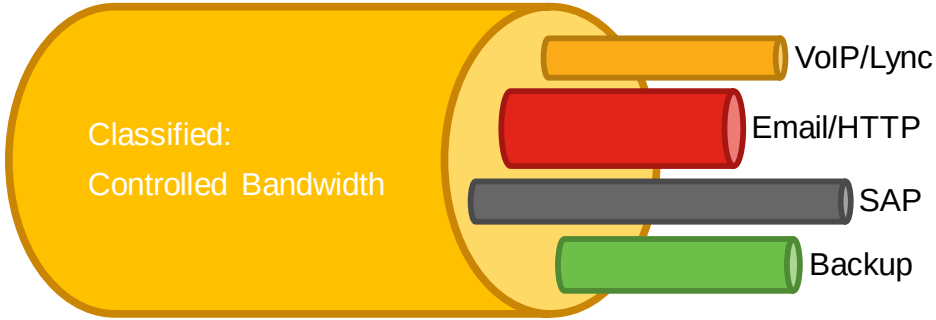
Control Bursts and Conform Traffic

Traffic Management Tools

- Classification
  - Traffic Categorisation based on traffic attributes
- Marking
  - Assigning different/new attribute (priority) to traffic
- Policing
  - Limit misbehaving flows

Classification and Marking – Two sides of a coin

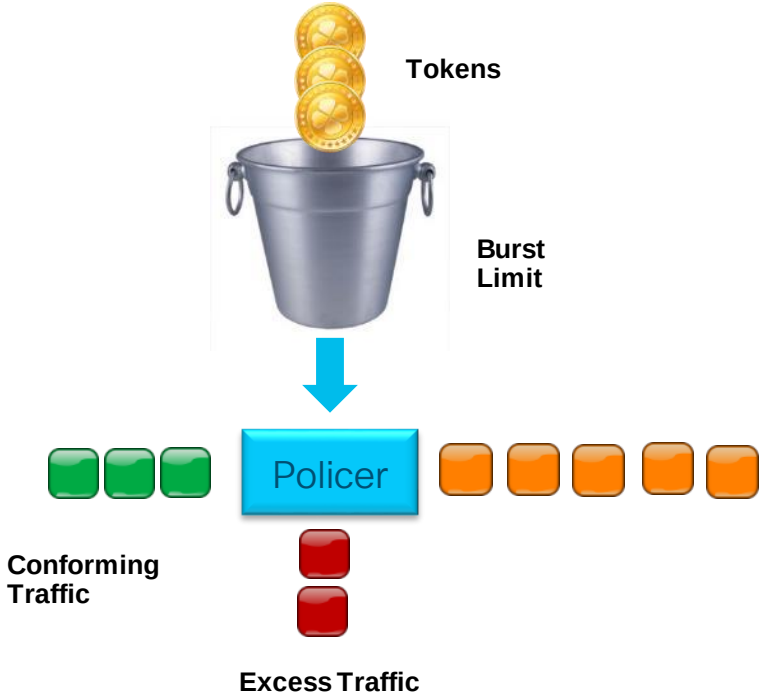
- Identify traffic
  - DSCP
  - IP PREC
  - CoS
  - ACLs
- Re-map Traffic
  - Like to Like (i.e. CoS to CoS)
  - Like to Unlike (i.e. DSCP to COS)
    - Needs mapping tables
    - Also called Mutation



Policing – Limit Misbehaving Traffic

- Single rate Two Color Policer
  - Conform Action (permit)
  - Exceed Action (drop)

- Two rate Three Color Policer
  - Conform Action (permit)
  - Exceed Action (markdown)
  - Violate Action (drop)

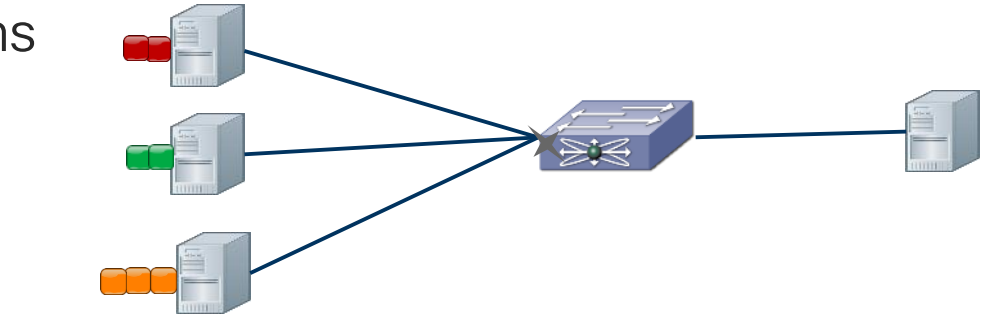


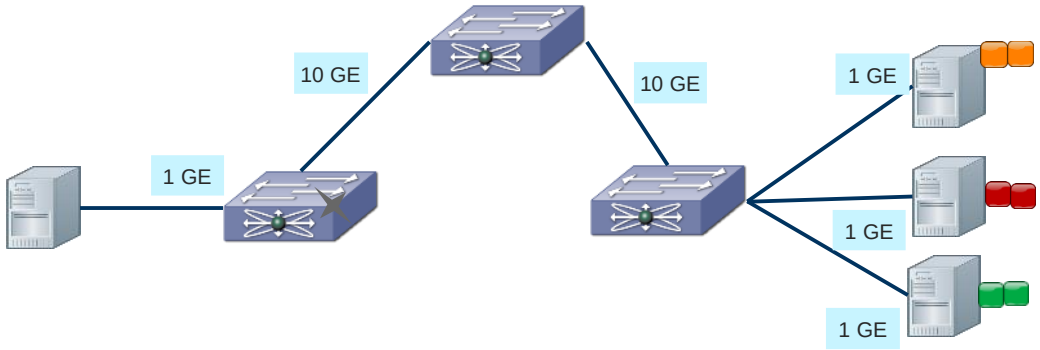
Congestion Management Tools

- Buffering
  - Storing packets in memory
- Queuing
  - Buffering packets according to traffic class
- Scheduling
  - Order of transmission of buffered packets
- Shaping
  - Smooth burst traffic

Buffering – Why do we need it?

- Many to One Conversations
  - Client to Server
  - Server to Storage
  - Aggregation Points
- Speed Mismatch
  - Client to WAN to Server





4 Class Queuing Model

- Matches most Service-Provider offerings
- Ready for No-Drop traffic like FCoE
- One Class left to place traffic above or below Best-Effort traffic priority
  - Special Application which is drop sensitive (above Best-Effort - Critical)
  - Non-Critical Bandwidth intensive application (below Best-Effort - Scavenger)

Class	CoS	Queues
Priority	5-7	PQ
No-Drop	3	Q2
Better or Worse then Best-Effort	1,2,4	Q1
Best-Effort	0	Default-Q

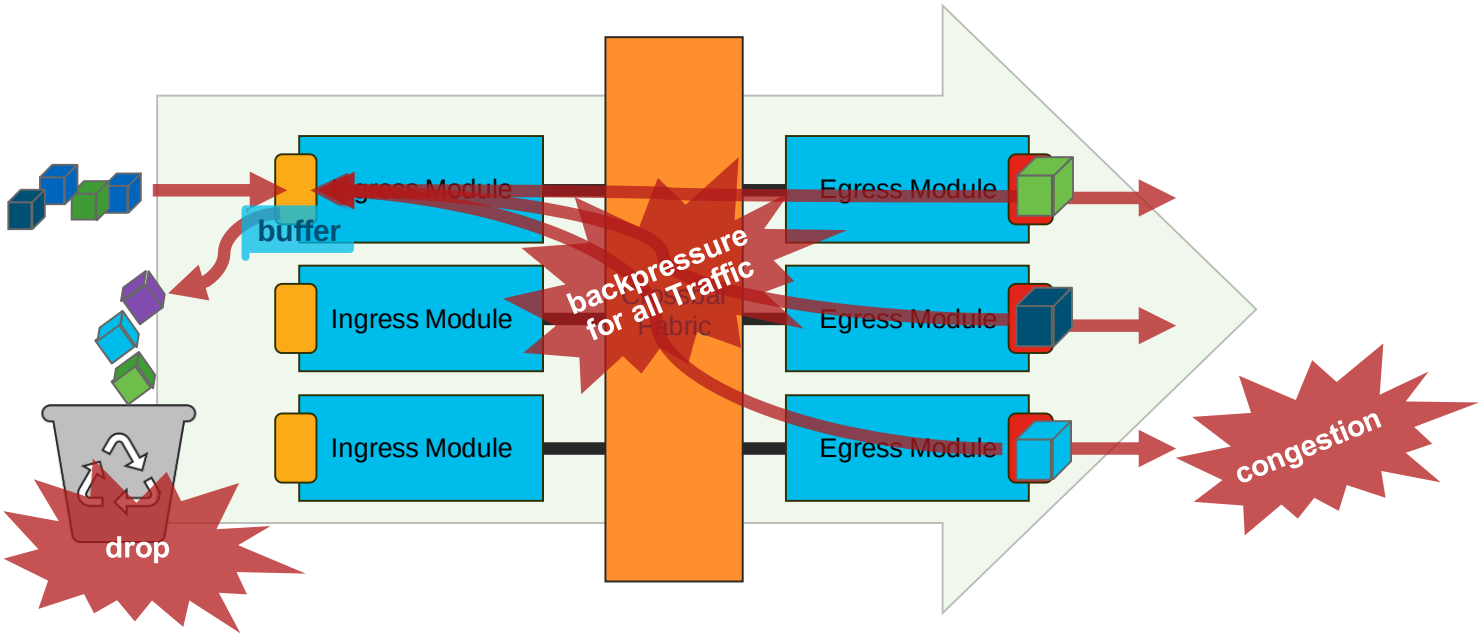
8 Class Queuing Model

- Matches often a Campus QoS concept
- DSCP to CoS derivation does NOT apply anymore
  - (Topmost 3-Bit mapping from DSCP to CoS)
- No-Drop still with CoS3 (DSCP 24-30 are “unusable”)
- Valid but most complex Classification to Marking implementation as per regards to No-Drop

Class	DSCP	Queues
Priority	CS6 (CS7)	PQ
Platinum	EF	
Gold	AF41	Q7
Silver	CS4	Q6
No-Drop	CoS3	Q5
Bronze	AF21	Q4
Management	CS2	Q3
Scavenger	AF11	Q2
Bulk Data	CS1	Q1
Best-Effort	0	Default-Q

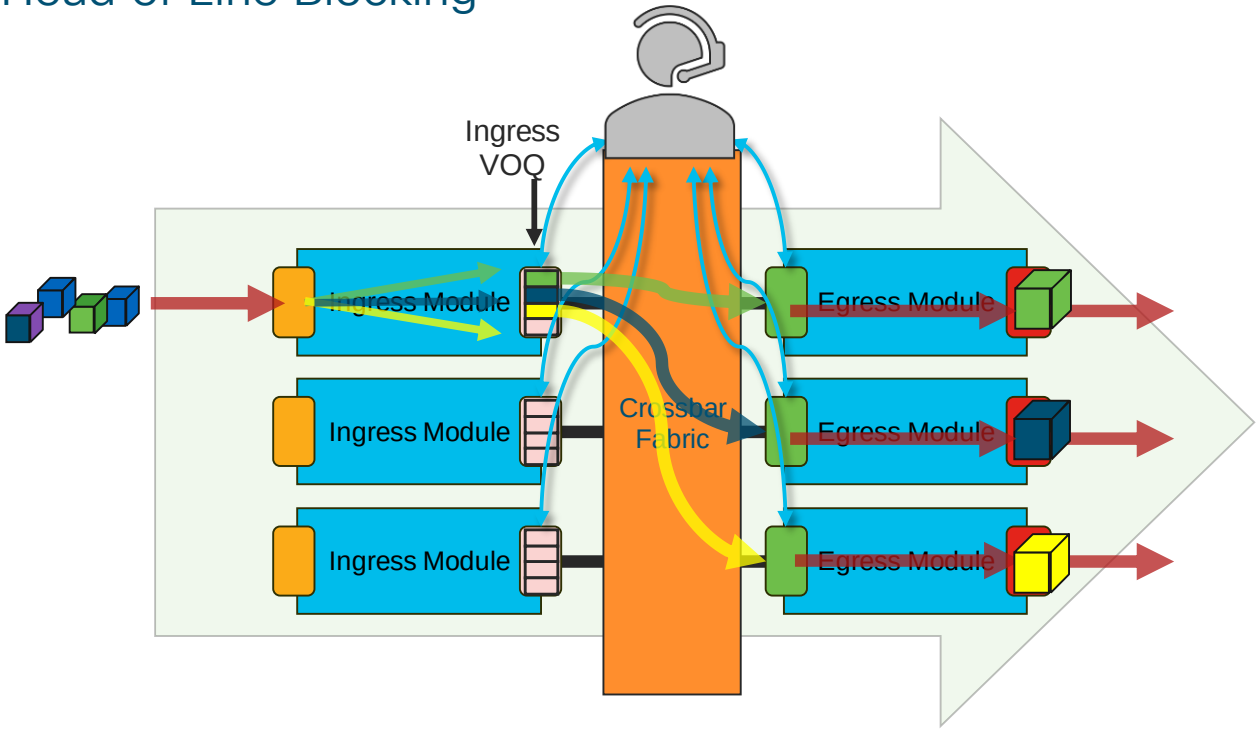
Head of Line Blocking

What is the Problem?



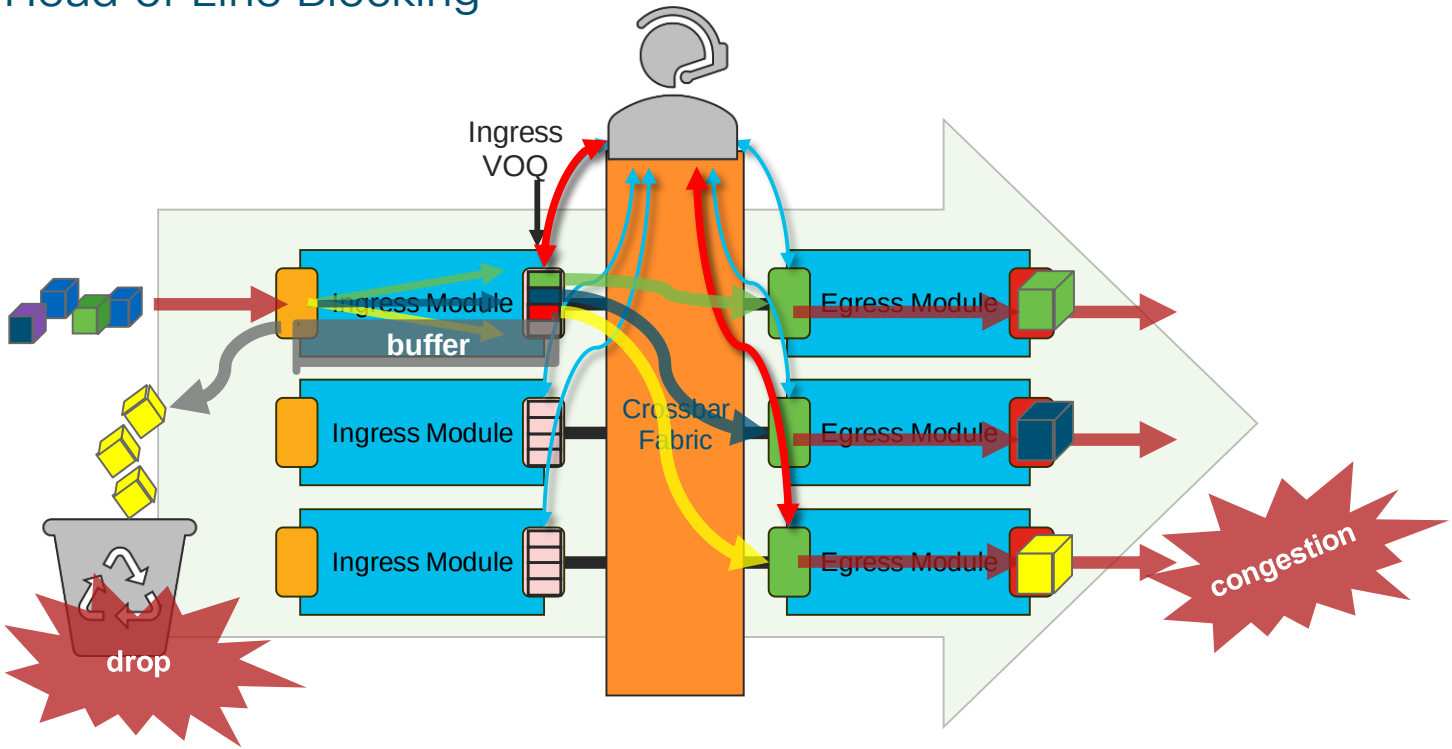
Virtual Output Queues

Avoid Head of Line Blocking

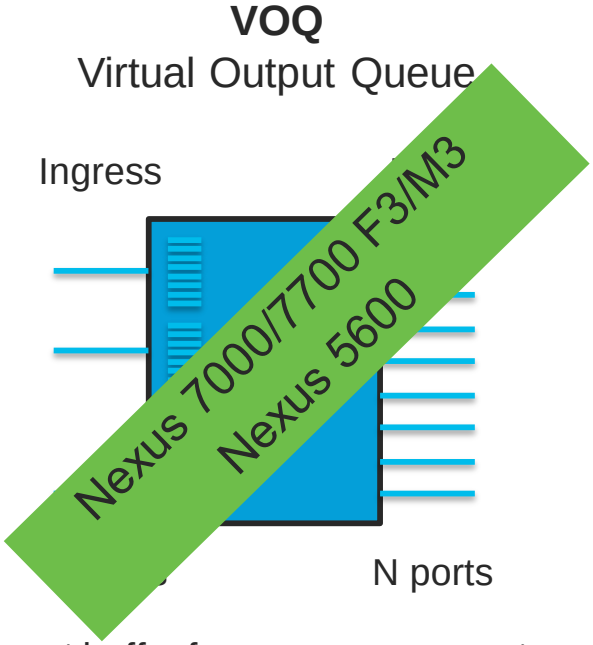


Virtual Output Queues

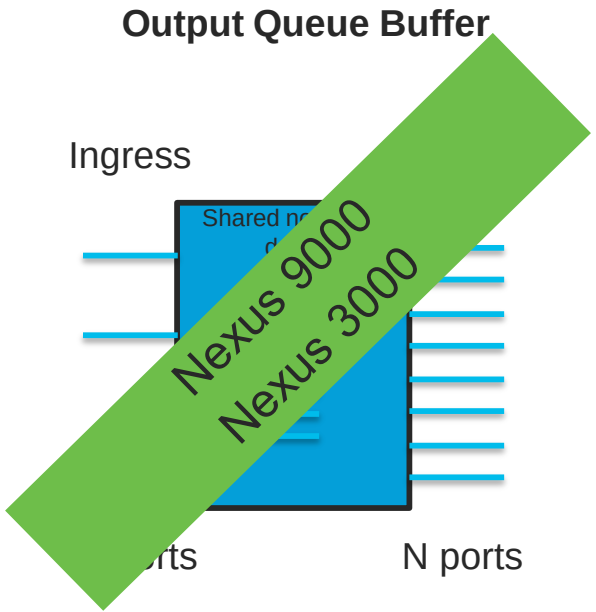
Avoid Head of Line Blocking



Buffering on Nexus Models compared



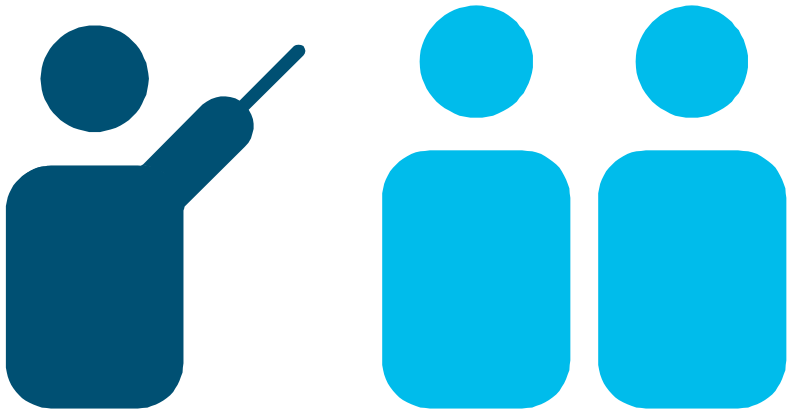
Input buffer for every egress port
NxN buffer size



Shared buffer for N egress ports
N buffer size

Scheduling – Who goes first?

- Defines Order of transmission
- The Priority-Queue always serviced first
- Normal Queues serviced only after Priority Queue empty
- Different Scheduling algorithms for normal queues



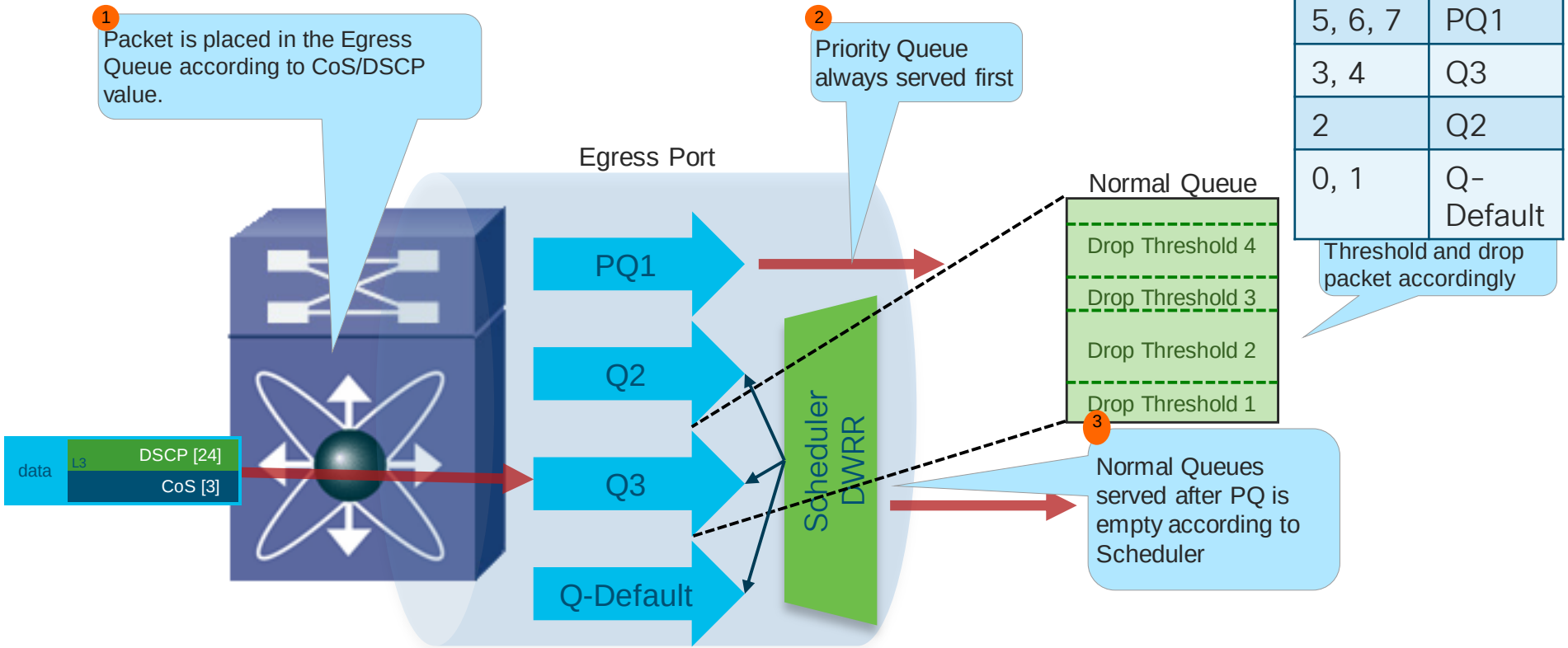
Common Scheduling Algorithms

- Round Robin (RR)
  - Simple and Easy to implement
  - Starvation-free
- Weighted Round Robin (WRR)
  - Serves n packets per non-empty queue
  - Assumes a mean packet size
- Deficit Weighted Round Robin
  - Variable sized packets
  - Uses a deficit counter
- Shaped Round Robin
  - More even distributed ordering
  - Weighted interleaving of flows

Congestion Avoidance Tools

- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue
- Weighted Random Early Drop (WRED)
  - One or more thresholds per queue
  - Threshold associated with DSCP or COS

Putting it all together!



COS	Queue
5, 6, 7	PQ1
3, 4	Q3
2	Q2
0, 1	Q-Default

Agenda

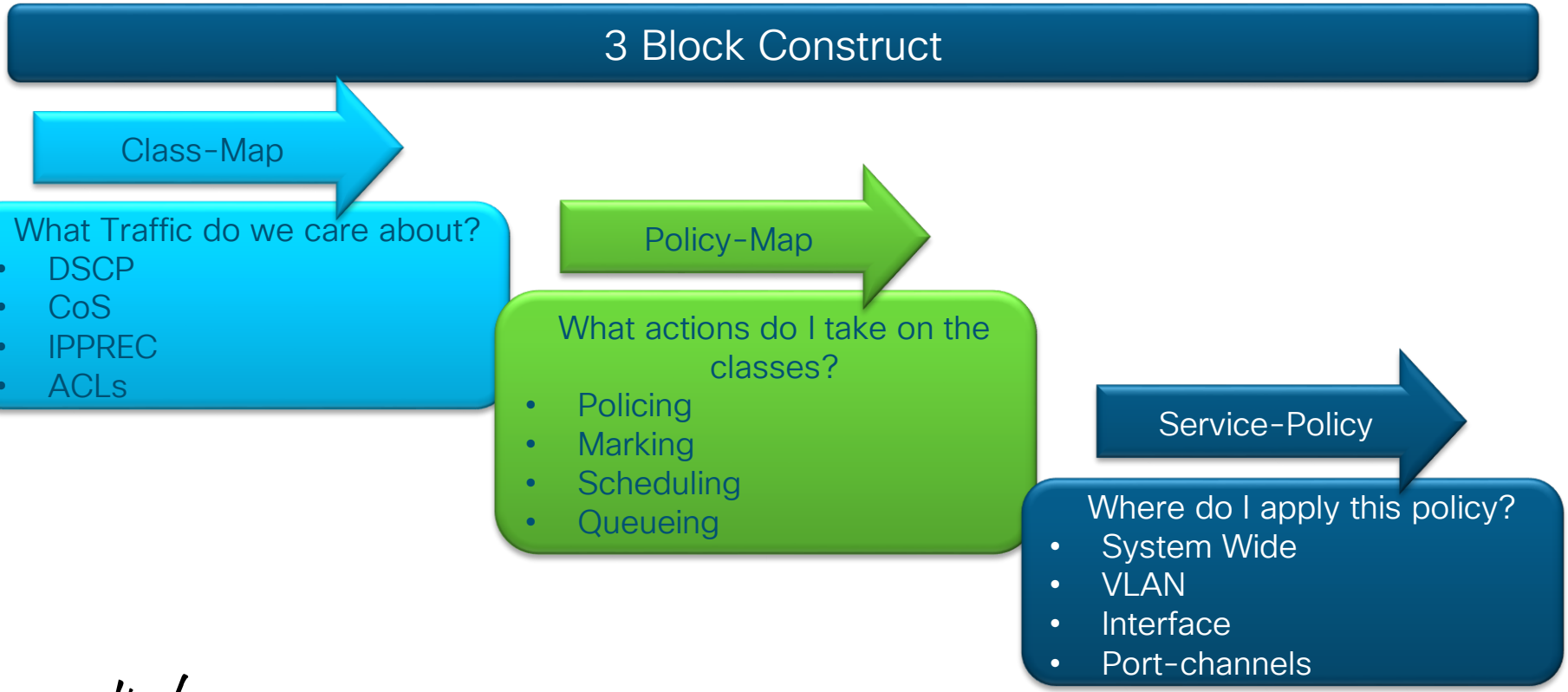
- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

QoS Implementation on Nexus

You make networking **possible**

Cisco live!

Nexus uses Modular QOS CLI (MQC)



Three Different Types

Class-map

QoS
- CoS
- DSCP
- PREC
- ACLs

Queuing
- CoS
- DSCP

Network-QoS
- CoS
- Protocol (FCoE)

Policy-map

QoS
- Marking
- Policing
- Mutation

Queuing
- Buffering
- Queuing
- Scheduling

Network-QoS
- Congestion-Control
- Pause / MTU per VL

Service-policy

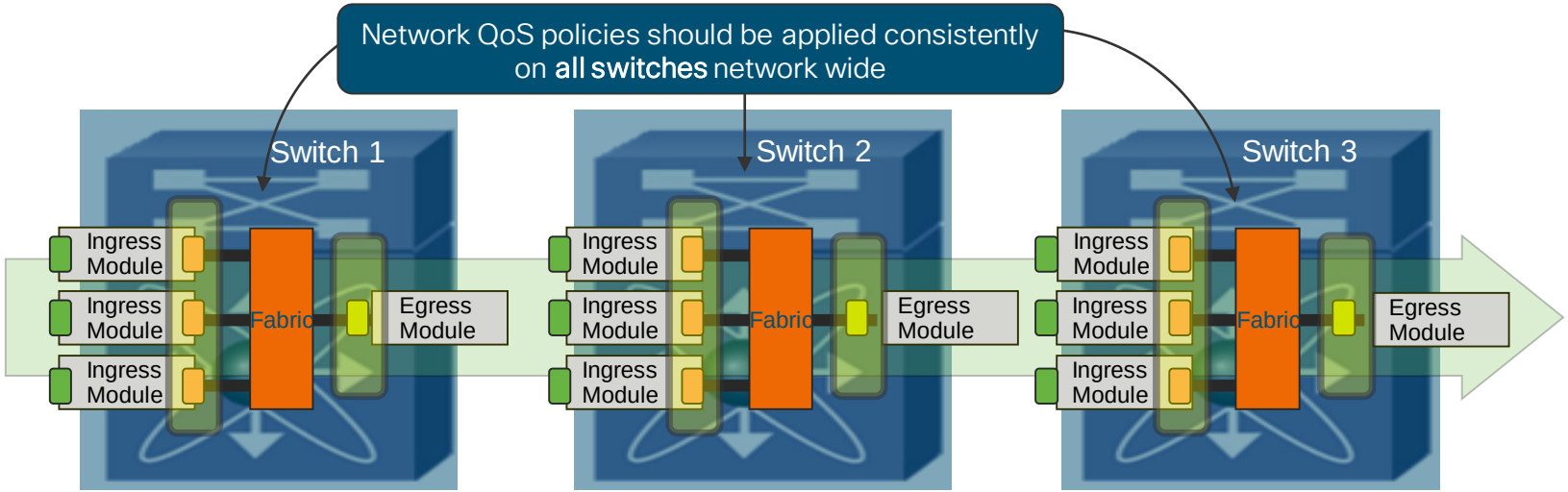
QoS
- Interfaces
- Vlans
- Port-channel
- System-qos

Queuing
- Interfaces
- Port-channels
- System-qos

Network-QoS
- System-qos

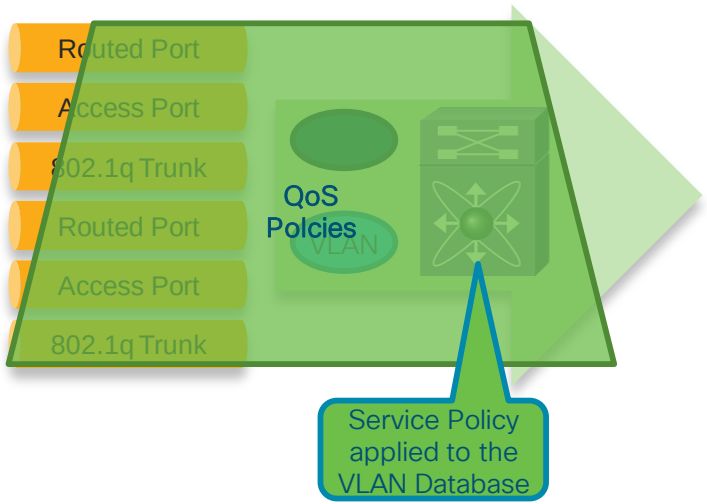
Network-QoS Policy

- Define global queuing and scheduling parameters for all interfaces in switch
  - Identify drop/no-drop classes, MTU and WRED/TD, etc.
- One network-QoS policy per system, applies to all ports
- Assumption is network-QoS policy defined/applied consistently network-wide



System based Policy attachment

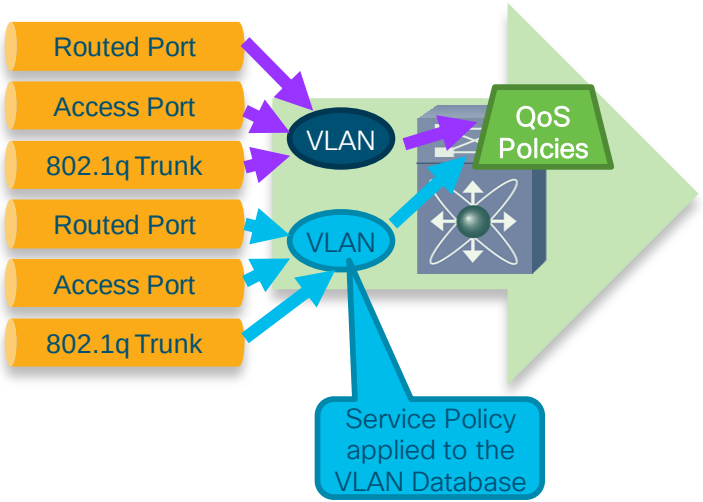
- System based QoS Policy gets globally applied to all interfaces and VLAN
- System based QoS Policy is configured in System QoS




```
Nexus(config)# system qos
Nexus(config-sys-qos)# service-policy input myPolicy
```

VLAN based QoS Policy attachment

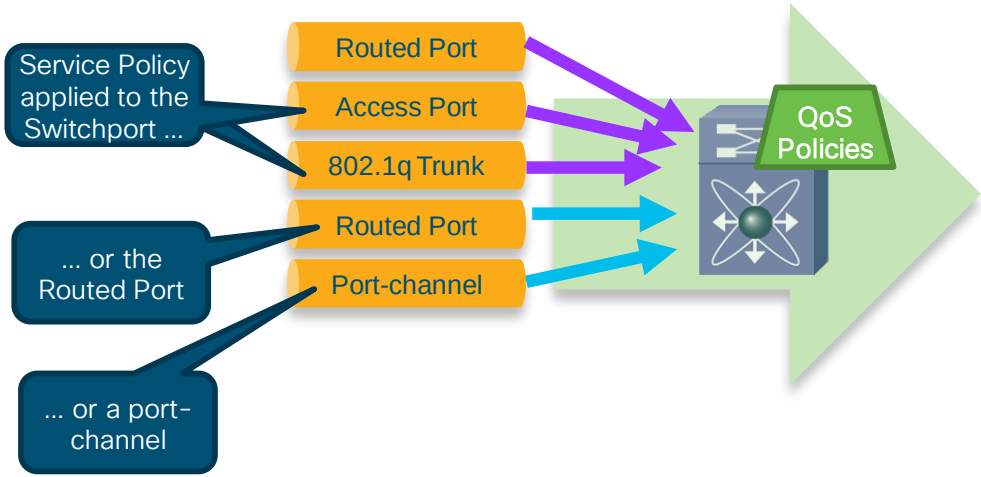
- VLAN based QoS Policy is configured in VLAN Database
- No SVI (aka L3 VLAN Interface) required




```
Nexus(config)# vlan configuration <vlan-id>
Nexus(config-vlan)# service-policy input myPolicy
```

Interface based QoS Policy attachment

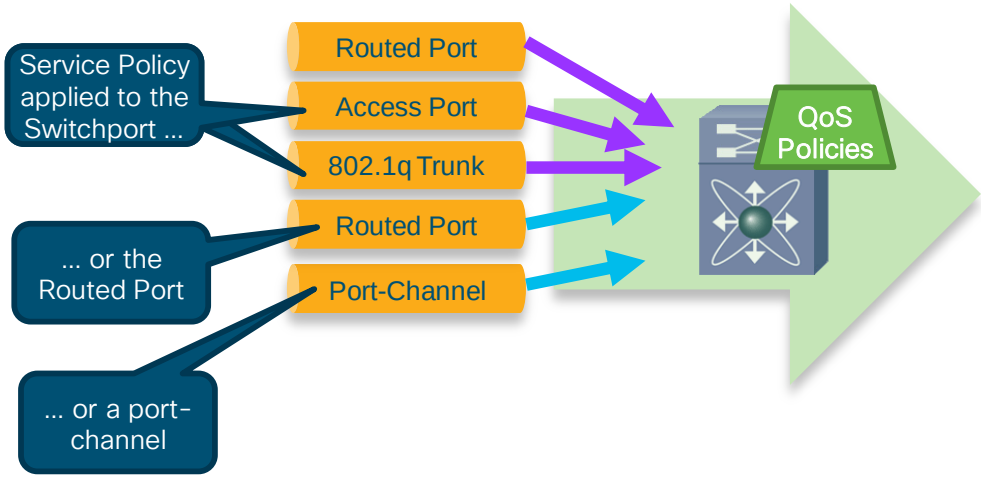
- Interface based QoS Policy takes precedence over VLAN
- Can also be attached to port-channel and applies to all member-ports
- No Egress QoS policies on L2 ports!




```
Nexus(config)# interface ethernet 1/1
Nexus(config-if)# service-policy input myPolicy
```

Interface based Queuing Policy attachment

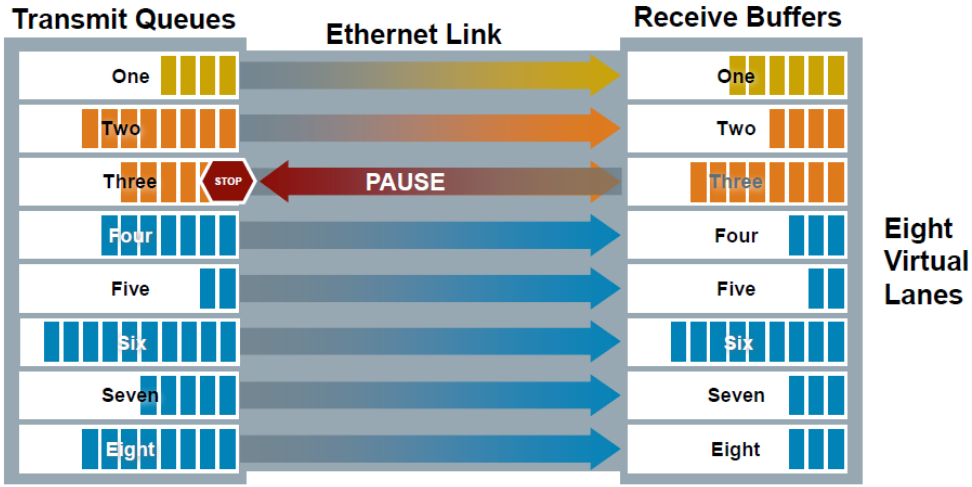
- Interface based QoS Policy takes precedence over VLAN
- Interface based QoS Policy is configured under the respective Interface
- Queuing Policy can be attached to port-channel also




```
Nexus(config)# interface ethernet 1/1
Nexus(config-if)# service-policy input myPolicy
```

New QoS Capabilities

- Priority Flow Control (802.1Qbb)
  - Enables Lossless Ethernet using per traffic class pause
  - During congestion, no-drop priority is paused
  - No effect on other priority values



DC QoS Capabilities

- DCBXP (802.1Qaz)
  - LLDP with new TLV Values
  - Negotiates capabilities (like PFC) with other devices
- ECN (Explicit Congestion Notification)
  - Congestion Notification without dropping packets
  - Uses two LSB bits in DiffServ field IP header

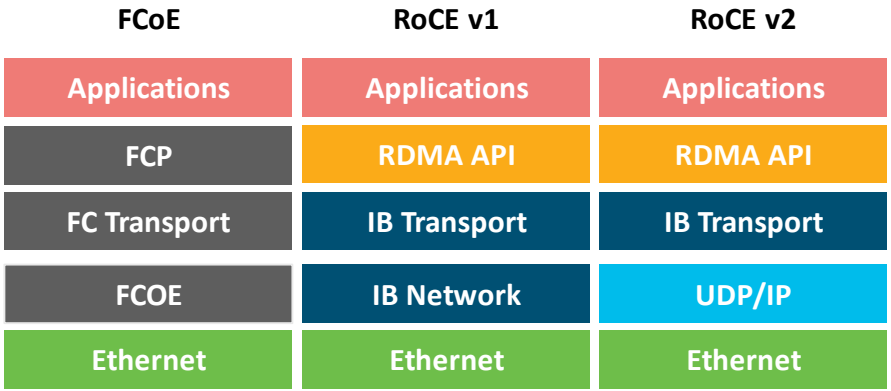
ECN	ECN Behavior
0x00	Non ECN Capable
0x10	ECN Capable Transport (0)
0x01	ECN Capable Transport (1)
0x11	Congestion Encountered

Data Centre Converged Infrastructure

- Simplification of the infrastructure by using Ethernet for data and storage traffic
- FCoE
  - Replaces Fibre Channel stack with Ethernet
- RoCE
  - RoCE extends RDMA capabilities over Ethernet

RoCE vs RoCEv2 (non-drop) FC/FCoE

- Requirement for FCoE and RoCEv1:
  - PFC
  - ETS
- Requirement for RoCEv2
  - PFC
  - ETS
  - ECN (optional)



To Trust or Not To Trust?

- Data Centre architecture provides a new set of trust boundaries
- Virtual Switch extends the trust boundary into the Hypervisor
- Nexus Switches always trust CoS and DSCP

Overlay QOS

You make security **possible**

Overlay QoS

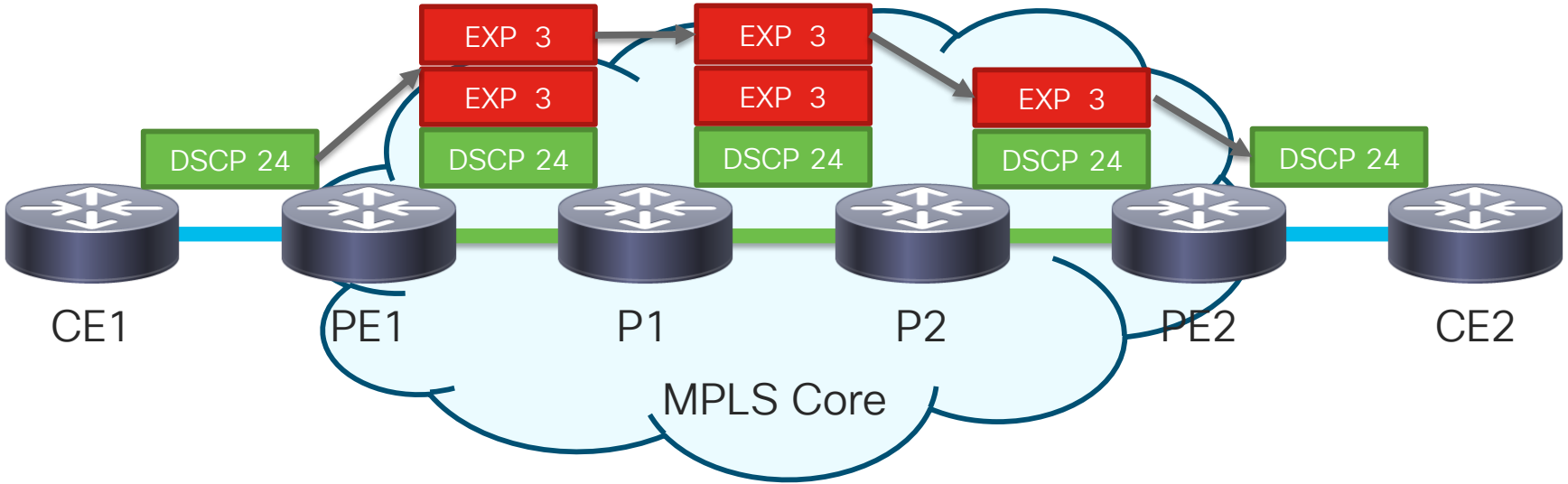
MPLS network

- Mapping between IP priorities to EXP on PE router
- Classification is done biased on COS, DSCP, IP precedence or ACL
- DiffServ Tunneling mode provides different QOS behavior in provider network
  - Uniform mode delivers overlay priority
  - Pipe mode extends underlay priority

EXP	COS	DSCP	IP pres
0	0	0	0
1	1	8	1
2	2	16	2
3	3	24	3
4	4	32	4
5	5	40	5
6	6	48	6
7	7	56	7

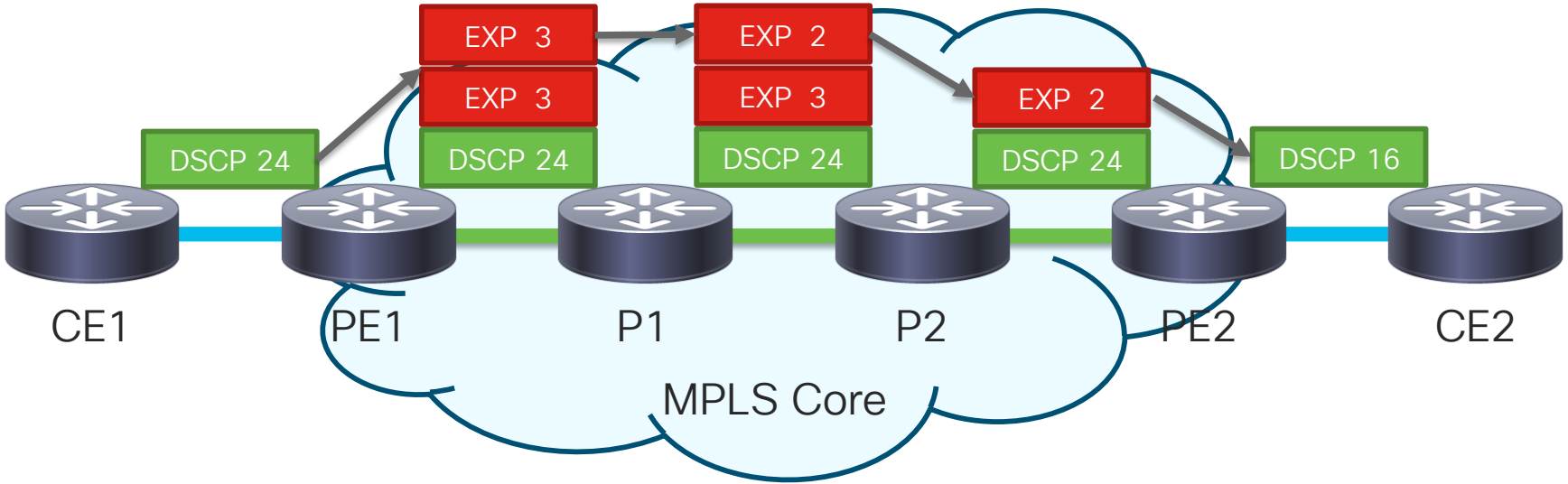
Overlay QOS

MPLS – Default Mode



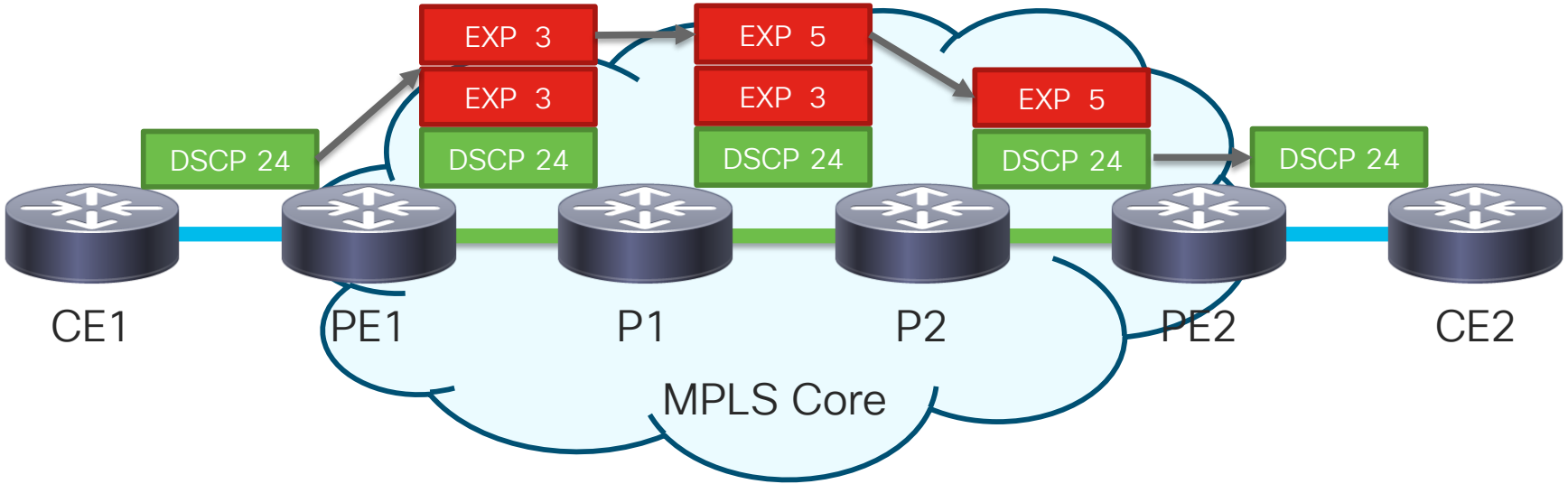
Overlay QOS

MPLS – Uniform Mode



Overlay QOS

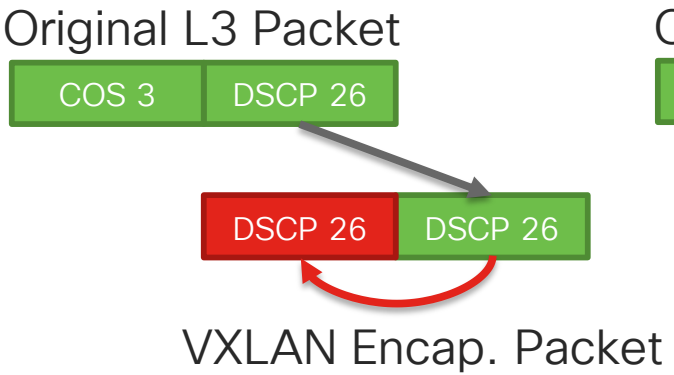
MPLS – Pipe Mode

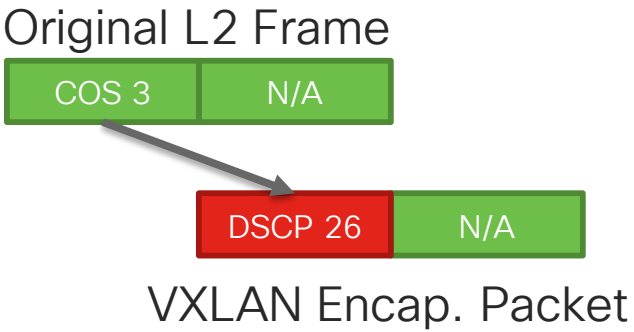


Overlay QoS

VXLAN EVPN – VXLAN Encapsulation

- Ingress L3 packet, original priority is mapped to outer header priority
- Ingress L2 frame, COS value will be mapped to outer priority
- VLAN header is not preserved in VXLAN tunnel



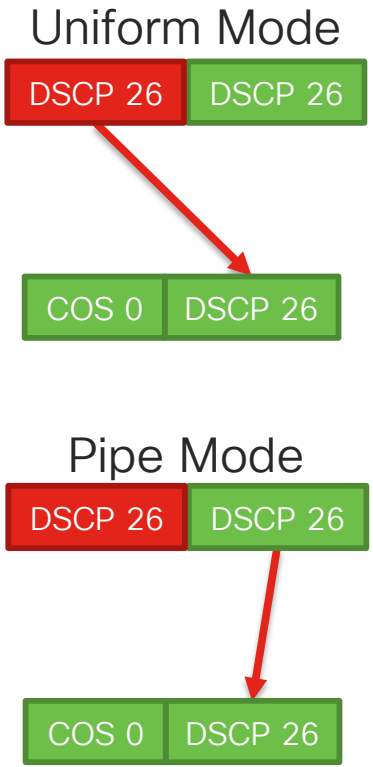


COS	DSCP
0	0
1	8
2	16
3	26
4	32
5	46
6	48
7	56

Overlay QoS

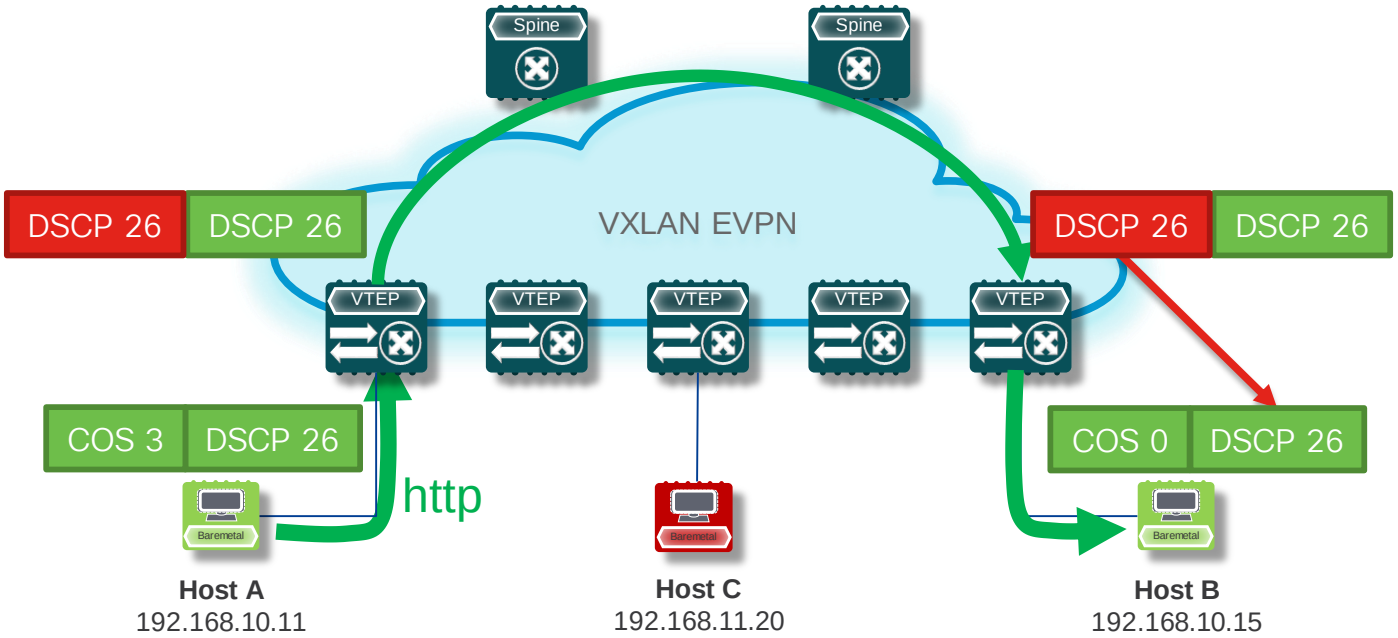
VXLAN EVPN – VXLAN Decapsulation

- DSCP value is derived based on a priority mode for L3 traffic:
  - Uniform mode: delivers overlay priority copying outer header to decapsulated frame
  - Pipe mode: extends original priority copying inner header to decapsulated frame
- Marking can be configure on the egress VTEP mark decapsulated traffic with priority (COS, DSCP)



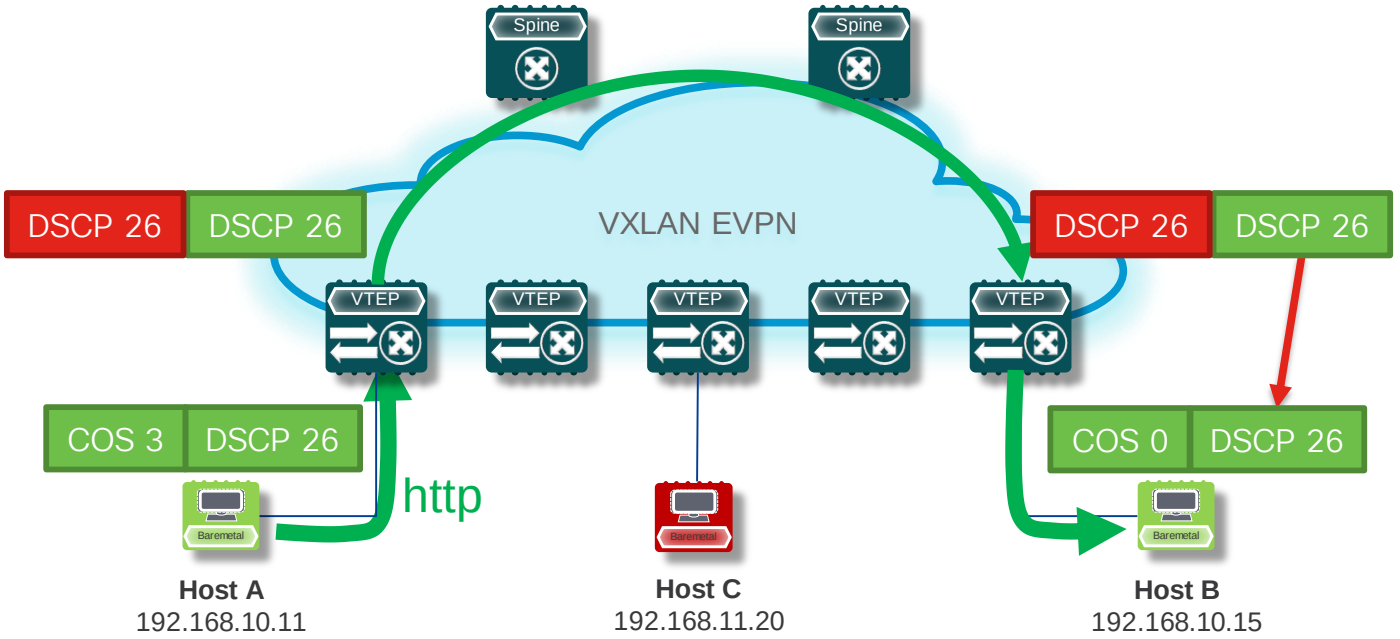
Overlay QoS

VXLAN – Uniform Mode



Overlay QoS

VXLAN – Pipe Mode



Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

Nexus 9000 QoS

You make networking **possible**

Cisco live!

Nexus 9000 Overview

- Modular and Fixed chassis
- Optimized for high density 10G/25G/40G/100G
- Standalone and ACI Mode
- Built with Cisco Silicon
  - Advanced QoS capabilities

Nexus 9000 - Cloud Scale

LSE

- 1.8T chip – 2 slices of 9 x 100G each
- X9700-EX modular linecards; 9300-EX TORs

LS1800FX

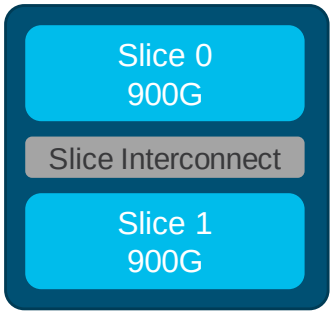
- 1.8T chip – 1 slice of 18 x 100G with MACSEC
- X9700-FX modular linecards; 9300-FX TORs

S6400

- 6.4T chip – 4 slices of 16 x 100G each
- E2-series fabric modules; 9364C TOR

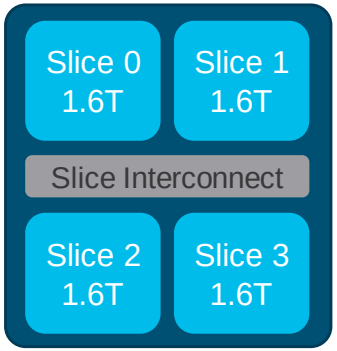
LS3600FX2

- 3.6T chip – 2 slices of 18 x 100G with MACSEC + CloudSec
- 9300-FX2 TORs

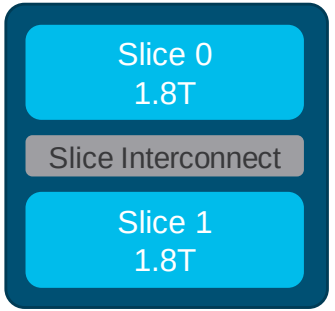


LSE – 18 x 100G

LS1800FX – 18 x 100G



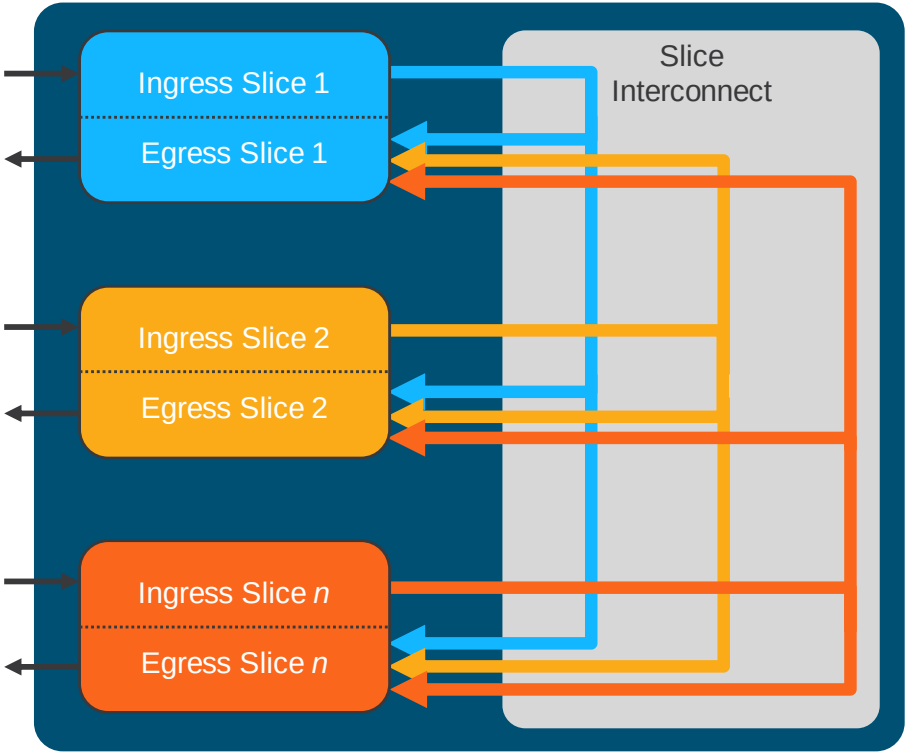
S6400 – 64 x 100G



LS3600FX2 – 36 x 100G

What Is a “Slice”?

- Self-contained forwarding complex controlling subset of ports on single ASIC
- Separated into Ingress and Egress functions
- Ingress of each slice connected to egress of all slices
- Slice interconnect provides non-blocking any-to-any interconnection between slices

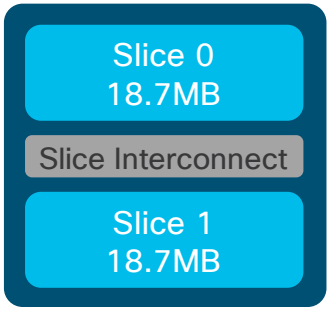


Cisco Nexus 9000 QoS Features

- Traffic classification
  - DSCP, CoS, IP Precedence and ACL
- Packet marking
  - DSCP, CoS, and ECN
- Strict Priority Queuing and DWRR
- Ingress and egress policing
- Tail Drop and WRED with ECN
- Shared buffer capability
- Egress Queuing

Buffering

- Cloud Scale platforms implement shared-memory egress buffered architecture
- Each ASIC slice has dedicated buffer – only ports on that slice can use that buffer
- Dynamic Buffer Protection adjusts max thresholds based on class and buffer occupancy
- Intelligent buffer options maximise buffer efficiency



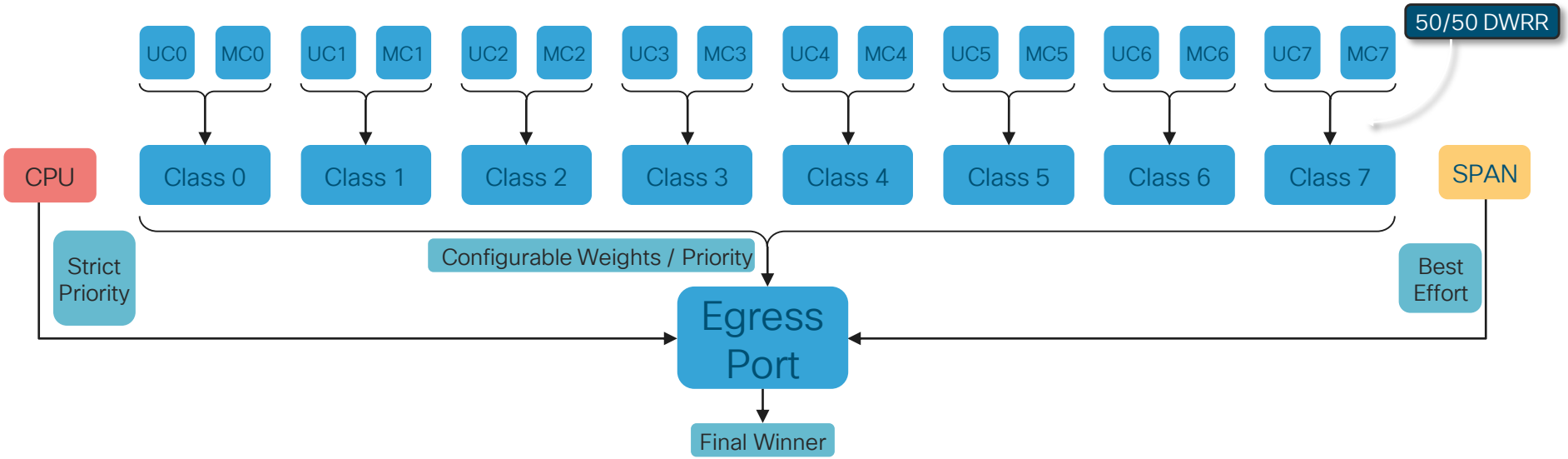
LSE
18.7MB/slice
(37.4MB total)

LS1800FX
40.8MB/slice
(40.8MB total)

S6400
10.2MB/slice
(40.8MB total)

LS3600FX2
20MB/slice
(40MB total)

Queuing and Scheduling

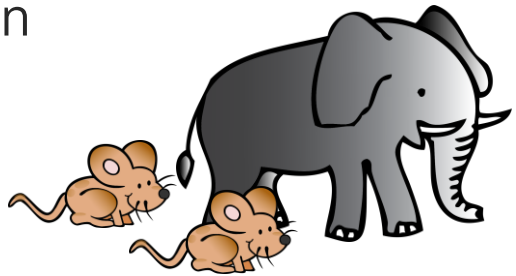


- 8 user classes and 16 queues per output port (8 unicast, 8 multicast)
- QOS-group drives class; egress queuing policy defines class priority and weights
- Dedicated classes for CPU traffic and SPAN traffic

Intelligent Buffering

Innovative Buffer Management for Cloud Scale switches

- **Dynamic Buffer Protection** (DBP) – Controls buffer allocation for congested queues in shared-memory architecture
- **Approximate Fair Drop** (AFD) – Maintains buffer headroom per queue to maximize burst absorption
- **Dynamic Packet Prioritization** (DPP) – Prioritizes short-lived flows to expedite flow setup and completion



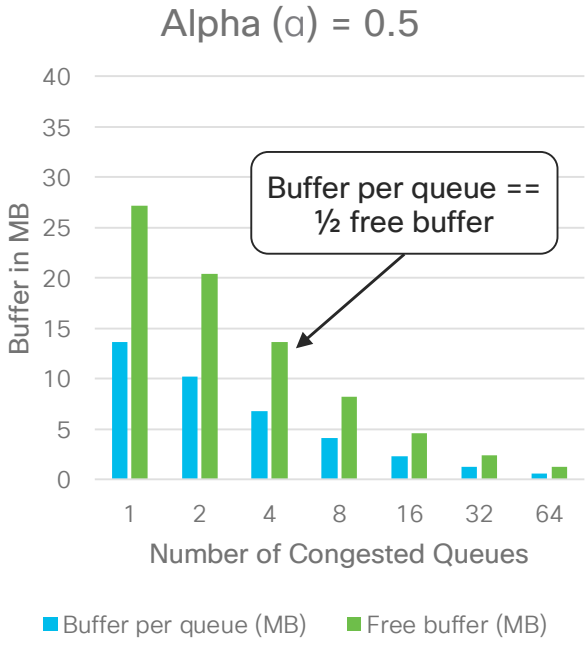
Miercom Report: Speeding Applications in Data Centre Networks
http://miercom.com/cisco-systems-speeding-applications-in-data-center-networks/

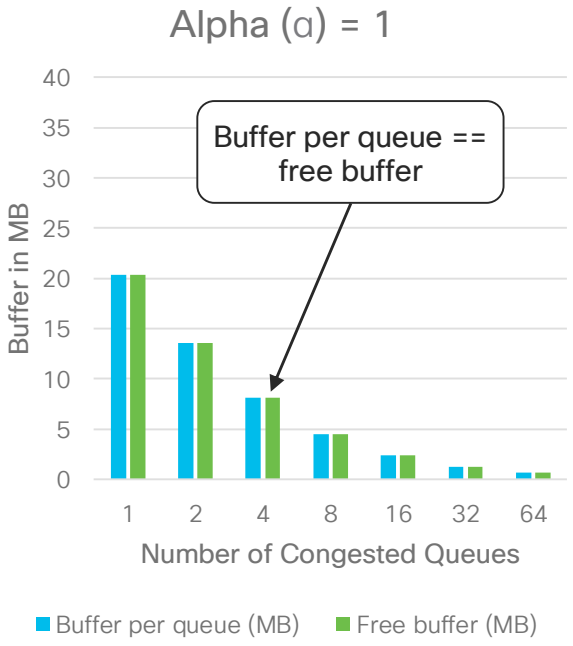
Dynamic Buffer Protection (DBP)

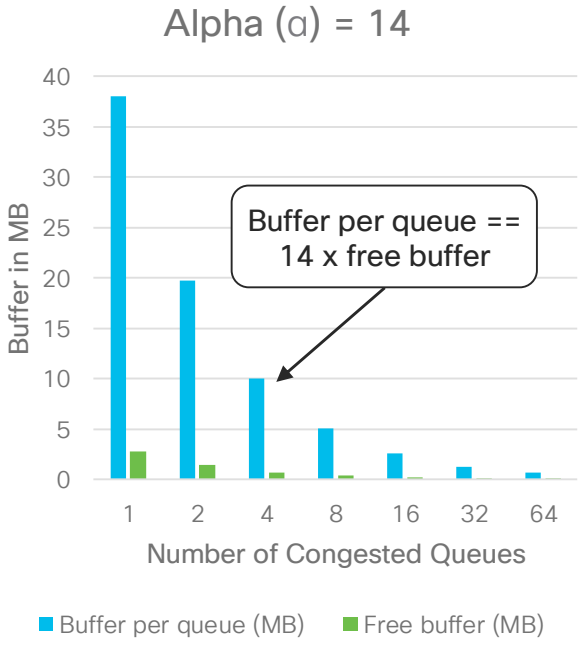
- Prevents any output queue from consuming more than its fair share of buffer in shared-memory architecture
- Defines dynamic max threshold for each queue
  - If queue length exceeds threshold, packet is discarded
  - Otherwise packet is admitted to queue and scheduled for transmission
- Threshold calculated by multiplying free memory by configurable, per-queue **Alpha** (α) value (weight)
  - Alpha controls how aggressively DBP maintains free buffer pages during congestion events

Alpha Parameter Examples

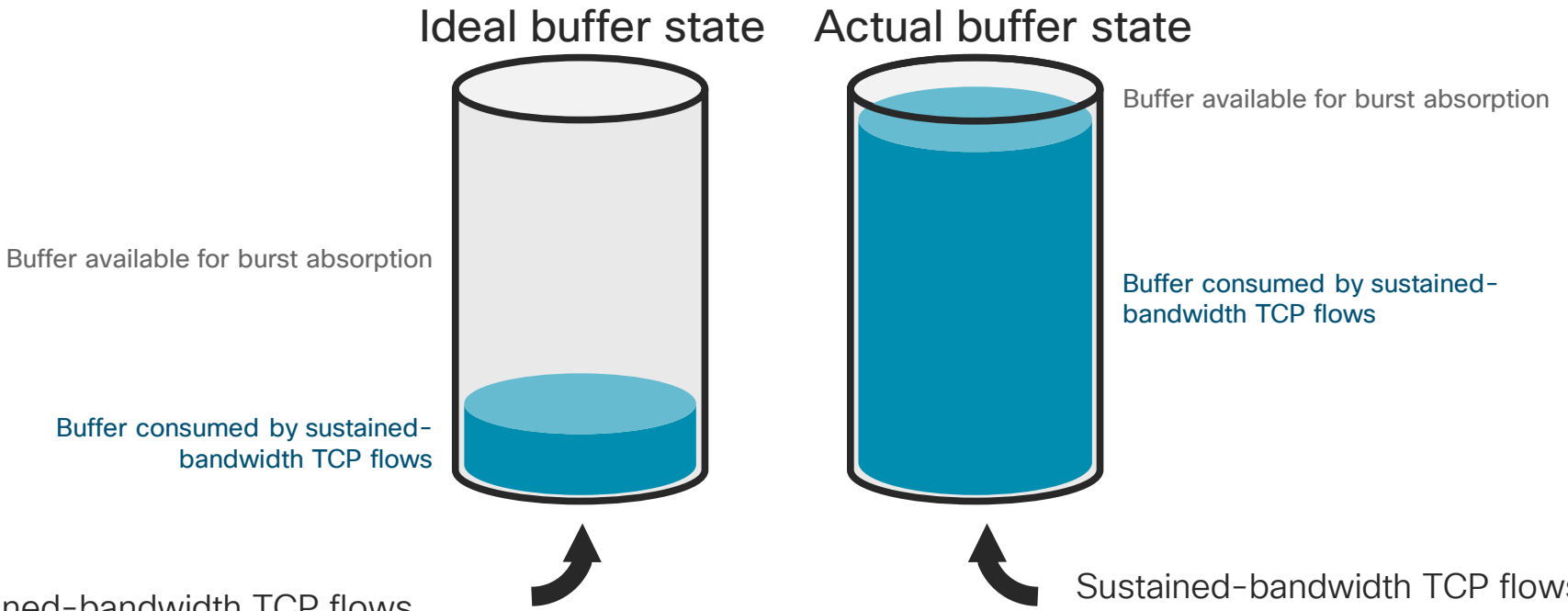
Default Alpha on Cloud Scale switches





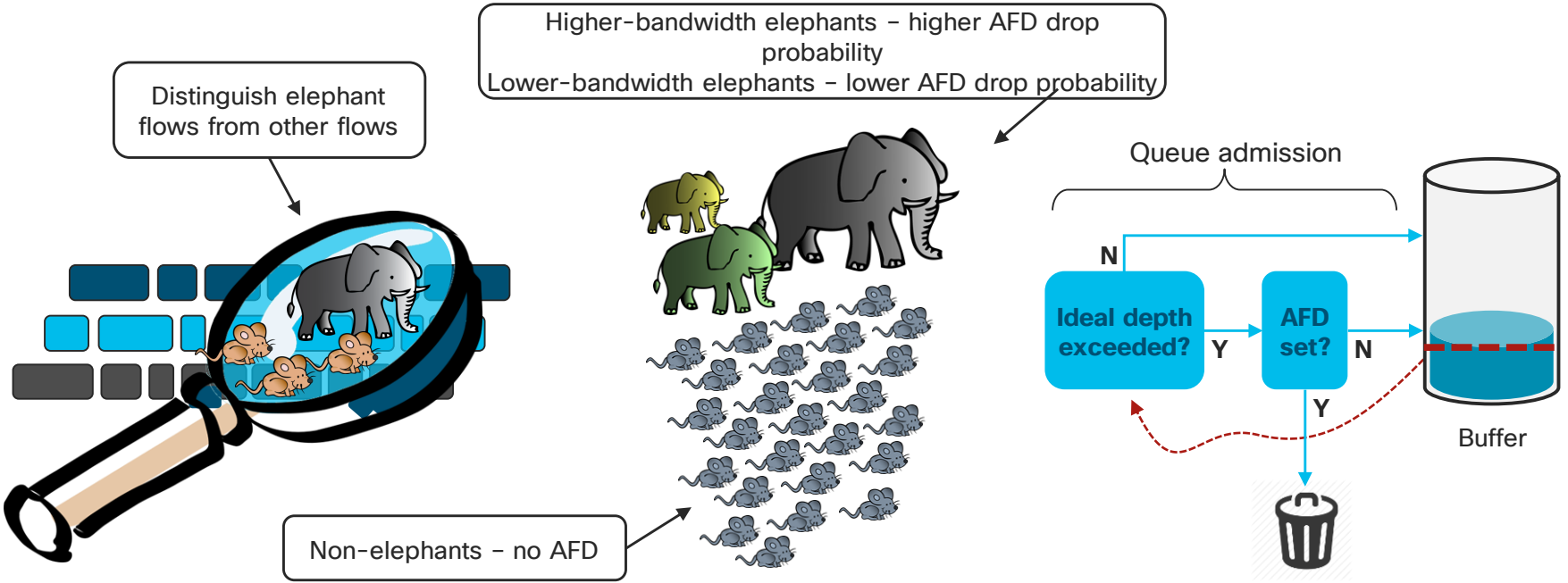


Buffering – Ideal versus Reality



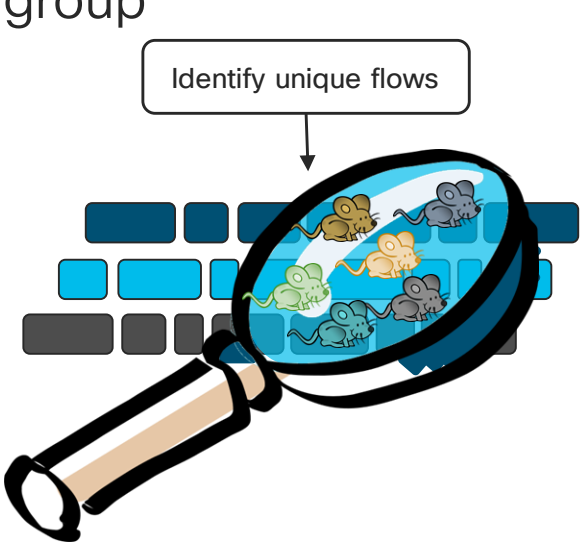
Approximate Fair Drop (AFD)

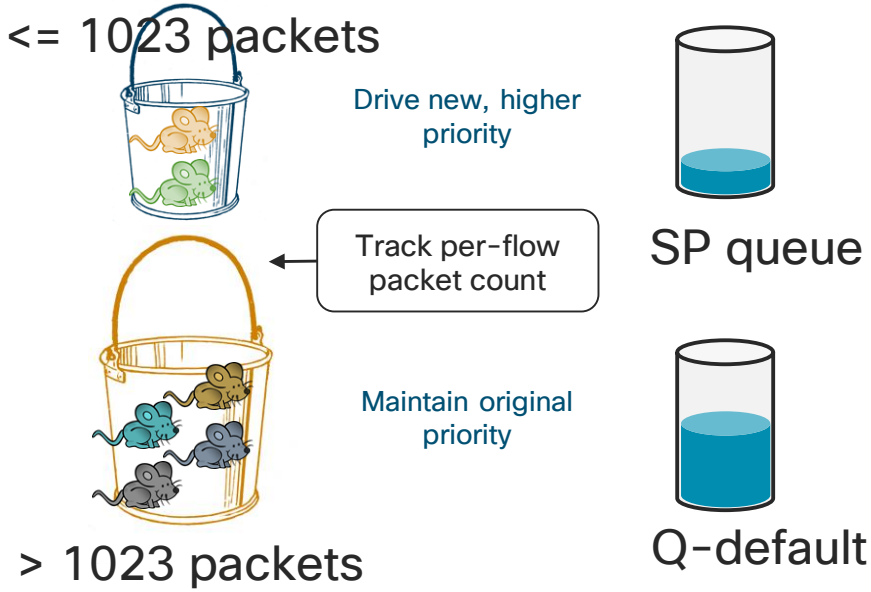
Maintain throughput while minimizing buffer consumption by elephant flows – **keep buffer state as close to the ideal as possible**



Dynamic Packet Prioritization (DPP)

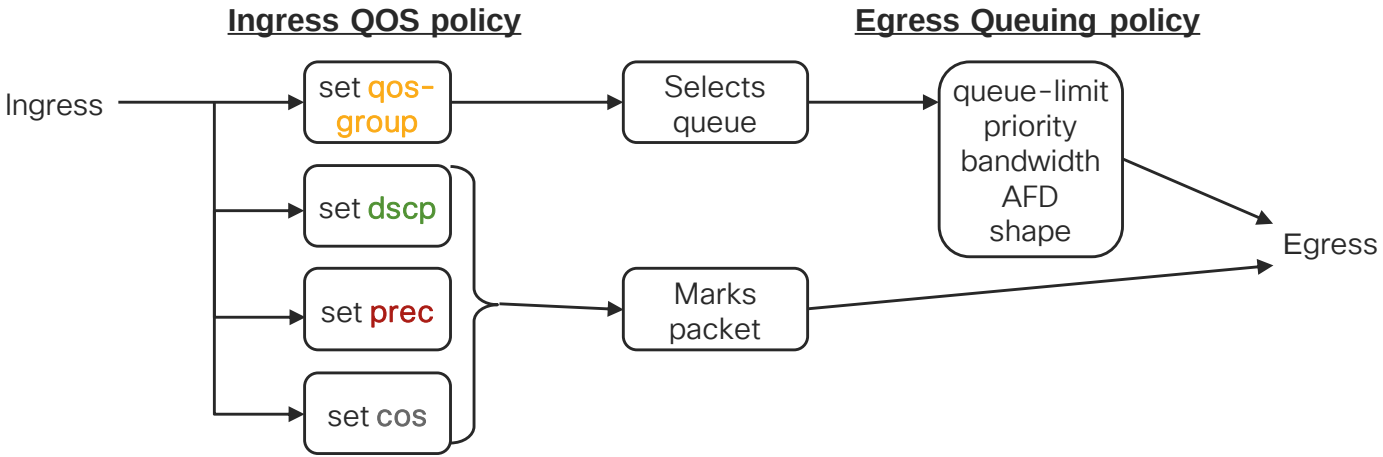
- Prioritize initial packets of new / short-lived flows
- Up to first 1023 packets of each flow assigned to higher-priority qos-group



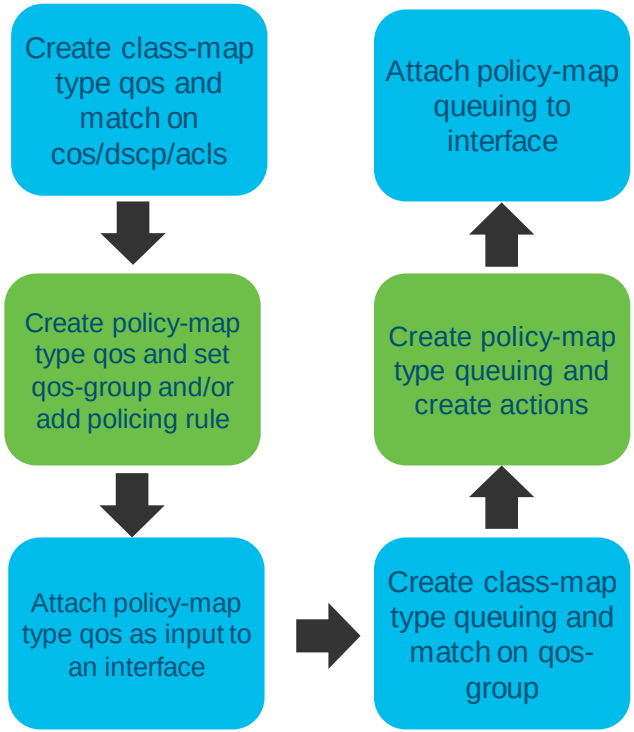


Ingress QOS / Egress Queuing Policies

- Default QOS behaviour:
  - **Trust** received QOS markings
  - All user data goes to q-default
- To select egress queue, use “set qos-group” in ingress QOS policy
- To set/change packet markings, use “set cos / precedence / dscp” in ingress QOS policy
- To change queuing behaviour, manipulate egress queuing policies



Putting it all together



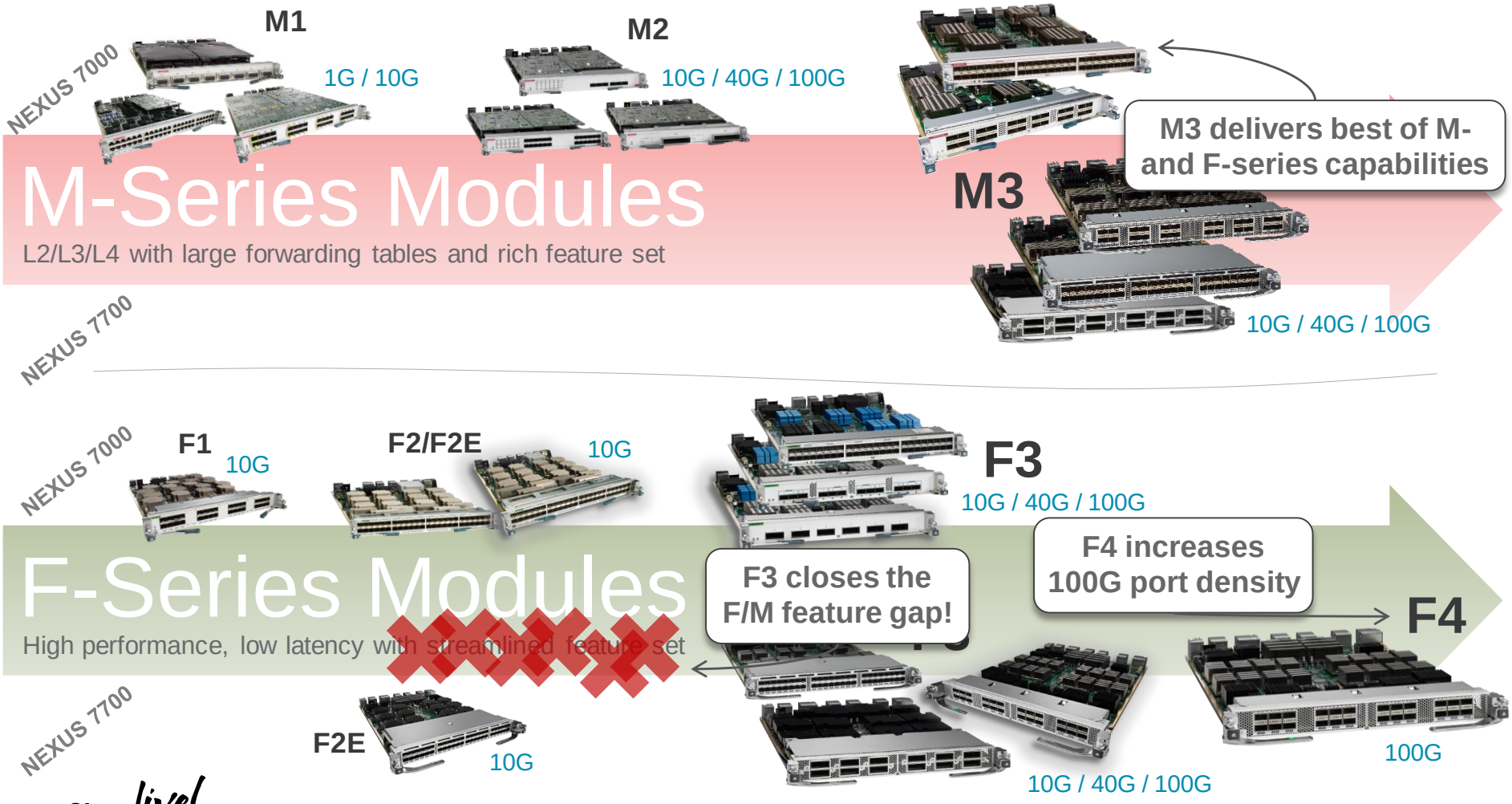

```
class-map type qos class_foo
    match cos 3-4

policy-map type qos pm1
    class type qos class_foo
        set qos-group 1
        police cir 20 mbytes conform transmit violate drop
    class type qos class-default
        set qos-group 0

interface ethernet 1/1
    service-policy type qos input pm1

class-map type queuing class-foo
    match qos-group 1

policy-map type queuing policy-foo
    class type queuing class-foo
        bandwidth percent 20
    class type queuing class-default
        bandwidth percent 80

interface ethernet 1/3
    service-policy type queuing output policy-foo
```

Nexus 9000 QoS Golden Rules

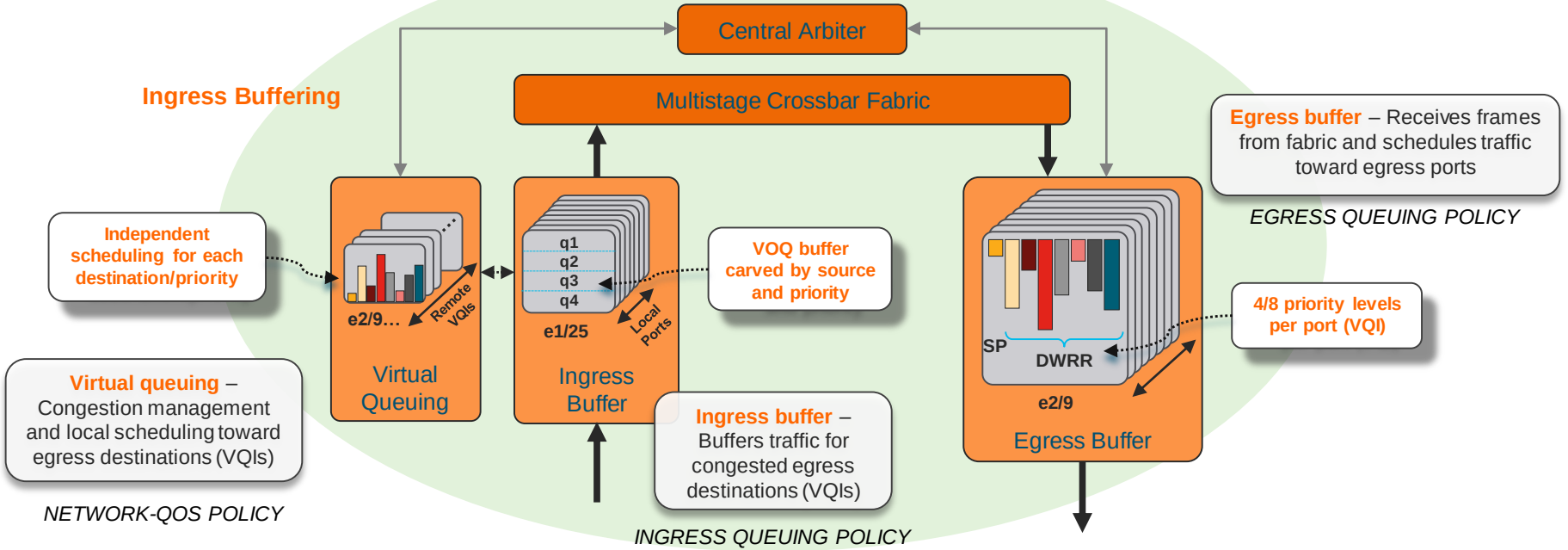
- QoS is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- Queuing and QoS policies are applied to a physical interface or at system level

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion



F3/M3/F4 – Ingress Buffered



F3/M3 I/O Module Buffering Capacity

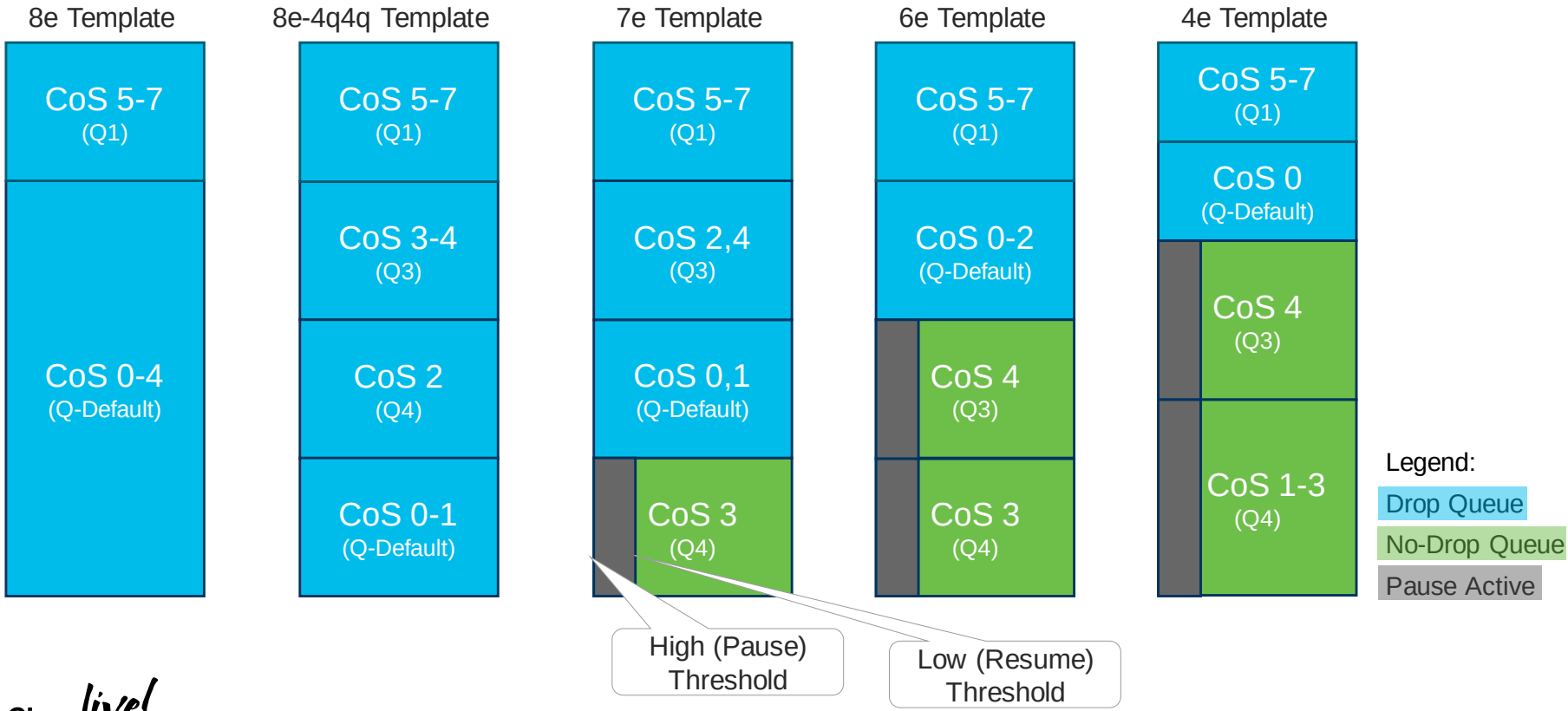
Ingress

Module	Total VOQ Buffer Per Module	Ingress Queue Structure	Ingress VOQ Buffer
M3 48-port 10G	1500MB	4q1t	31.25MB / port
F3 48-port 10G	72MB	4q1t	1.5MB / port
M3 24-port 40G	3000MB	4q1t	125MB / port
F3 24-port 40G	144MB	4q1t	6MB / port

Egress

Module	Egress VOQ Structure	Egress VOQ Buffer (Credited)	Egress VOQ Buffer (Uncredited)
M3 48-port 10G	1p7q1t	512KB / port	4MB / 24 ports
F3 48-port 10G	1p7q1t	295KB / port	512KB / 8 ports
M3 24-port 40G	1p7q1t	2MB / port	4MB / 6 port
F3 24-port 40G	1p7q1t	1.1MB / port	512KB / 2 ports

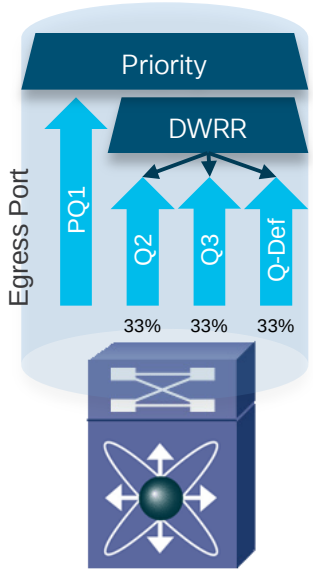
Ingress Queuing – Template View



Egress Queuing – Logical View

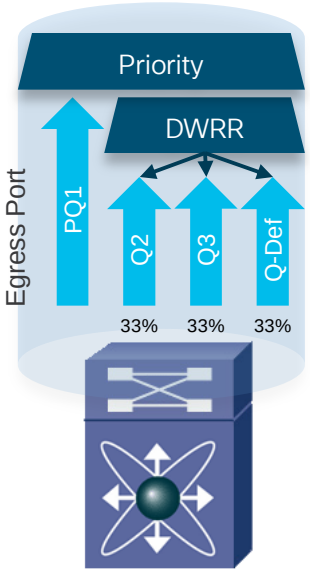
default-4q-8e-out-policy

PQ1 (5,6,7) **Q2** (3,4) **Q3** (2) **Q-Def.** (0,1)



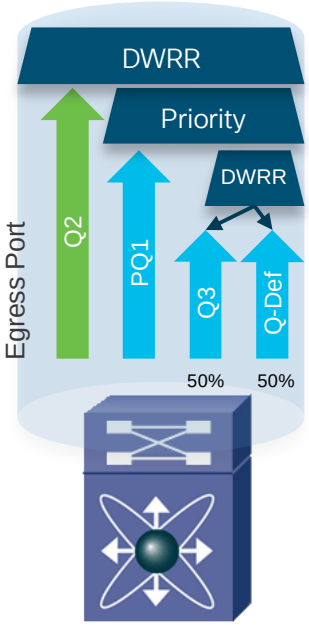
default-4q4q-8e-out-policy

PQ1 (5,6,7) **Q2** (3,4) **Q3** (2) **Q-Def.** (0,1)



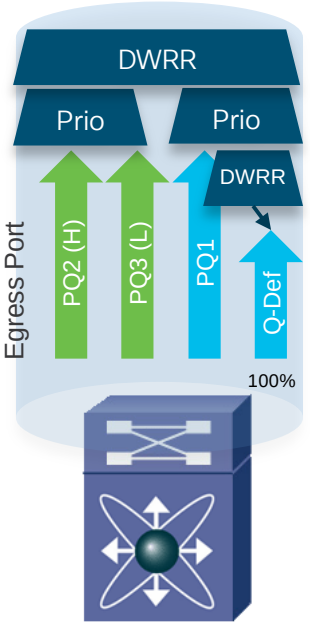
default-4q-7e-out-policy

Q2 (3) **PQ1** (5,6,7) **Q3** (2,4) **Q-Def.** (0,1)



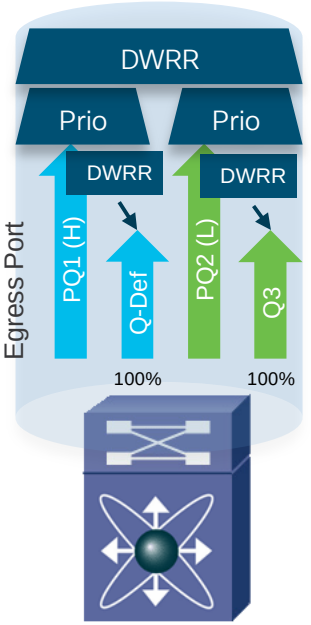
default-4q-6e-out-policy

PQ2. (4) **PQ3** (3) **PQ1** (5,6,7) **Q-Def.** (0-2)



default-4q-4e-out-policy

PQ1 (5,6,7) **Q-Def.** (0) **PQ2** (4) **Q3** (1,2,3)



Green indicates no-drop

DSCP to CoS / CoS to DSCP – Mapping Tables

```
N7k# show table-map | grep -a 2 dscp-cos-map
  Table-map dscp-cos-map
    default copy
```

```
N7k# show system internal ipqos global-defaults | grep -a 12 cos-dscp-map
table-map: cos-dscp-map (len: 12)
        default copy
        Bit array:
        Values set:

              0  8 16 24 32 40 48 56
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
             -- -- -- -- -- -- -- --
```

CoS **2** mapped to DSCP **16-23**

```
N7k# show table-map | grep -a 2 cos-dscp-map
  Table-map cos-dscp-map
    default copy
```

```
N7k# show system internal ipqos global-defaults | grep -a 12 dscp-cos-map
table-map: dscp-cos-map (len: 12)
        default copy
        Bit array:
        Values set:

              0  0  0  0  0  0  0  0
              1  1  1  1  1  1  1  1
              2  2  2  2  2  2  2  2
              3  3  3  3  3  3  3  3
              4  4  4  4  4  4  4  4
              5  5  5  5  5  5  5  5
              6  6  6  6  6  6  6  6
              7  7  7  7  7  7  7  7
```

DSCP **24-31** mapped to CoS **3**

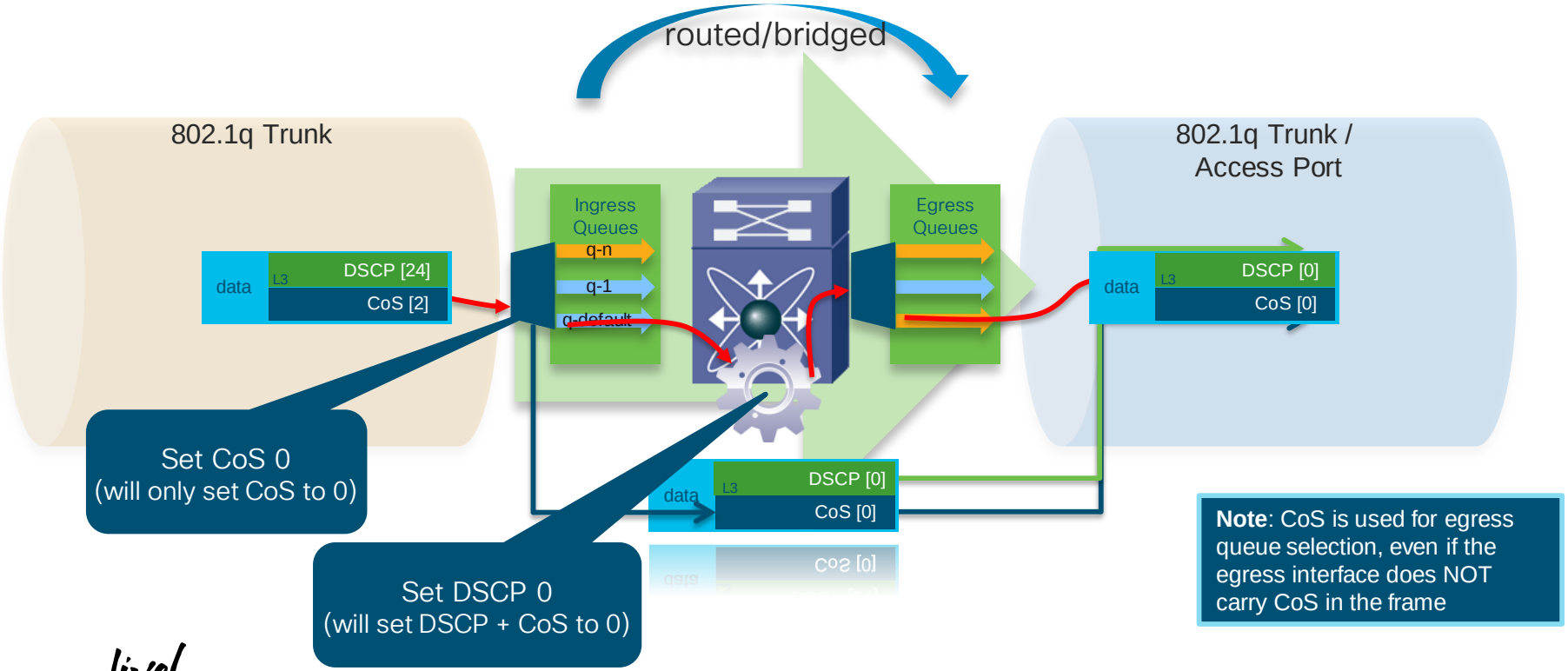
Note: Output taken from Nexus 7000

CoS or DSCP to Queue Mapping

- Default CoS to Queue Mapping for Nexus 7000/7700 (F- and M-Series I/O Module)
  - Ingress: CoS to Queue
  - Egress: CoS to Queue
- DSCP to Queue Mapping for Nexus 7000/7700 (F- and M-Series I/O Module)
  - Ingress: DSCP to Queue
  - **Egress: CoS to Queue**
- Global Configuration (Admin/Default VDC) required to enable DSCP to Queue Mapping:

```
N7k(config)# hardware qos dscp-to-queue ingress module type {all | f-series | m-series}
```

Changing the Default Trust



Default Rules Summary

Routed Traffic

- If CoS and DSCP is present
  - CoS is used for ingress queue selection
  - DSCP is preserved and rewrites CoS (top most 3bit)
  - CoS is used for egress queue selection
- If only DSCP is present
  - No CoS gets treated as CoS 0 on ingress
  - DSCP is preserved and rewrites CoS (top most 3bit)
  - CoS (derived from DSCP) drives egress queue selection

Bridged Traffic

- If CoS and DSCP is present
  - CoS is used for ingress queue selection
  - CoS is preserved
  - DSCP is unmodified
  - CoS is used for egress queue selection
- If only DSCP is present
  - No CoS gets treated as CoS 0 on ingress
  - CoS 0 is used for ingress and egress queue selection
  - DSCP is unmodified

Nexus 7000 QoS Golden Rules

- QoS is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Default Queuing and QoS policies are applied to all physical interfaces across all VDCs
- For bridged traffic, CoS is preserved, DSCP is unmodified
- For routed traffic, DSCP is copied to CoS (first 3 bits)
  - Ex: DSCP 40 (b101000) becomes CoS 5 (b101)

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

Nexus 5000 Series Overview

10G\40G

Nexus 5600

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

Nexus 5500

Flexibility
10G Uplinks

Nexus 5010/5020

Scalability (Large Buffers and Tables)

Fabric Innovations

VXLAN

Density

Programmability

Network Visibility

LAN/SAN Convergence

FabricPath

FEX Architecture

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

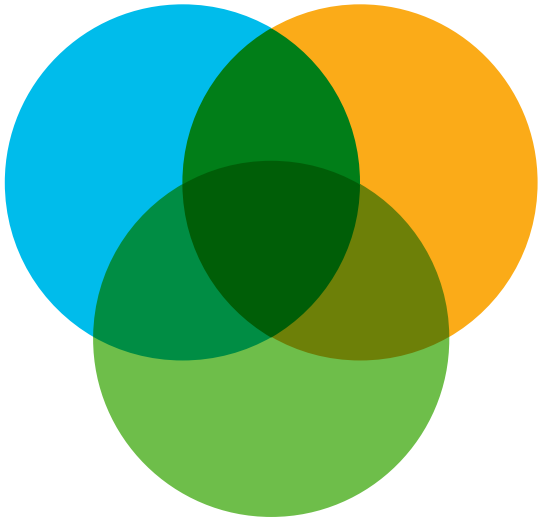
CUSTOMER VALUE

CISCO INNOVATION

Key Concepts – Common Points

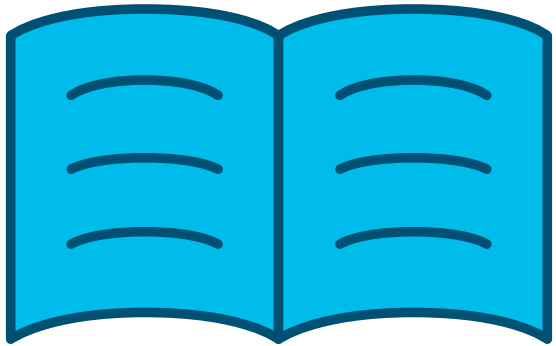
Nexus 7000 compared to Nexus 5000 QoS

- Nexus 5000/6000 and Nexus 7000 F-Series I/O Modules share the Ingress Buffer Model
- Ingress buffering and queuing occur at VOQ of each ingress port
- Egress scheduling enforced by egress port
- No Egress QOS Policies

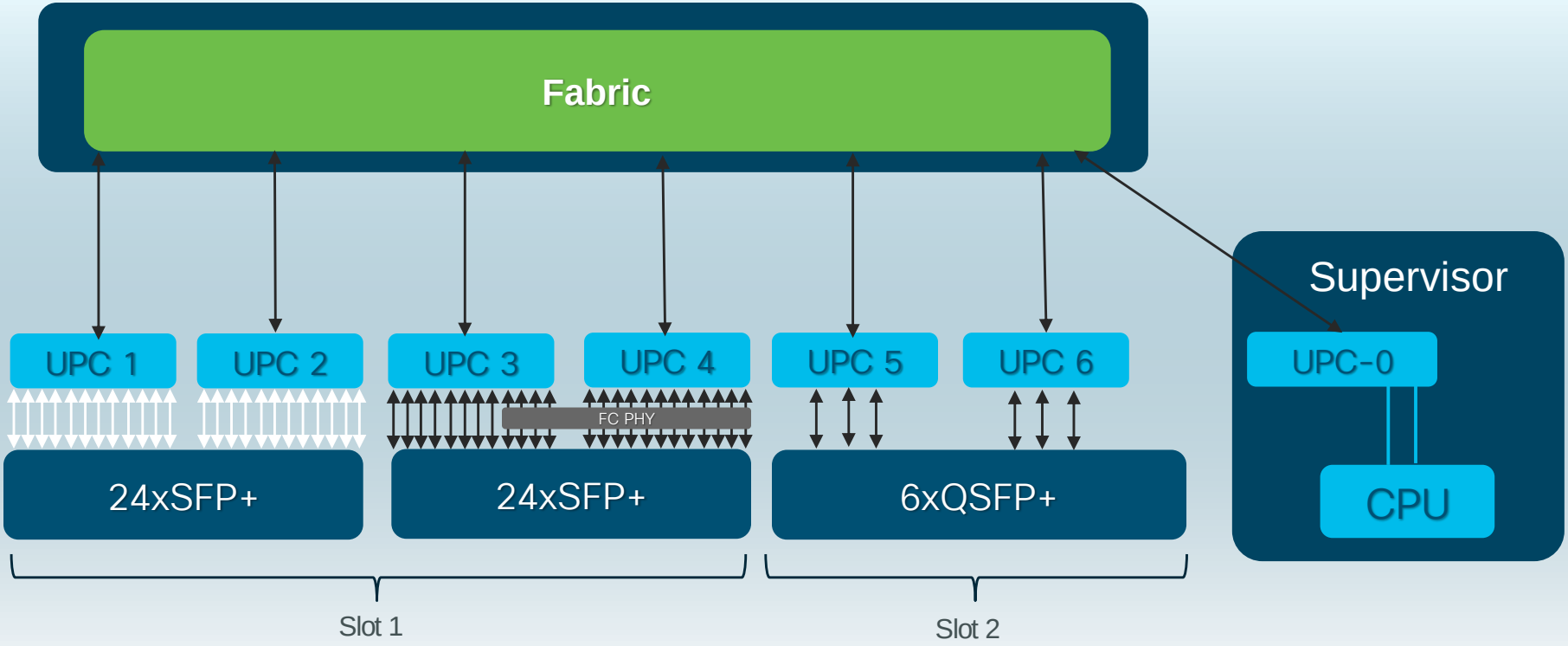


Cisco Nexus 5600 QoS Features

- 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)
- Flexible buffer management

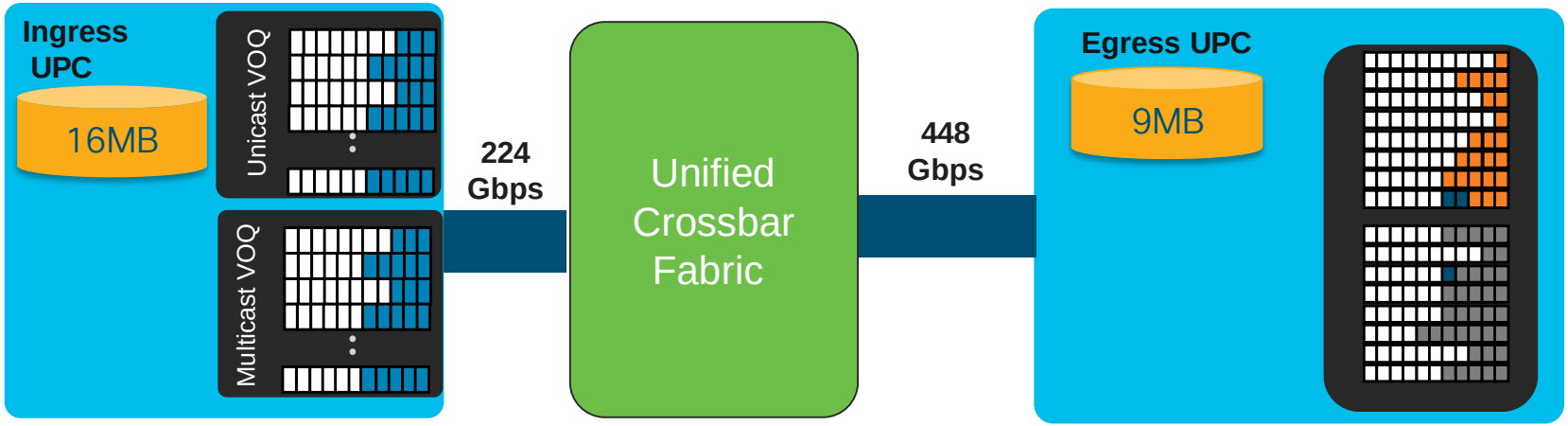


Cisco Nexus 5672UP Internal Architecture



Packet Buffering

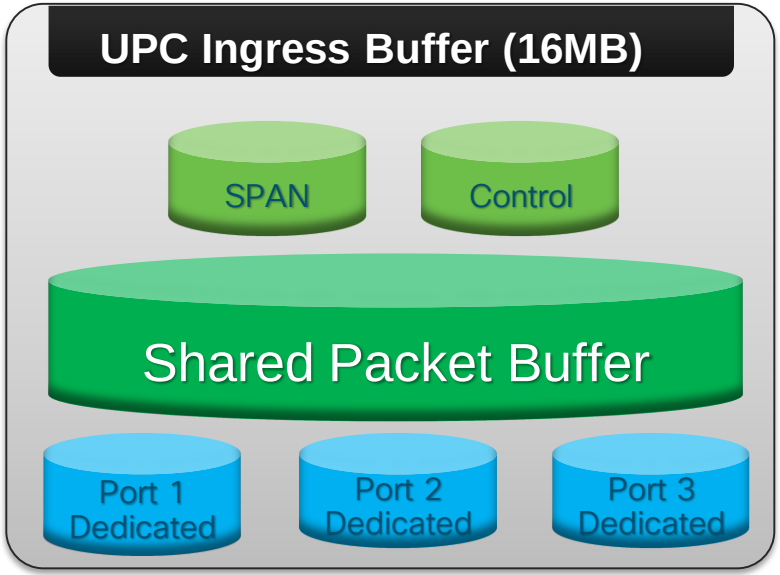
- 25MB packet buffer is shared by every three 40 GE ports or twelve 10 GE ports.
- Buffer is 16MB at ingress and 9MB at egress.
- Unicast packet can be buffered at both ingress and egress.
- Multicast Buffered at egress only



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

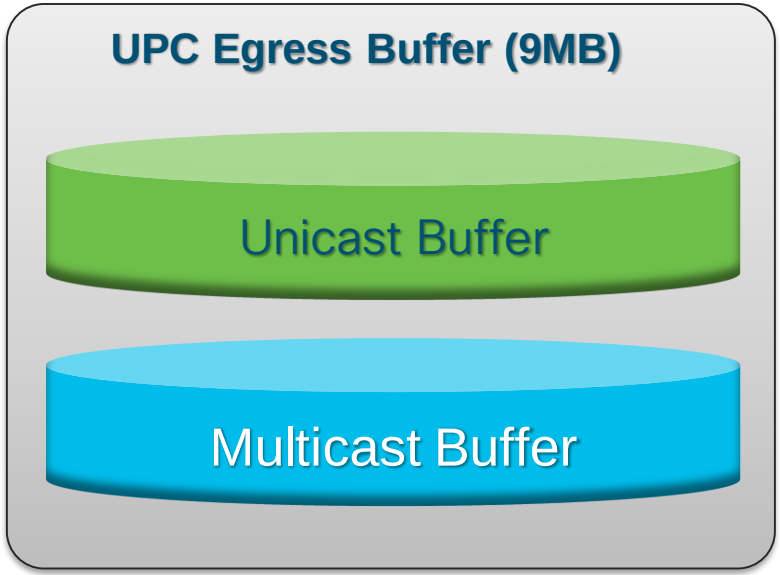
- “queue-limit” under “network-qos” policy specifies the dedicated buffer for each port and each class. The dedicated buffer can be used by the port for only that class of service.
- Without “queue-limit” each class of service will get 100 KB of dedicated buffer.
- The size of dedicated buffer can be 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 from unicast to multicast.
- Unicast traffic can be buffered at egress and ingress.
- Multicast is buffered at egress in case of interface oversubscription.



Default Egress Buffer Allocation

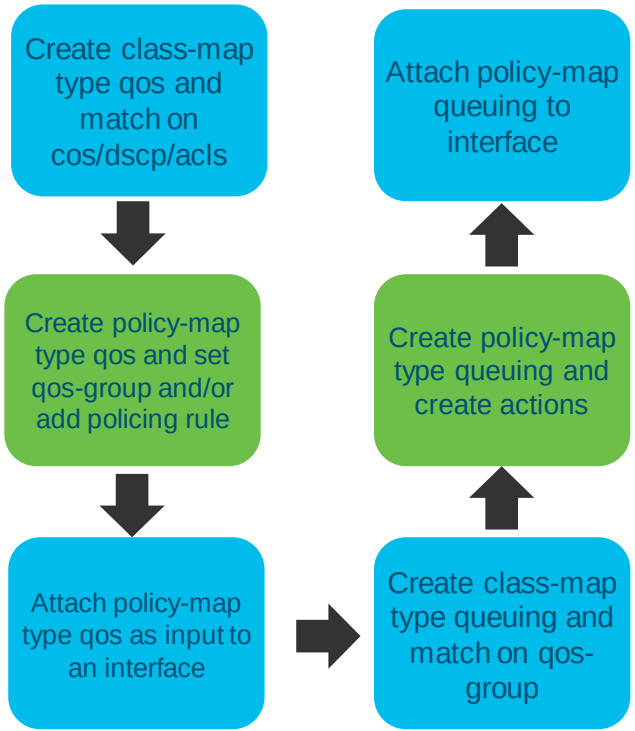
- Software provides CLI to tune the egress buffer allocation.
- At egress, unicast buffer is allocated on a per-port basis. For multicast, the egress buffer is shared among all ports.
- Use ”hardware multicast-buffer-tune” to assign unicast buffer to multicast pool on egress

Buffer pool	10GE Port	40GE 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

Nexus 5600/6000 QoS Configuration Model

- Uses QOS-Groups to tie together QoS, Queuing and Network-QoS policies
- QoS-Group has no direct relation with priority values
- QoS-Groups defined (set) in policy-map type qos.
- QoS-groups referenced (match) in policy type queuing and policy-map type network-qos

Putting it all together




```
class-map type qos class_foo
    match cos 3-4

policy-map type qos pm1
    class type qos class_foo
        set qos-group 1
    class type qos class-default
        set qos-group 0

interface ethernet 1/1
    service-policy type qos input pm1

class-map type queuing class-foo
    match qos-group 1

policy-map type queuing policy-foo
    class type queuing class-foo
        bandwidth percent 20
    class type queuing class-default
        bandwidth percent 80

interface ethernet 1/3
    service-policy type queuing input policy-foo
```

Buffering Capacity

Ingress

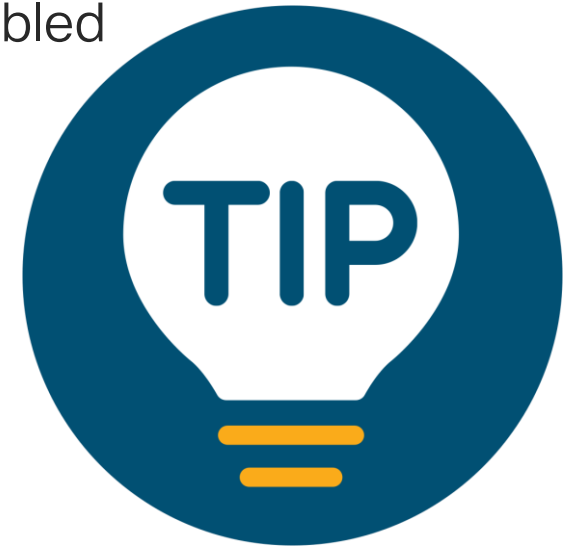
Traffic Type	Ingress Queue Structure	10 GE Port	40 GE Port
Control traffic (per port)	6q1t	64 KB	67 KB
Span Traffic (per Port)	6q1t	38.4 KB	154 KB
Class Default (per Port)	6q1t	100 KB	100 KB
Shared Buffer	6q1t	13.2 MB	14.7 MB

Egress

Traffic Type	Egress Queue Structure	10 GE Port	40 GE Port
Unicast	1p5q0t	363 KB	650 KB with 10GB Fabric Mode 635 KB with 40GB Fabric Mode
Multicast	1p5q0t	4.3MB	6.6 MB

Nexus 5600 QoS Golden Rules

- WRED is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- No Egress QOS policies



Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

Nexus 3000 Series Switches

Nexus 3100

- ToR Leaf
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

Nexus 3200

- Fixed High Density
- High throughput and performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

Nexus 3600

- Deep Buffer
- High route scale
- Video and Drop sensitive deployments
- Based on Jericho ASIC family

Nexus 3400

- Programmable pipeline
- Support for P4-INT
- Enable custom use cases
- Includes Tofino and Teralynx ASICs

Nexus 3500

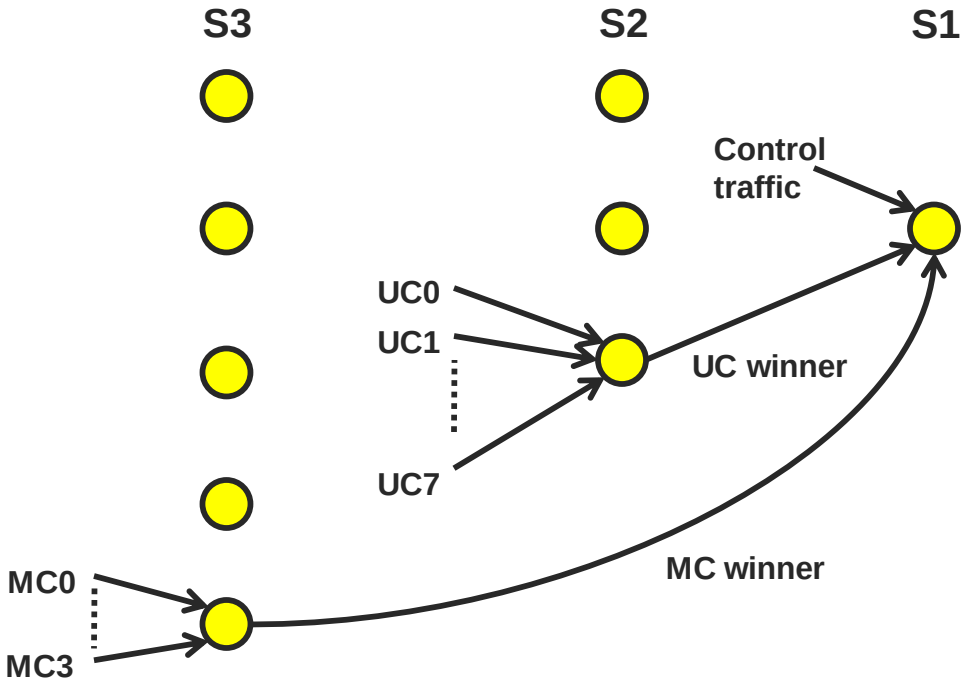
- Ultra Low Latency
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

Cisco Nexus 3000 QoS Features

- Traffic classification
  - DSCP, CoS, IP Precedence and ACL
- Packet marking
  - DSCP, CoS, and ECN
- Strict Priority Queuing and DWRR
- Tail Drop and WRED with ECN
- Shared buffer capability
- Egress Queuing
- 3-level hierarchical scheduling

Hardware Scheduler Implementation

- 3 level scheduling hierarchy



Dynamic Buffer Protection

- Buffer is shared dynamically any queue can use shared buffer
- Dynamic Buffer Protection prevents any queue unfair use shared buffer
- The basic algorithm uses dynamic queue length threshold, and account for usage of unicast and multicast

Nexus 3000 QoS Golden Rules

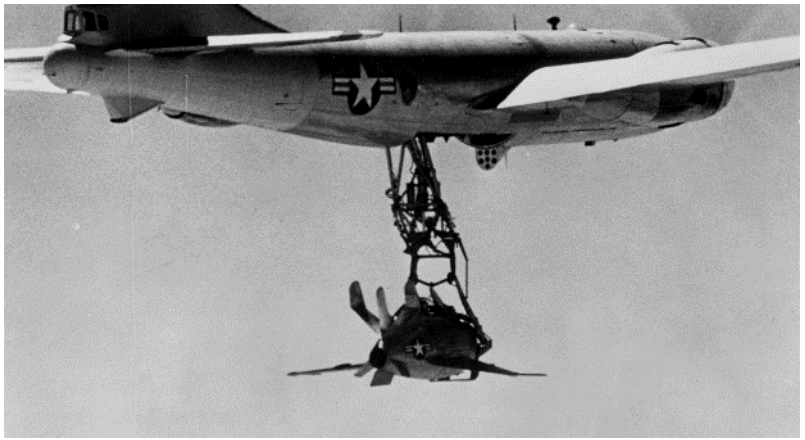
- QoS is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- Queuing and QoS policies are applied to a physical interface or at system level

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

FEX Overview

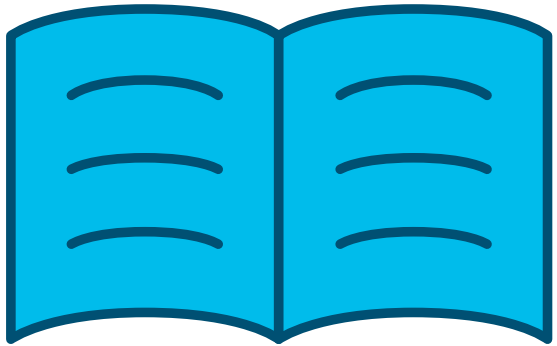
- Scalable and Extensible Fabric
- Single point of management
- Homogeneous and consistent policies



By Author listed as "U.S. Air Force photo" [Public domain], via Wikimedia Commons

Cisco Nexus 2000 QoS Features

- Traffic classification
  - DSCP, CoS
  - ACL classification (FEX offload) on Nexus 5600/6000
- Strict Priority Queuing and DWRR
- Priority Flow Control
- Queue-limit Carving



FEX Policy Offload (Nexus 5600/6000 only)

- TCAM resources on a FEX to perform ACL-based classification
- The feature is disabled by default
- By default, a FEX classifies packets on CoS value
- Both system level and interface level policies are offloaded to the FEX

```
switch# configure terminal
fex chassis_ID
hardware card-type qos-policy-offload
```

FEX Policy with Nexus 9000 as parent

- The FEX QoS policy is applied to the hardware resources of the fabric port associated with the FEX HIF port
- Classification is based on the COS value.
- System level input queueing for DWRR and Strict priority scheduling for HIF to NIF traffic and for NIF to HIF traffic
- Queuing:
  - 4 queues are present on the FEX
  - The scheduling is done per port and each port has its own scheduler.

FEX Queuing Policies – Nexus 7000

- On Nexus 7000 with FEX + M-Series parent modules, network-qos and F-series ingress queuing class-maps drive FEX queuing configuration
- Ingress queuing class-maps drive:
  - Both ingress and egress COS/DSCP-to-queue mapping
- Enabling DSCP-to-queue on parent switch enables DSCP-to-queue on FEX
  - DSCP-to-queue only active in the HIF→NIF direction
  - NIF→HIF direction always uses COS-to-queue mapping, based on COS transmitted by parent switch to FEX

FEX Queue-Limit – Nexus 7000

- Provides FEX queue-limit configuration option
- Manages buffer thresholds on FEX based on platform capabilities
- Default has queue-limit enabled
- Configuration applied per-VDC (on Nexus 7000/7700)
- Different FEX models have different capabilities

Nexus 2000 QoS Golden Rules

- FEX QOS classification on COS or DSCP unless FEX offload enabled
- FEX queuing driven implicitly by parent switch queuing configuration
- No support for per-queue shaping, policing or marking
- Drop thresholds are tail-drop only, no WRED support

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

What do we want to achieve?

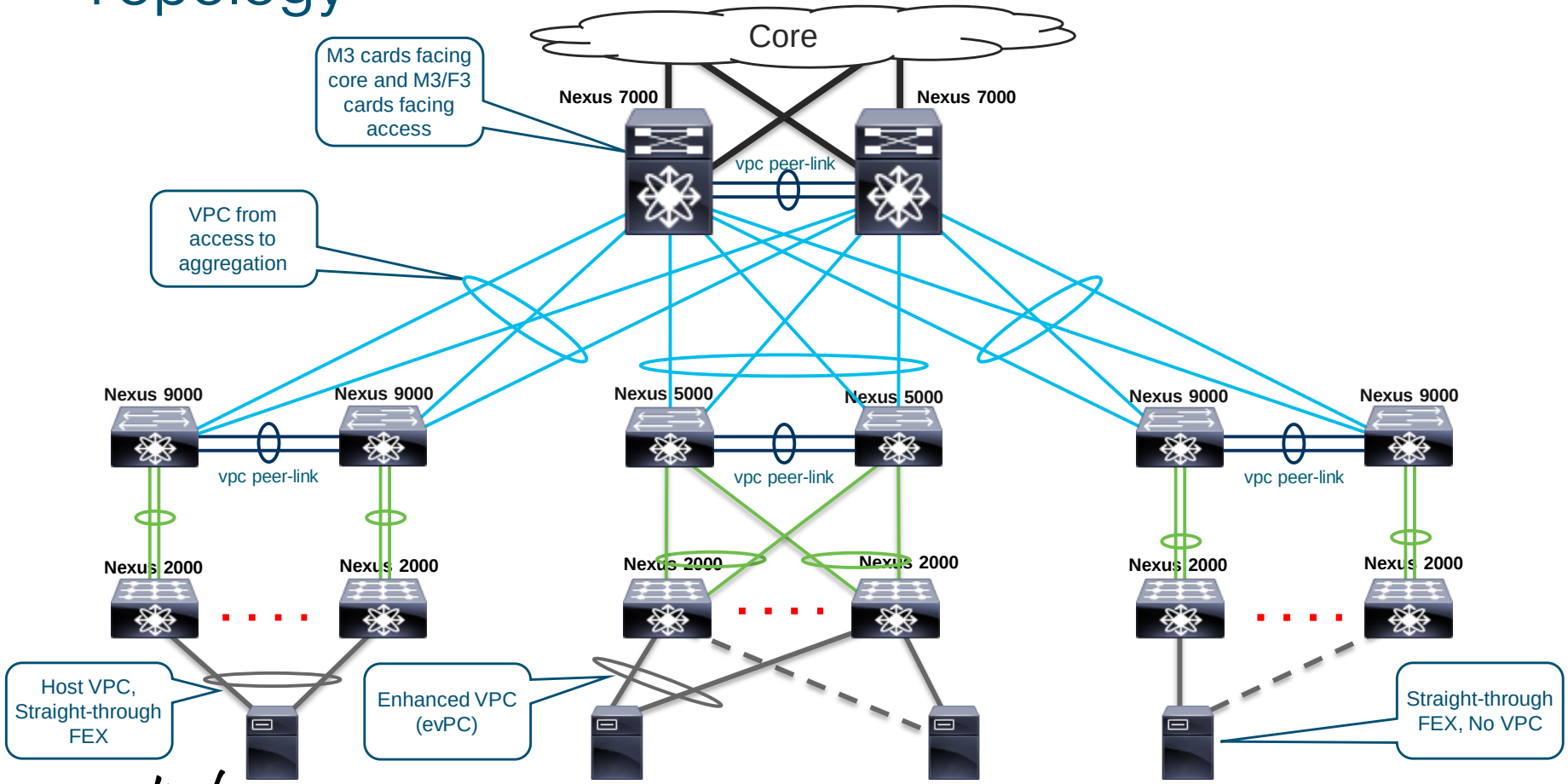
Company XYZ's Business Goals

- Make sure no disruption in network services
  - *Put control traffic in priority queue*
- Video/voice hosting also an business objective
  - *Put voice traffic in priority queue*
  - *Dedicated bandwidth to video traffic*
- Flexibility in moving applications across servers
  - *Dedicated bandwidth to vmotion/mobility*
  - *Everything else best-effort*

Translating to the language of QoS

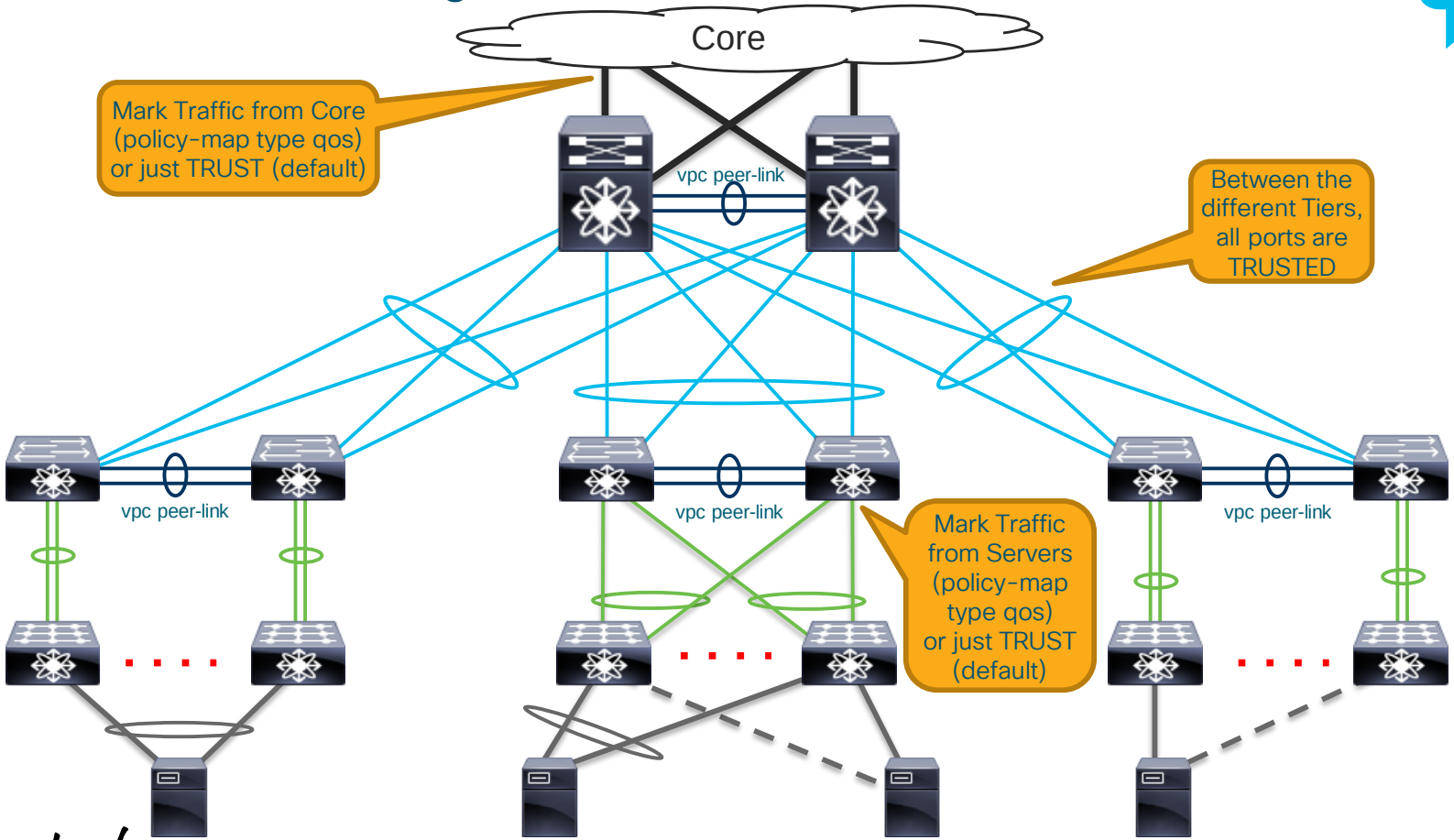
Application	CoS	Queuing (Scheduling)	Queue-Limit (Buffer)	Character
Best Effort	0, 1	BW remaining 50%	60%	High Volume / Less Important
vMotion / Live Migration	2	BW remaining 20%	10%	Medium Volume / Important
Multimedia	3, 4	BW remaining 30%	20%	Medium Volume Very Important
Strict Priority	5	Priority Queue	10%	Low Volume / Important / Delay Sensitive
Network Control	6,7			Low Volume / Very important

Topology



Classification, Marking and Trust on Nexus 5000/7000/9000

Type: QoS



Type:
QoS

Classification and Marking: Nexus 7000

```
ip access-list ACL_QOS_LOWPRIO
  10 permit …
ip access-list ACL_QOS_VMOTION
  10 permit …
ip access-list ACL_QOS_MULTIMEDIA
  10 permit …
ip access-list ACL_QOS_STRICTPRIO
  10 permit …
!
class-map type qos match-any CM_QOS_LOWPRIO_COS1
  match access-group name ACL_QOS_LOWPRIO
!
class-map type qos match-any CM_QOS_VMOTION_COS2
  match access-group name ACL_QOS_VMOTION
!
class-map type qos match-any CM_QOS_MULTIMEDIA_COS4
  match access-group name ACL_QOS_MULTIMEDIA
!
class-map type qos match-any CM_QOS_STRICTPRIO_COS5
  match access-group name ACL_QOS_STRICTPRIO
```

```
policy-map type qos PM_QOS_MARK_COS_IN
  class CM_QOS_STRICTPRIO_COS5
    set cos 5
  class CM_QOS_MULTIMEDIA_COS4
    set cos 4
  class CM_QOS_VMOTION_COS2
    set cos 2
  class CM_QOS_LOWPRIO_COS1
    set cos 1
!
interface Ethernet1/1
  service-policy type qos input PM_QOS_MARK_COS_IN
!
vlan configuration 100
  service-policy input PM_QOS_MARK_COS_IN
```

Type:
QoS

Classification and Marking: Nexus 5600 (1)

```
ip access-list ACL_QOS_LOWPRIO
  10 permit …
ip access-list ACL_QOS_VMOTION
  10 permit …
ip access-list ACL_QOS_MULTIMEDIA
  10 permit …
!
class-map type qos match-any CM_QOS_LOWPRIO_COS1
  match access-group name ACL_QOS_LOWPRIO
!
class-map type qos match-any CM_QOS_VMOTION_COS2
  match access-group name ACL_QOS_VMOTION
!
class-map type qos match-any CM_QOS_MULTIMEDIA_COS4
  match access-group name ACL_QOS_MULTIMEDIA
!
class-map type qos match-any CM_QOS_STRICTPRIO_COS5
  match cos 5
```

```
policy-map type qos PM_QOS_MARK_COS_IN
  class CM_QOS_STRICTPRIO_COS5
    set qos-group 5
  class CM_QOS_MULTIMEDIA_COS4
    set qos-group 4
  class CM_QOS_VMOTION_COS2
    set qos-group 3
  class CM_QOS_LOWPRIO_COS1
    set qos-group 2
!
system qos
  service-policy type qos input PM_QOS_MARK_COS_IN
```

QoS-Group # is mapping between Slide 1 and Slide 2

Classification and Marking: Nexus 5600 (2)

Type:
Networ-QoS

```
class-map type network-qos CM_N-QOS_MATCH_QG2_COS1
  match qos-group 2
class-map type network-qos CM_N-QOS_MATCH_QG3_COS2
  match qos-group 3
class-map type network-qos CM_N-QOS_MATCH_QG4_COS4
  match qos-group 4
class-map type network-qos CM_N-QOS_MATCH_QG5_COS5
  match qos-group 5
```

```
policy-map type network-qos PM_N-QOS_SYSTEM
  class type network-qos CM_N-QOS_MATCH_QG2_COS1
    set cos 1
  class type network-qos CM_N-QOS_MATCH_QG3_COS2
    set cos 2
  class type network-qos CM_N-QOS_MATCH_QG4_COS4
    set cos 4
  class type network-qos CM_N-QOS_MATCH_QG5_COS5
    set cos 5
    queue-limit 20480 bytes
!
system qos
  service-policy type network-qos PM_N-QOS_SYSTEM
```

QoS-Group # is mapping between Slide 1 and Slide 2

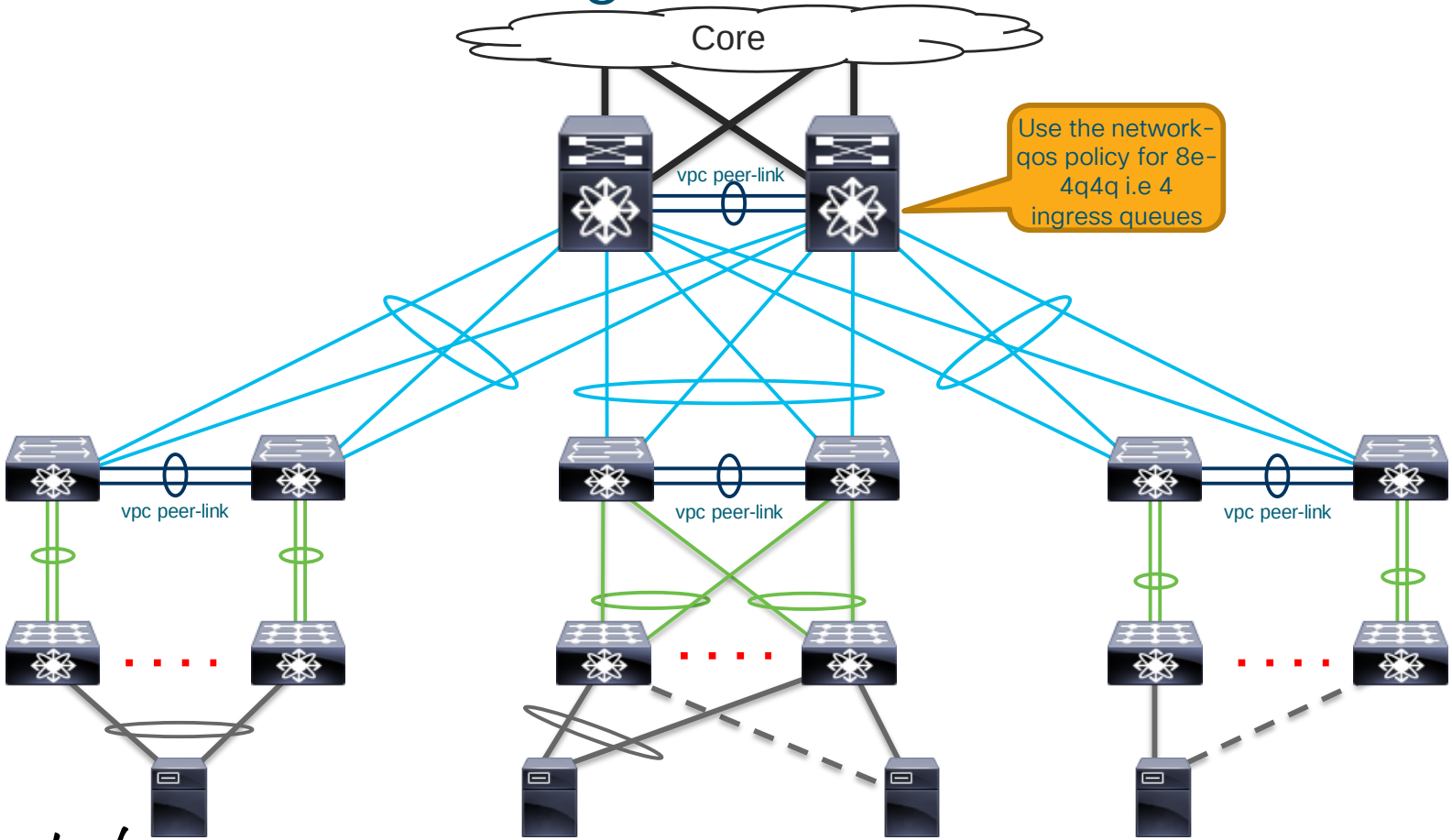
Classification and Marking: Nexus 9000

Type: QoS

```
ip access-list ACL_QOS_LOWPRIO
  10 permit …
ip access-list ACL_QOS_VMOTION
  10 permit …
ip access-list ACL_QOS_MULTIMEDIA
  10 permit …
!
class-map type qos match-any CM_QOS_LOWPRIO_COS1
  match access-group name ACL_QOS_LOWPRIO
!
class-map type qos match-any CM_QOS_VMOTION_COS2
  match access-group name ACL_QOS_VMOTION
!
class-map type qos match-any CM_QOS_MULTIMEDIA_COS4
  match access-group name ACL_QOS_MULTIMEDIA
!
class-map type qos match-any CM_QOS_STRICTPRIO_COS5
  match cos 5
```

```
policy-map type qos PM_QOS_MARK_COS_IN
  class CM_QOS_STRICTPRIO_COS5
    set qos-group 5
    set cos 5
  class CM_QOS_MULTIMEDIA_COS4
    set qos-group 4
    set cos 4
  class CM_QOS_VMOTION_COS2
    set qos-group 3
    set cos 2
  class CM_QOS_LOWPRIO_COS1
    set qos-group 2
    set cos 1
!
system qos
  service-policy type qos input PM_QOS_MARK_COS_IN
```

Network-QoS Configuration on M3/F3-Series

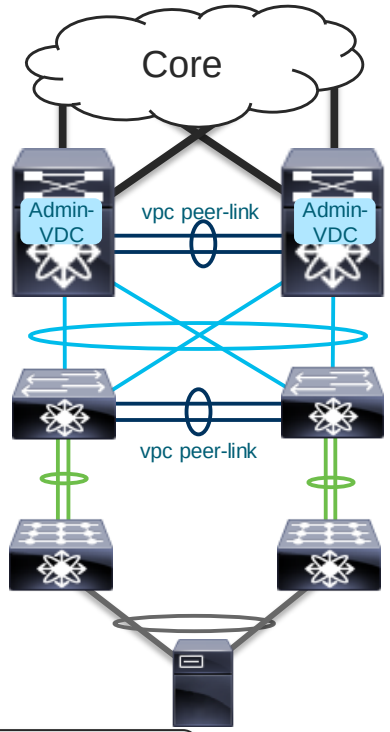


Type:
Network-QoS

Network-QoS Configuration –M3/F3 cards

Example (Admin- / Default-VDC)

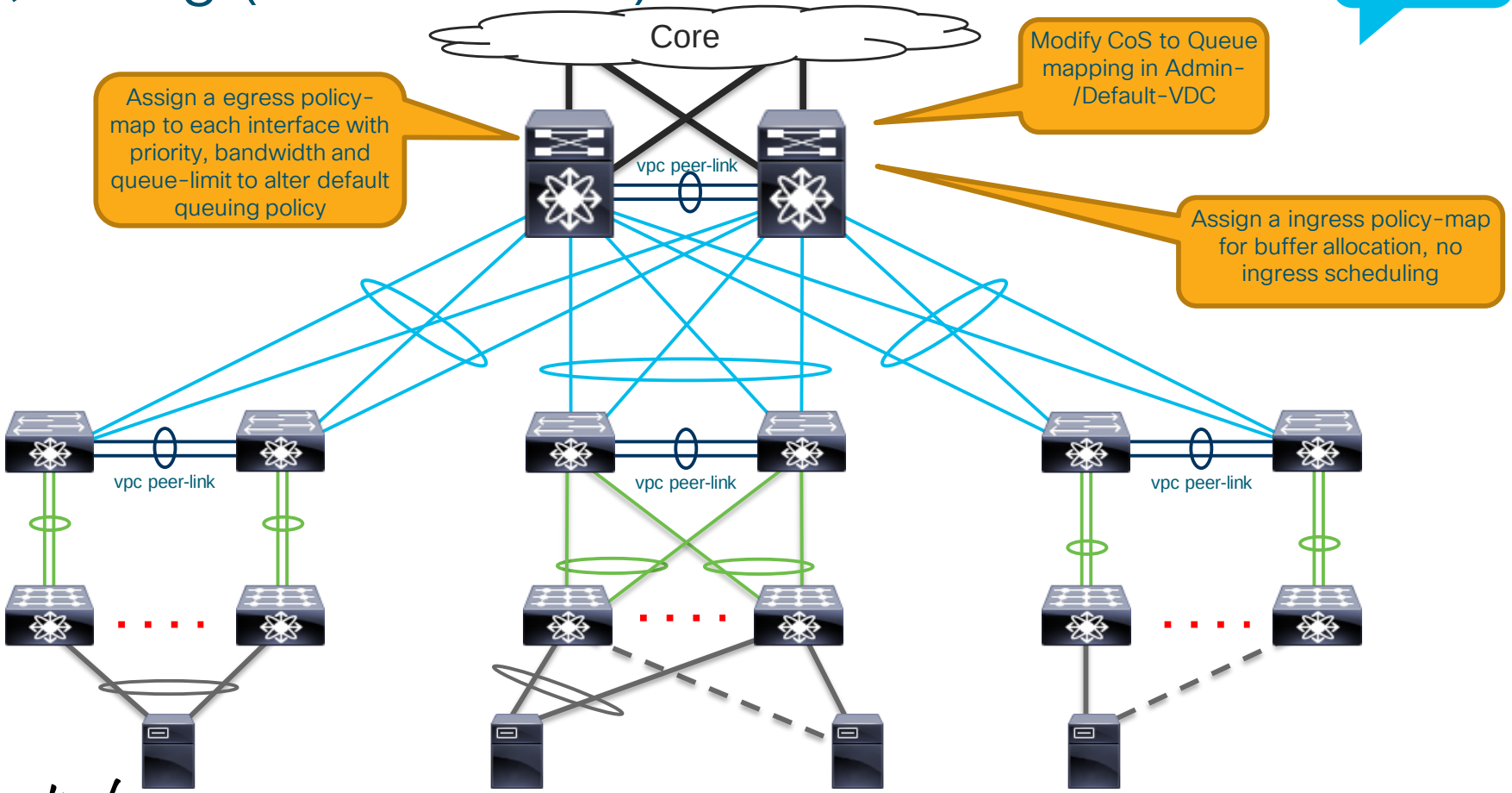
```
system qos
     service-policy type network-qos default-nq-8e-4q4q-policy

policy-map type network-qos default-nq-8e-4q4q-policy template 8e-4q4q
   class type network-qos c-nq-8e-4q4q
     match cos 0-7
     congestion-control tail-drop
     mtu 1500
```



Changes apply to ALL ports of specified type in ALL VDCs
Changes are traffic disruptive for ports of specified type

Queuing (M3/F3 cards)

Type: Queuing



CoS to Queue Mapping – M3/F3 I/O Module

Example

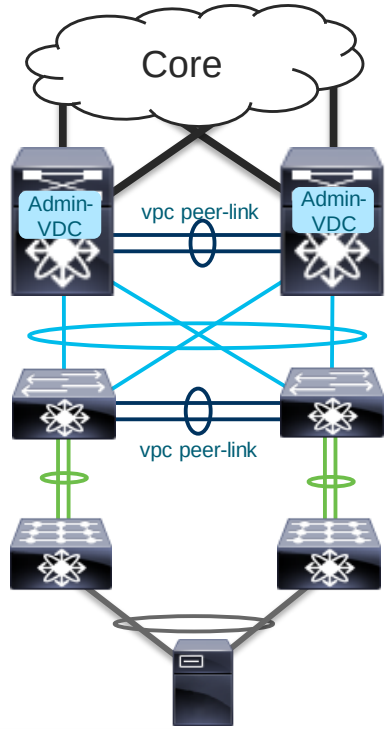
Application	CoS	Queuing (Scheduling)-egress	Queue-Limit (Buffer)-ingress	Queue (Ingress/Egress)	Character
Best Effort	0,1	BW remaining 50%	50%	4q1t-8e-4q4q-in-q-default / 1p3q1t-8e-4q4q-out-q-default	High Volume / Less Important
vMotion / Live Migration	2	BW remaining 20%	10%	4q1t-8e-4q4q-in-q4 / 1p3q1t-8e-4q4q-out-q3	Medium Volume / Important
Multimedia	3, 4	BW remaining 30%	30%	4q1t-8e-4q4q-in-q3 / 1p3q1t-8e-4q4q-out-q2	Medium Volume Very Important
Strict Priority	5	Priority Queue	10%	4q1t-8e-4q4q-in-q1 / 1p3q1t-8e-4q4q-out-pq1	Low Volume / Important / Delay Sensitive
Network Control	6/7				Low Volume / Very important

CoS to Queue Configuration –M3/F3 slides

Example (Admin- / Default-VDC)

Type: Queuing

```
class-map type queuing match-any 4q1t-8e-4q4q-in-q1
  match cos 5-7
class-map type queuing match-any 4q1t-8e-4q4q-in-q-default
  match cos 0-1
class-map type queuing match-any 4q1t-8e-4q4q-in-q3
  match cos 3-4
class-map type queuing match-any 4q1t-8e-4q4q-in-q4
  match cos 2


class-map type queuing match-any 1p3q1t-8e-4q4q-out-pq1
  match cos 5-7
class-map type queuing match-any 1p3q1t-8e-4q4q-out-q2
  match cos 3-4
class-map type queuing match-any 1p3q1t-8e-4q4q-out-q3
  match cos 2
class-map type queuing match-any 1p3q1t-8e-4q4q-out-q-default
  match cos 0-1
```



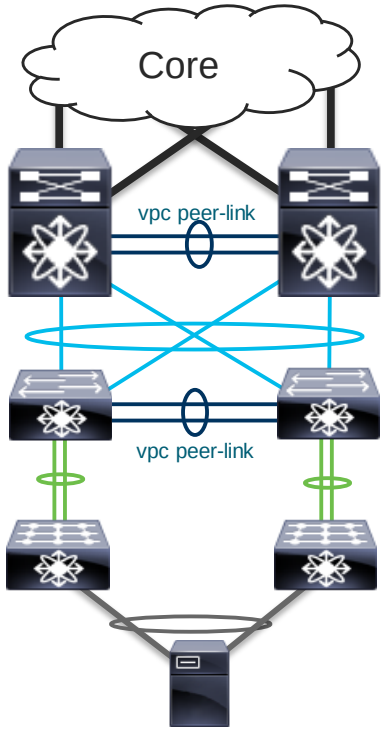
Changes apply to ALL ports of specified type in ALL VDCs
Changes are traffic disruptive for ports of specified type

Ingress Queuing Configuration for M3/F3 cards

Type: Queuing

Example (Payload-VDC)

```
qos copy policy-map type queuing default-8e-4q4q-in-policy prefix
Custom-

 policy-map type queuing Custom-8e-4q4q-in
    class type queuing 4q1t-8e-4q4q-in-q1
      queue-limit percent 10
      bandwidth percent 25
    class type queuing 4q1t-8e-4q4q-in-q-default
      queue-limit percent 50
      bandwidth percent 25
    class type queuing 4q1t-8e-4q4q-in-q3
      queue-limit percent 30
      bandwidth percent 25
    class type queuing 4q1t-8e-4q4q-in-q4
      queue-limit percent 10
      bandwidth percent 25

interface Ethernet1/1
   service-policy type queuing input Custom-8e-4q4q-in
```



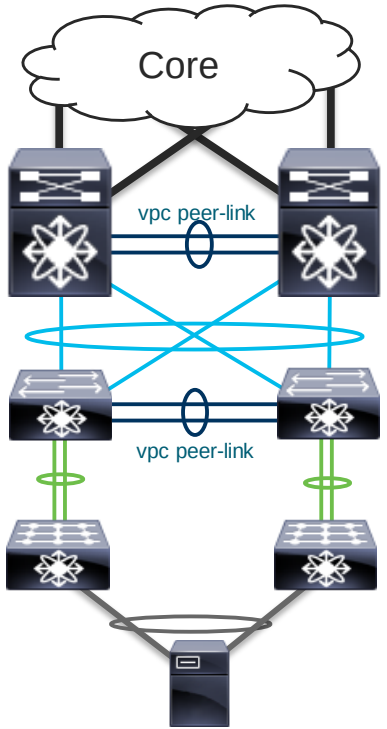
All Policy-Map and Service-Policy are done in relevant Payload-VDC and only affect the interface to which they get applied

Egress Queuing Configuration for M3/F3 cards

Type: Queuing

Example (Payload-VDC)

```
qos copy policy-map type queuing default-8e-4q4q-out-policy prefix
Custom-

policy-map type queuing Custom-8e-4q4q-out
    class type queuing 1p3q1t-8e-4q4q-out-pq1
      priority level 1
    class type queuing 1p3q1t-8e-4q4q-out-q2
      bandwidth remaining percent 30
    class type queuing 1p3q1t-8e-4q4q-out-q3
      bandwidth remaining percent 20
    class type queuing 1p3q1t-8e-4q4q-out-q-default
      bandwidth remaining percent 50
 !

interface Ethernet1/1
   service-policy type queuing output Custom-8e-4q4q-out
```



All Policy-Map and Service-Policy are done in relevant Payload-VDC and only affect the interface to which they get applied

CoS to Queue Mapping – Nexus 9000

Example

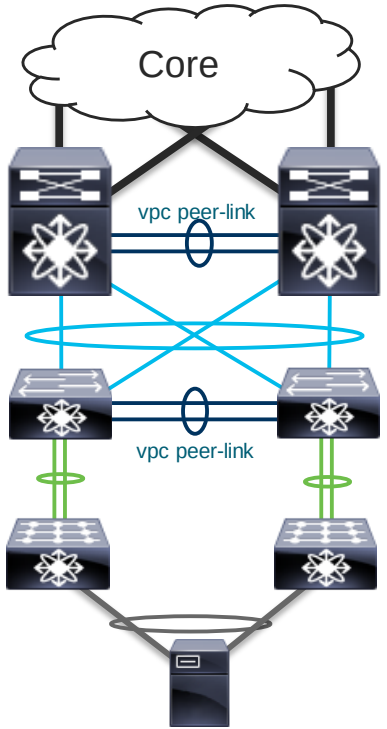
Application	CoS	Queuing (Scheduling)	Queue limit (Alpha)	Queue (6q1t / 1p6q0t)	Character
Best Effort	0,1	BW percent 40%	Default (9)	qos-group 0 (default)	High Volume / Less Important
vMotion / Live Migration	2,3	BW percent 20%	Default (9)	qos-group 3	Medium Volume / Important
Multimedia	4	BW percent 30%	Default (9)	qos-group 4	Medium Volume Very Important
Strict Priority	5	BW percent 10%	Default (9)	qos-group5 / priority	Low Volume / Important / Delay Sensitive
Network Control	6,7				Low Volume / Very important

Type: Queuing

Egress Queuing Configuration: Nexus9000

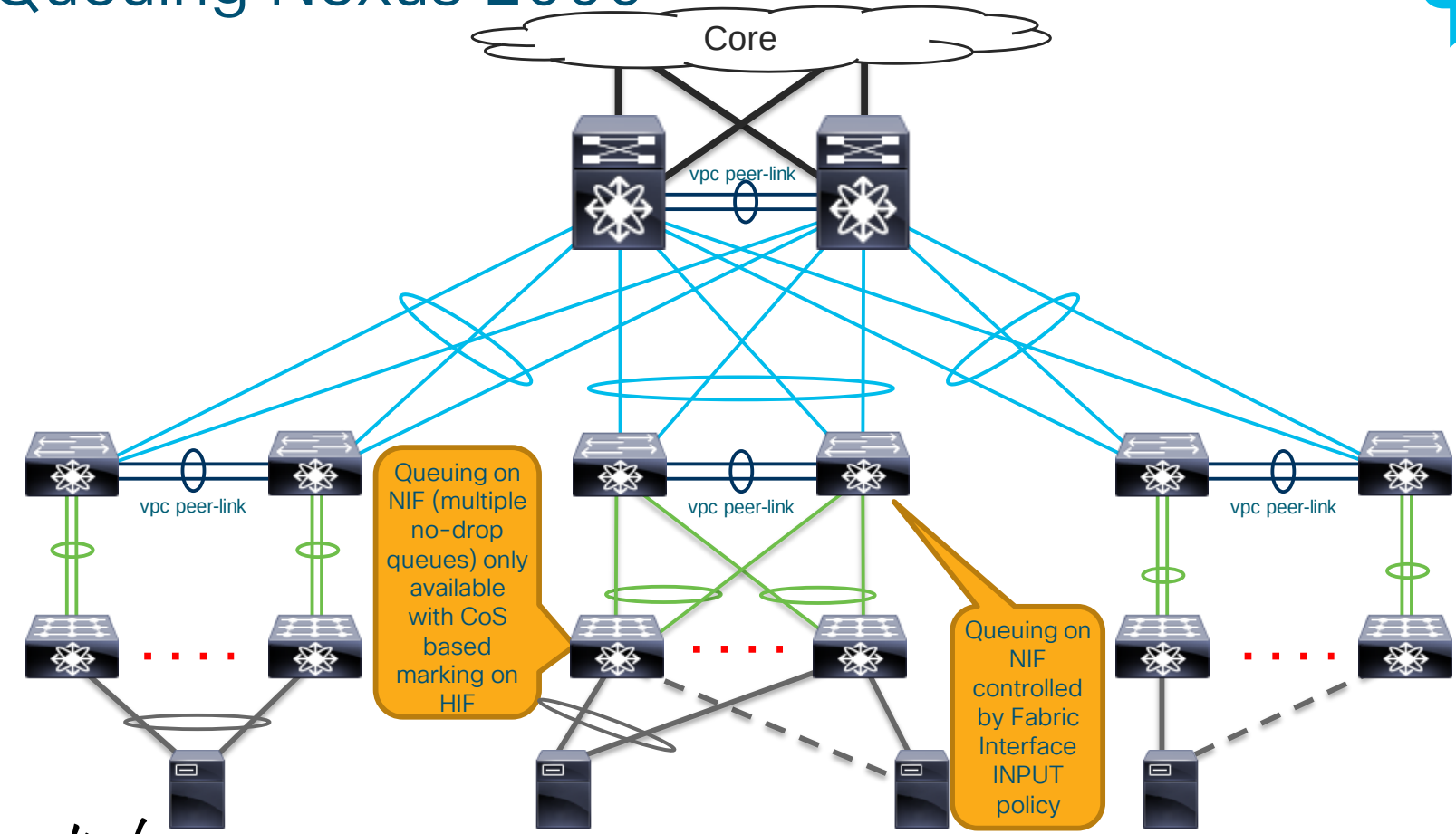
Example

```
class-map type queuing CM_Q_MATCH_QG3_COS2
  match qos-group 3
class-map type queuing CM_Q_MATCH_QG4_COS4
  match qos-group 4
class-map type queuing CM_Q_MATCH_QG5_COS5
  match qos-group 5
!
policy-map type queuing PM_QUEUING_SYSTEM_OUT
class type queuing CM_Q_MATCH_QG3_COS2
    bandwidth percent 20
  class type queuing CM_Q_MATCH_QG4_COS4
    bandwidth percent 30
  class type queuing CM_Q_MATCH_QG5_COS5
    priority
  class type queuing class-default
    bandwidth percent 50
```



Queuing Nexus 2000

Type: Queuing

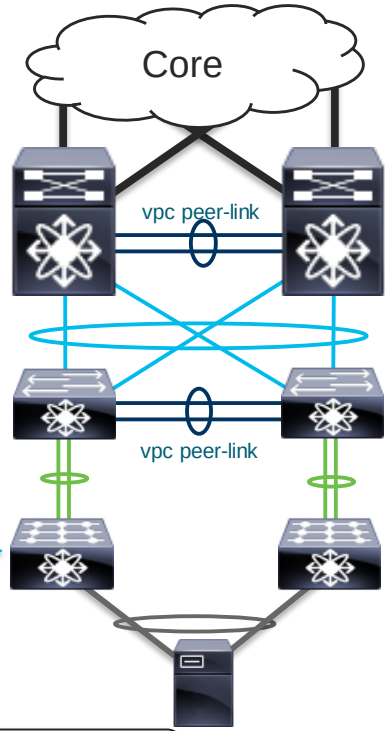


Type: Queuing

Queuing Configuration (Nexus 2000)

Example

```
class-map type queuing CM_Q_MATCH_QG3_COS2
  match qos-group 3
class-map type queuing CM_Q_MATCH_QG4_COS4
  match qos-group 4
class-map type queuing CM_Q_MATCH_QG5_COS5
  match qos-group 5
!
policy-map type queuing PM_QUEUING_SYSTEM_N2K
class type queuing CM_Q_MATCH_QG3_COS2
    bandwidth percent 20
  class type queuing CM_Q_MATCH_QG4_COS4
    bandwidth percent 30
  class type queuing CM_Q_MATCH_QG5_COS5
    priority
class type queuing class-default
    bandwidth percent 40
```



Amount of Queues depend on FEX (Nexus 2000) Model

Agenda

- Introduction
- QoS and Queuing Basics
- QoS Implementation on Nexus
- Nexus 9000 QoS
- Nexus 7000/7700 QoS
- Nexus 5600 QoS
- Nexus 3000 QoS
- Nexus 2000 QoS
- Real World Configuration Examples
- Conclusion

Why QoS in the Data Centre?

Assign Colour to Traffic

Manage Congestion

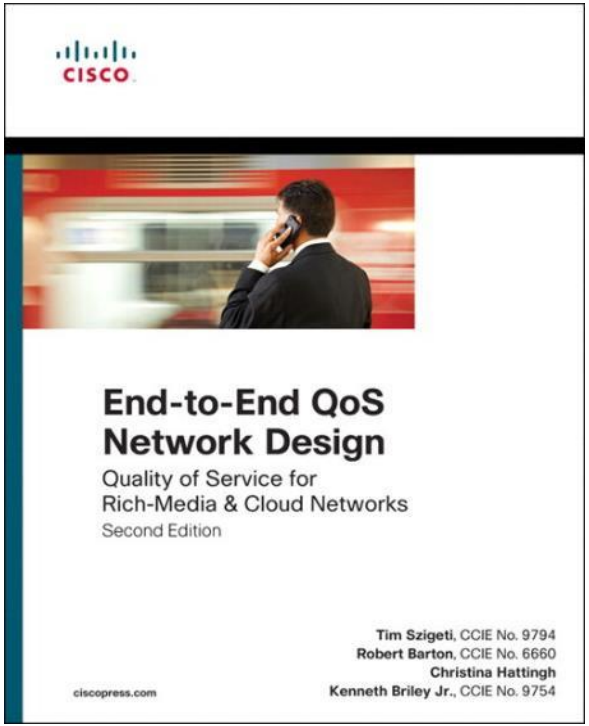
Maximise Throughput

Maximise Throughput and Manage Congestion!

Recommended Reading

- End-to-End QoS Network Design: Quality of Service for Rich-Media and Cloud Networks, 2nd Edition
  - Tim Szigeti
  - Christina Hattingh
  - Robert Barton
  - Kenneth Briley

- ISBN-10: 1-58714-369-0
- ISBN-13: 978-1-58714-369-4



With some help of my friends

I would like to thank all the people, who started the QoS journey and contributed to it:

- Lukas Krattiger, Principal Engineer
- Tim Stevenson, Distinguished Technical Marketing Engineer
- Matthias Wessendorf, Technical Marketing Engineer

Bonus Slides

You make networking **possible**

Cisco live!

FYI

FEX QoS Configuration Examples

You make networking **possible**

Cisco live!

FYI

Fex QoS Policy Configuration Example

```
policy-map type qos fex-qos
  class fex-qos-class-1
    set dscp 10
  class fex-qos-class-2
    set dscp 18
  class fex-qos-class-3
    set dscp 26
!
interface Ethernet101/1/1
  service-policy type qos input fex-qos
```

Marking policy

Policy applied on ingress of FEX HIF

FYI

Nexus 7000 Network-QoS Configuration Example #1

- Applying 8e-4q4q template to enable 4 ingress/egress queues on FEX with COS to queue mapping (also enables 4 ingress queues on F-series modules, if present)

```
system qos
  service-policy type network-qos default-nq-8e-4q4q-policy
```

Default 8e-4q4q template applied to "system qos" target

- FEX output ("show queuing interface"):

```
Queuing:
  queue    qos-group    cos            priority  bandwidth mtu
  -------+------------+--------------+---------+---------+----
  ctrl-hi  n/a          7               PRI         0      2400
  ctrl-lo  n/a          7               PRI         0      2400
  2        0            0 1             WRR        30      1600
  3        1            2               WRR        30      1600
  4        2            5 6             WRR        10      1600
  5        3            3 4             WRR        30      1600
```

8e4q4q configuration (4 data traffic queues)

FYI

Nexus 7000 Network-QoS Configuration Example #2

- Applying custom 8e-4q4q-based template with new MTU

```
policy-map type network-qos custom-nq-8e-4q4q template 8e-4q4q
  class type network-qos c-nq-8e-4q4q
    congestion-control tail-drop
    mtu 9216
system qos
  service-policy type network-qos custom-nq-8e-4q4q
```

Custom network-qos policy with new MTU

Custom template applied to "system qos" target

- FEX output ("show queuing interface") after MTU change:

```
Queuing:
  queue    qos-group    cos            priority  bandwidth mtu
  -------+------------+--------------+---------+---------+----
  ctrl-hi  n/a          7               PRI         0      2400
  ctrl-lo  n/a          7               PRI         0      2400
  2        0            0 1             WRR        30      9280
  3        1            2               WRR        30      9280
  4        2            5 6             WRR        10      9280
  5        3            3 4             WRR        30      9280
```

MTU increased on data traffic queues

FYI

Modifying CoS- or DSCP-to-Queue Mappings

- Changing CoS- or DSCP-to-queue mappings in parent switch F-type ingress queuing class-maps modifies mappings on FEX
- Queuing class-maps modified only in default/admin VDC (apply to entire system)

```
class-map type queuing match-any 4q1t-8e-4q4q-in-q1
  match cos 1-3
  match dscp 8-31
class-map type queuing match-any 4q1t-8e-4q4q-in-q-default
  match cos 0
  match dscp 0-7
class-map type queuing match-any 4q1t-8e-4q4q-in-q3
  match cos 4-5
  match dscp 32-47
class-map type queuing match-any 4q1t-8e-4q4q-in-q4
  match cos 6-7
  match dscp 48-63
```

Non-default F-series ingress queuing class-maps (COS and DSCP match statements modified)

```
Queuing:
 queue    qos-group    cos            priority  bandwidth mtu
 --------+------------+--------------+---------+---------+----
 ctrl-hi  n/a          7              PRI         0       2400
 ctrl-lo  n/a          7              PRI         0       2400
 2        0            0              WRR        30       1600
 3        1            6              WRR        30       1600
 4        2            1 2 3          WRR        10       1600
 5        3            4 5            WRR        30       1600
<…>

 queue          DSCPs
 -----          -----
 02             0-7,
 04             8-31,
 03             48-63,
 05             32-47,
```

FEX queue mappings reflect changes

FYI

Enabling FEX Queue Limits

- Example #1 – N2K-C2248TP-1GE

```
fex 101
  hardware N2248T queue-limit 50000
```

- Example #2 – N2K-C2232TM-E-10GE

```
fex 102
   hardware N2232TM-E queue-limit 50000
```

- FEX output ("show queuing interface") **before**:

```
Queue limit: Disabled
```

- FEX output ("show queuing interface") **after** (configured queue-limit rounded to nearest hardware supported value):

```
Queue limit: 51200 bytes
```

FYI

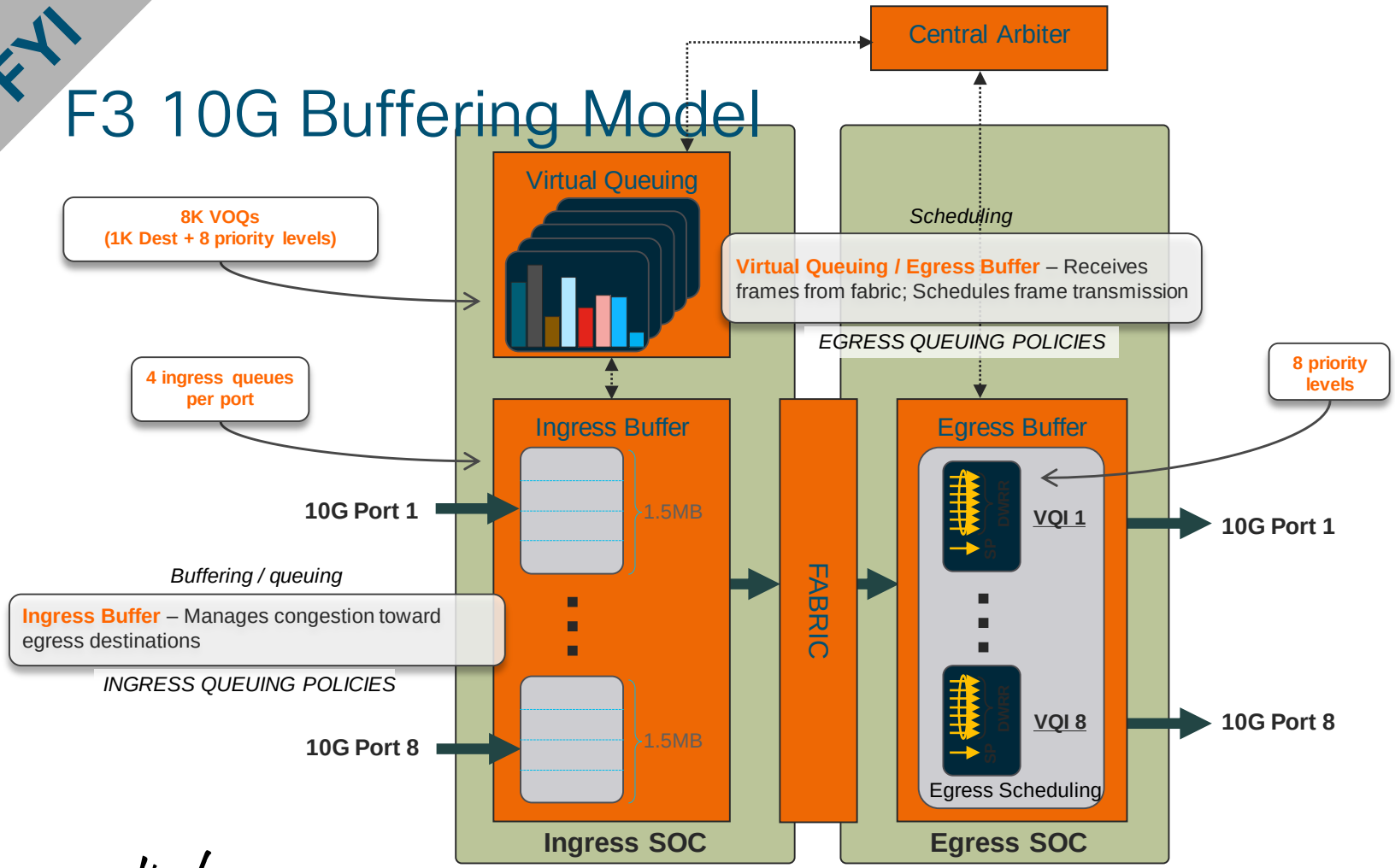
F3 Queuing Configuration Examples

You make networking **possible**

Cisco live!

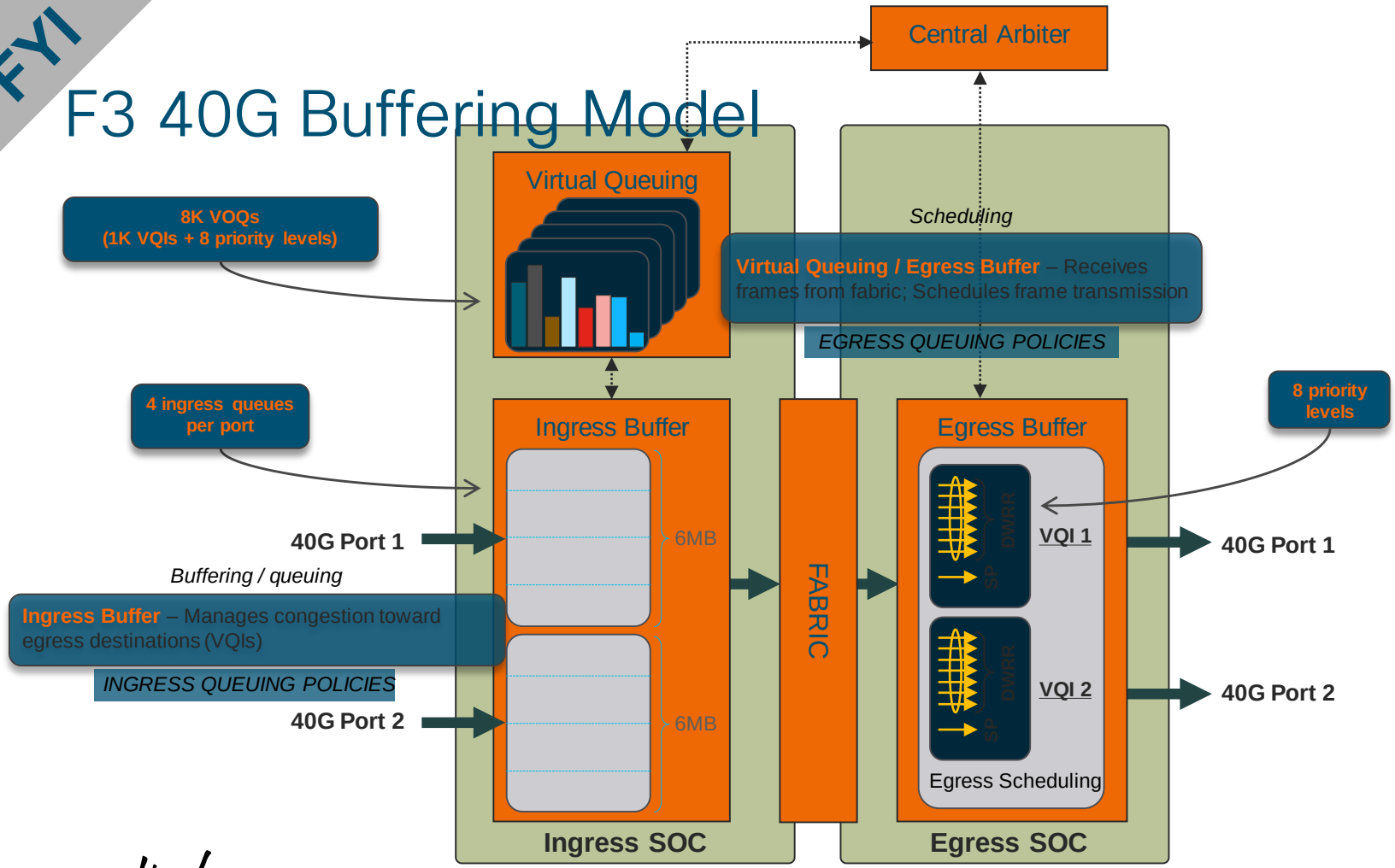
FYI

F3 10G Buffering Model



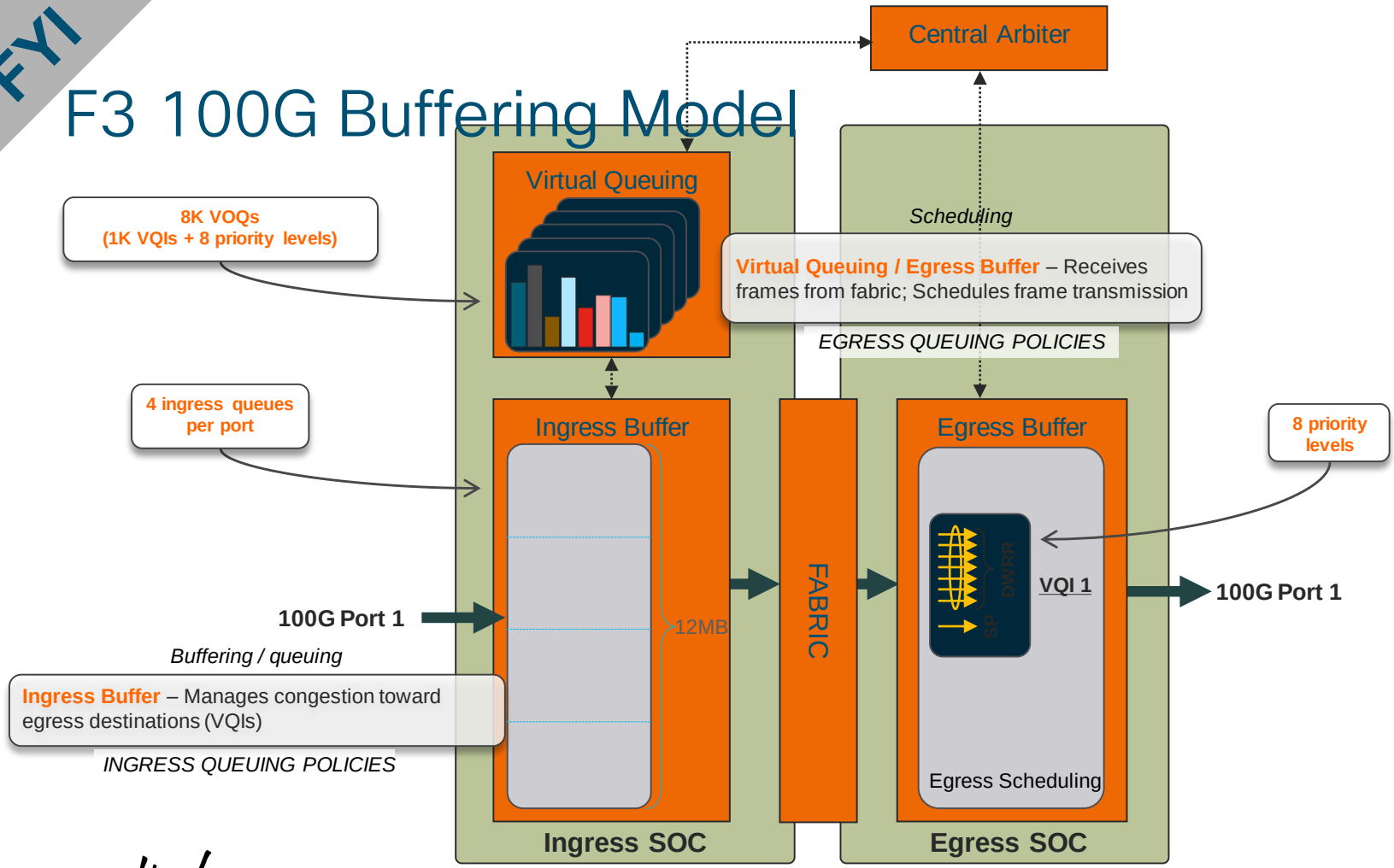
FYI

F3 40G Buffering Model



FYI

F3 100G Buffering Model



FYI

Network-QoS and Default Queuing (global)

- default-nq-8e-policy is default `network-qos` policy and attached to `system qos` in Admin-/Default-VDC
- The system queuing policy applied by default can be overridden on a per port basis.

2 ingress queues with buffer ratio 1:9 and DWRR wieights 1:1

4 egress queues with one priority queue and DWRR wieights 1:1:1

```
N7k# show policy-map type queuing | beg default 4q-8e

policy-map type queuing default-4q-8e-in-policy
    class type queuing 2q4t-8e-in-q1
      queue-limit percent 10
      bandwidth percent 50
    class type queuing 2q4t-8e-in-q-default
      queue-limit percent 90
      bandwidth percent 50

policy-map type queuing default-4q-8e-out-policy
    class type queuing 1p3q1t-8e-out-pq1
      priority level 1
    class type queuing 1p3q1t-8e-out-q2
      bandwidth remaining percent 33
    class type queuing 1p3q1t-8e-out-q3
      bandwidth remaining percent 33
    class type queuing 1p3q1t-8e-out-q-default
      bandwidth remaining percent 33
```

Note: show policy-map system does display similar output

FYI

Modifying Queuing and Scheduling Behaviour on F3 Modules

I want to...	Steps to follow
...remap COS/DSCP values from one queue to another queue without activating additional queues	1. Modify the type queuing class-map(s) for the desired queue(s)
...change queuing behaviour without changing COS-or DSCP-to-queue mapping	1. Define new type queuing policy-map (you cannot modify the default policies) 2. Modify class-map parameters 3. Apply new policy-map to interfaces
...activate additional queues and remap COS/DSCP values	1. Define new type queuing policy-map 2. Modify COS-to-queue mapping for target port type 3. Apply new policy-map to interfaces
...shape the SP queue	1. (Optional) Clone the default egress queuing policy 2. Shape the SP queue in the new (cloned) policy 3. Apply the new queuing policy to the target interfaces

FYI

Modifying Queuing Behaviour

Remap Some COS/DSCP Values from One Queue to Another Queue without Activating Additional Queues

Modify “type queuing” class-map(s) for desired queue(s)

Remap COS- or DSCP-to-queue mapping for given queue(s)

Important: changing COS- or DSCP-to-queue mapping takes effect immediately and is disruptive to all ports

Modifying Queuing Behaviour

Remap Some COS/DSCP Values from One Queue to Another Queue without Activating Additional Queues

FYI

Example: remap COS 4 and DSCP 32-39 to ingress queue “q1”:

```
n77# show class-map type queuing 8q2t-in-q1
  Type queuing class-maps
  ========================
    class-map type queuing match-any 8q2t-in-q1
      Description: Classifier for ingress queue 1 of type 8q2t
      match cos 5-7
      match dscp 40-63
n77# configure terminal
Enter configuration commands, one per line.  End with CNTL/Z.
n77(config)# ! Modify ingress queue q1
n77(config)# class-map type queuing match-any 8q2t-in-q1
n77(config-cmap-que)# ! Change COS- and DSCP-to-queue mapping for this queue
n77(config-cmap-que)# match cos 4
n77(config-cmap-que)# match dscp 32-39
n77(config-cmap-que)# show class-map type queuing 8q2t-in-q1
  Type queuing class-maps
  ========================
    class-map type queuing match-any 8q2t-in-q1
      Description: Classifier for ingress queue 1 of type 8q2t
      match cos 4-7
      match dscp 32-63
n77(config-cmap-que)#
```

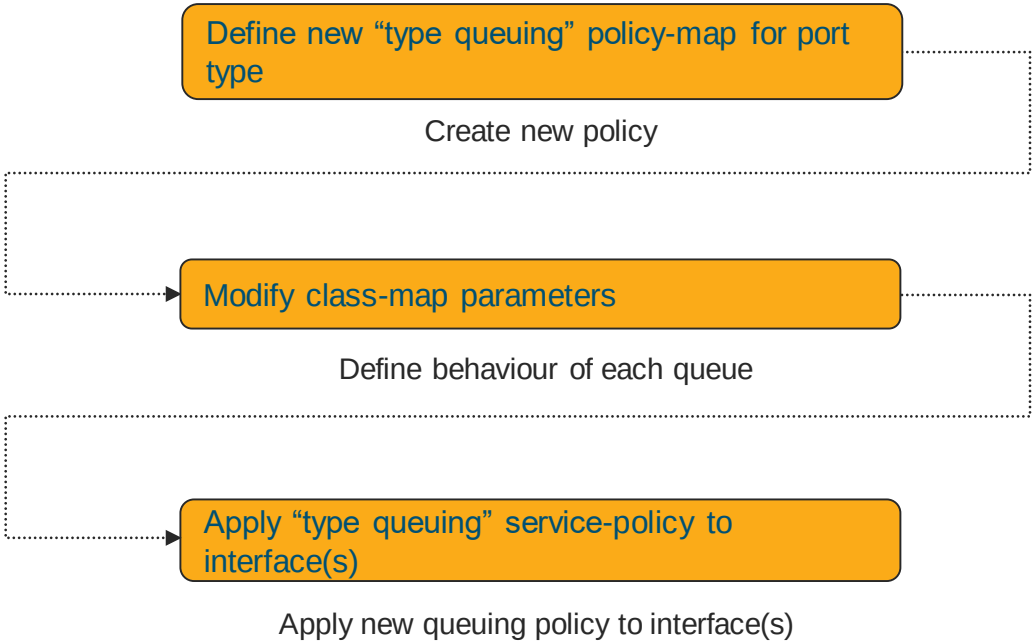
Show current mapping

Configure new mapping

Show new mapping

Modifying Queuing Behaviour

Changing Default Queuing Behaviour without Changing COS- or DSCP-to-Queue



Important: applying new queuing policy takes effect immediately and is disruptive to any ports to which the policy is applied

Modifying Queuing Behaviour

Changing Default Queuing Behaviour without Changing COS- or DSCP-to-Queue

FYI

Example: Resize ingress queues without modifying COS- or DSCP-to-queue mapping

```
n77# configure terminal
Enter configuration commands, one per line.  End with CNTL/Z.
n77(config)# ! Define new "type queuing" policy
n77(config)# policy-map type queuing new-f3-ingress
n77(config-pmap-que)# ! Define behavior for F3 ingress q-default
n77(config-pmap-que)# class type queuing 8e-4q8q-in-q-default
n77(config-pmap-c-que)# ! Resize this queue
n77(config-pmap-c-que)# queue-limit percent 74
n77(config-pmap-c-que)# ! Define behavior for F3 ingress queue 1
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q1
n77(config-pmap-c-que)# ! Resize this queue
n77(config-pmap-c-que)# queue-limit percent 24
n77(config-pmap-c-que)# ! Policy must include all queues (even inactive)
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q3
n77(config-pmap-c-que)# ! Must give at least 1% to inactive queues
n77(config-pmap-c-que)# queue-limit percent 1
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q4
n77(config-pmap-c-que)# queue-limit percent 1
n77(config-pmap-c-que)# interface e 2/1-48
n77(config-if-range)# ! Apply the new policy to F3 interfaces
n77(config-if-range)# service-policy type queuing input new-f3-ingress
n77(config-if-range)#
```

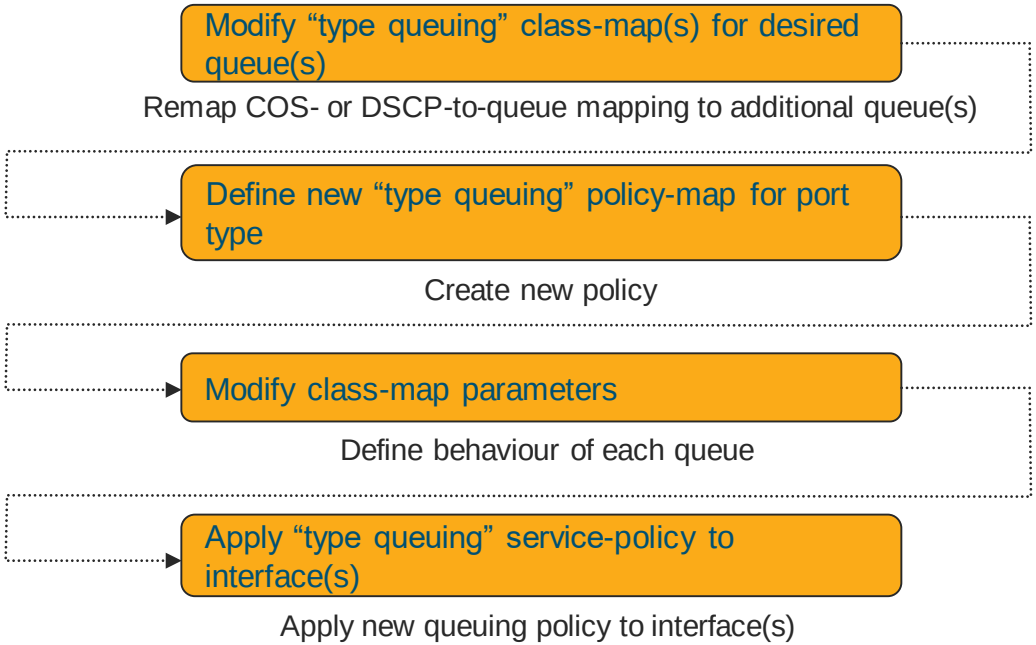
Create new queuing policy

Modify class-map parameters (resize queues)

Apply policy to interface(s)

FYI

Modifying Queuing Behaviour
Activate Additional Queues and Remap COS/DSCP Values



Important: changing COS/DSCP-to-queue mapping and takes effect immediately and is disruptive to all ports; applying new queuing policy takes effect immediately and is disruptive to any ports to which the policy is applied

Modifying Queuing Behaviour

Activate Additional Queues and Remap COS/DSCP Values

Example: Enable one additional ingress queue and map COS/DSCP values to all active queues

```
n77# configure terminal
Enter configuration commands, one per line.  End with CNTL/Z.
n77(config)# ! Modify ingress queue q3
n77(config)# class-map type queuing match-any 8q2t-in-q3
n77(config-cmap-que)# ! Map COS and DSCP values to this queue
n77(config-cmap-que)# match cos 3-4
n77(config-cmap-que)# match dscp 24-39
n77(config-cmap-que)# ! Define new "type queuing" policy
n77(config-cmap-que)# policy-map type queuing new-f3-ingress
n77(config-pmap-que)# ! Define behavior for F3 ingress q-default
n77(config-pmap-que)# class type queuing 8e-4q8q-in-q-default
n77(config-pmap-c-que)# ! Resize this queue
n77(config-pmap-c-que)# queue-limit percent 60
n77(config-pmap-c-que)# ! Define behavior for F3 ingress queue 1
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q1
n77(config-pmap-c-que)# ! Resize this queue
n77(config-pmap-c-que)# queue-limit percent 25
n77(config-pmap-c-que)# ! Define behavior for F3 ingress queue 3
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q3
n77(config-pmap-c-que)# ! Resize this queue
n77(config-pmap-c-que)# queue-limit percent 14
n77(config-pmap-c-que)# ! Policy must include all queues (even inactive)
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q4
n77(config-pmap-c-que)# queue-limit percent 1
n77(config-pmap-c-que)# int e 2/1-48
n77(config-if-range)# ! Apply the new policy to F3 interfaces
n77(config-if-range)# service-policy type queuing input new-f3-ingress
```

Map COS/DSCP values to inactive queue

Create new queuing policy

Modify class-map parameters (resize queues)

Apply policy to interface(s)

FYI

FYI

Important!

- If you change the COS- or DSCP-to-queue mapping for a port type, make sure **all** ports of that type in **all** VDCs have a queuing policy applied that defines behaviour for **all** queues with COS/DSCP values mapped
- For example, if you do THIS…

```
n77(config)# class-map type queuing match-any 8e-4q8q-in-q-default
n77(config-cmap-que)# match cos 0-1
n77(config-cmap-que)# match dscp 0-15
n77(config-cmap-que)# class-map type queuing match-any 8e-4q8q-in-q4
n77(config-cmap-que)# match cos 2-4
n77(config-cmap-que)# match dscp 16-39
n77(config-cmap-que)# class-map type queuing match-any 8e-4q8q-in-q3
n77(config-cmap-que)# match cos 6-7
n77(config-cmap-que)# match dscp 48-63
n77(config-cmap-que)# class-map type queuing match-any 8e-4q8q-in-q1
n77(config-cmap-que)# match cos 5
n77(config-cmap-que)# match dscp 40-47
```

Changes the default COS/ DSCP-to-queue mapping

FYI

Important!

- ... then make sure you do THIS...

```
n77(config)# policy-map type queuing new-f3-ingress
n77(config-pmap-que)# class type queuing 8e-4q8q-in-q-default
n77(config-pmap-c-que)# queue-limit percent 50
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q1
n77(config-pmap-c-que)# queue-limit percent 20
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q3
n77(config-pmap-c-que)# queue-limit percent 20
n77(config-pmap-c-que)# class type queuing 8e-4q8q-in-q4
n77(config-pmap-c-que)# queue-limit percent 10
n77(config-pmap-c-que)#
```

Defines a new queuing policy that defines behaviour of all queues that COS/DSCP values have been mapped to

FYI

Important!

- … and then do THIS:

Maps the new policy to **ALL** interfaces in the system (do this on **ALL** ports in EVERY VDC!!)

```
n77(config)# int e 2/1-48
n77(config-if-range)# service-policy type queuing input new-f3-ingress
n77(config-if-range)#
```

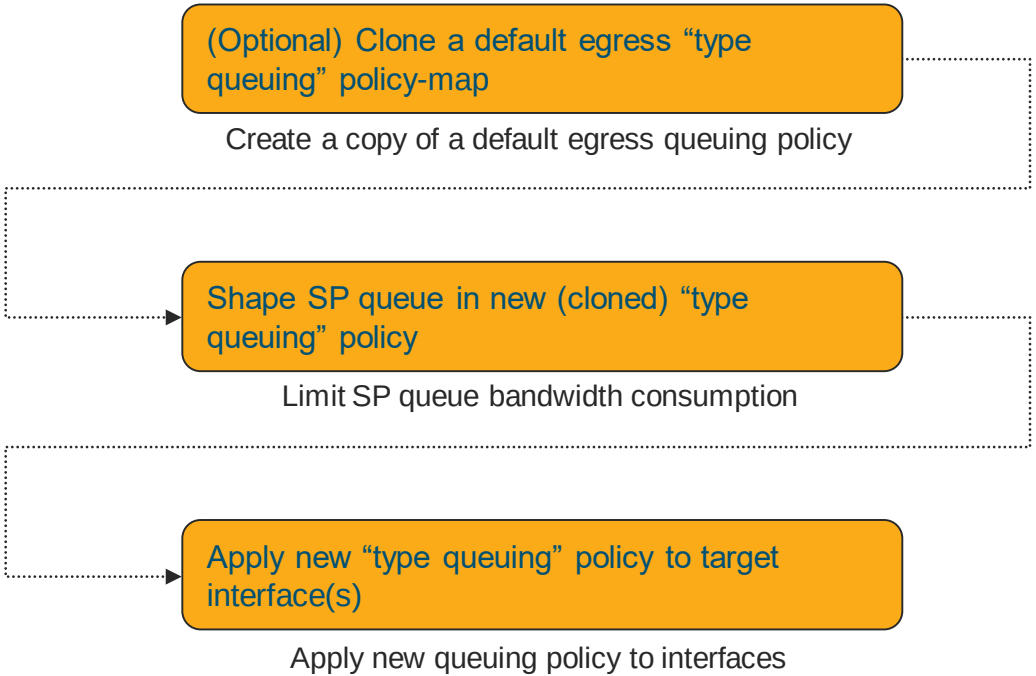
- If you DON'T, traffic arriving on ports with default policy (i.e., without all queues activated that have COS/DSCP values mapped) will suffer – packet drops, poor performance, etc.

- Of course, you can have *different* non-default policies on different sets of interfaces, but all interfaces in the system must use some policy that defines all activated queues!

FYI

Modifying Queuing Behaviour

Shape the SP Queue



Important: applying new queuing policy takes effect immediately and is disruptive to any ports to which the policy is applied

Modifying Queuing Behaviour

Shape the SP Queue

Example: Shape the SP queue to 2Gbps on a 10G interface, using a queuing policy cloned from the default “8e4q4q” egress queuing policy

```
n77# ! Clone the 8E egress queuing policy
n77# qos copy policy-map type queuing default-8e-4q8q-out-policy prefix new-
n77# conf
Enter configuration commands, one per line.  End with CNTL/Z.
n77(config)# ! Modify new queuing policy
n77(config)# policy-map type queuing new-8e-4q8q-out
n77(config-pmap-que)# ! Modify egress queue q1
n77(config-pmap-que)# class type queuing 8e-4q8q-out-q1
n77(config-pmap-c-que)# ! Make this queue strict priority
n77(config-pmap-c-que)# priority level 1
n77(config-pmap-c-que)# ! Shape the queue to 20% (2G on 10G port)
n77(config-pmap-c-que)# shape percent 20
n77(config-pmap-c-que)# int e 2/1-48
n77(config-if-range)# ! Apply the new policy to target interfaces
n77(config-if-range)# service-policy type queuing output new-8e-4q8q-out
n77(config-if-range)#
n77#
```

Clone the default egress queuing policy

Modify the cloned policy

Make q1 Strict Priority and shape to 20% (2G)

Apply new policy to target interfaces

FYI

Changing The Default Trust (M-Series I/O Module)

- You can make an interface untrusted (CoS and DSCP)
  - CoS for bridged traffic
  - DSCP for routed traffic
- You need two Policies
  - A "type queuing" policy to set the CoS to 0
  - A "type qos" policy to set the DSCP to 0

```
policy-map type queuing Reset-CoS
  class type queuing 8q2t-in-q-default
    set cos 0
    bandwidth percent 100
    queue-limit percent 100
!
policy-map type qos Reset-DSCP
  class class-default
    set dscp 0
!
! Tie to an interface:
interface Ethernet1/1
  service-policy type queuing input Reset-CoS
  service-policy type qos input Reset-DSCP
```

- Set DSCP will set the CoS value for Bridged traffic as well.

FYI

Changing The Default Trust (F-Series I/O Module)

```
qos copy policy-map type queuing default-4q-8e-in-policy prefix UNTRUSTED-
!
policy-map type queuing untrusted-4q-8e-in
  class type queuing 2q4t-8e-in-q1
    queue-limit percent 1
  class type queuing 2q4t-8e-in-q-default
   queue-limit percent 99
   set cos 0
!
policy-map type qos UNTRUSTED
  class class-default
    set dscp 0
!
! Tie to an interface:
interface Ethernet1/1
  service-policy type queuing input untrusted-4q-8e-in
  service-policy type qos input UNTRUSTED
```

Complete your online session evaluation

- 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 water bottle.
- All surveys can be taken in the Cisco Live Mobile App or by logging in to the Session Catalog on ciscolive.cisco.com/us.

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

Continue your education

Demos in the Cisco campus

Walk-in labs

Meet the engineer 1:1 meetings

Related sessions

Thank you

You make **possible**