The bridge to possible

End-to-End QoS Implementation and Operation with Nexus

Nemanja Kamenica,
Technical Marketing Engineer

BRKDCN-3953

cisco Live!

#CiscoLive

Cisco Webex App

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-3953

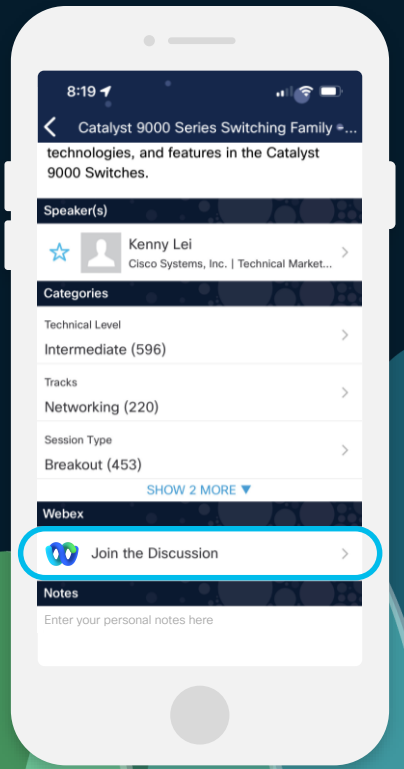
Questions?

Use Cisco Webex App 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 the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 7, 2024.



Session Objectives

- Provide a refresh of QoS Basics
- Understand QOS implementation on Nexus Operating System
- Provide a detailed understanding of QoS on Nexus Nexus 9000 Cloud Scale and Nexus 9800 platforms
- Learn how to configure QOS on Nexus 9000 devices through real-world configuration examples

Session Non-Objectives

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

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale QoS
- Nexus 9800 QOS
- Real World Configuration Examples
- Conclusion

Congestion Happens Everyday!

Why QoS in the Data Centre?

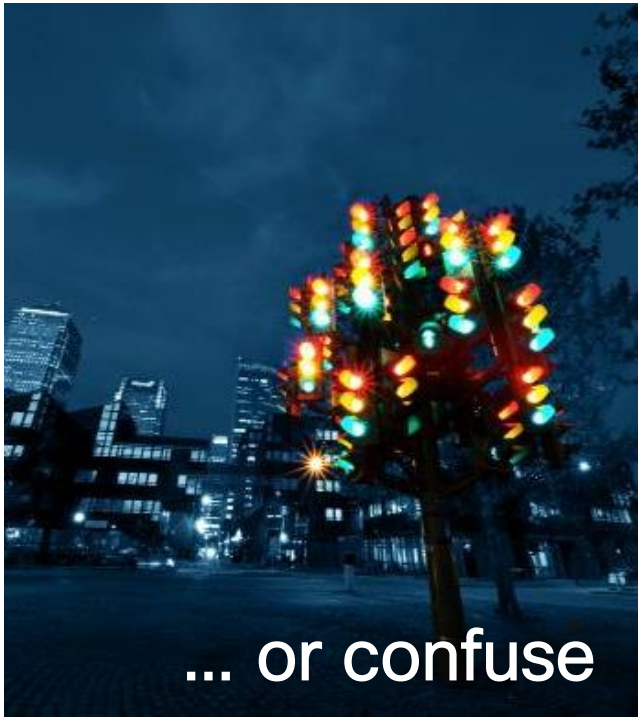
Assign
Color to Traffic

Manage
Congestion

Maximize
Throughput

Maximize Throughput and Manage Congestion!

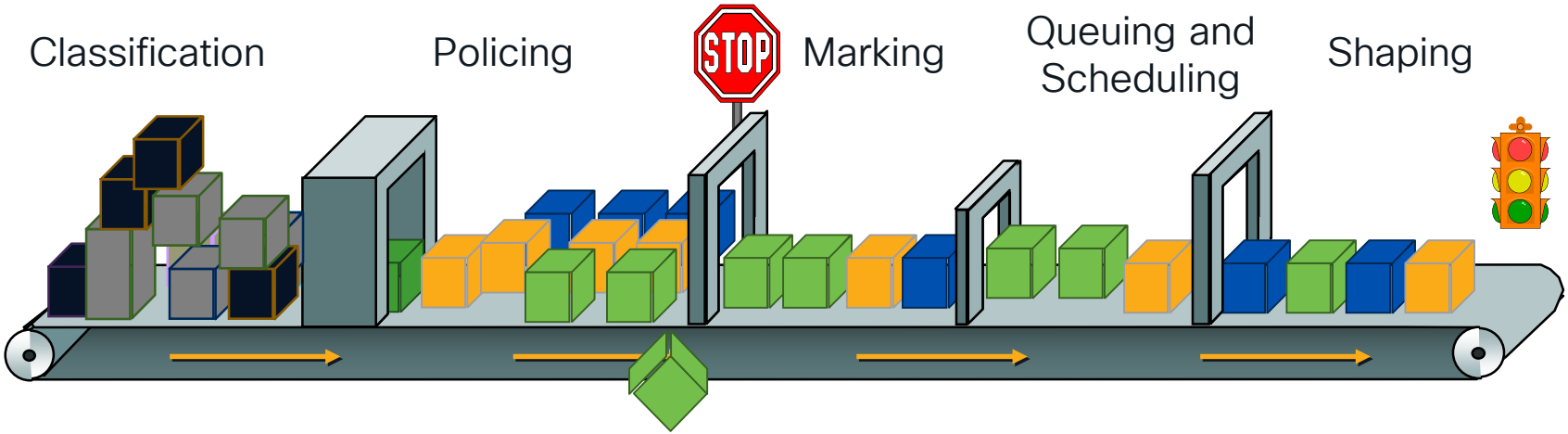
Can Traffic Control help ...



Agenda

- **Introduction**
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale QoS
- Nexus 9800 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 Behavior and Business Policies

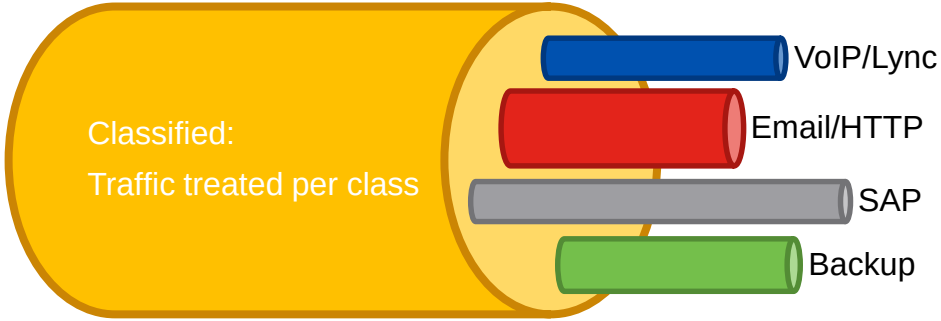
Prioritize, Protect and Isolate Traffic Based on Markings

Control Bursts and Conform Traffic

Classification and Marking

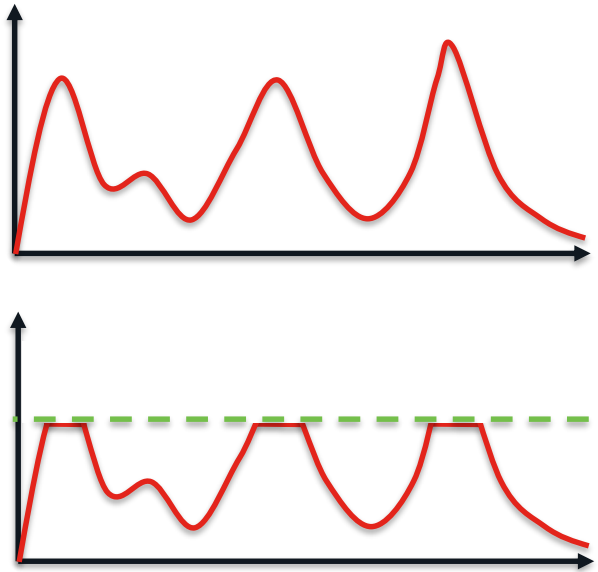
Two sides of a coin

- Classification – Identify and separate traffic in classes
- Identify traffic
  - ACLs
  - CoS
  - DSCP
  - IP PREC
- Marking – Mark traffic with QoS priority value
- Marking Traffic
  - With new priority value (i.e. CoS or DSCP)
  - Changing Like to Like (i.e. CoS to CoS)
  - Like to Unlike (i.e. DSCP to CoS)



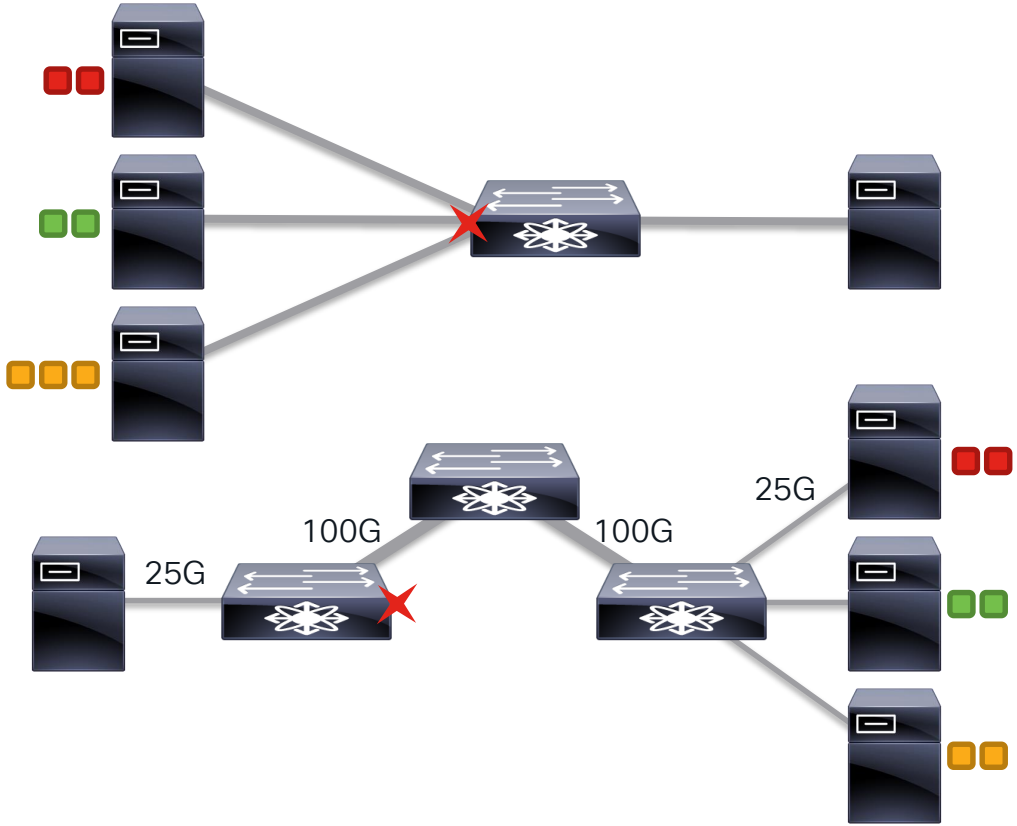
Policing – Limit Misbehaving Traffic

- Policing – Protecting other classes by dropping traffic in misbehaving class
- 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)



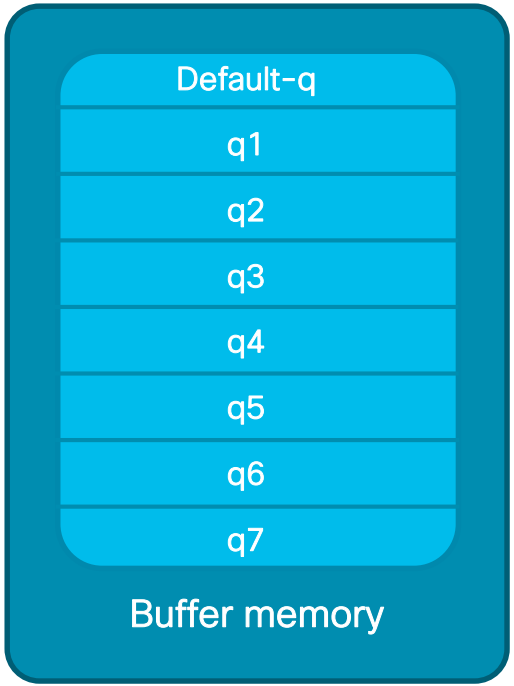
Buffering – Why do we need it?

- Buffering – Storing data packets in memory
- Many to One Conversations
  - Client to Server
  - Server to Storage
  - Aggregation Points
- Speed Mismatch
  - Client to WAN to Server



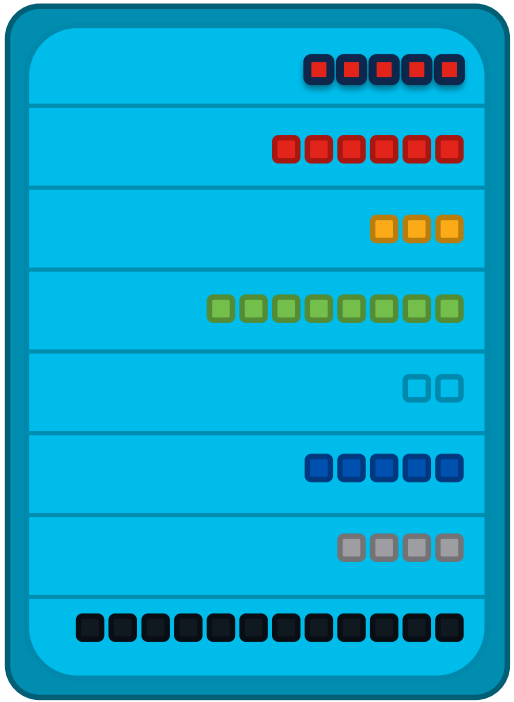
Queueing

- Traffic in buffer is divided logically in the queues
- Queueing provide dedicated buffer for packets of different priority
- Traffic separation allows multiple traffic classes to be mapped to same or different queue
- Traffic in a queue can be treated differently from other queues



Scheduling

- Scheduling – defines order of transmission of traffic out the queues
- Different types of queue are server differently
  - Strict Priority Queue – always serviced first
  - Normal Queues – served only after priority queue is empty
- Normal queues can have different algorithms

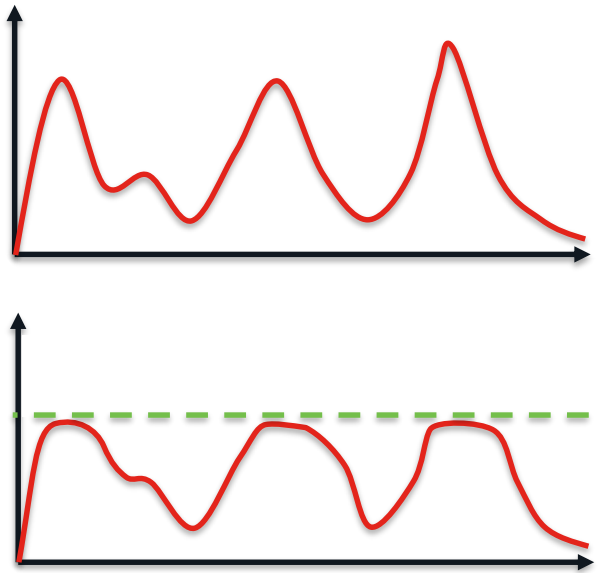


Common Scheduling Algorithms

- 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

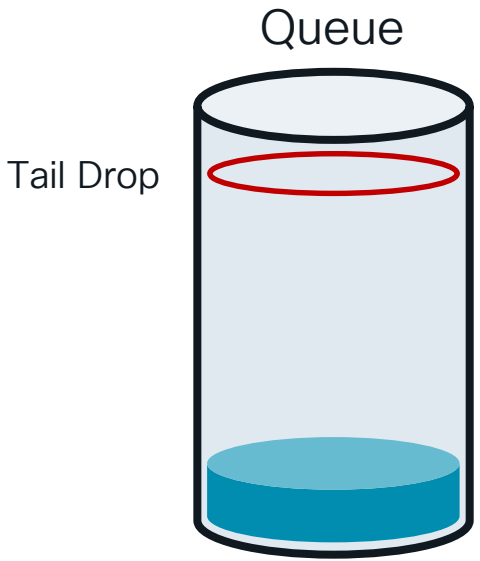
Shaping

- Shaping – Smooth out traffic peaks, microburst, with preserving all traffic
- Usually in egress direction to limit traffic toward ISP



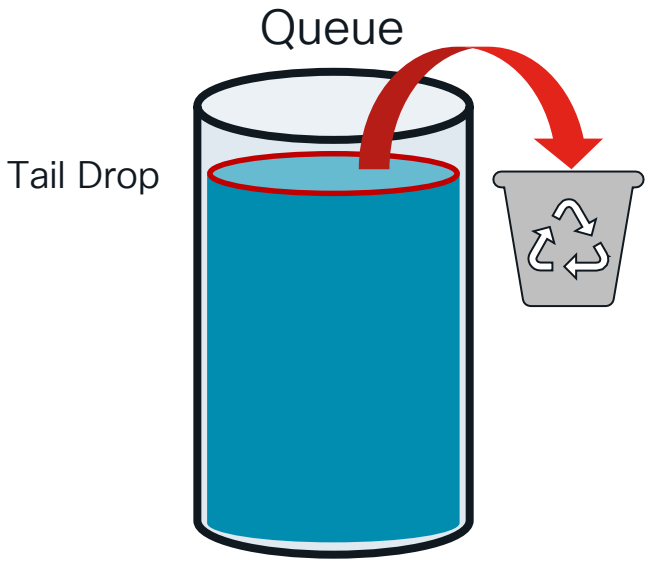
Congestion Avoidance Tools

- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue



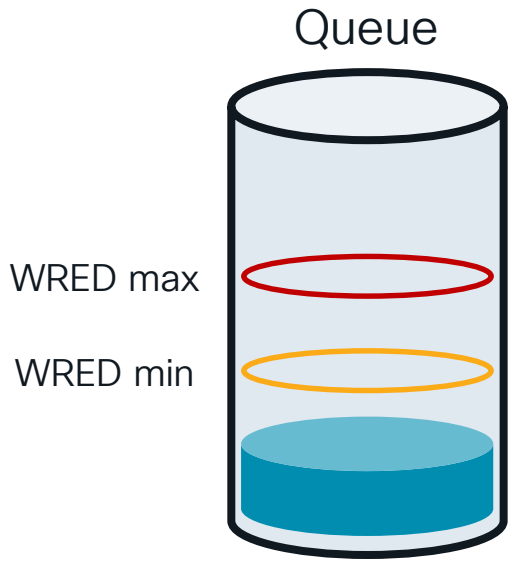
Congestion Avoidance Tools

- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue



Congestion Avoidance Tools

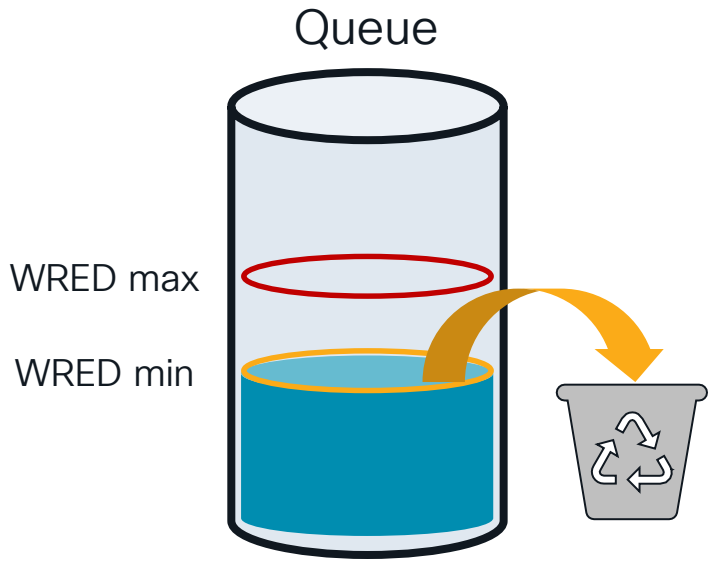
- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue
- Weighted Random Early Drop (WRED)
  - One or more thresholds per queue
  - Threshold associated with priority
  - Buffer usage below threshold no affect



Congestion Avoidance Tools

- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue

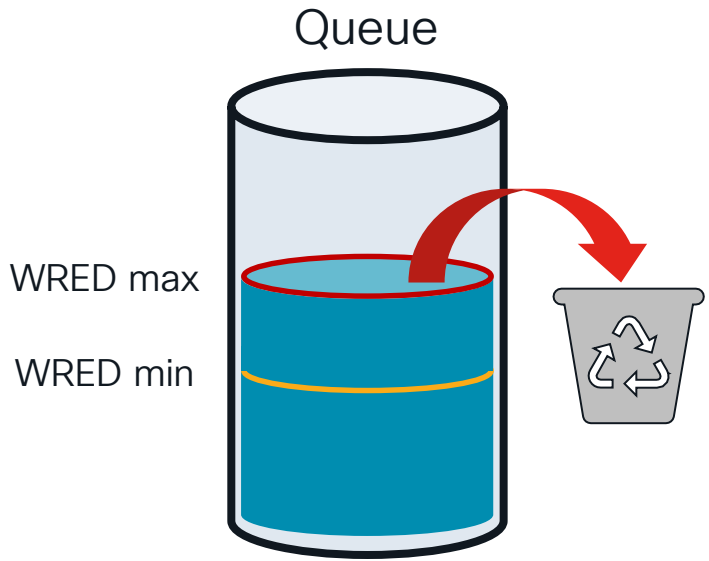
- Weighted Random Early Drop (WRED)
  - One or more thresholds per queue
  - Threshold associated with priority
  - Buffer usage below threshold no affect
  - Buffer usage over min threshold = random drops



Congestion Avoidance Tools

- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue

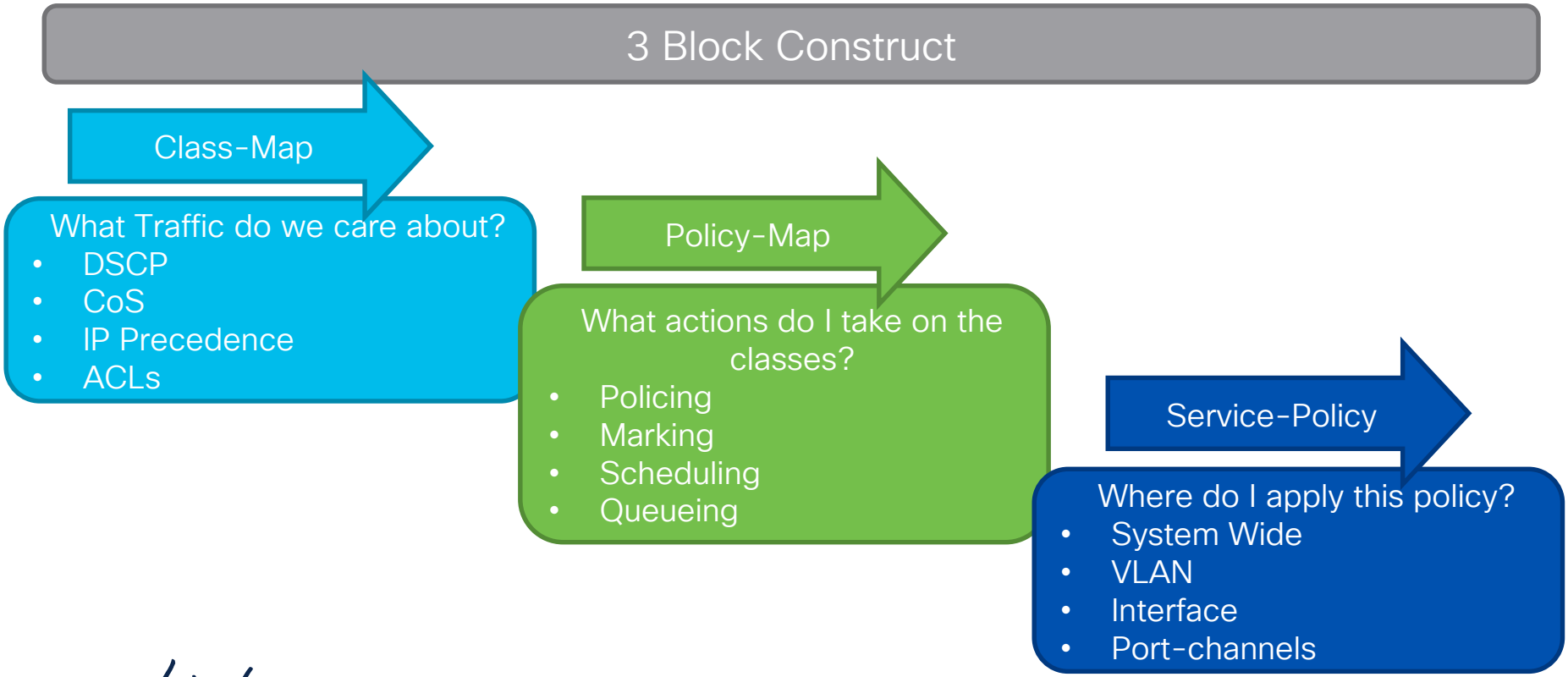
- Weighted Random Early Drop (WRED)
  - One or more thresholds per queue
  - Threshold associated with priority
  - Buffer usage below threshold no affect
  - Buffer usage over min threshold = random drops
  - Buffer usage over max threshold = all traffic drop



Agenda

- Introduction
- **QoS Basics**
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale QoS
- Nexus 9800 QOS
- Real World Configuration Examples
- Conclusion

Nexus uses Modular QOS CLI (MQC)



Three Different Types

Class-map

Type QoS
- CoS
- DSCP
- PREC
- ACLs

Type Queuing
- qos-group

Type Network-QoS
- qos-group

Policy-map

Type QoS
- Classification
- Marking
- Policing

Type Queuing
- Queuing
- Scheduling
- Shaping

Type Network-QoS
- MTU
- Non-drop

Service-policy

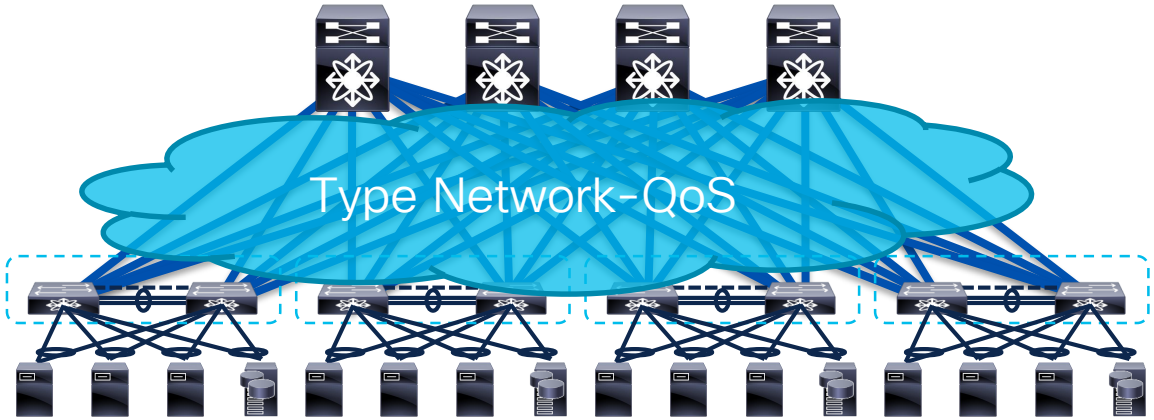
Type QoS
- Interface
- Port-channel
- VLAN

Type Queuing
- Interface
- System-qos

Type Network-QoS
- System-qos

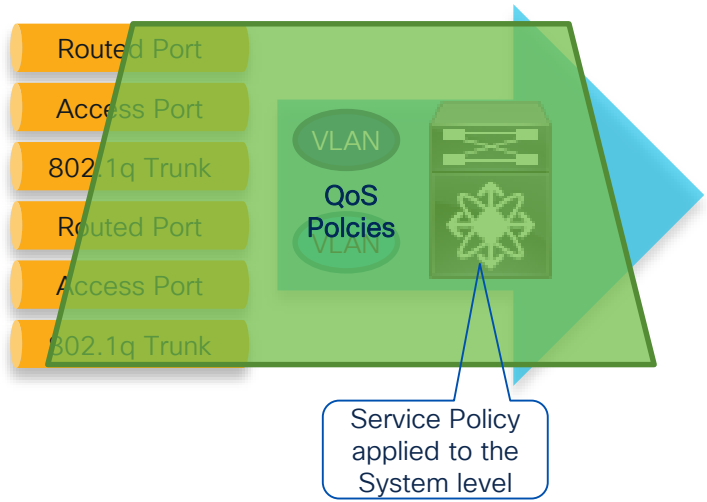
Type Network-QoS Policy

- 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

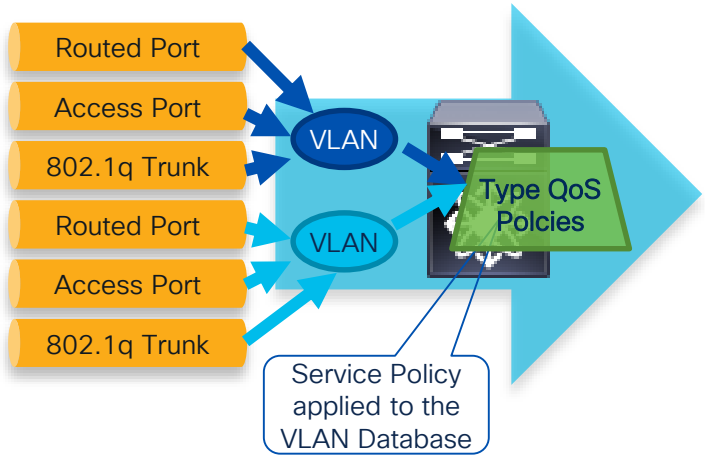
- System based QoS Policy gets globally applied to a system (to all interfaces)
- System based QoS Policy is configured in System QoS
- Type Queueing can be attached to the system level
- Type Network-QoS is mandatory to be attached to the system level




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

VLAN Based QoS Policy Attachment

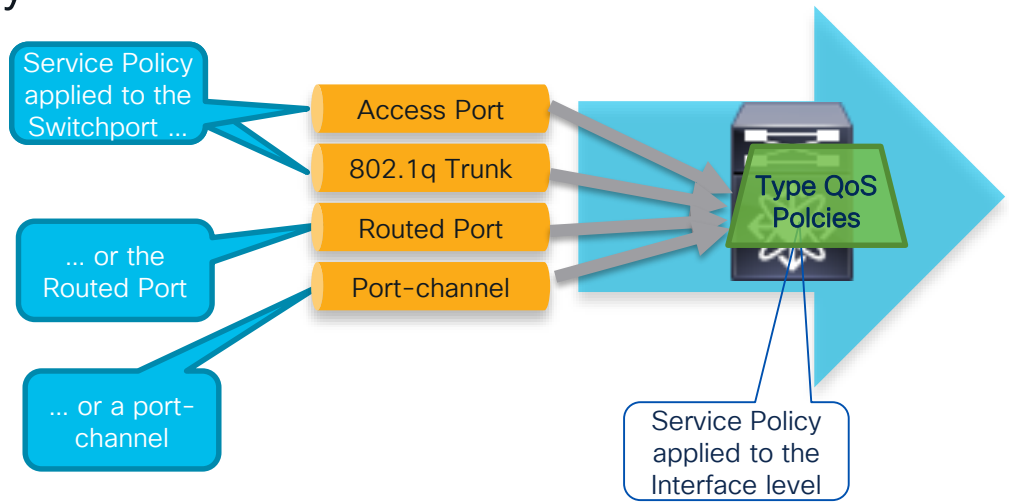
- VLAN based QoS Policy is configured in VLAN Database
- No SVI (aka L3 VLAN Interface) required




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

Interface based Type QOS Policy attachment

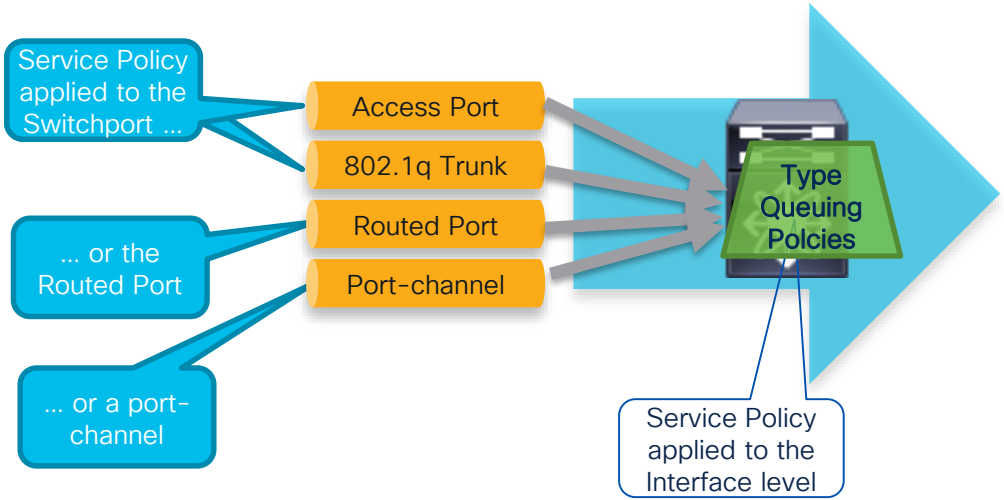
- Interface based type qos Policy takes precedence over VLAN
- Can also be attached to port-channel and applies to all member-ports




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

Interface based Type Queuing Policy attachment

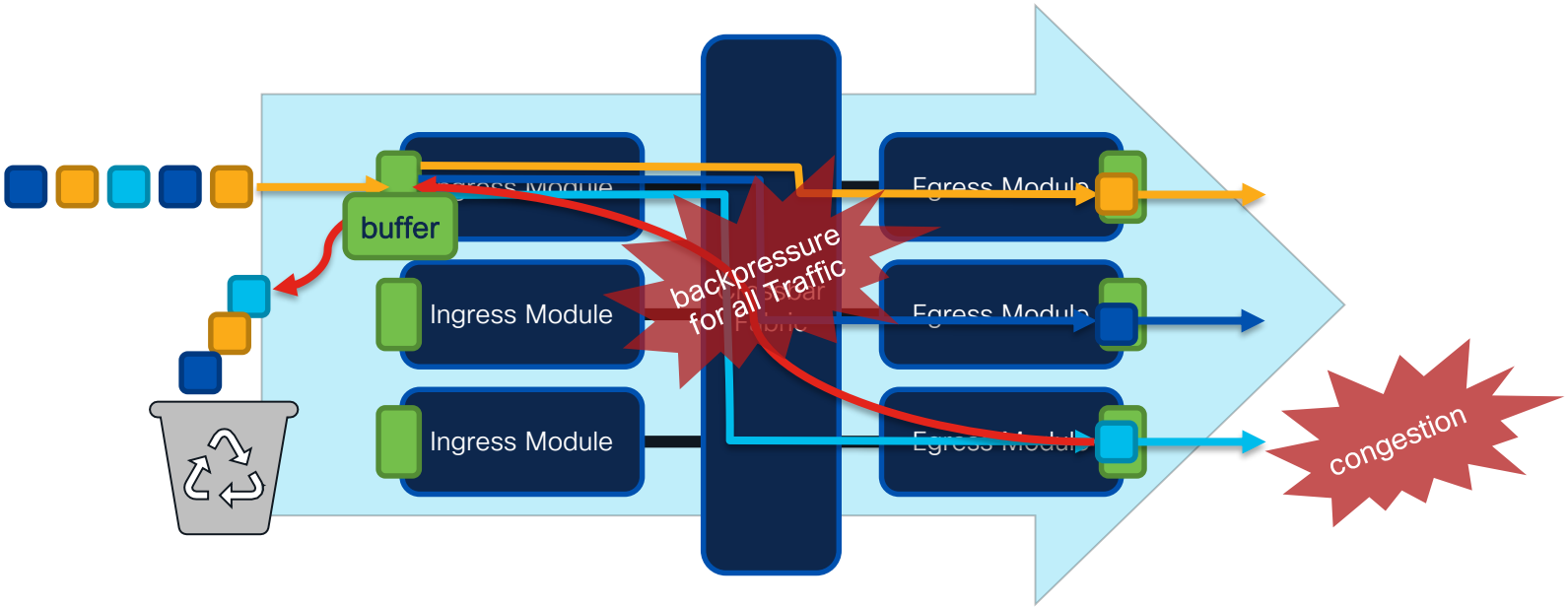
- Type Queuing has to be attached to a physical interface or system level
- Queuing Policy can be attached to port-channel and all member ports



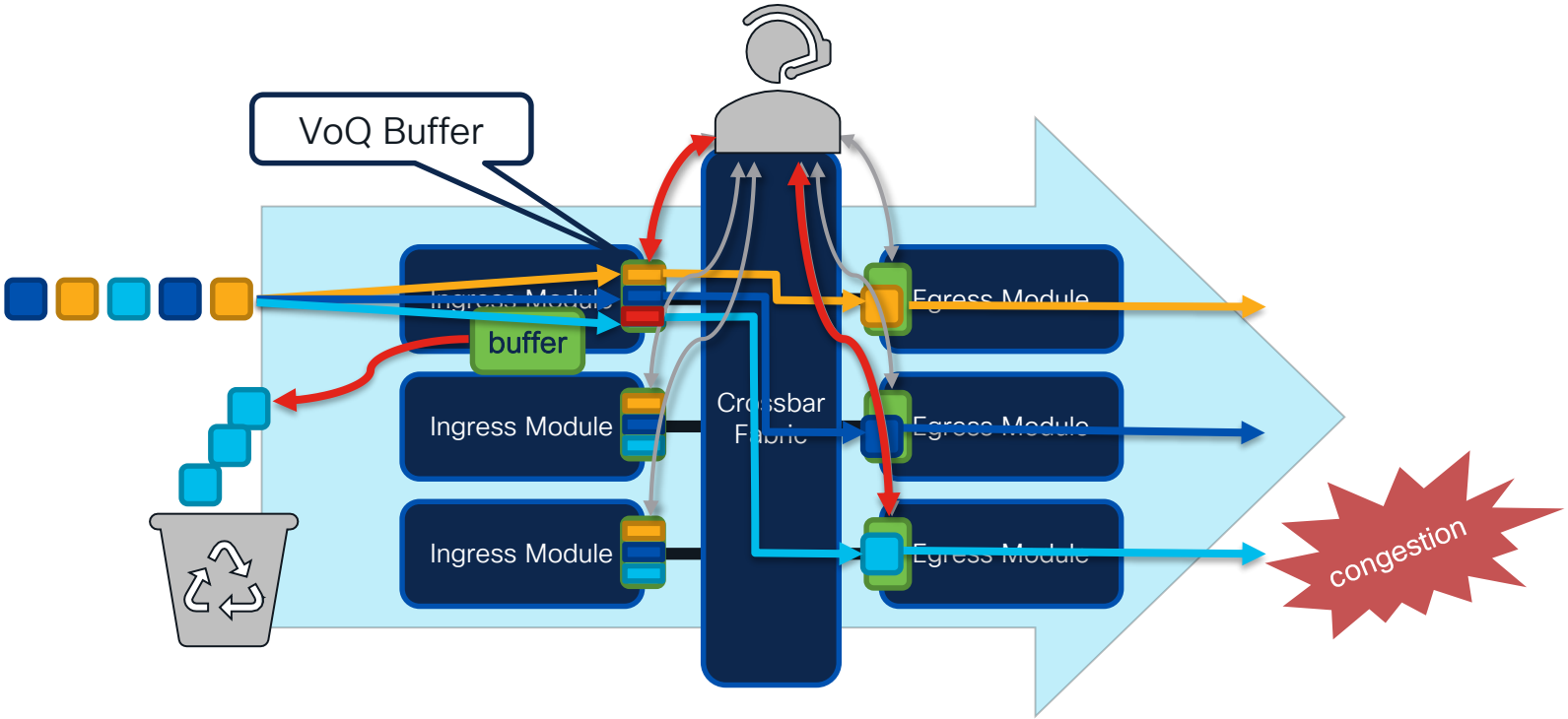

```
Nexus(config)# interface ethernet 1/1
Nexus(config-if)# service-policy type queueing output myPolicy
```

Buffer types – Head of Line Blocking

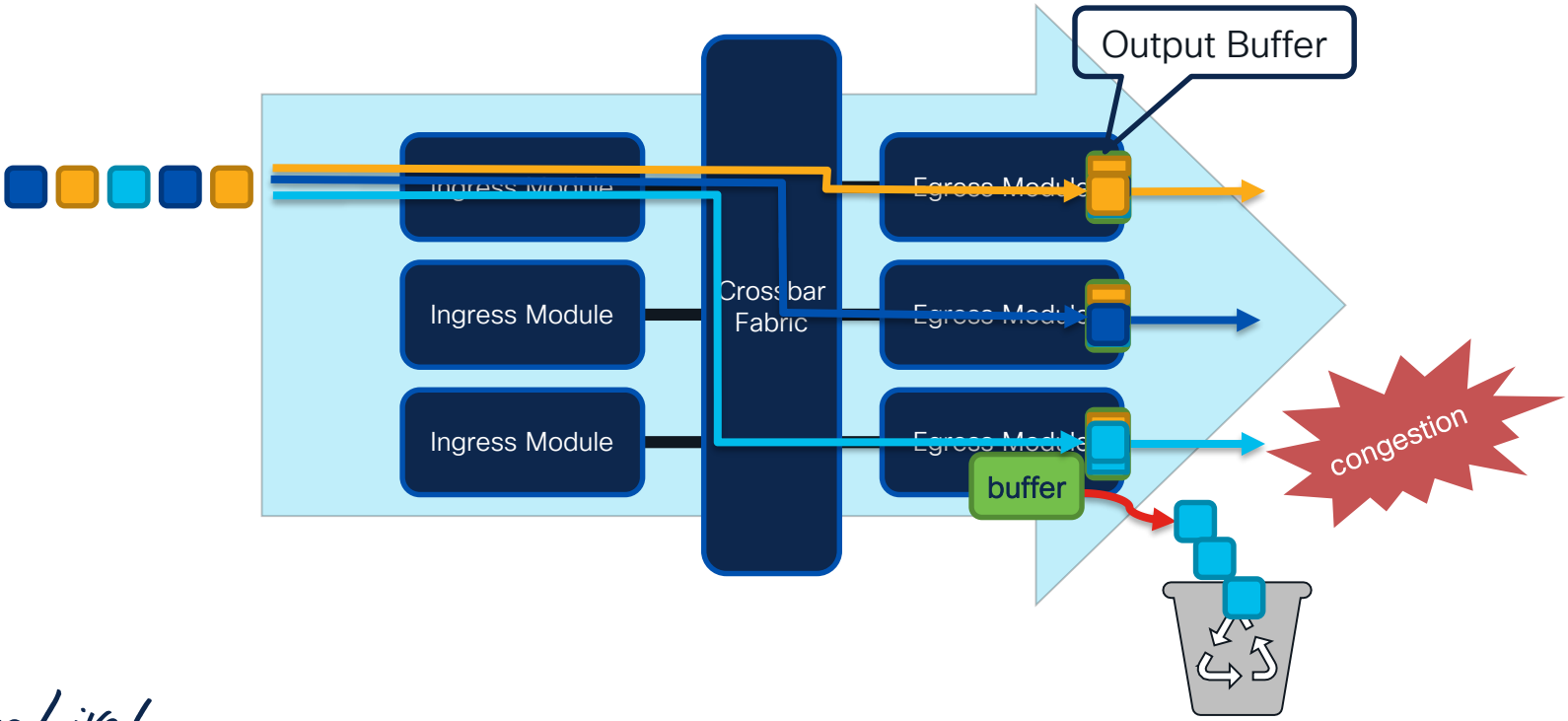
What is the Problem?



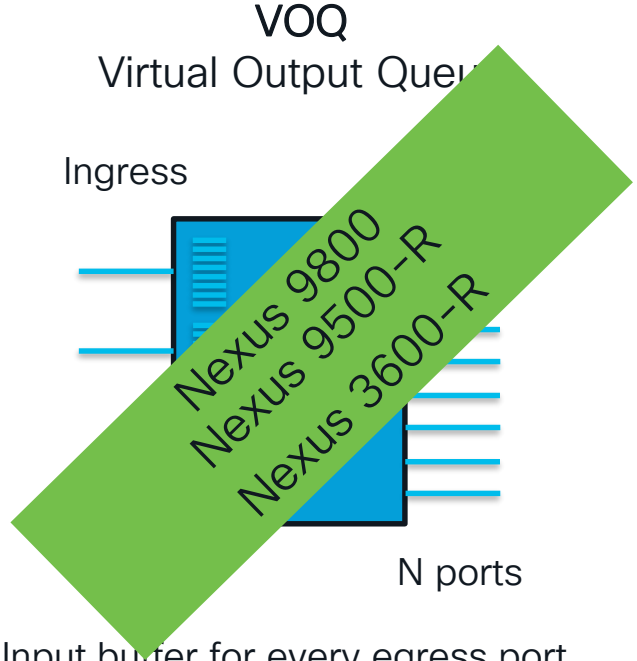
Virtual Output Queuing



Output Queuing

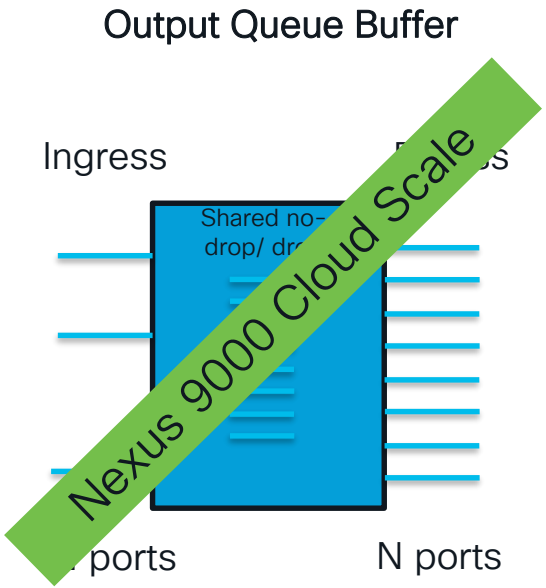


Buffering on Nexus Models



Input buffer for every egress port

NxN buffer size



Shared buffer for N egress ports

N buffer size

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 than Best-Effort	1,2,4	Q1
Best-Effort	0	Default-Q

8 Class Queuing Model

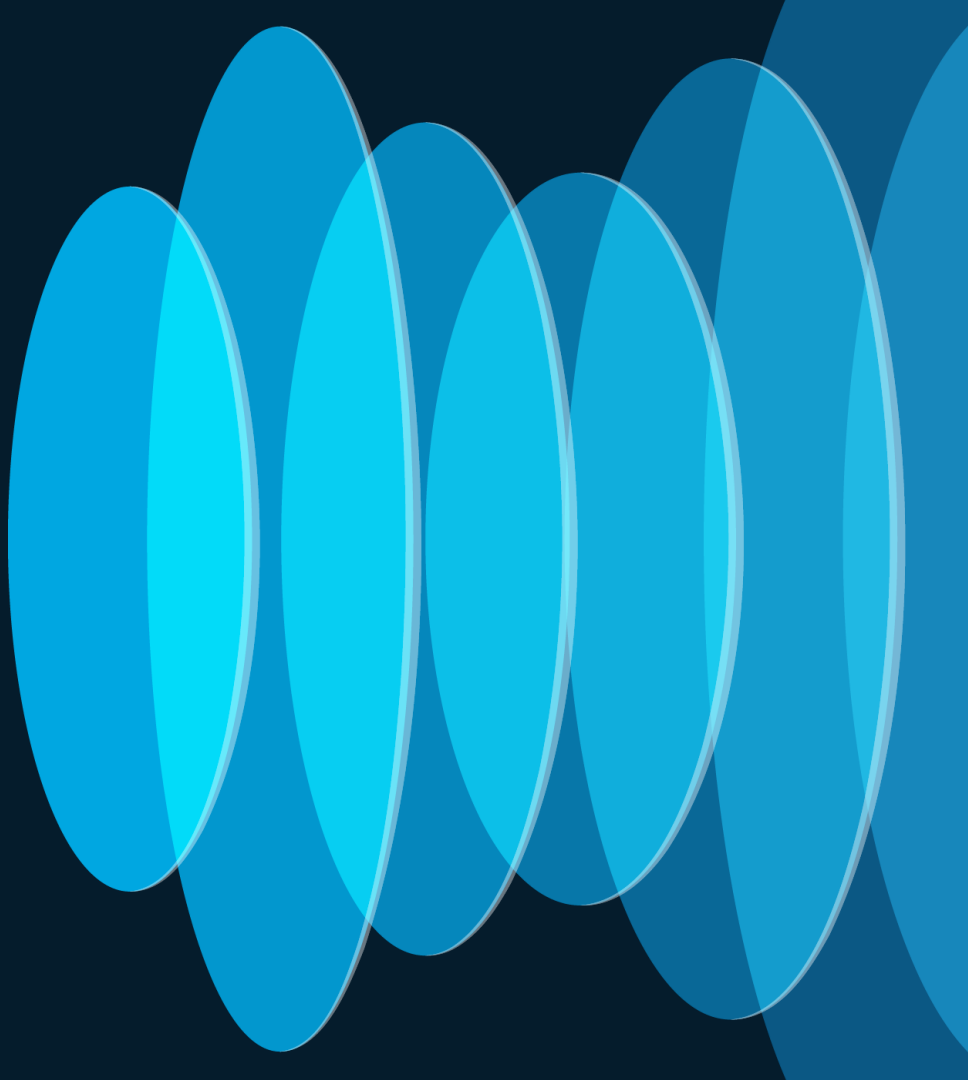
- Matches often a Campus QoS concept
- DSCP to CoS derivation does NOT apply anymore
  - (Topmost 3-Bit mapping from DSCP to CoS)
- No-Drop still with CoS3
- DSCP 24-30 are usable for IP storage traffic (RoCEv2)

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

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

Data Center QoS Capabilities



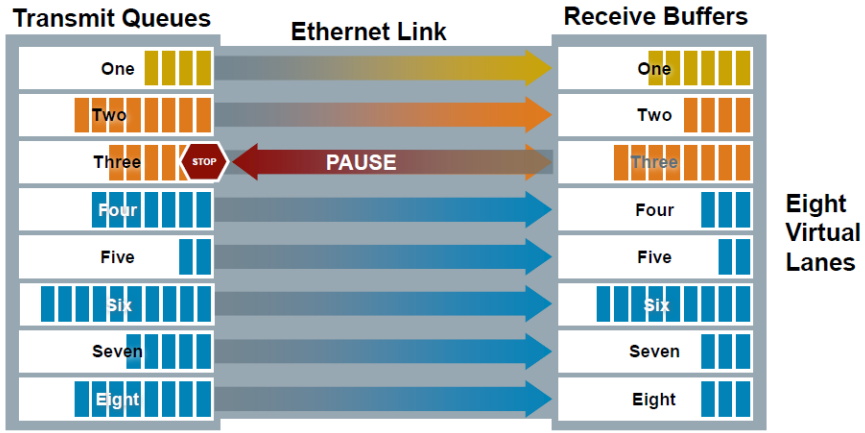
Data Centre Converged Infrastructure

- Enable, sensitive to drop, storage traffic to use Ethernet
- Simplification of the infrastructure by using Ethernet for data and storage traffic
- Data Center QoS capabilities, enabling new transport:
  - PFC – Priority Flow Control
  - ETS – Enhanced Transmission Selection
  - DCBX – Data Center Bridging Exchange
  - ECN – Explicit Congestion Notification

Priority Flow Control

Flow Control Mechanism – 802.1Qbb

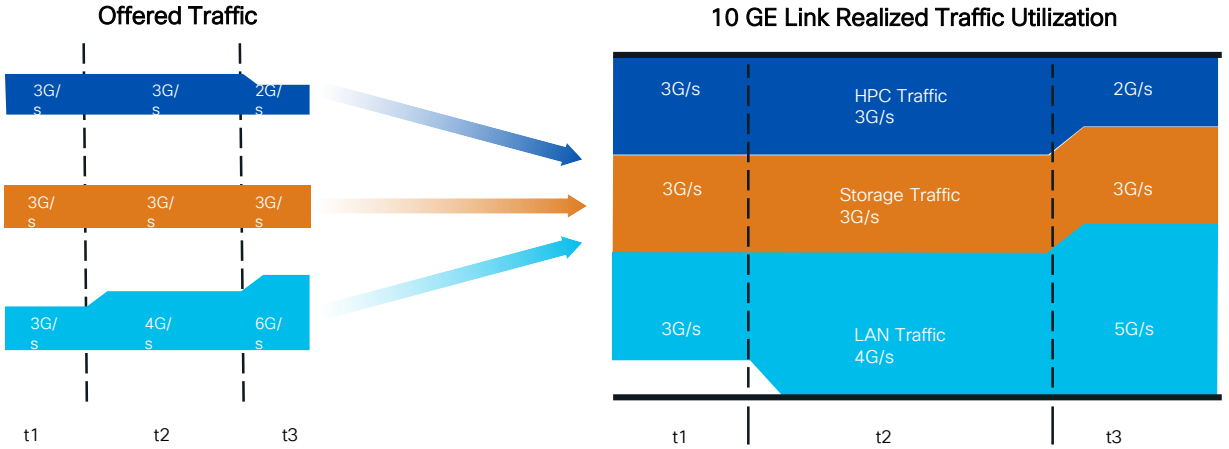
- A.k.a "Lossless Ethernet”
- PFC enables Flow Control on a Per-Priority basis
- Therefore, we have the ability to have lossless and lossy priorities at the same time on the same wire
- Allows traffic to operate over a lossless priority independent of other priorities
- Other traffic assigned to other priority will continue to transmit and rely on upper layer protocols for retransmission



Enhanced Transmission Selection

(ETS) Bandwidth Management – 802.1Qaz

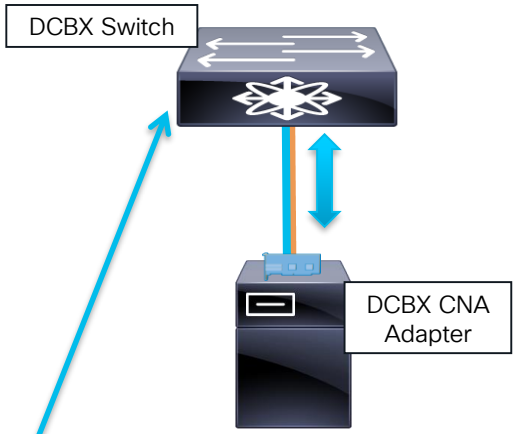
- Prevents a single traffic class of “hogging” all the bandwidth and starving other classes
- When a given load doesn’t fully utilize its allocated bandwidth, it is available to other classes
- Helps accommodate for classes of a “bursty” nature



Data Center Bridging Exchange Protocol

DCBX Overview – 802.1Qaz

- Negotiates Ethernet capability's PFC, ETS, CoS values between DCB capable peer devices
- Simplifies Management allows for configuration and distribution of parameters from one node to another
- DCBX is LLDP with new TLV fields



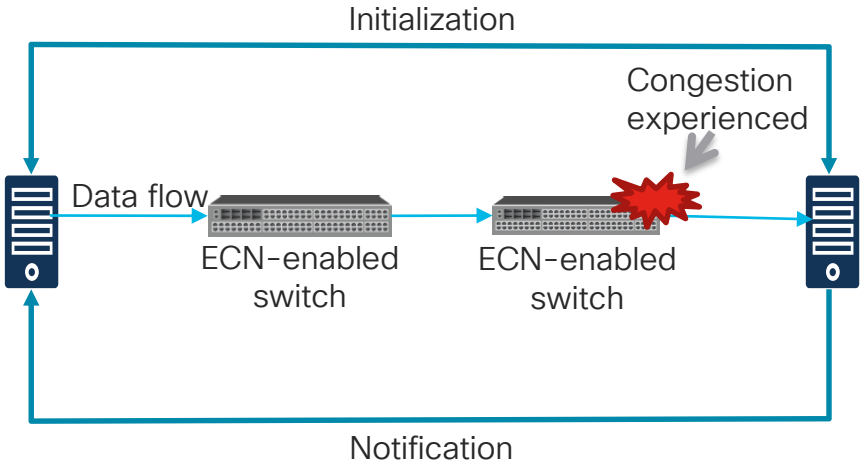

```
dc11-5020-3# sh lldp dcbx interface eth 1/40

Local DCBXP Control information:
Operation version: 00  Max version: 00  Seq no: 7  Ack no: 0
Type/
Subtype    Version    En/Will/Adv Config
006/000     000        Y/N/Y      00
<snip>
```

https://www.cisco.com/en/US/netsol/ns783/index.html

Explicit Congestion Notification (ECN)

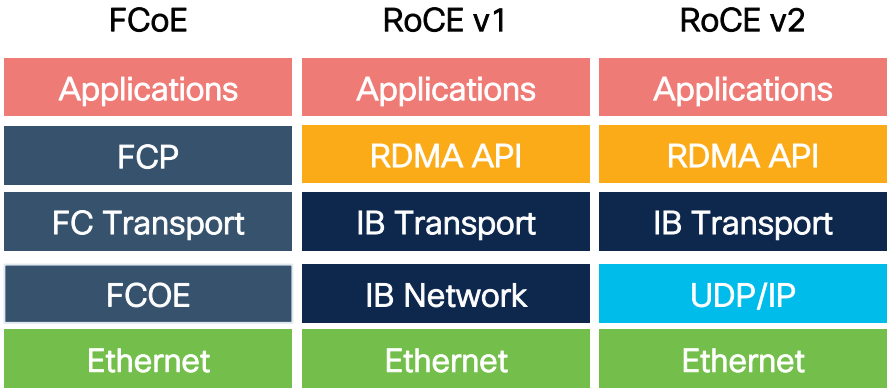
- IP Explicit Congestion Notification (ECN) is used for congestion notification.
- ECN enables end-to-end congestion notification between two endpoints on a IP network
- In case of congestion, ECN gets transmitting device to reduce transmission rate until congestion clears, without pausing traffic.
- ECN uses 2 LSB of Type of Service field in IP header



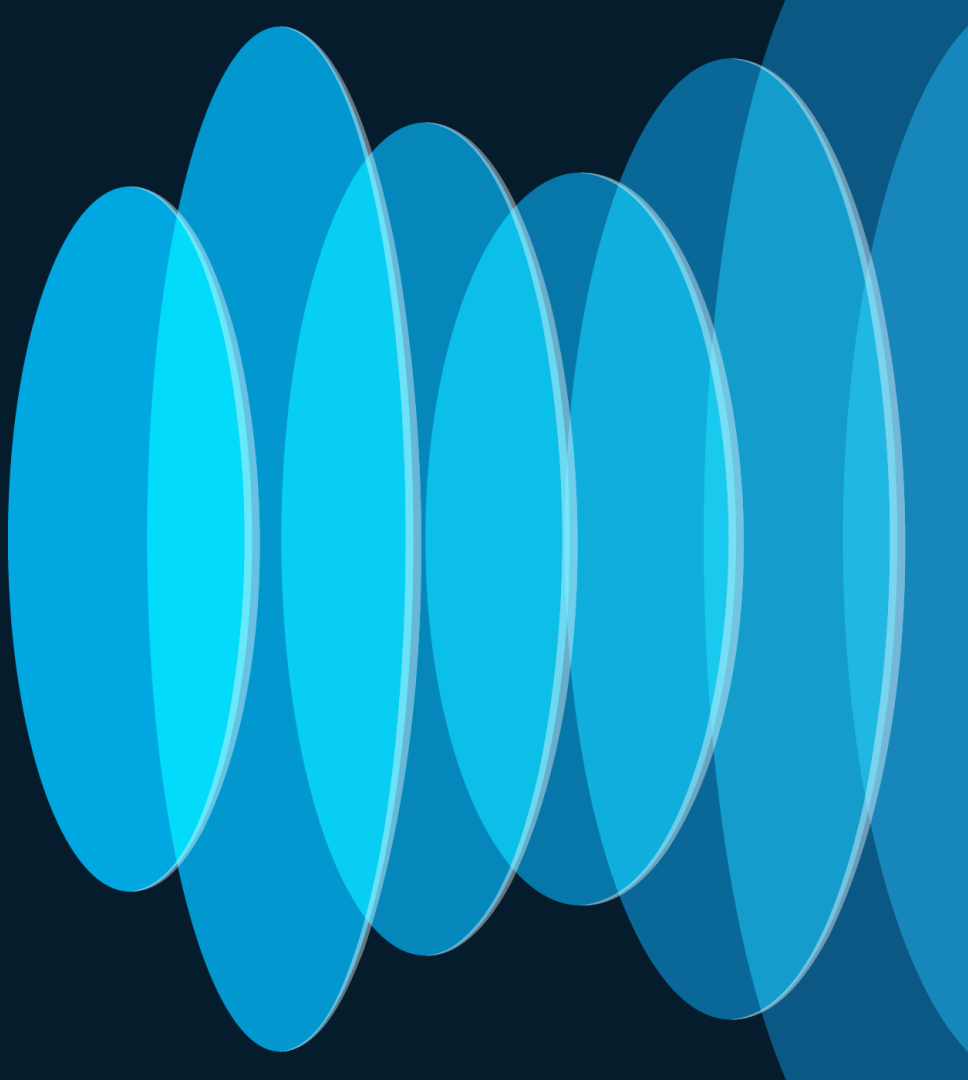
ECN	ECN Behavior
0x00	Non ECN Capable
0x10	ECN Capable Transport (0)
0x01	ECN Capable Transport (1)
0x11	Congestion Encountered

IP Storage Transports in Data Center

- Converged storage Protocols:
- Requirement for FCoE and RoCEv1:
  - PFC
  - ETS
- Requirement for RoCEv2
  - PFC
  - ETS
  - ECN



Overlay QOS



Overlay QoS

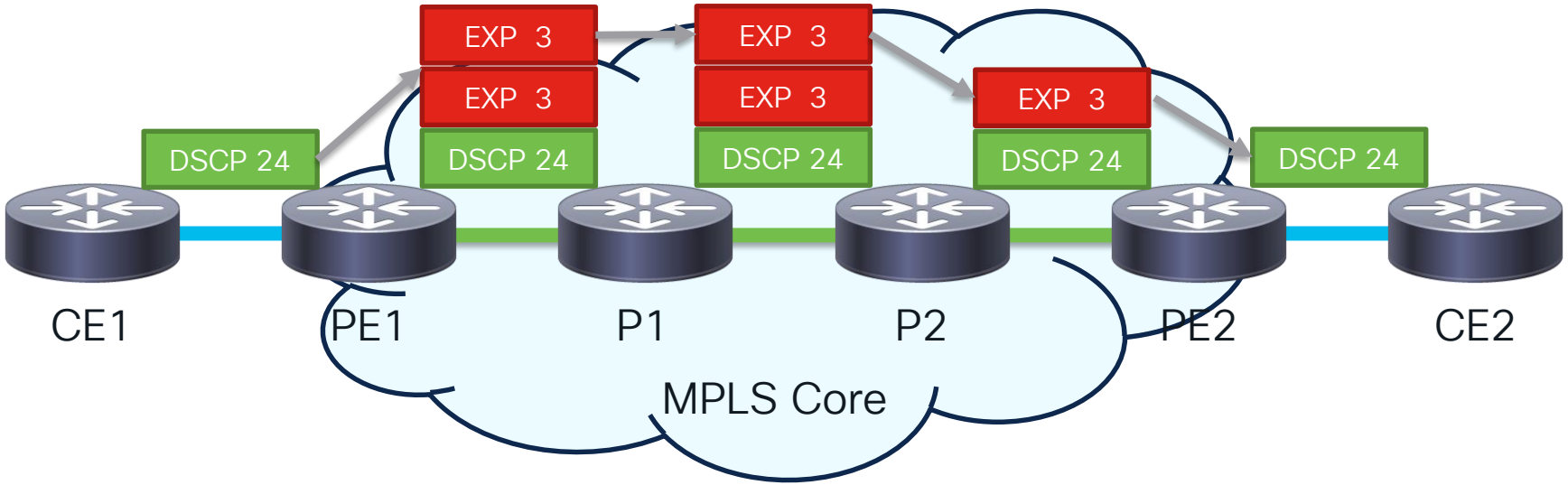
MPLS network

- 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

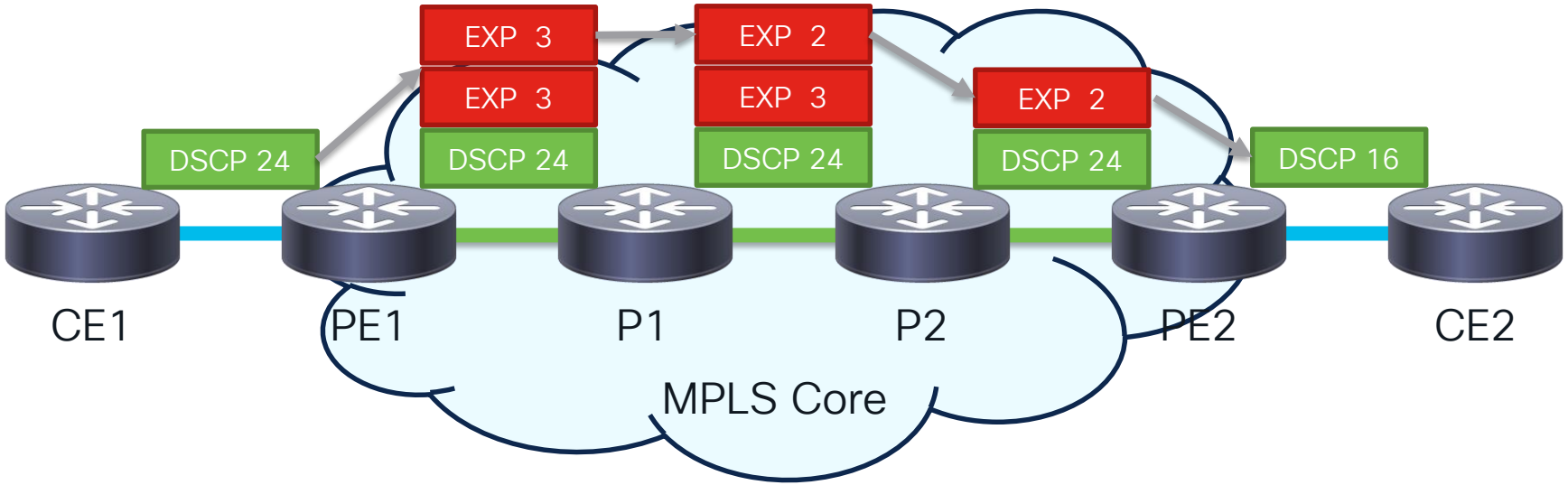
Overlay QOS

MPLS – Default Mode



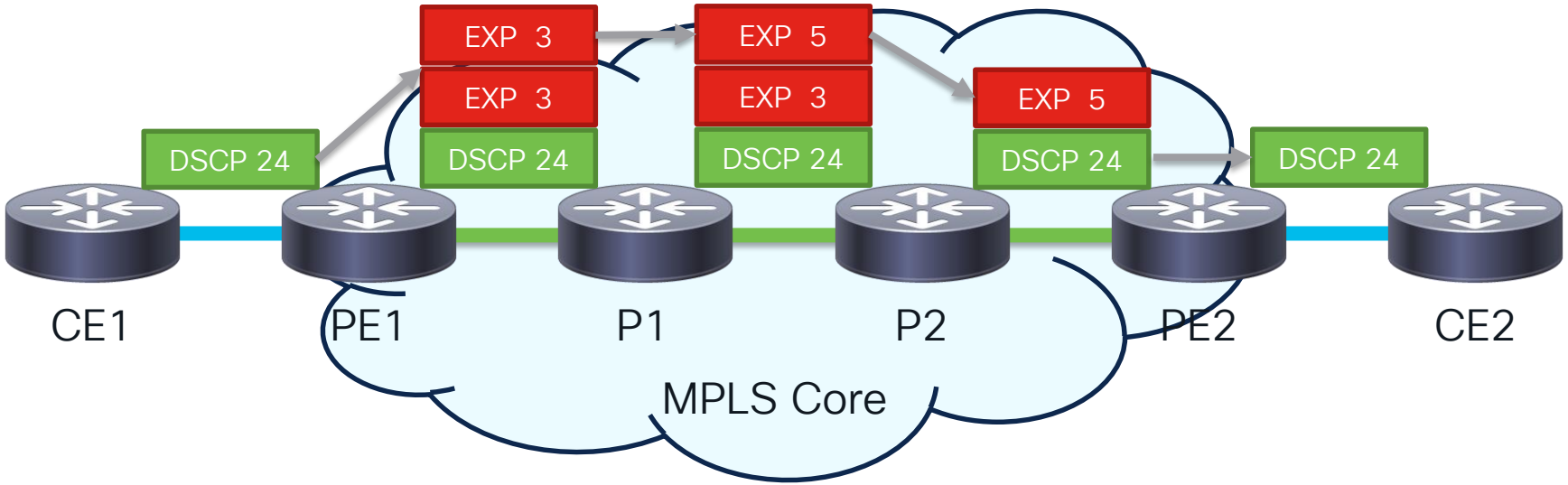
Overlay QOS

MPLS – Uniform Mode



Overlay QOS

MPLS – Pipe Mode

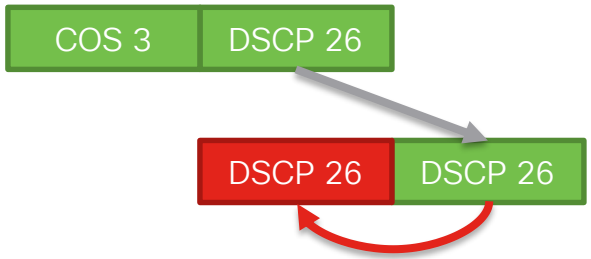


Overlay QoS

VXLAN EVPN – VXLAN Encapsulation

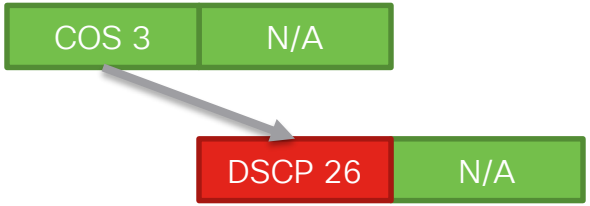
- Ingress L3 packet, original priority is mapped to outer header priority
- Ingress L2 frame, COS value will be mapped to outer priority
- User can choose to mark outer priority independently
- VLAN header is not preserved in VXLAN tunnel

Original L3 Packet



VXLAN Encap. Packet

Original L2 Frame



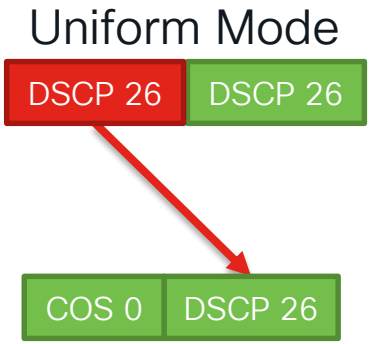
VXLAN Encap. Packet

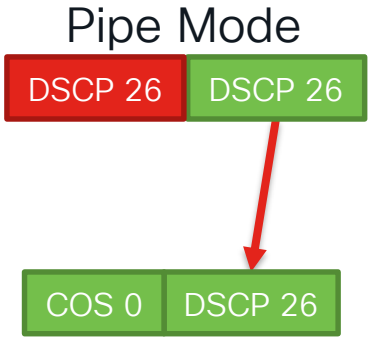
COS	DSCP
0	0
1	8
2	16
3	26
4	32
5	46
6	48
7	56

Overlay QoS

VXLAN EVPN – VXLAN Decapsulation

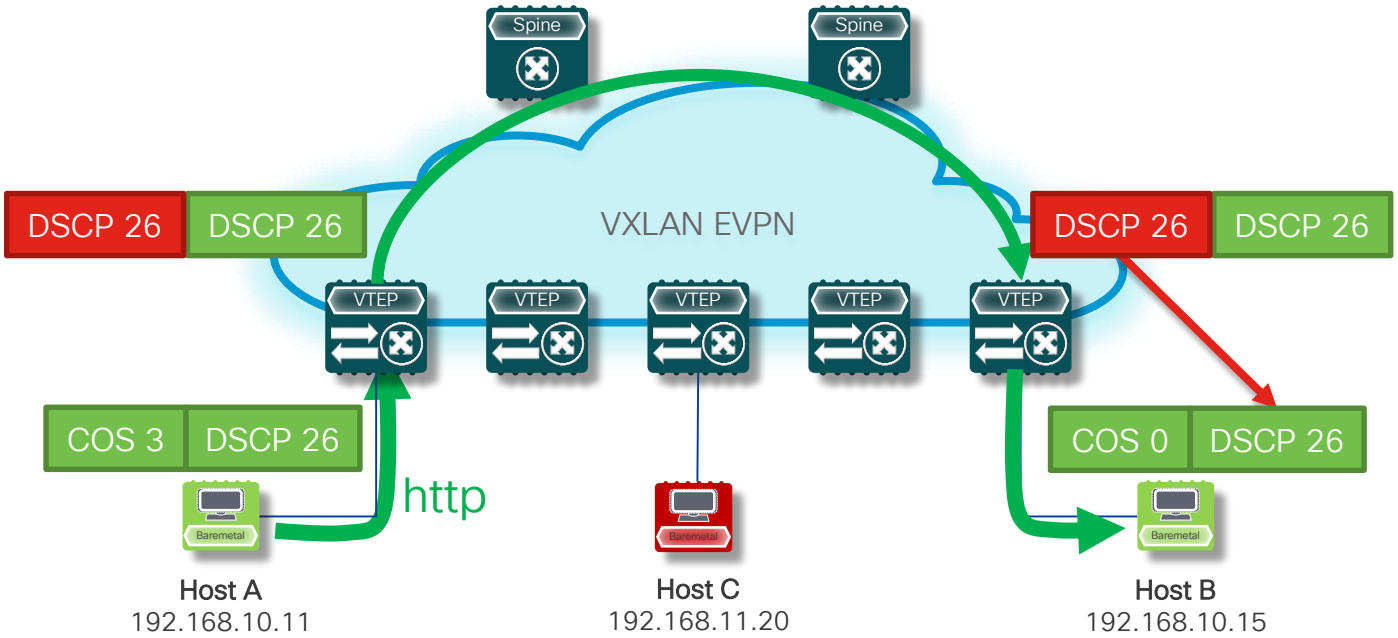
- DSCP value is derived based on a priority mode for L3 traffic:
  - Uniform mode: delivers overlay priority copying outer header to decapsulated frame
  - Pipe mode: extends original priority copying inner header to decapsulated frame
- User can classify traffic based on outer or inner priority
- Marking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





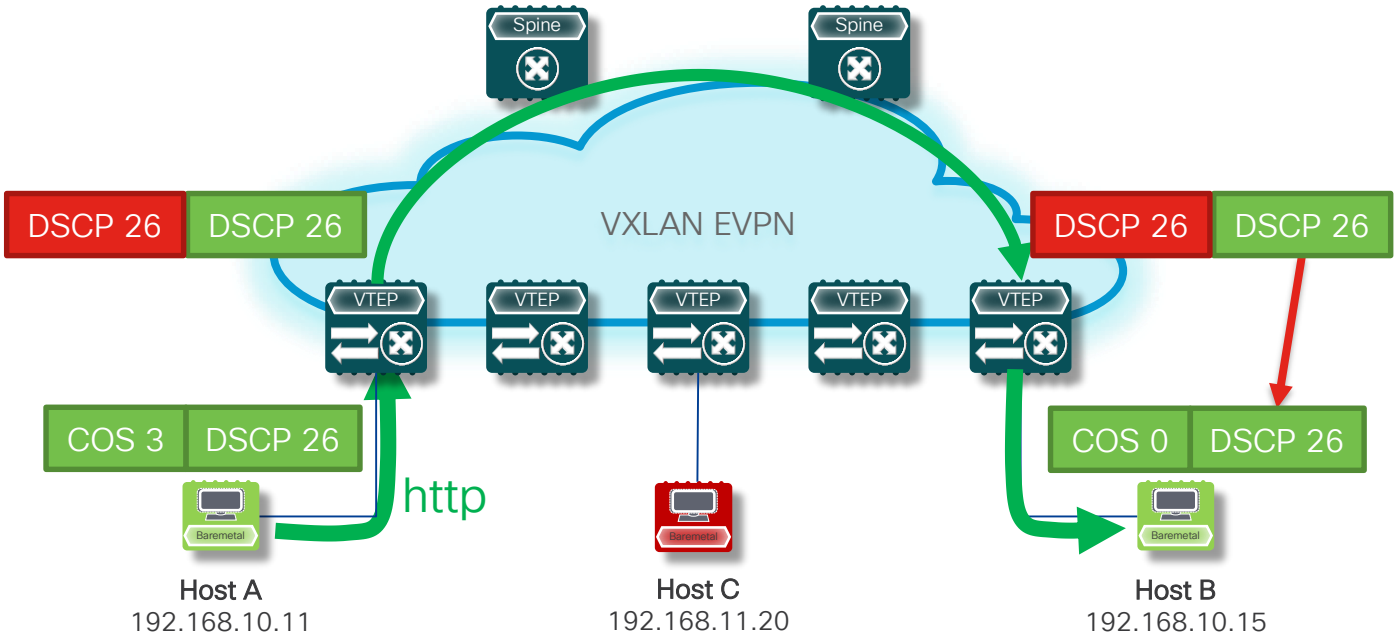
Overlay QoS

VXLAN – Uniform Mode



Overlay QoS

VXLAN – Pipe Mode



Agenda

- Introduction
- QoS Basics
- **QoS Implementation on Nexus**
- Nexus 9000 Cloud Scale QoS
- Nexus 9800 QOS
- Real World Configuration Examples
- Conclusion

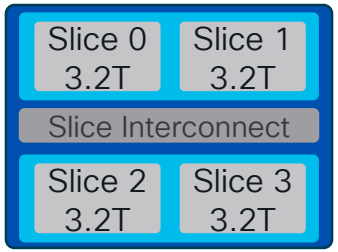
Nexus 9000 Overview

- Modular and Fixed chassis
- Optimized for high density 10G/25G/40G/50G/100G/ 200G/400G
- Standalone and ACI Mode
- Cisco Silicon – Cloud Scale
  - Advanced QoS capabilities

Key Cloud Scale Family Members

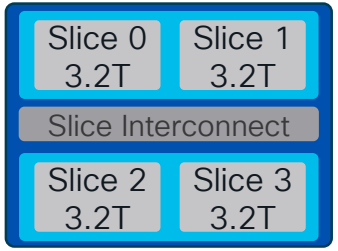
LS25600GX2A – 64 x 400G

25.6T chip – 4 slice pairs of 8 x 400G
9300-GX2A TORs; 9408 centralized modular TOR



LS12800GX2B – 32 x 400G

12.8T chip – 2 slice pairs of 8 x 400G
9300-GX2B TOR



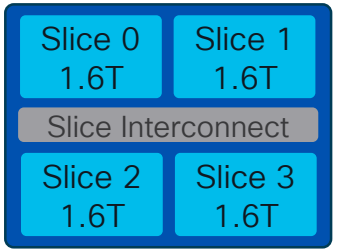
LS12800 H2R – 32 x 400G

12.8T chip – 2 slice pairs of 8 x 400G, 8GB HBM
9300-H2R TOR

LS6400H1 – 16 x 400G

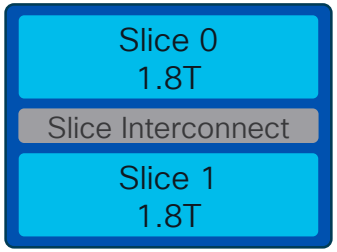
3.6T chip – 2 slices of 8 x 400G
9300-H1 TORs

Key Cloud Scale Family Members



LS6400GX – 16 x 400G

6.4T chip – 4 slices of 4 x 400G
X9700-GX modular linecards; 9300-GX TORs



LS3600FX2 – 36 x 100G

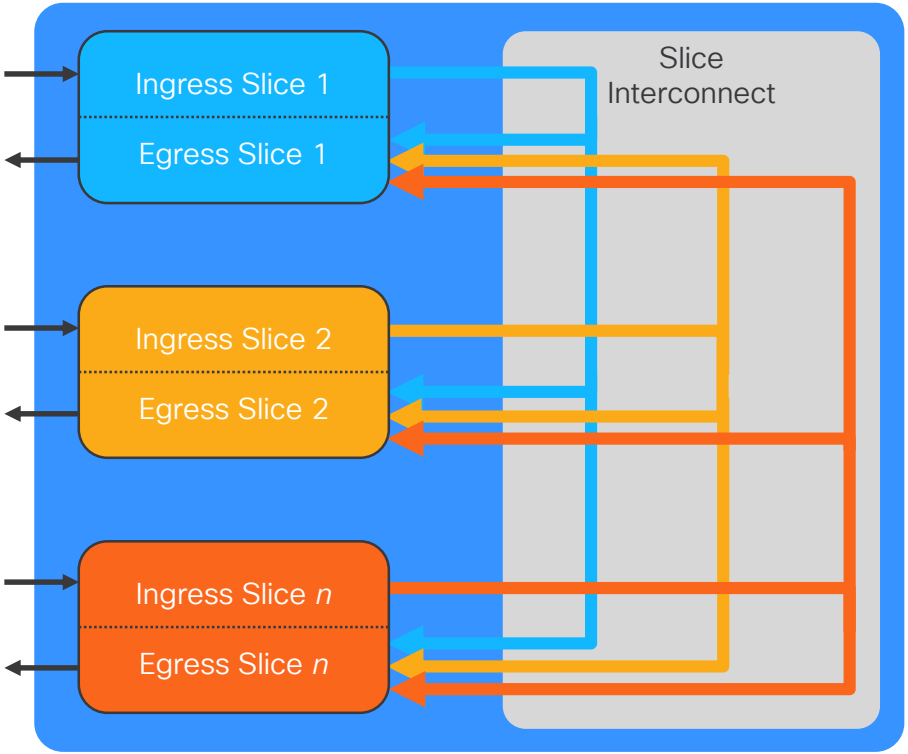
3.6T chip – 2 slices of 18 x 100G with MACSEC
9300-FX2 TORs

LS1800FX/FX3 – 18 x 100G

1.8T chip – 1 slice of 18 x 100G with MACSEC
X9700-FX modular linecards; 9300-FX/FX3 TORs

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

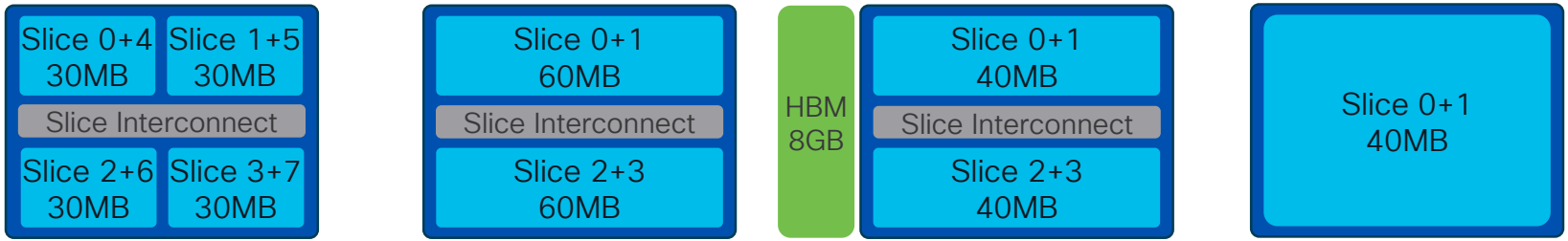


Cisco Nexus 9000 – Cloud Scale QoS Features

- Classification based on:
  - ACL
  - DSCP, CoS, and IP Precedence
- Marking traffic with:
  - DSCP
  - CoS
  - IP Precedence
- Policing:
  - 1R2C and 2R3C
  - Ingress and Egress
- Buffering/Queueing:
  - Shared egress buffer; 8 Egress Queues
- Scheduling:
  - Strict Priority Queuing and DWRR
- Shaping:
  - Egress per queue shaper
- Congestion Avoidance:
  - Tail Drop
  - WRED with ECN

Buffering

- Cloud Scale platforms implement shared-memory egress buffered architecture
- Slices share pool of buffer – ports on a slice pairs can use that buffer
- Dynamic Buffer Protection adjusts max thresholds based on class and buffer occupancy
- Intelligent buffer options maximize buffer efficiency



LS25600GX2A
30MB/slice pair
(120MB total)

L12800GX2B
60MB/slice pair
(120MB total)

LS12800 H2R
60MB/slice pair
(120MB total)

LS6400H1
40MB/slice pair
(40MB total)

Buffering

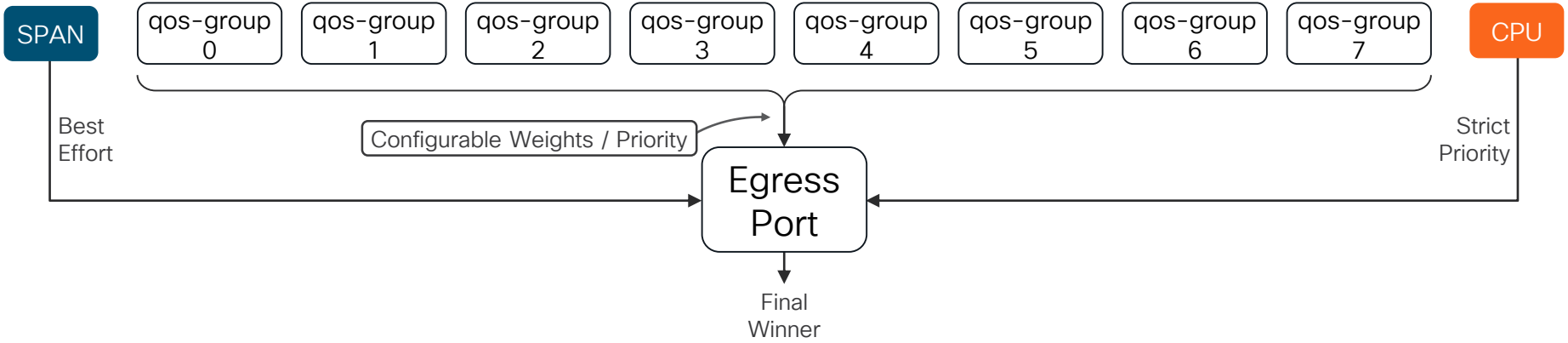
- Cloud Scale platforms implement shared-memory egress buffered architecture
- Each ASIC slice has dedicated buffer – only ports on that slice can use that buffer
- Dynamic Buffer Protection adjusts max thresholds based on class and buffer occupancy
- Intelligent buffer options maximize buffer efficiency

LS6400GX
20MB/slice
(80MB total)

LS3600FX2
20MB/slice
(40MB total)

LS1800FX3
40MB/slice
(40MB total)

Queuing and Scheduling

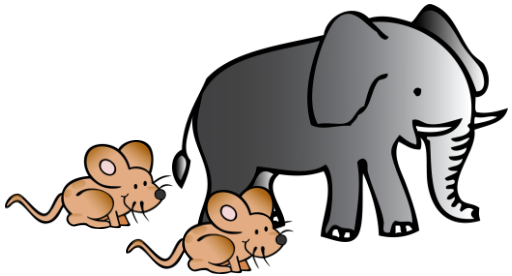


- 8 qos-groups per output port
- Egress queuing policy defines priority and weights
- Dedicated classes for CPU traffic and SPAN traffic

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



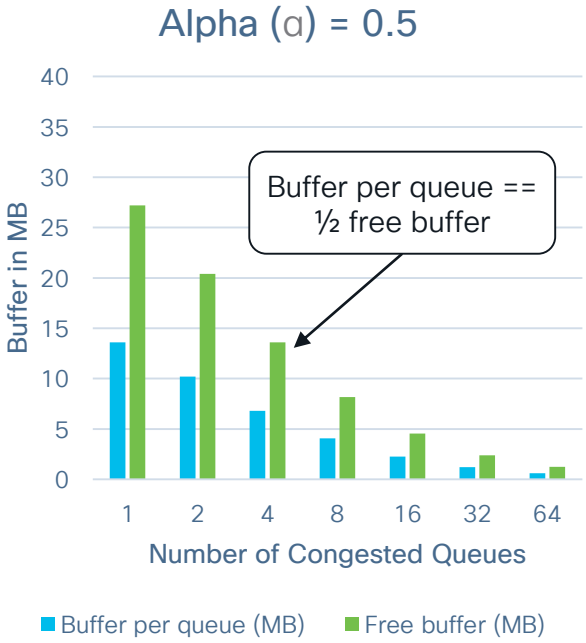
Miercom Report: Speeding Applications in Data Centre Networks
http://miercom.com/cisco-systems-speeding-applications-in-data-center-networks/

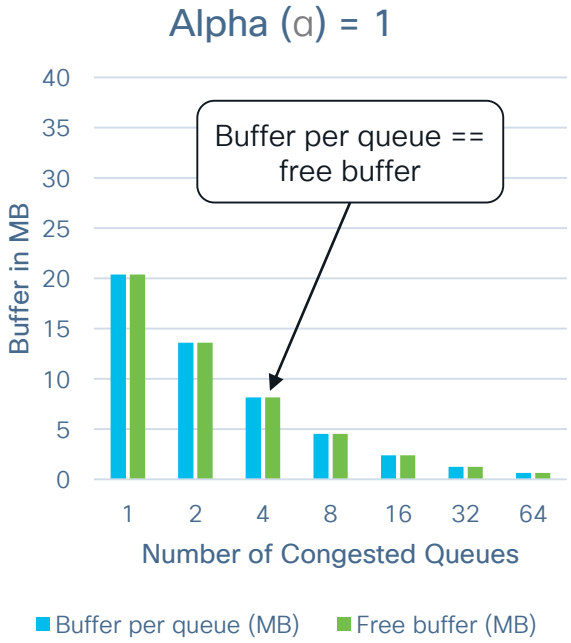
Dynamic Buffer Protection (DBP)

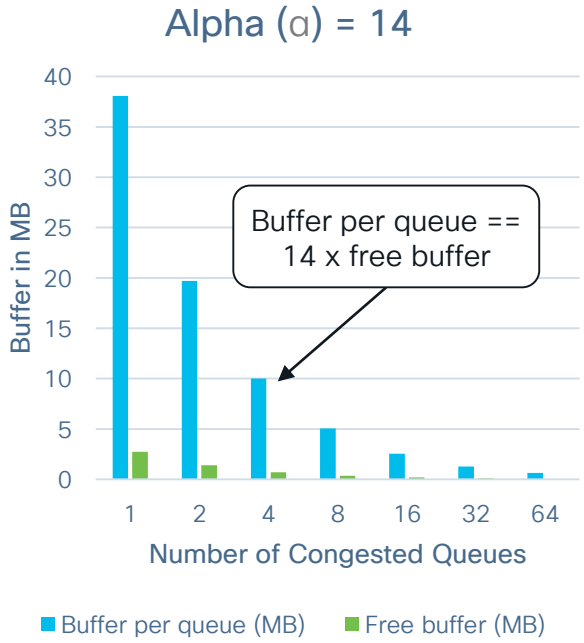
- Prevents any output queue from consuming more than its fair share of buffer in shared-memory architecture
- Defines dynamic max threshold for each queue
  - If queue length exceeds threshold, packet is discarded
  - Otherwise packet is admitted to queue and scheduled for transmission
- Threshold calculated 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

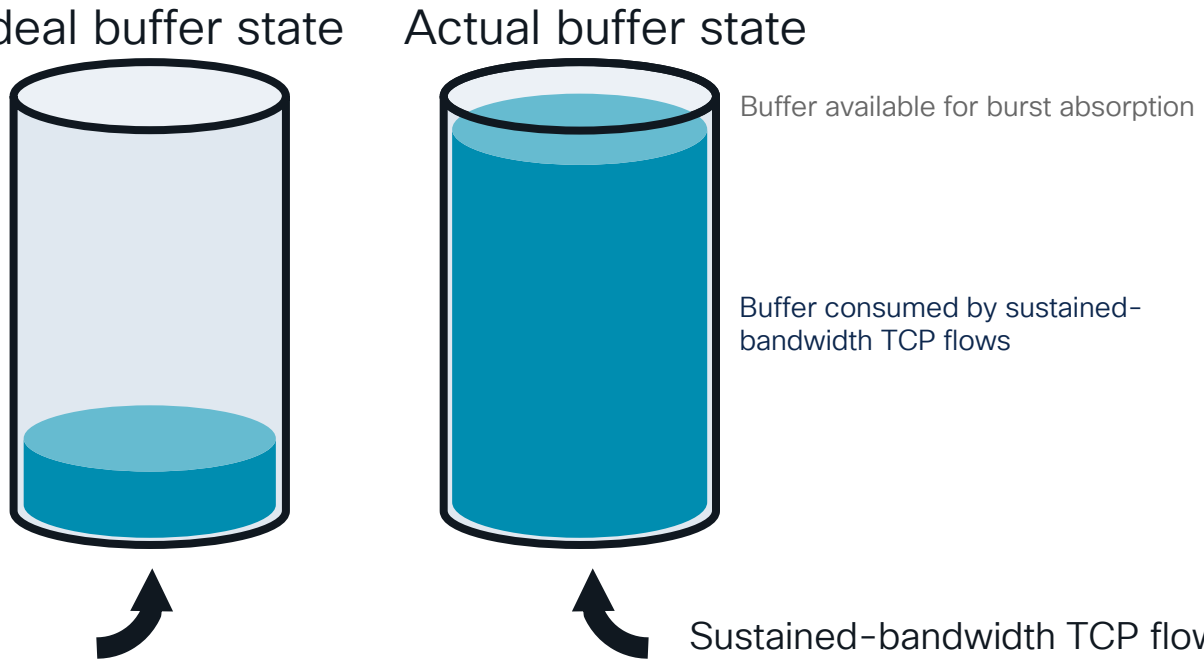
Default Alpha on
Cloud Scale switches





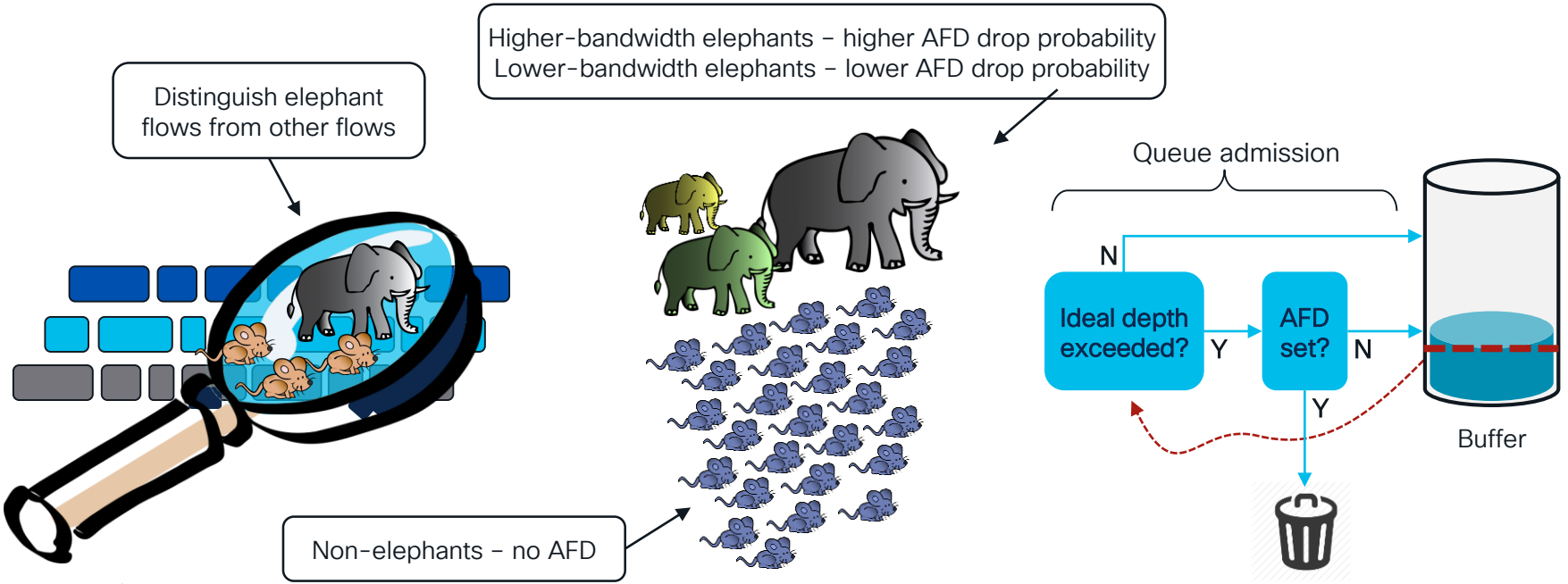


Buffering – Ideal versus Reality



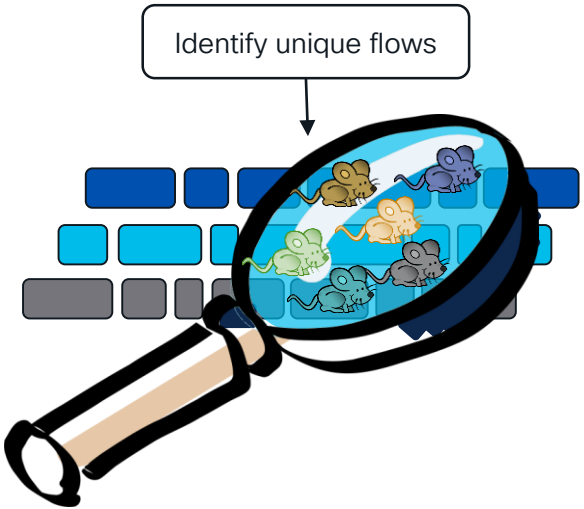
Approximate Fair Drop (AFD)

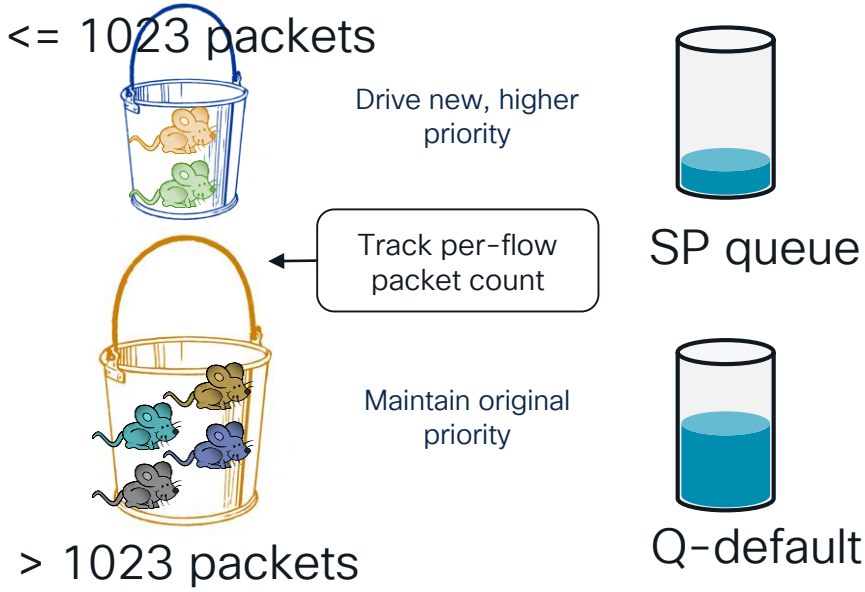
Maintain throughput while minimizing buffer consumption by elephant flows – keep buffer state as close to the ideal as possible



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





Configuration – Class-Map Type QoS

- Class-map type qos used to classify traffic based on
  - Access List
  - Priority (CoS, DSCP, IP Precedence)
- Match by singe criteria or match all criteria under class-map:
  - match-any: Traffic needs to match any criteria under class map

```
class-map type qos match-any class-q1
 match access-group HTTP
 match cos 1
 match dscp 8
```

Configuration – Policy-Map Type QoS

- Policy-map type qos used to take action on class-map traffic
  - Set new priorities (COS, DSCP, IP Precedence)
  - Set a policer
- The policy-map sets qos-group

```
policy-map type qos Classification-Marking
  class class-q1
    set cos 1
    police cir 1000 mbps bc 200 ms conform transmit violate drop
    set qos-group 1
```

QoS-Group

- QoS group is used to reference classification for all the types class-maps
  - Class-map type queueing and type network qos have class-maps referencing qos-groups
  - Class-maps are present in system by default, no user configuration required
- Default class-map type queueing for Q1:

```
class-map type queuing match-any c-out-8q-q1
  match qos-group 1
```

- Default class-map type network-qos for Q1

```
class-map type network-qos c-8q-nq1
  description Default class on qos-group 1
  match qos-group 1
```

Configuration – Policy-Map Type Queuing

- Policy-map type queueing define queuing and scheduling options
  - Define queue limit – change alpha value
  - Define scheduling options, strict priority and weight for DWRR queues
- Default Queueing policy cannot be changed
  - User needs to define custom policy
- Shaping defined per queue in queueing policy

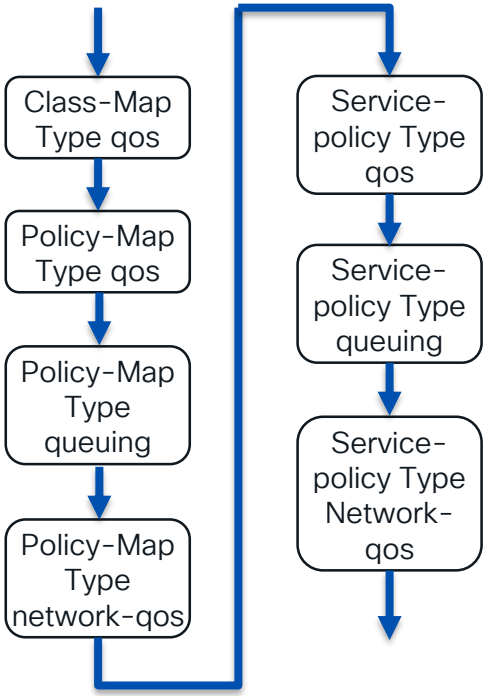
```
policy-map type queuing custom-8q-out-policy
  class type queuing c-out-8q-q7
    priority level 1
  class type queuing c-out-8q-q6
    bandwidth remaining percent 0
  class type queuing c-out-8q-q5
    bandwidth remaining percent 0
  class type queuing c-out-8q-q4
    bandwidth remaining percent 0
  class type queuing c-out-8q-q3
    bandwidth remaining percent 0
  class type queuing c-out-8q-q2
    bandwidth remaining percent 0
  class type queuing c-out-8q-q1
    bandwidth remaining percent 50
  class type queuing c-out-8q-q-default
    bandwidth remaining percent 50
```

Configuration – Policy-Map Type Network-QoS

- Policy-map type network-qos define:
  - Non-drop queue
  - End to end queueing policy (8 queue or 4 queue)
- Default Network-QoS policy cannot be changed
  - User needs to define custom policy

```
policy-map type network-qos custom-8q-nq-
policy
  class type network-qos c-8q-nq7
    mtu 1500
  class type network-qos c-8q-nq6
    mtu 1500
  class type network-qos c-8q-nq5
    mtu 1500
  class type network-qos c-8q-nq4
    mtu 1500
  class type network-qos c-8q-nq3
    mtu 1500
  class type network-qos c-8q-nq2
    mtu 1500
  class type network-qos c-8q-nq1
    mtu 1500
  class type network-qos c-8q-nq-default
    mtu 1500
```

Configuration – Putting it all together




```
class-map type qos match-any class-q1
 match access-group HTTP

policy-map type qos Classification-Marking
  class class-q1
    set cos 1
    set qos-group 1

policy-map type queuing custom-8q-out-policy
<snip>
  class type queuing c-out-8q-q1
    bandwidth remaining percent 50
  class type queuing c-out-8q-q-default
    bandwidth remaining percent 50

policy-map type network-qos custom-8q-nq-policy
<snip>
  class type network-qos c-8q-nq1
    mtu 1500
  class type network-qos c-8q-nq-default
    mtu 1500

interface Ethernet 1/1
  service-policy type qos input Classification-Marking

system qos
  service-policy type network-qos custom-8q-nq-policy
  service-policy type queuing output custom-8q-out-policy
```

Nexus 9000 QoS Golden Rules

- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- Nexus 9000 Cloud Scale – Egress Buffer
  - Queuing/scheduling policy attached in egress direction

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- **Nexus 9000 Cloud Scale QoS**
- Nexus 9800 QOS
- Real World Configuration Examples
- Conclusion

Nexus 9800 Overview

- Modular chassis switches
- Optimized for high density 100G/200G/400G/800G
- Standalone and ACI Mode*
- Cisco Silicon – Silicon One
  - VOQ architecture
  - HBM external buffer

* Future release in ACI

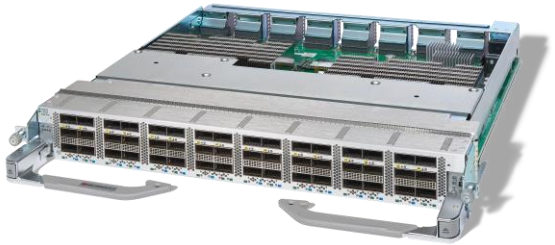
Nexus 9800 – Line Cards

Nexus 9836DM-A Line Card

- 36 x 400G ports
- 3 x 8GB HBM packet buffer

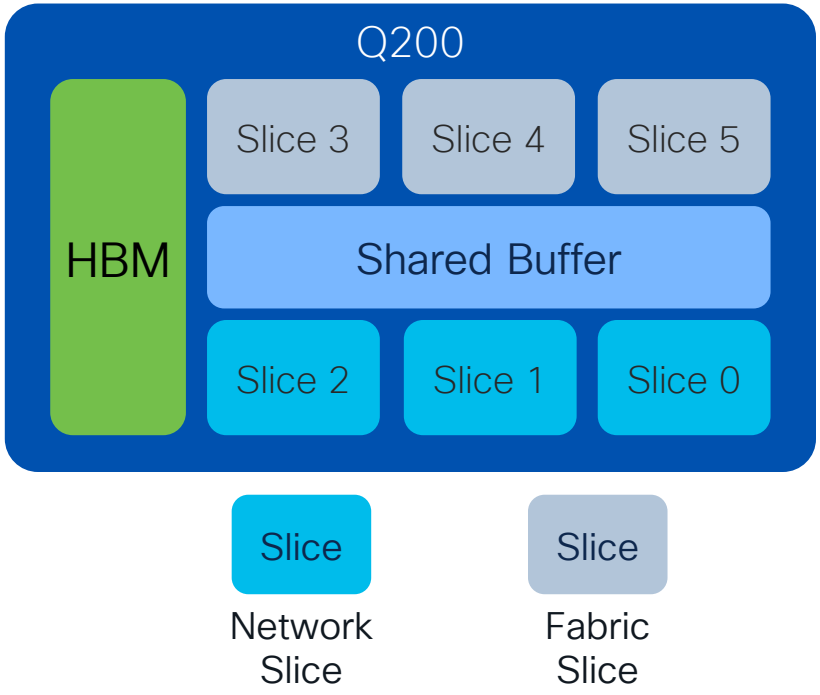
Nexus 98900CD-A Line Card

- 48 ports – 34 x 100G ports and 14 x 400G ports
- 2 x 8GB HBM packet buffer



Nexus 9800 – Q200 ASIC

- Q200 Silicon One ASIC
  - 6 Slices
    - 3 slices for network connectivity
    - 3 slices for fabric connectivity
  - Nexus 9836DM-A Line Card
    - 3 Q200 per line card
  - Nexus 98900CD-A Line Card
    - 2 Q200 per line card

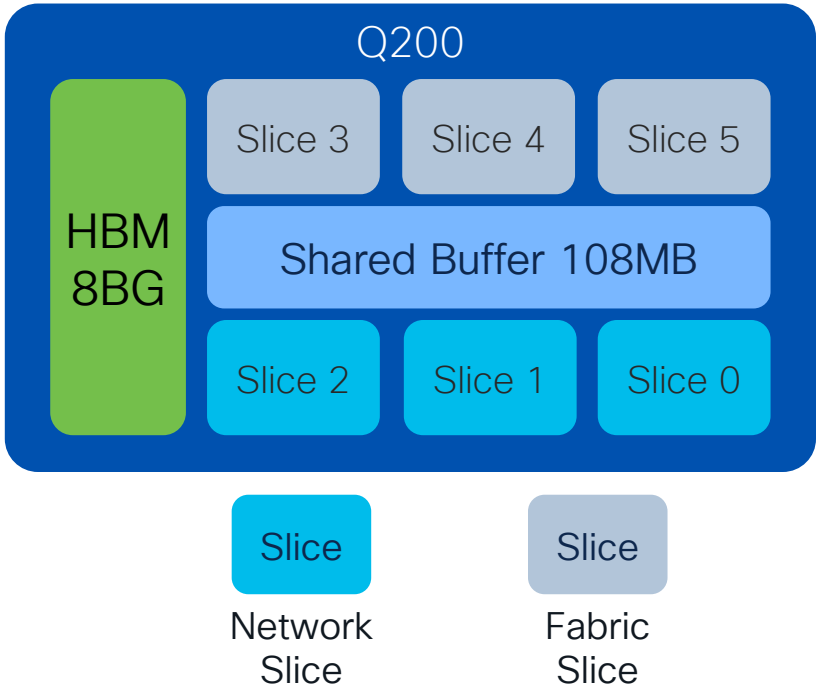


Cisco Nexus 9800 – QoS Features

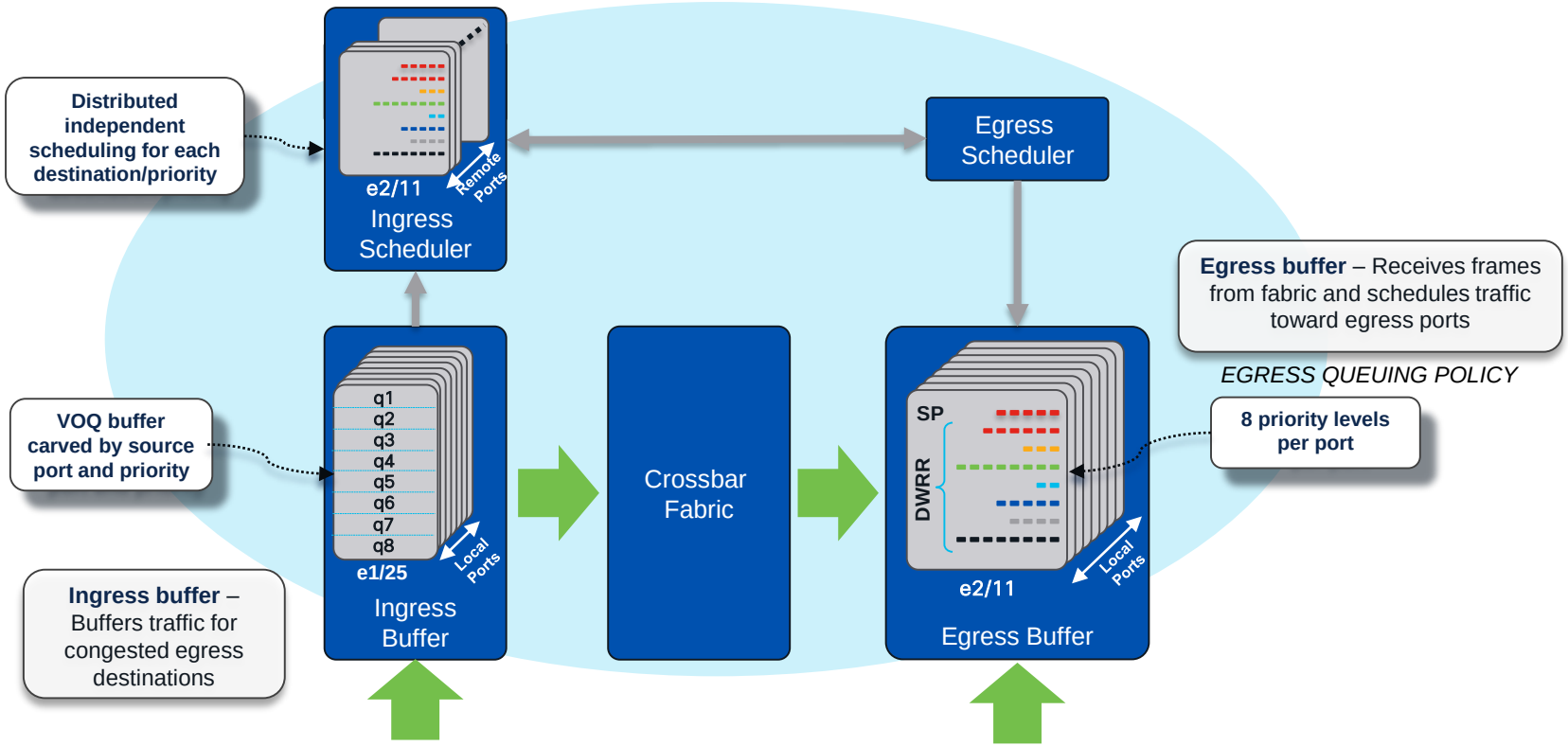
- Classification based on:
  - ACL
  - DSCP, CoS, and IP Precedence
- Marking traffic with:
  - DSCP
  - CoS
  - IP Precedence
- Policing:
  - 1R2C and 2R3C
  - Ingress
- Buffering/Queueing:
  - Fully Shared VoQ; 8 Egress Queues
- Scheduling:
  - Strict Priority Queuing and DWRR
- Shaping:
  - Egress per queue shaper
- Congestion Avoidance:
  - Tail Drop
  - WRED with ECN

Nexus 9800 – Buffering

- Q200 Silicon One ASIC
  - Primary Shared Buffer
    - Absorbs initial congestion
    - All slices have access to share buffer
  - HBM Buffer
    - Absorbs sever congestion



Queuing and Scheduling



Configuration

- On Nexus 9800 series, uses the same QoS configuration as Nexus 9000 Cloud Scale
- Type QOS used for Classification/ Marking/Policing, and association to QoS-Group
- Type Queueing used for Queueing/Scheduling adjustments
- Type Network-QoS used to accommodate different queueing model (4 Queue or 8 Queue), and non-drop queueing properties

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale QoS
- **Nexus 9800 QOS**
- Real World Configuration Examples
- Conclusion

What do we want to achieve?

Company XYZ's Business Goals

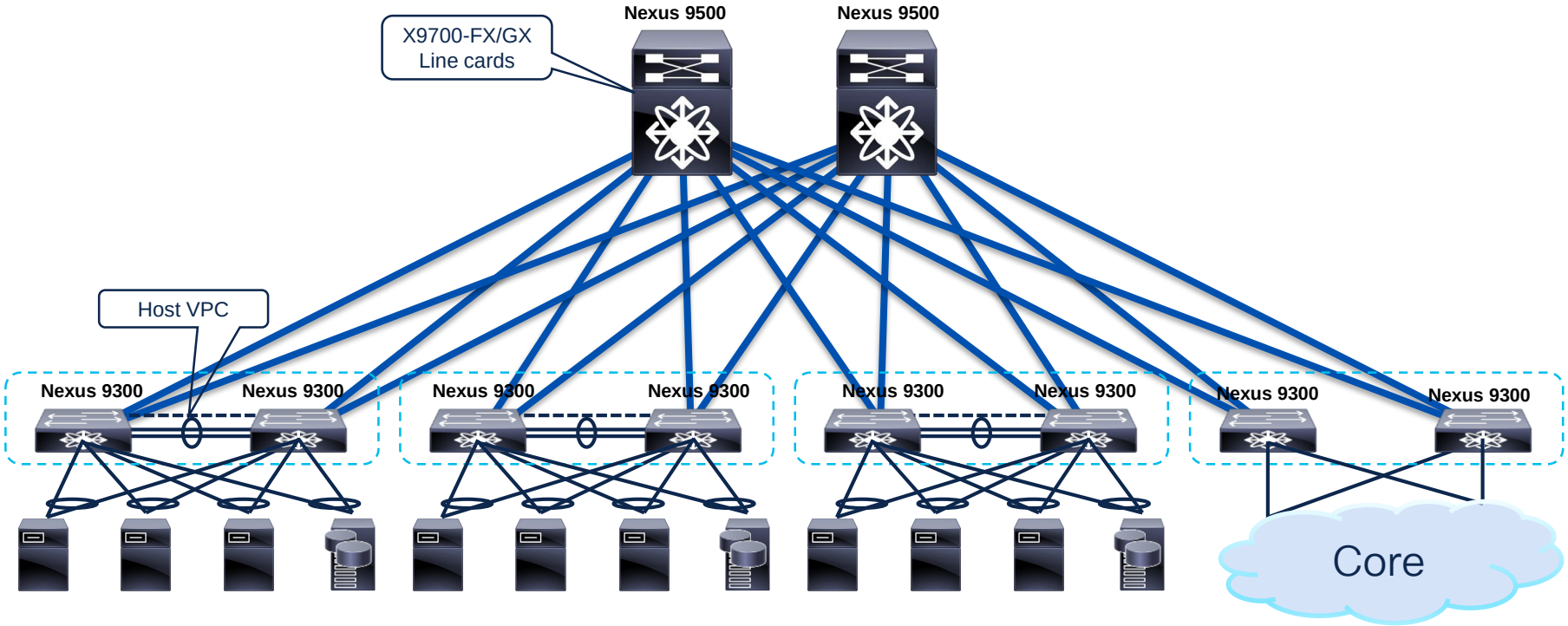
- Make sure no disruption in network services
  - Put control traffic in priority queue
- Video/voice hosting also a business objective
  - Put voice traffic in priority queue
  - Dedicated bandwidth to video traffic
- Flexibility in moving applications across servers
  - Dedicated bandwidth to vmotion/mobility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

Translating to the language of QoS

Application	CoS	DSCP	Queuing (Scheduling)	Character
Best Effort	0, 1	0, 8	BW remaining 40%	High Volume / Less Important
vMotion / Live Migration	2	N/A*	BW remaining 20%	Medium Volume / Important
Multimedia (Video)	3, 4	24, 32	BW remaining 30%	Medium Volume Very Important
Strict Priority (Voice)	5	46	Priority Queue	Low Volume / Important / Delay Sensitive
Network Control	6,7	48, 56		Low Volume / Very important

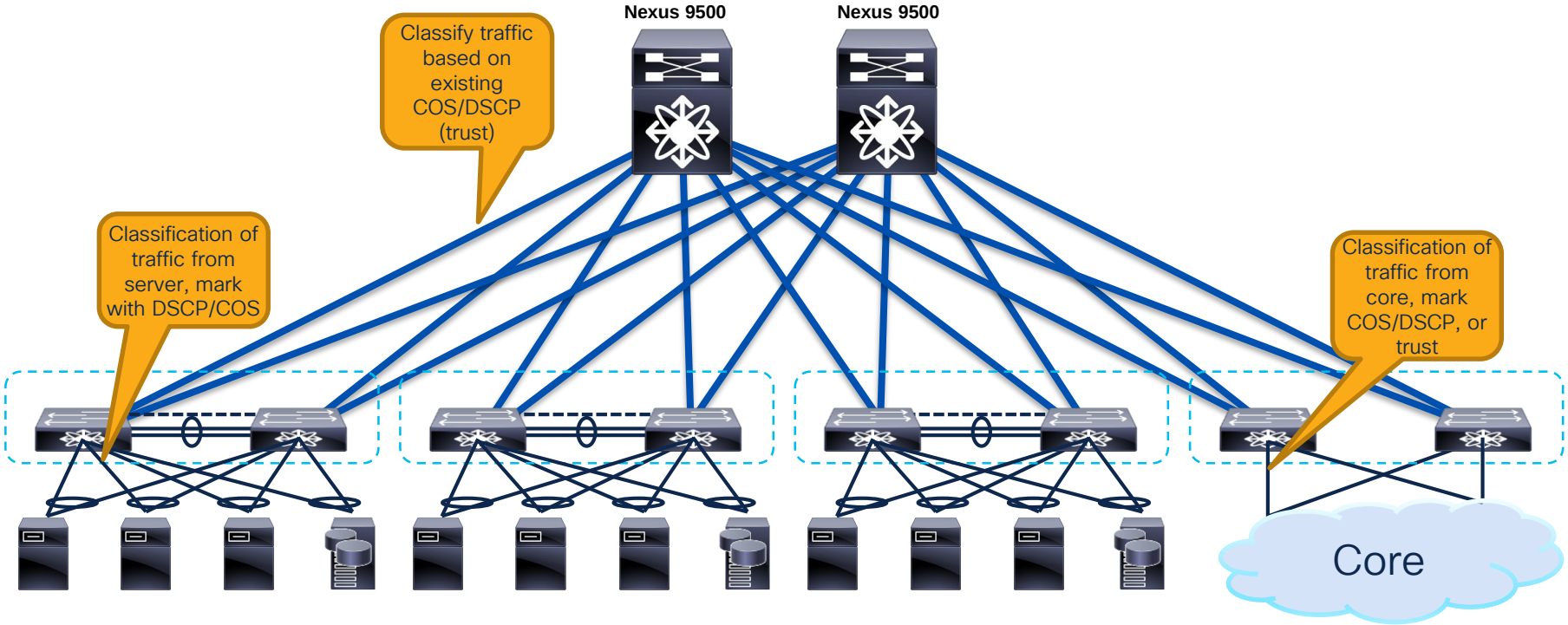
* Layer 2 traffic without IP header

Topology



Type:
QoS

Classification, Marking and Trust



Marking Definition

Application	CoS	DSCP	Character
Best Effort	0, 1	0, 8	High Volume / Less Important
vMotion / Live Migration	2	N/A*	Medium Volume / Important
Multimedia (Video)	3, 4	24, 32	Medium Volume Very Important
Strict Priority (Voice)	5	46	Low Volume / Important / Delay Sensitive
Network Control	6,7	48, 56	Low Volume / Very important

Classification and Marking

Nexus 9300 Leaf (Host Interfaces)

```
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
  march dscp 46
```

```
policy-map type qos PM_QOS_MARK_COS_IN
  class CM_QOS_STRICTPRIO_COS5
    set qos-group 7
    set cos 5
    set dscp 46
  class CM_QOS_MULTIMEDIA_COS4
    set qos-group 4
    set cos 4
    set dscp 32
  class CM_QOS_VMOTION_COS2
    set qos-group 2
    set cos 2
class CM_QOS_LOWPRIO_COS1
    set qos-group 0
    set cos 1
    set dscp 8
!
interface Ethernet 1/1
  service-policy type qos input PM_QOS_MARK_COS_IN
!
vlan configuration 100
  service-policy input PM_QOS_MARK_COS_IN
```

Classification and Marking

Nexus 9300 Leaf (Uplink Interfaces) and Nexus 9500 (Spine Interfaces)

```
class-map type qos match-any CM_QOS_LOWPRIO_COS1
  match dscp 8
!
class-map type qos match-any CM_QOS_VMOTION_COS2
  match dscp 16
!
class-map type qos match-any CM_QOS_MULTIMEDIA_COS4
  match dscp 32
!
class-map type qos match-any CM_QOS_STRICTPRIO_COS5
  match dscp 46
```

```
policy-map type qos PM_QOS_MARK_COS_IN
  class CM_QOS_STRICTPRIO_COS5
    set qos-group 7
class CM_QOS_MULTIMEDIA_COS4
    set qos-group 4
class CM_QOS_VMOTION_COS2
    set qos-group 2
class CM_QOS_LOWPRIO_COS1
    set qos-group 0
!
interface Ethernet 1/1
  service-policy type qos input PM_QOS_MARK_COS_IN
```

Queueing and Scheduling

Nexus 9300, and 9500

Application	CoS	DSCP	Queuing (Scheduling)	Queue limit (Alpha)	Queue	Character
Best Effort	1	8	BW percent 30%	Default (9)	qos-group 0	High Volume / Less Important
vMotion / Live Migration	2,3	16	BW percent 30%	Default (9)	qos-group 2	Medium Volume / Important
Multimedia (Video)	4	24, 32	BW percent 40%	Default (9)	qos-group 4	Medium Volume Very Important
Strict Priority (Voice)	5	46	Priority Queue	Default (9)	qos-group 7	Low Volume / Important / Delay Sensitive
Network Control	6,7	48, 56		Default (9)		Low Volume / Very important

Queueing and Scheduling

Nexus 9300, and 9500

- Class-maps type queueing are predefined
- Class-maps referring to qos-groups

```
policy-map type queuing custom-8q-out-policy
  class type queuing c-out-8q-q7
    priority level 1
  class type queuing c-out-8q-q6
    bandwidth remaining percent 0
  class type queuing c-out-8q-q5
    bandwidth remaining percent 0
  class type queuing c-out-8q-q4
    bandwidth remaining percent 40
  class type queuing c-out-8q-q3
    bandwidth remaining percent 0
  class type queuing c-out-8q-q2
    bandwidth remaining percent 30
  class type queuing c-out-8q-q1
    bandwidth remaining percent 0
  class type queuing c-out-8q-q-default
    bandwidth remaining percent 30

system qos
  service-policy type queuing output custom-8q-out-policy
```

Network-QoS

- Keep default Network-QoS:
  - Default 8 Queue model
  - No configuration for non-drop queue

Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale QoS
- **Real World Configuration Examples**
- Conclusion

Why QoS in the Data Centre?

Assign Colour to Traffic

Manage Congestion

Maximise Throughput

Complete Your Session Evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to **win 1 of 5 full conference passes** to Cisco Live 2025.

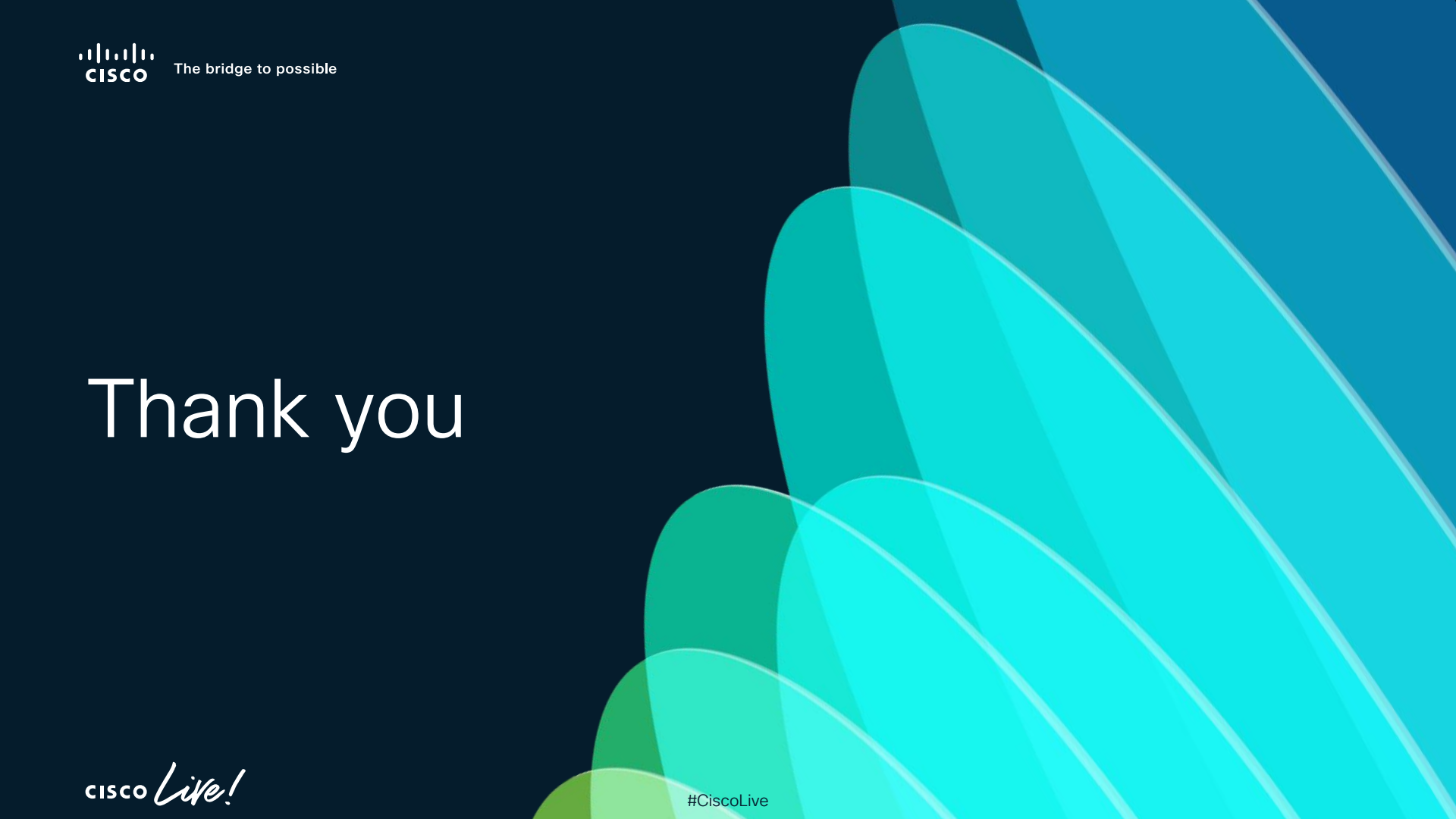
Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn **exclusive prizes!**

Complete your surveys in the **Cisco Live mobile app.**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand


The bridge to possible
Thank you
cisco Live!
#CiscoLive